TRENDS IN NEUROERGONOMICS: A COMPREHENSIVE OVERVIEW

EDITED BY: Klaus Gramann, Stephen H. Fairclough, Thorsten O. Zander and Hasan Ayaz

PUBLISHED IN: Frontiers in Human Neuroscience



frontiers Research Topics



frontiers

Frontiers Copyright Statement

© Copyright 2007-2017 Frontiers Media SA. All rights reserved.

All content included on this site, such as text, graphics, logos, button icons, images, video/audio clips, downloads, data compilations and software, is the property of or is licensed to Frontiers Media SA ("Frontiers") or its licensees and/or subcontractors. The copyright in the text of individual articles is the property of their respective authors, subject to a license granted to Frontiers.

The compilation of articles constituting this e-book, wherever published, as well as the compilation of all other content on this site, is the exclusive property of Frontiers. For the conditions for downloading and copying of e-books from Frontiers' website, please see the Terms for Website Use. If purchasing Frontiers e-books from other websites or sources, the conditions of the website concerned apply.

Images and graphics not forming part of user-contributed materials may not be downloaded or copied without permission.

Individual articles may be downloaded and reproduced in accordance with the principles of the CC-BY licence subject to any copyright or other notices. They may not be re-sold as an e-book.

As author or other contributor you grant a CC-BY licence to others to reproduce your articles, including any graphics and third-party materials supplied by you, in accordance with the Conditions for Website Use and subject to any copyright notices which you include in connection with your articles and materials.

All copyright, and all rights therein, are protected by national and international copyright laws.

The above represents a summary only. For the full conditions see the Conditions for Authors and the Conditions for Website Use.

ISSN 1664-8714
ISBN 978-2-88945-203-3
DOI 10.3389/978-2-88945-203-3

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view.
By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: **researchtopics@frontiersin.org**

TRENDS IN NEUROERGONOMICS: A COMPREHENSIVE OVERVIEW

Topic Editors:
Klaus Gramann, Berlin Institute of Technology (TU Berlin), Germany & University of California, San Diego, USA
Stephen H. Fairclough, Liverpool John Moores University, UK
Thorsten O. Zander, Berlin Institute of Technology (TU Berlin), Germany
Hasan Ayaz, Drexel University & University of Pennsylvania & Children's Hospital of Philadelphia, USA

This Research Topic is dedicated to Raja Parasuraman who unexpectedly passed on March 22nd 2015. Raja Parasuraman's pioneering work led the emergence of Neuroergonomics as a new scientific field. He combined his research interests in the field of Neuroergonomics which he defined as the study of the human brain in relation to performance at work and everyday settings. Raja Parasuraman was a pioneer, a truly exceptional researcher and an extraordinary person. He made significant contributions to a number of disciplines, from human factors to cognitive neuroscience. His advice to young researchers was to be passionate in order to develop theory and knowledge that can guide the design of technologies and environments for people. His legacy, the field of Neuroergonomics, will live on in countless faculties and students whom he advised and inspired with unmatched humility throughout the whole of his distinguished career. Raja Parasuraman was an impressive human being, a very kind person, and an absolutely inspiring individual who will be remembered by everyone who had the chance to meet him.

About this Research Topic

Since the advent of neuroergonomics, significant progress has been made with respect to methodology and tools for the investigation of the brain and behavior at work. This is especially the case for neuroscientific methods where the availability of ambulatory hardware, wearable sensors and advanced data analyses allow for imaging of brain dynamics in humans in applied environments. Methods such as: electroencephalography (EEG), functional near-infrared spectroscopy (fNIRS), and stimulation approaches like transcranial direct-currrent stimulation (tDCS) have made significant progress in both recording and altering brain activity while allowing full body movements outside laboratory environments.

For neuroergonomics, the application of brain imaging in real-world scenarios is highly relevant. Traditionally, brain imaging experiments in human factors research tend to avoid active behavior for fear of artifacts and a contaminated data set that would provide limited insight into brain dynamics in real working environments. To overcome these problems new analyses approaches have to be developed that identify artifacts resulting from hostile recording environments and movement-related non-brain activity stemming from eye-, head, and full-body movements. The application of methodology from the field of Brain-Computer Interfacing

(BCI) for neuroergonomics is one approach that has significant potential to enhance ambulatory monitoring and applied testing. Passive BCIs allow for assessing aspects of the user state online, such that systems can automatically adapt to their user. This neuroadaptive technology could lead to highly efficient working environments, to auto-adaptive experimental paradigms and to a continuous tracking of cognitive and affective aspects of the user state. Hence, deployment of portable neuroimaging technologies to real time settings could help assess cognitive and motivational states of personnel assigned to perform critical tasks.

This Research Topic gathers submissions that cover new approaches in neuroergonomics. Different article type cover advanced neuroscience methods and neuroergonomics techniques as well as analysis approaches to investigate brain dynamics in working environments. The selection of papers provides insights into new neuroergonomic research approaches that demonstrate significant advances in brain imaging technologies that become more and more mobile, Moreover, a strong trend for new analyses approaches and paradigms investigating real work settings can be seen. Together, this unique collection of latest research papers provides a comprehensive overview on the latest developments in neuroergonomics.

Citation: Gramann, K., Fairclough, S. H., Zander, T. O., Ayaz, H., eds. (2017). Trends in Neuroergonomics: A Comprehensive Overview. Lausanne: Frontiers Media.
doi: 10.3389/978-2-88945-203-3

Table of Contents

Chapter 0: Introduction

07 ***Editorial: Trends in Neuroergonomics***
Klaus Gramann, Stephen H. Fairclough, Thorsten O. Zander and Hasan Ayaz

Chapter 1: Electroencephalography (EEG)
Section 1: Electroencephalography (EEG) in Neuroergonomics

11 ***Mobile Brain/Body Imaging (MoBI) of Physical Interaction with Dynamically Moving Objects***
Evelyn Jungnickel and Klaus Gramann

26 ***Age-Sensitive Effects of Enduring Work with Alternating Cognitive and Physical Load. A Study Applying Mobile EEG in a Real Life Working Scenario***
Edmund Wascher, Holger Heppner, Sven O. Kobald, Stefan Arnau, Stephan Getzmann and Tina Möckel

40 ***Benefits of Instructed Responding in Manual Assembly Tasks: An ERP Approach***
Pavle Mijović, Vanja Ković, Maarten De Vos, Ivan Mačužić, Branislav Jeremić and Ivan Gligorijević

53 ***Pre-Trial EEG-Based Single-Trial Motor Performance Prediction to Enhance Neuroergonomics for a Hand Force Task***
Andreas Meinel, Sebastián Castaño-Candamil, Janine Reis and Michael Tangermann

70 ***Evaluation of a Dry EEG System for Application of Passive Brain-Computer Interfaces in Autonomous Driving***
Thorsten O. Zander, Lena M. Andreessen, Angela Berg, Maurice Bleuel, Juliane Pawlitzki, Lars Zawallich, Laurens R. Krol and Klaus Gramann

Section 2: Brain-Computer Interfaces (BCI) in Neuroergonomics

86 ***An Intelligent Man-Machine Interface—Multi-Robot Control Adapted for Task Engagement Based on Single-Trial Detectability of P300***
Elsa A. Kirchner, Su K. Kim, Marc Tabie, Hendrik Wöhrle, Michael Maurus and Frank Kirchner

107 ***Perception and Cognition of Cues Used in Synchronous Brain–Computer Interfaces Modify Electroencephalographic Patterns of Control Tasks***
Luz María Alonso-Valerdi, Francisco Sepulveda and Ricardo A. Ramírez-Mendoza

120 ***The Brain Is Faster than the Hand in Split-Second Intentions to Respond to an Impending Hazard: A Simulation of Neuroadaptive Automation to Speed Recovery to Perturbation in Flight Attitude***
Daniel E. Callan, Cengiz Terzibas, Daniel B. Cassel, Masa-aki Sato and Raja Parasuraman

141 ***Efficient Workload Classification based on Ignored Auditory Probes: A Proof of Concept***
Raphaëlle N. Roy, Stéphane Bonnet, Sylvie Charbonnier and Aurélie Campagne

153 ***Gaussian Process Regression for Predictive But Interpretable Machine Learning Models: An Example of Predicting Mental Workload across Tasks***
Matthew S. Caywood, Daniel M. Roberts, Jeffrey B. Colombe, Hal S. Greenwald and Monica Z. Weiland

172 ***Evaluation of an Adaptive Game that Uses EEG Measures Validated during the Design Process as Inputs to a Biocybernetic Loop***
Kate C. Ewing, Stephen H. Fairclough and Kiel Gilleade

Section 3: EEG in Multimodel Recordings

185 ***Neural Mechanisms of Inhibitory Response in a Battlefield Scenario: A Simultaneous fMRI-EEG Study***
Li-Wei Ko, Yi-Cheng Shih, Rupesh Kumar Chikara, Ya-Ting Chuang and Erik C. Chang

200 ***Exploring Neuro-Physiological Correlates of Drivers' Mental Fatigue Caused by Sleep Deprivation Using Simultaneous EEG, ECG, and fNIRS Data***
Sangtae Ahn, Thien Nguyen, Hyojung Jang, Jae G. Kim and Sung C. Jun

214 ***Steering Demands Diminish the Early-P3, Late-P3 and RON Components of the Event-Related Potential of Task-Irrelevant Environmental Sounds***
Menja Scheer, Heinrich H. Bülthoff and Lewis L. Chuang

Chapter 2: Functional Near-Infrared Spectroscopy (fNIRS)

229 ***Toward a Wireless Open Source Instrument: Functional Near-infrared Spectroscopy in Mobile Neuroergonomics and BCI Applications***
Alexander von Lühmann, Christian Herff, Dominic Heger and Tanja Schultz

243 ***Why a Comprehensive Understanding of Mental Workload through the Measurement of Neurovascular Coupling Is a Key Issue for Neuroergonomics?***
Kevin Mandrick, Zarrin Chua, Mickaël Causse, Stéphane Perrey and Frédéric Dehais

248 ***Acute Supramaximal Exercise Increases the Brain Oxygenation in Relation to Cognitive Workload***
Cem Seref Bediz, Adile Oniz, Cagdas Guducu, Enise Ural Demirci, Hilmi Ogut, Erkan Gunay, Caner Cetinkaya and Murat Ozgoren

259 ***Prefrontal Cortex Activation Upon a Demanding Virtual Hand-Controlled Task: A New Frontier for Neuroergonomics***
Marika Carrieri, Andrea Petracca, Stefania Lancia, Sara Basso Moro, Sabrina Brigadoi, Matteo Spezialetti, Marco Ferrari, Giuseppe Placidi and Valentina Quaresima

272 ***Into the Wild: Neuroergonomic Differentiation of Hand-Held and Augmented Reality Wearable Displays during Outdoor Navigation with Functional Near Infrared Spectroscopy***
Ryan McKendrick, Raja Parasuraman, Rabia Murtza, Alice Formwalt, Wendy Baccus, Martin Paczynski and Hasan Ayaz

287 ***Processing Functional Near Infrared Spectroscopy Signal with a Kalman Filter to Assess Working Memory during Simulated Flight***
Gautier Durantin, Sébastien Scannella, Thibault Gateau, Arnaud Delorme and Frédéric Dehais

Chapter 3: Stimulation Methods

296 ***Commentary: Cumulative effects of anodal and priming cathodal tDCS on pegboard test performance and motor cortical excitability***
Pierre Besson, Stephane Perrey, Wei-Peng Teo and Makii Muthalib

299 Simultaneous tDCS-fMRI Identifies Resting State Networks Correlated with Visual Search Enhancement
Daniel E. Callan, Brian Falcone, Atsushi Wada and Raja Parasuraman

311 Transcranial Direct Current Stimulation Modulates Neuronal Activity and Learning in Pilot Training
Jaehoon Choe, Brian A. Coffman, Dylan T. Bergstedt, Matthias D. Ziegler and Matthew E. Phillips

336 Does a Combination of Virtual Reality, Neuromodulation and Neuroimaging Provide a Comprehensive Platform for Neurorehabilitation? – A Narrative Review of the Literature
Wei-Peng Teo, Makii Muthalib, Sami Yamin, Ashlee M. Hendy, Kelly Bramstedt, Eleftheria Kotsopoulos, Stephane Perrey and Hasan Ayaz

351 Corrigendum: Does a Combination of Virtual Reality, Neuromodulation and Neuroimaging Provide a Comprehensive Platform for Neurorehabilitation? – A Narrative Review of the Literature
Wei-Peng Teo, Makii Muthalib, Sami Yamin, Ashlee M. Hendy, Kelly Bramstedt, Eleftheria Kotsopoulos, Stephane Perrey and Hasan Ayaz

Chapter 4: Eye Movement Methods

352 High Working Memory Load Impairs Language Processing during a Simulated Piloting Task: An ERP and Pupillometry Study
Mickaël Causse, Vsevolod Peysakhovich and Eve F. Fabre

366 The impact of expert visual guidance on trainee visual search strategy, visual attention and motor skills
Daniel R. Leff, David R. C. James, Felipe Orihuela-Espina, Ka-Wai Kwok, Loi Wah Sun, George Mylonas, Thanos Athanasiou, Ara W. Darzi and Guang-Zhong Yang

377 The Role of Cognitive and Perceptual Loads in Inattentional Deafness
Mickaël Causse, Jean-Paul Imbert, Louise Giraudet, Christophe Jouffrais and Sébastien Tremblay

Chapter 5: Overview on Neuroscience Methods in Automation Research

389 From Trust in Automation to Decision Neuroscience: Applying Cognitive Neuroscience Methods to Understand and Improve Interaction Decisions Involved in Human Automation Interaction
Kim Drnec, Amar R. Marathe, Jamie R. Lukos and Jason S. Metcalfe

frontiers in Human Neuroscience

EDITORIAL
published: 05 April 2017
doi: 10.3389/fnhum.2017.00165

Editorial: Trends in Neuroergonomics

Klaus Gramann[1,2]*, *Stephen H. Fairclough*[3], *Thorsten O. Zander*[1,4] *and Hasan Ayaz*[5,6,7]

[1] Department of Psychology and Ergonomics, Berlin Institute of Technology (TU Berlin), Berlin, Germany, [2] Center for Advanced Neurological Engineering, University of California, San Diego, San Diego, CA, USA, [3] School of Natural Sciences and Psychology, Liverpool John Moores University, Liverpool, UK, [4] Team PhyPA, Berlin Institute of Technology (TU Berlin), Berlin, Germany, [5] School of Biomedical Engineering, Science and Health Systems, Drexel University, Philadelphia, PA, USA, [6] Department of Family and Community Health, University of Pennsylvania, Philadelphia, PA, USA, [7] Division of General Pediatrics, Children's Hospital of Philadelphia, Philadelphia, PA, USA

Keywords: neuroergonomics, human-machine interaction, brain-computer interface, mobile brain/body imaging, physiological computing, EEG/ERP, NIRS/fNIRS, tDCS

Editorial on the Research Topic

Trends in Neuroergonomics

NEW METHODS IN NEUROERGONOMICS

This Research Topic is dedicated to Professor Raja Parasuraman who unexpectedly passed on March 22nd 2015.

Raja Parasuraman's pioneering work led to the emergence of Neuroergonomics as a new scientific field. Neuroergonomics is defined as the study of the human brain in relation to performance at work and everyday settings (Parasuraman, 2003; Parasuraman and Rizzo, 2008). Since the advent of Neuroergonomics, significant progress has been made with respect to methodology and tools for the investigation of the brain and behavior at work. This is especially the case for neuroscientific methods where the availability of ambulatory hardware, wearable sensors, and advanced data analyses allow for imaging of brain dynamics in humans in applied environments.

For neuroergonomics, the application of brain imaging in real-world scenarios is highly desirable as an investigation tool. Traditionally, brain imaging experiments in human factors research tend to avoid active behavior for fear of artifacts contaminating the signal of interest. Here, the development of new data analyses techniques as well as the combination of different methods providing complementary insights into brain and behavioral dynamics allow for new insights into the human-machine interaction. To overcome the problem of artifactual data in mobile recordings and to allow analyses of brain activity in real working environments new portable sensors and improved analyses approaches have to be developed. Hence, deployment of portable neuroimaging technologies to real time settings could help assess cognitive and motivational states of personnel assigned to perform critical tasks.

THE "TRENDS IN NEUROERGONOMICS" RESEARCH TOPIC: A BRIEF INTRODUCTION

The eBook of this Frontiers Research Topic is divided into four sections, defined by the primary research methods used to address a variety of neuroergonomic research questions. The scientific topics range from air traffic control and automation, over mental load detection and the use of brain activity to control a system (brain computer interfaces, BCI), to physical work, rehabilitation, and training. Across the diverse research areas, the majority of studies in this Research Topic used electroencephalography (EEG), followed by functional near infrared spectroscopy (fNIRS),

OPEN ACCESS

Edited and reviewed by:
Mikhail Lebedev,
Duke University, USA

***Correspondence:**
Klaus Gramann
klaus.gramann@tu-berlin.de

Received: *07 February 2017*
Accepted: *21 March 2017*
Published: *05 April 2017*

Citation:
Gramann K, Fairclough SH, Zander TO and Ayaz H (2017) Editorial: Trends in Neuroergonomics. Front. Hum. Neurosci. 11:165. doi: 10.3389/fnhum.2017.00165

reflecting the constantly growing use of these methods in neuroergonomics. At the same time this eBook clearly demonstrates a trend to investigate physical and cognitive activity outside standard laboratory settings, moving neuroergonomics "into the wild." In addition, traditional methods like the measurement of eye movements, pupil metrics, (ECG), and established imaging approaches like functional magnetic resonance imaging (fMRI) are used in combination with other methods to better understand the physiological responses to cognitive or physical tasks and their coupling to hemodynamic changes. The different sections include original research articles, but also reviews and opinion pieces. Here, we provide short summaries as an orientation for the interested reader.

The first section in this eBook thus comprises studies that used EEG to investigate ergonomic research questions with three studies using EEG in a mobile setting. The first of these by Jungnickel and Gramann uses a mobile brain/body imaging approach (MoBI; Makeig et al., 2009; Gramann et al., 2012, 2014) to investigate the brain and behavioral dynamics of human participants interacting with dynamically moving objects. The results indicate increased activity in parietal regions when active physical behavior as compared to standard laboratory button press behavior was required to respond to relevant changes in the environment. The findings point to changes in brain dynamic states dependent on the behavioral state. The study by Wascher et al. demonstrates that mobile EEG allows for a non-obtrusive assessment of mental fatigue in natural working situations. The authors investigate EEG variations time-locked to eye-blinks as a new tool to unobtrusively monitor cognitive processing in real-life environments. Mijović et al. also use EEG in a naturalistic work environment to show that instructed responses can increase attention as reflected in brain dynamics without changes in response parameters. The results point to the possibility to use instructed responses to increase attentional processing without compromising work performance in manual assembly tasks. Again, explicitly allowing movement of participants, Meinel et al. demonstrate that EEG can be used to improve motor rehabilitation approaches. Better performance in movement tasks can be achieved by identifying comodulation of different sources in the EEG before a movement is executed. The results provide participant-specific prediction of performance fluctuations that could be used to enhance neuroergonomic and rehabilitation scenarios. The last study by Zander et al. investigates how well a passive brain-computer Interface can work in an autonomous driving scenario using a dry EEG system. The results reveal comfort issues but acceptable usability of the tested EEG system and sufficient signal quality for use in an autonomous driving context.

A number of EEG-studies use the recorded brain electrical signals for system interaction through brain-computer interfaces (BCIs, Zander and Kothe, 2011). In this context, Kirchner et al. demonstrate that event-related potentials can be used on a single trial level to infer task-engagement of an operator controlling multiple robots and to adapt the man-machine interface to the individual operator. The results could be used to adapt the task load to operators with different qualifications or capabilities to avoid mental overload. Alonso-Valerdi et al. suggest that a wider variety of control commands in motor imagery-based BCIs might lead to an accelerate brain-computer communication while Callan et al. increase response speed in flight simulation by using a passive BCI based on MEG. The former study provides insights into the use of control commands to increase BCI-based system communication while the latter study demonstrates the potential to decode motor intention faster than manual control in response to hazardous change in the system interaction cycle. Roy et al. investigate mental workload based on auditory evoked potentials. The authors present a new minimal intrusive paradigm that paves the way to monitor operators' mental state in real-life settings to allow adaption of the user interface without interfering with the primary task. Caywood et al. increase the interpretability of BCI models by using the approach of Gaussian Process Regression for assessing cognitive workload. Ewing et al. describe the development of an adaptive game system that measures spontaneous EEG activity in real-time in order to adjust the difficulty level of the game. In two studies the concept of a biocybernetic control loop (Fairclough, 2009) is introduced in detail with a particular emphasis on validating EEG measures experimentally prior to their incorporation into an adaptive game system.

Studies using EEG in combination with other methods show that different physiological parameters can lead to an improved understanding of the construct under investigation. Ko et al. demonstrate the advantage of integrating the high resolution in the time and spatial domain for EEG and fMRI, respectively, in a stop-signal paradigm. Their results from multimodal recordings provide new insights into the complex brain networks underlying inhibitory control in naturalistic environments. Using EEG in combination with ECG and fNIRS, Ahn et al. investigate mental fatigue in drivers. They show that a combination of different physiological measures substantially improves the classification of sleep deprived or well-rested drivers. Scheer et al. investigate the demands on mental resources during a closed-loop steering task in simulated car driving scenarios. The results indicate an impact of steering demands on event-related EEG activity for task-irrelevant distractor probes allowing for an evaluation of mental workload in steering environments.

The second section summarizes the use of fNIRS in traditional stationary settings but also in new mobile applications. The first study of these, Von Lühmann et al. describe the development of a wireless and low-cost open source fNIRS hardware with details on system concept, hardware, software and mechanical implementation. The proposed system was tested in a mental arithmetic BCI experiment. Mandrick et al. discuss electrocortical and neurovascular measures with respect to the measurement of mental workload. They propose that EEG and fNIRS are complementary methods in the context of applied testing in the sense that the weakness of each approach, e.g., poor spatial and temporal resolution, respectively, is counteracted by the strength of the other. They argue for a combined fNIRS-EEG approach to index neurovascular coupling during the assessment of mental workload.

In Bediz et al. the effects of supramaximal exercise on cognitive task related oxygenation changes are investigated. Performance in a working memory task before and after

exercise indicated higher task related activation changes in prefrontal cortex post-exercise and higher cognitive task related brain activation increase in high-performing participants. The study by Carrieri et al. investigates the neural correlates of a cognitive/motor task in a virtual reality (VR) environment. The findings support the use of VR in combination with fNIRS as a very good platform for neuroergonomic studies to objectively evaluate cortical hemodynamic activity. Mckendrick et al. utilize ultra-portable, wearable and miniaturized fNIRS sensors on participants walking outdoors in the open air. The battery-operated miniaturized system implementation was described by Ayaz et al. (2013). The results of a spatial navigation task indicated greater mental capacity reserves for users with head mounted displays but also unwanted attention capture and cognitive tunneling as indicated by hemodynamics measures. The final contribution in this section by Durantin et al. provides evidence for using Kalman filter as a suitable approach for real-time noise removal for fNIRS signals in ecological situations and the development of BCI. The findings from working memory tasks indicate Kalman filter increased the performance of the classification of task load levels based on brain signal.

Section three comprises studies using stimulation methods alone or in combination with other methods to investigate stimulation-based improvement of motor or cognitive functions. The introductory commentary by Besson et al. prepares the stage for this section by providing a critical commentary on existing transcranial direct current stimulation (tDCS) protocols to enhance neuroplasticity and enhance performance in real-world settings. They argue that priming tDCS protocols have significant potential to improve learning and motor performance. Callan et al. demonstrate the simultaneous use of tDCS and fMRI to investigate the effect of neurostimulation on resting state functional connectivity and behavioral performance. The results reveal greater spontaneous resting state activity for the tDCS group with higher resting state functional connectivity for participants demonstrating performance improvement in a visual search task. Choe et al. investigate the use of tDCS along with multimodal neuroimaging (EEG and fNIRS) demonstrating that tDCS stimulation to the dorsolateral prefrontal or left motor cortex during flight simulation training enhances behavioral performance and changes neurophysiological measurements (EEG and fNIRS) indicating improved skill acquisition consistent with previous studies (Ayaz et al., 2012). The final contribution in this section is the review of Teo et al. providing a discussion of the theoretical framework underlying the use of VR in combination with neuroimaging and neuromodulation as a therapeutic intervention for neurorehabilitation. The authors provide evidence for the use of VR in treating motor and mental disorders such as cerebral palsy, Parkinson's disease, stroke, schizophrenia, anxiety disorders, and other emerging clinical areas.

The fourth and last section in this eBook then covers research approaches using eye movement measures including pupilometric methods and other peripheral physiological methods like ECG. The first study in this section by Causse et al. uses EEG and pupilometry to investigate the impact of high working memory load on language processing during piloting. The results demonstrate high working memory load to disrupt visual and language processing and a subtle effect of congruency that was observable only at an electrophysiological level. In the study by Leff et al. the authors use a collaborative gaze channel (CGC) to detect and display trainer gaze behavior to trainees in surgery tasks. The results of a simulated robotic surgery task imply liberation of attentional resources with the use of CGC potentially improving the capability of trainees to attend to additional safety critical events during the procedure. Causse et al. then demonstrate that the pupil diameter correlates with inattentional deafness in an air traffic control task with varying perceptual and cognitive load.

The final contribution to this Research Topic is a review on neuroscientific methods in automation research by Drnec et al. The authors provide a comprehensive overview on neuroscientific methods in trust in automation research and summarize how neuroscience can improve interaction design.

CONCLUDING REMARKS

Neuroergonomics has demonstrated an incredible development since the introduction of the field by Raja Parasuraman. This Research Topic demonstrates how different methods can be used to better understand the mind, body, and brain at work and to create and design systems that are better adapted to and make use of the human information processing structures, including the body and the brain. This research is important, because it allows for a human-centered design of environments that include natural behaviors.

AUTHOR CONTRIBUTIONS

All authors listed, have made substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

This Research Topic gathers, among other contributions, papers presented at the 11th Berlin Workshop Human-Machine Systems, 2016 supported by a grant from the German Research Foundation (DFG GR2627/6-1) awarded to KG.

REFERENCES

Ayaz, H., Onaral, B., Izzetoglu, K., Shewokis, P. A., McKendrick, R., and Parasuraman, R. (2013). Continuous monitoring of brain dynamics with functional near infrared spectroscopy as a tool for neuroergonomic research: Empirical examples and a technological development. *Front. Hum. Neurosci.* 7, 1–13. doi: 10.3389/fnhum.2013.00871

Ayaz, H., Shewokis, P. A., Bunce, S., Izzetoglu, K., Willems, B., and Onaral, B. (2012). Optical brain monitoring for operator training and mental workload assessment. *Neuroimage* 59, 36–47. doi: 10.1016/j.neuroimage.2011.06.023

Fairclough, S. H. (2009). Fundamentals of physiological computing. *Interact. Comput.* 21, 133–145. doi: 10.1016/j.intcom.2008.10.011

Gramann, K., Ferris, D. P., Gwin, J., and Makeig, S. (2014). Imaging natural cognition in action. *Int. J. Psychophysiol.* 91, 22–29. doi: 10.1016/j.ijpsycho.2013.09.003

Gramann, K., Gwin, J. T., Ferris, D. P., Oie, K., Jung, T.-P., Lin, C.-T., et al. (2012). Cognition in action: Imaging brain/body dynamics in mobile humans. *Rev. Neurosci.* 22, 593–608. doi: 10.1515/RNS.2011.047

Makeig, S., Gramann, K., Jung, T.-P., Sejnowski, T. J., and Poizner, H. (2009). Linking brain, mind and behavior. *Int. J. Psychophysiol.* 73, 95–100. doi: 10.1016/j.ijpsycho.2008.11.008

Parasuraman, R. (2003). Neuroergonomics: research and practice. *Theor. Issues Ergon. Sci.* 4, 5–20. doi: 10.1080/14639220210199753

Parasuraman, R., and Rizzo, M. (2008). *Neuroergonomics*: *The Brain at Work*. Oxford: Oxford University Press.

Zander, T. O., and Kothe, C. (2011). Towards passive brain–computer interfaces: applying brain–computer interface technology to human–machine systems in general. *J. Neural Eng.* 8:025005. doi: 10.1088/1741-2560/8/2/025005

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Gramann, Fairclough, Zander and Ayaz. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 27 June 2016
doi: 10.3389/fnhum.2016.00306

Mobile Brain/Body Imaging (MoBI) of Physical Interaction with Dynamically Moving Objects

Evelyn Jungnickel[1]* *and Klaus Gramann*[1,2]

[1] Department of Psychology and Ergonomics, Biological Psychology and Neuroergonomics, Institute of Psychology and Ergonomics, Berlin Institute of Technology, Berlin, Germany, [2] Center for Advanced Neurological Engineering, University of California, San Diego, CA, USA

The non-invasive recording and analysis of human brain activity during active movements in natural working conditions is a central challenge in Neuroergonomics research. Existing brain imaging approaches do not allow for an investigation of brain dynamics during active behavior because their sensors cannot follow the movement of the signal source. However, movements that require the operator to react fast and to adapt to a dynamically changing environment occur frequently in working environments like assembly-line work, construction trade, health care, but also outside the working environment like in team sports. Overcoming the restrictions of existing imaging methods would allow for deeper insights into neurocognitive processes at workplaces that require physical interactions and thus could help to adapt work settings to the user. To investigate the brain dynamics accompanying rapid volatile movements we used a visual oddball paradigm where participants had to react to color changes either with a simple button press or by physically pointing towards a moving target. Using a mobile brain/body imaging approach (MoBI) including independent component analysis (ICA) with subsequent backprojection of cluster activity allowed for systematically describing the contribution of brain and non-brain sources to the sensor signal. The results demonstrate that visual event-related potentials (ERPs) can be analyzed for simple button presses and physical pointing responses and that it is possible to quantify the contribution of brain processes, muscle activity and eye movements to the signal recorded at the sensor level even for fast volatile arm movements with strong jerks. Using MoBI in naturalistic working environments can thus help to analyze brain dynamics in natural working conditions and help improving unhealthy or inefficient work settings.

Keywords: mobile brain/body imaging, EEG, embodied cognition, independent component analysis, P300, oddball paradigm, MoBI

OPEN ACCESS

Edited by:
Lutz Jäncke,
University of Zurich, Switzerland

Reviewed by:
Edmund Wascher,
Leibniz Research Centre for Working Environment and Human Factors, Germany
Irene Sturm,
Berlin School of Mind and Brain, Germany

***Correspondence:**
Evelyn Jungnickel
evelyn.jungnickel@tu-berlin.de

Received: *08 January 2016*
Accepted: *03 June 2016*
Published: *27 June 2016*

Citation:
Jungnickel E and Gramann K (2016) Mobile Brain/Body Imaging (MoBI) of Physical Interaction with Dynamically Moving Objects. Front. Hum. Neurosci. 10:306. doi: 10.3389/fnhum.2016.00306

INTRODUCTION

Studying human brain dynamics accompanying natural cognition (Gramann et al., 2014) works best by studying the brain under naturalistic conditions. The embodied cognition paradigm claims that the body's interactions with the world are an essential root of cognitive processes (Wilson, 2002). Thus it appears that perception and action should both be considered when studying

cognitive processes and their neural basis. However, conventional neuroimaging studies consider electrical potentials generated by eye movement or muscle activity during physical movements as artifacts that have to be avoided not to contaminate the signal of interest. This view led to experimental setups that restrict participants' mobility and require them to sit still or lie even in tasks that would require standing or moving (Makeig et al., 2009; Gramann et al., 2011, 2014). These constraints are changing the way information is perceived and processed by the human agent as becomes obvious, for example, with respect to the integration of proprioceptive and vestibular information (Gramann, 2013). This kind of idiothetic information is absent when movement is restricted or altered in case the body orientation differs from its natural state for a particular task. Following the embodied cognition approach those alterations will change the concurring cognitive processes and thereby lead to different brain activity.

Neuroergonomics as the scientific study of the human brain in relation to performance at work and everyday settings (Parasuraman, 2003) is faced with the challenge to investigate the brain dynamics in environments that require physical interaction of the operator with a system. New insights into brain activity during physical human-machine interaction allow for the improvement of systems to adapt to the operators' physical and cognitive resources (see e.g., Wascher et al., 2016; Mijovic et al., 2016). However, traditional brain imaging approaches do not allow for any kind of movement (Makeig et al., 2009; Gramann et al., 2011). Mobile brain/body imaging (MoBI), in contrast, is a general research approach that embraces a variety of (the best fitting) hardware and software solutions to record and analyze brain dynamics in actively behaving participants. Lightweight and mobile sensors like electroencephalography (EEG) or near infrared spectroscopy (fNIRS) agree with experimental paradigms using a MoBI approach to study the brain and body dynamics that accompany natural cognition and behaviors including physical interaction with an environment (Mehta and Parasuraman, 2013; Gramann et al., 2014). While fNIRS provides relatively high spatial resolution of a restricted cortical surface, this methods lacks the high temporal resolution that is desirable when investigating fast cognitive processes. EEG provides the necessary temporal resolution but has only limited spatial resolution. However, recent investigations using MoBI have demonstrated that equivalent dipole reconstruction of independent components (ICs) as decomposed by independent component analysis (ICA) allow for reconstructing the origin of EEG activity with reasonable spatial accuracy (Gramann et al., 2010a; Acar and Makeig, 2013). In conclusion, mobile EEG allows for an investigation of cognitive processes in working environments with high temporal resolution and with sufficient spatial resolution to allow for conclusions regarding the underlying cortical sources and their neuroanatomical function. Such a MoBI approach no longer considers eye movements and muscle activity as artifacts but as aspects of cognitive activity associated with the accomplishment of a task (Gramann et al., 2010a). By using high density EEG recordings synchronized with motion tracking of participant's movements and data-driven analyses methods it overcomes existing imaging restrictions and enables participants to behave more naturally (Makeig et al., 2009; Gramann et al., 2011, 2014).

First MoBI studies investigated participants walking and running on a treadmill and clearly demonstrated that brain activity can be analyzed under such conditions (Gramann et al., 2010a; Gwin et al., 2010, 2011). However, walking is a highly symmetric recurrent behavior that does not include fast movements associated with jerk. Stereotyped movements like walking further allow for extracting templates of artifacts based on recurrent movement patterns (Gwin et al., 2010). It is important to investigate to what extent MoBI can be used to measure and analyze brain dynamics during nonstereotyped and aperiodic behaviors that include sudden orientation movements or manual interaction with dynamic systems. First, such an approach could be used to determine how much traditional brain imaging results restricting participants' movements deviate from results in actively moving, more naturally behaving participants. Secondly, if feasible, such an approach would significantly increase the number of conceivable neurocognitive studies especially in the fields of physical ergonomics and in human-machine interaction that require physical manipulation. Insights gained from MoBI studies comprising natural recurrent and non-stereotyped movements would thus open up new vistas for investigating cognition and action within the field of Neuroergonomics and beyond.

This study investigated the feasibility of MoBI during physical interaction with a dynamic system based on non-stereotypical fast movements. The setup mimicked real-world working environments that require physical interaction in a dynamically changing system. Dynamic changes in the system were simulated using a three-stimulus visual oddball paradigm (Grillon et al., 1990) with participants reacting either by simple button presses or by pointing at the moving stimulus. We examined whether it is possible to record and analyze an event-related P3 component during rapid pointing movements that include strong eye movement and neck muscle activities. To this end we compared event-related potentials (ERPs) at the sensor level with ERPs back projected from ICs that decomposed the sensor data into maximally statistically independent time source series using ICA. By separating brain processes from activity generated by muscles and eye movement and comparing these to the scalp recorded potential allowed for a direct comparison and evaluation of the feasibility of standard sensor based analyses approaches during active pointing movements. In addition, isolation of brain related activity patterns and their contribution to the surface signal allowed for a quantification of how much certain ICs representing brain processes contributed to the surface signal.

MATERIALS AND METHODS

Participants

Data was collected from 15 healthy right-handed adult volunteers (7 females, 8 males) with a mean age of 26.1 years (σ = 2.9). All participants had normal or corrected to normal vision,

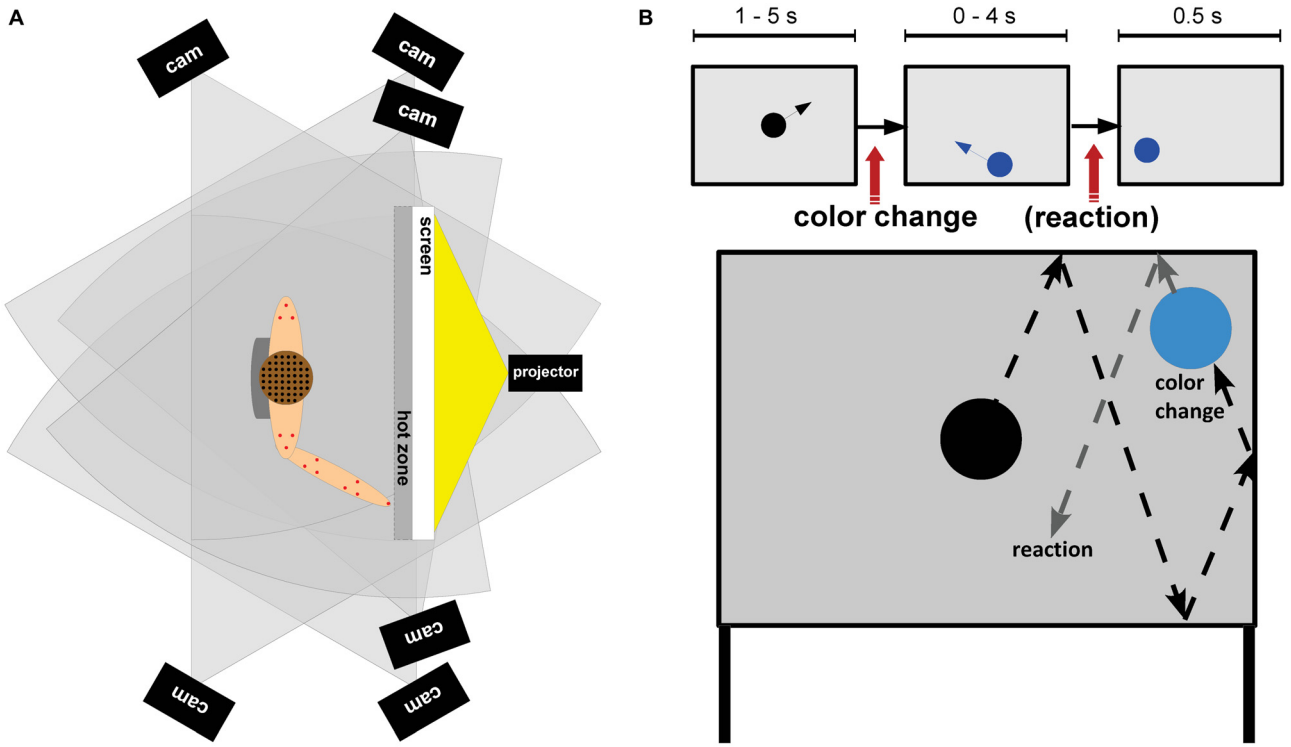


FIGURE 1 | (A) Experimental setup: top view of a participant standing in front of the screen. The setup comprised a motion capture system with six cameras (black rectangles) and 16 emitters (red dots), an EEG system with 156 wireless actively amplified electrodes (black dots), and the transmission system placed in a backpack (gray). **(B)** Task design. Top row: a black sphere moved over the projection screen and bounced off the walls of the screen. After changes to the target color participants responded according to the response condition either with a button press (button press condition) or a pointing movement towards the sphere (physical pointing condition). After a response or 4 s after a color change the sphere stopped and remained on the screen for 500 ms. Subsequently the next trial started. Bottom row: example of a trajectory of the sphere changing to the target color.

none reported a history of neurological disease and all provided written informed consent before the experiment in compliance with the standards as defined in the Declaration of Helsinki. The study was approved by the local ethics committee of the Institute of Psychology and Ergonomics of the Berlin Institute of Technology according to the guidelines of the German Psychological Society. Volunteers were compensated 12 €/h for their participation. Due to technical issues the behavioral data of three participants had to be excluded from further analysis and all results reported are based on the final group of 12 participants.

Experimental Design and Procedure

Participants stood in front of a projection screen (W × H: 1.2 m × 1.0 m) with a light gray background placed one arm length in front of them (**Figure 1**). Participants had to attend to a three-stimulus visual oddball paradigm and were asked to react to color changes of a moving sphere by either pointing to the stimulus with their right index finger (physical pointing condition) or pressing a response button (button press condition) on a Bluetooth remote (Logitech wireless presenter R400, Logitech, Apples, Switzerland). The response conditions were blocked and block order was counterbalanced across participants. Each response condition consisted of five blocks with 50 trials each. Breaks between blocks within each response condition were adapted to the participants needs.

Every trial began with a black sphere (ø14 cm) moving from the middle of the screen in a randomly chosen direction and being reflected from the borders of the projection screen. Color changes took place uniformly randomized between 1 and 5 s after onset of a trial. A change from black to blue indicated a target stimulus (15%), a change to green indicated a distractor stimulus (15%), and a change to yellow indicated a standard stimulus (70%). Participants were instructed to react as fast and correct as possible to the onset of the target color. After a response, or after 4 s in case no response was given, the sphere stopped moving and remained on the screen for 500 ms. Thus, the trial duration for correct non-target trials ranged from 5.5 to 9.5 s with an average duration of 7.5 s. For target trials the mean trial duration was shorter because button presses or pointing movements were executed before the 4 s time window closed. Thus, the duration of target trials depended on the response onset, movement speed and movement path. Altogether the experiment lasted about 1 h.

EEG Recording

The EEG was recorded from 156 active electrodes referenced to Cz with a sampling rate of 500 Hz and band-passed from

0.016 Hz to 250 Hz (BrainAmps and Move System, Brain Products, Gilching, Germany). To allow for recording of neck muscle activity resulting from participants' head movements, 28 electrodes were placed around the neck using a custom neck band (EASYCAP, Herrsching, Germany). The remaining 128 electrodes were placed on the head using an elastic cap with a custom design (EASYCAP, Herrsching, Germany). Electrode impedances were brought below 7 kΩ. Due to a technical problem the neck EEG data of one participant was not recorded. Individual electrode locations were recorded using an optical tracking system (Polaris Vicra, NDI, Waterloo, ON, Canada).

Motion Capture Recordings

Motion was captured using six cameras tracking the position of 16 red active LEDs (Impulse X2 System, PhaseSpace Inc., San Leandro, CA, USA) placed on the shoulders, the chest, and the right arm as well as the right index finger of the participants. The motion tracking system generated a data stream containing *x*, *y*, and *z* location and a reliability value for each LED with a sampling rate of 480 Hz. Before each data acquisition the screen position and orientation was calibrated to align with the motion capture coordinate system.

All data streams, namely EEG, motion capture, events from the experimental protocol, and behavioral data, were synchronized and recorded using the Lab Streaming Layer Software (Kothe, 2014).

Behavioral Analysis

In the physical pointing condition, online tracking of the LED on the participants' right index finger allowed to stop sphere movement as soon as the distance between the LED and the projection screen was smaller than 10 cm (labeled "hot zone" in **Figure 1**). This information was also used to create corresponding event markers. The LED was placed 5 cm apart from the fingertip approximately over the proximal phalanx of the index finger. The distance of 10 cm was chosen to avoid damage to the setup due to impact of the participants' finger with the screen. Because of occlusions the position of the finger LED was not recorded correctly in some trials and event markers were generated that did not match the movement profile of the participant. For the statistical analyses only trials with consistent event markers and motion tracking data of the right index finger were considered. This led to the exclusion of about 34.4% of the trials per participant in the physical pointing condition (range: 5.2–69.9%, σ = 21.7%) with the highest percentage of removals in standard ($\bar{x}$ = 38.0%) and distractor trials ($\bar{x}$ = 38.3%) that required no response. In these cases event markers indicated a movement even though in most cases the velocity profile did not indicate a response. In case of target trials on average only 13.1% were rejected.

To calculate velocity profiles from the motion capture data the MATLAB toolbox MoBILAB (Ojeda, 2011) was used. Occluded samples for each LED were interpolated by using spline interpolation and the data stream was smoothed by applying a 6 Hz low-pass zero phase distortion FIR filter before computing the velocity data. Subsequently the velocity profiles in the *z*-dimension of the LED placed on the index finger were analyzed with custom MATLAB scripts detecting pointing movements in the physical pointing condition on the basis of velocity peaks. To identify response movements, only the *z*-axis of the motion capture data was used indicating motions towards or away from the screen. This excluded smaller movements not related to the response. Based on velocity peaks defined as maximum positive deflections preceding and being followed by lower values, the onset, and offset of the corresponding movement were defined. For each color change the time window from 200 to 1800 ms after stimulus onset was selected to exclude movements unrelated to the stimulus response. As estimated from visual inspection only peaks with a velocity of at least 22% of the participants' maximum finger velocity in the physical pointing condition were regarded. This excluded smaller jerks and other movements not related to the pointing behavior. The definition of the movement onset is important in this context since its time-lag to the color change was taken as response time and used for further statistical analysis. The earliest movement onset was defined as the time point with a velocity of 5% of the subsequent peak velocity. To allow for a more conservative comparison of response times in the physical pointing condition with response times in the button press condition, increasing percentage values (>5%) of the subsequent maximum peak velocity were analyzed. The resulting movement onset distributions were then compared to response time distributions in the button press condition where no velocity profiles or force time-series could be derived. Response time statistics were calculated by means of a one-way analysis of variance (ANOVA) with subsequent correction for multiple comparisons using honestly significant difference (HSD) contrasts (Tukey, 1949).

EEG Analysis

EEG Data Preprocessing

Data analysis was done by custom Matlab scripts based on the open source EEGLAB toolbox[1] (Delorme and Makeig, 2004) . **Figure 2** shows a flow chart explaining the whole data processing pipeline. The data was filtered using a high-pass filter (1 Hz) and a low-pass filter (120 Hz) and subsequently down sampled to 250 Hz. Single channels and time periods containing artifacts were removed by visual inspection of the data. Eye movements were not considered as artifacts. Artifact rejection was performed with an EEGLAB function automatically removing channels in case they contained zero activity for more than 5 s or revealed a correlation value below 0.6 with neighboring channels and time windows containing more than 30% noisy channels. On average, 132 EEG channels remained for further analyses (range: 114–142; σ = 8.1).

In a next step the data was re-referenced to an average reference and then parsed into maximally temporally independent and spatially fixed components (ICs; Makeig et al., 1996) using an adaptive ICA mixture model algorithm (AMICA; Palmer et al., 2006, 2008) which is a

[1] http://www.sccn.ucsd.edu/eeglab

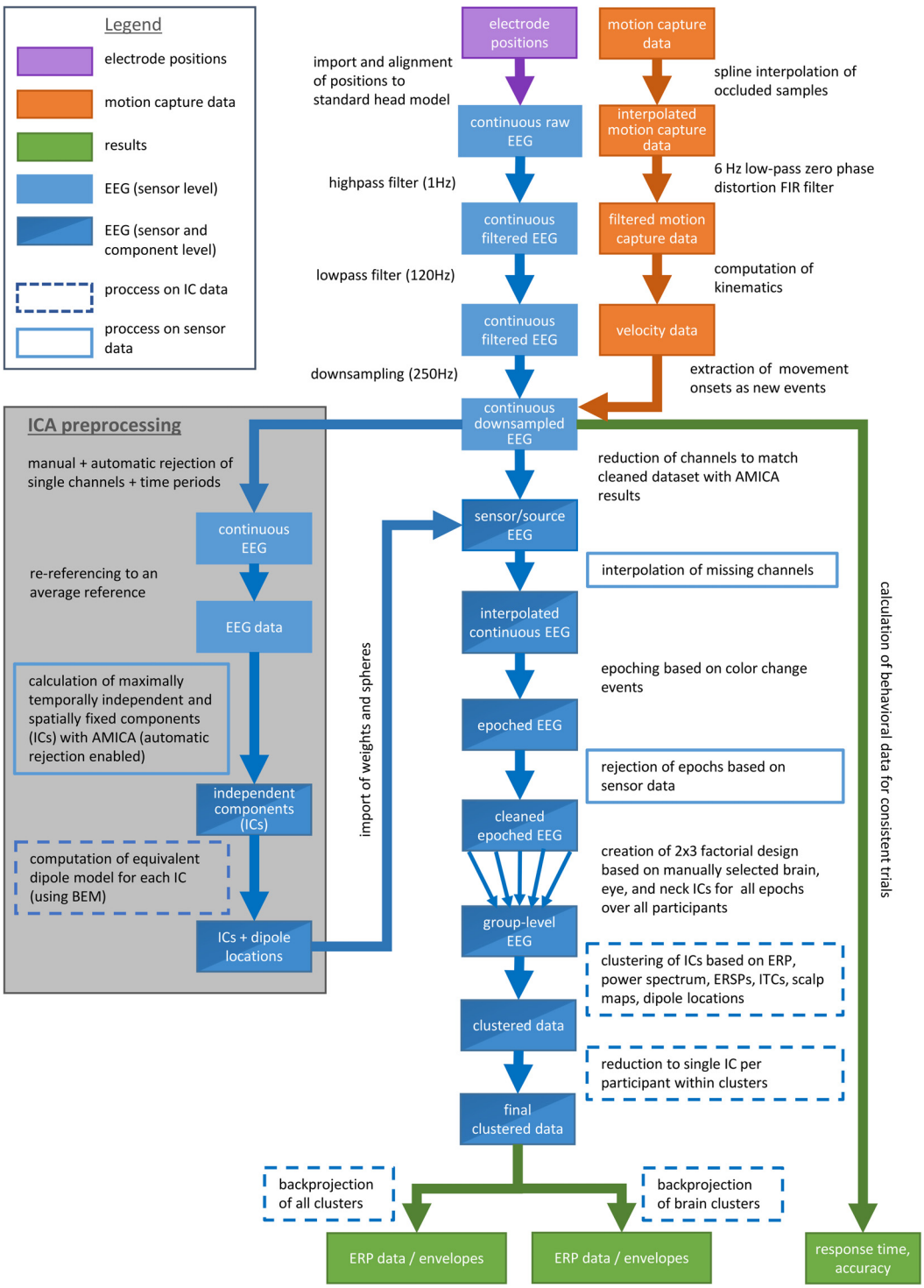


FIGURE 2 | Flow chart explaining the data processing pipeline.

generalization of former ICA approaches as the infomax (Bell and Sejnowski, 1995; Lee et al., 1999a) and the multiple mixture approach (Lee et al., 1999b; Lewicki and Sejnowski, 2000). After the first iteration the model was trained for 10 iterations rejecting time windows with a likelihood below 4 standard deviations (SDs). For the remaining parameters the default settings were used (Palmer, 2016).

For each IC an equivalent dipole model was computed using a boundary element head model (BEM) based on the MNI brain (Montreal Neurological Institute, MNI, Montreal, QC, Canada) as implemented by DIPFIT routines (Oostenveld and Oostendorp, 2002). To this end corresponding landmarks (nasion, ion, vertex and ears) were aligned by rotating and rescaling each individually measured electrode montage. The use of an average head model decreases the accuracy of source localization and thus we refer to the approximation of the spatial origin of surface activity using the description "in or near" a specific structure. ICs primarily accounting for brain, eye or neck muscle activity were selected for further analysis based on their time courses, spectra, and scalp topographies as well as the location and residual variance of their corresponding dipoles. Dipoles placed outside of the head model were not further considered. This resulted in 594 remaining ICs for all participants with an average of 49.5 ICs per subject (range: 31–92, σ = 17.3, $\sum$ = 594). The weights and spheres returned from the AMICA decomposition were copied to the down sampled, high- and low-pass filtered continuous EEG data excluding the same channels that were excluded for ICA decomposition. Missing channels were interpolated.

EEG Group Level Analyses

The continuous data was epoched into 3 s long epochs with onset of a color change including a 1 s pre-stimulus baseline. Only epochs with correct responses were included in the study. Artifactual epochs containing fluctuations above 1000 μV or data values outside of 5 SDs on the sensor level were rejected in an iterative fashion keeping at least 95% of the total trial numbers per iteration. The remaining epochs ($\bar{x}$ = 370.5 per participant, σ = 53.1) were subsequently combined into a study. The study comprised a 2 (response condition) × 3 (stimulus type) factorial design providing main effects for the two independent variables as well as their interaction.

Distances between all ICs were calculated with the weighted measures of ERP, power spectrum (for a frequency range of 3–75 Hz), event-related spectral perturbations (ERSPs), inter-trial coherences (ITCs), the components' scalp maps and their equivalent dipole model locations using the EEGLAB preclustering function. For all measures (except dipole location with only three dimensions) a principal component analysis (PCA) reduced the dimensionality to the first 10 principle components. The resulting measures were normalized, weighted and combined into cluster position vectors. Dipole locations were weighted by a factor of 25 to promote spatially tight clusters and to compensate for its low dimensionality. ERSPs were weighted with a factor of 10 as they were assumed to express the most relevant time-varying information regarding the task. All other measures were weighted with the standard weighting of 1. Subsequently a PCA restricted the resulting cluster position vectors to a 10-dimensional subspace.

Clustering was done via a K-means algorithm implemented in EEGLAB with the number of clusters set to 36. By default, ICs with a distance of more than 3 SDs to the mean of any cluster centroid in joint measure space were assigned to an outlier cluster. The same was done manually for ICs if a cluster contained more than one IC of a participant relying on the same measures as for the calculation of the cluster position vectors. The residual variance of the equivalent dipole models of the remaining ICs was about 10.5% for all ICs representing brain processes (range: 1.3–47.7%, σ = 7.4%) and about 23.7% for all other ICs (range: 2.8–69.1%, σ = 14.3%). Overall, 302 ICs were assigned to the outlier cluster and 292 ICs were assigned to the other clusters (range: 20–30, $\bar{x}$ = 24.3, σ = 2.6 ICs per participant). Of those 292 ICs, 106 ICs revealed equivalent dipole locations within the gray matter of the head model (range: 7–11, $\bar{x}$ = 8.8, σ = 1.4 ICs per participant).

RESULTS

Behavioral Data

An exemplary velocity profile for one physical pointing response with corresponding events derived from the velocity profile and the system generated markers is displayed in **Figure 3** illustrating a typical pointing movement. In most cases, movements towards the screen were faster than the subsequent backward movements to the initial position.

Response times were significantly faster in the physical pointing condition ($\bar{x}$ = 383.1 ms, σ = 40.7 ms) than in the button press condition ($\bar{x}$ = 515.8 ms, σ = 52.9 ms) when response onsets in the physical pointing condition were defined as starting at 5% of the subsequent peak velocity ($p < 0.001$). The means for each condition and participant are shown in **Figure 4**. Significant differences in response onsets between the

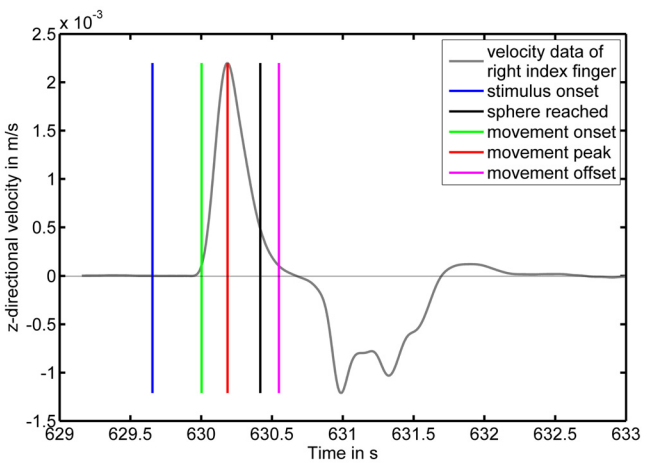


FIGURE 3 | Pointing movement velocity profile as a function of time with corresponding markers. The *y*-axis displays the *z*-component of the velocity in m/s with positive values corresponding to motion towards the screen. The blue vertical line indicates a color change of the moving sphere to the target color. The green and magenta vertical lines indicate the movement onset and offset, respectively. The red vertical line indicates the velocity peak. The black vertical line indicates a distance between LED and projection screen below 10 cm.

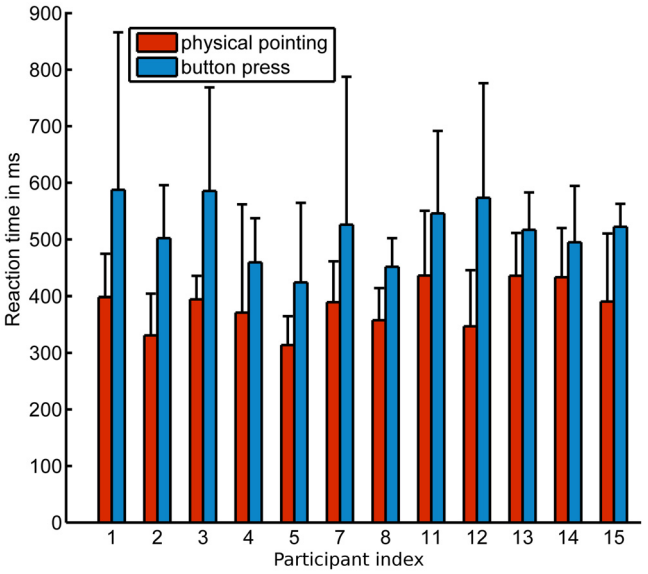


FIGURE 4 | Mean onset response times (threshold criterion 5% of subsequent max. velocity) for all participants in the physical pointing and button press condition. *x*-axis displays the participant index, *y*-axis the response time in ms. Error bars show standard deviation.

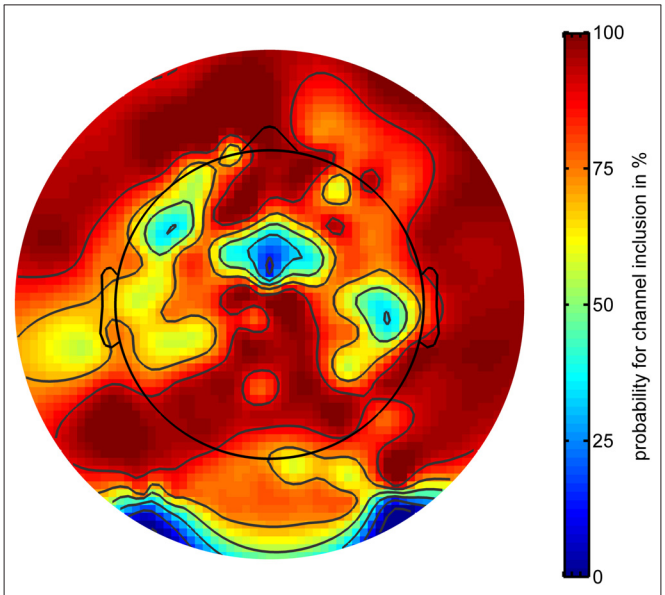


FIGURE 5 | Probability of electrodes to be included in subsequent analyses as a function of electrode location. Warm (red) colors indicate higher probabilities.

two response conditions were observed up to a threshold of 53% of the subsequent peak velocity ($p < 0.05$; $\bar{x} = 474.3$ ms, $\sigma = 41.5$ ms).

Response accuracies were very high with an average of only 0.24% and 7.99% false alarms to color changes indicating a standard stimulus in the button press and physical pointing condition, respectively. Incorrect responses to distractors revealed comparable tendencies with 1.13% and 7.47% false alarms for button presses and physical pointing responses, respectively. In cases of color changes indicating a target stimulus only 0.20% misses were observed for the button press condition and no incorrect responses at all (0%) in the physical pointing condition. While for both standard and distractor stimuli more incorrect responses were observed in the physical pointing condition, target stimuli were associated with less incorrect responses when participants had to point at the moving object. However, only 3 out of 12 participants committed errors in the physical pointing condition while eight participants committed errors in the button press condition. Due to the absence of incorrect responses in the majority of participants no further statistical analyses was conducted. **Table 1** displays mean and standard deviations of response errors in all conditions.

TABLE 1 | Means and standard deviations of response errors for all conditions.

	Physical pointing	Button press
Standard	$\bar{x} = 7.99\%$, $\sigma = 0.2155$	$\bar{x} = 0.24\%$, $\sigma = 0.0036$
Distractor	$\bar{x} = 7.47\%$, $\sigma = 0.2407$	$\bar{x} = 1.13\%$, $\sigma = 0.0175$
Target	$\bar{x} = 0.00\%$, $\sigma = 0.0000$	$\bar{x} = 0.20\%$, $\sigma = 0.0067$

EEG Data

Rapid volatile pointing movements were associated with increasing artifactual activity stemming from both physiological and mechanical sources. To correct for artifactual activity, the EEG signal was cleaned in the time and channel domain (see "Materials and Methods" Section). Cleaning in the channel domain revealed a specific topography for channels with a high probability to be removed. **Figure 5** displays the probability for each channel to be included in the analysis plotted with respect to its scalp position.

Channels were most likely to be removed in five different regions of the montage with the highest likelihood of removal for channels located to the left and right posterio-inferior locations in the montage. One position over the midline located near Cz and two lateralized areas around FT7 and TP8 also showed a high likelihood of channel removal. On average, a subset of 24 channels were removed from the montage before further data analyses (range: 14–42).

Event-Related Potentials on the Sensor Level

Changes in the color of the moving sphere were associated with ERPs including a late positive complex at parietal sensors in the time range of the P3. **Figure 6** displays ERPs with onset of color changes indicating standard, distractor, and target stimuli for the button press and the physical pointing condition for the electrode closest to the parieto-central electrode of the international 10–20-system (referred to as Pz' in the following). To investigate differences in the P3 component measured at the scalp, mean amplitudes in the time range from 400 to 800 ms after a color change were submitted to a 2 (response condition) × 3 (stimulus type) repeated measures ANOVA.

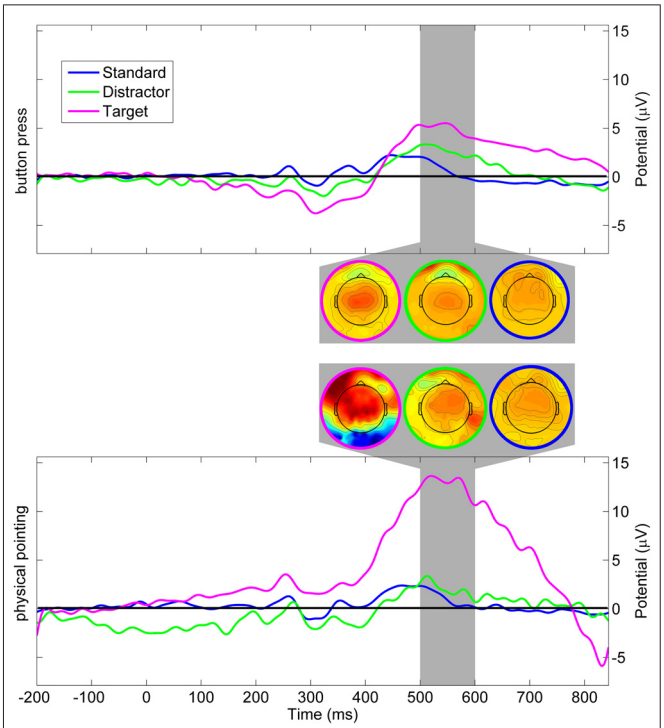


FIGURE 6 | Grand average event-related potential (ERP) at Pz'. Upper panel displays scalp potentials in the button press condition, lower panel displays scalp potentials in the physical pointing condition. Blue: standard stimuli, green: distractor stimuli, magenta: target stimuli. The dark gray area displays the latency range of the mean topographic EEG maps for the different stimuli and response conditions.

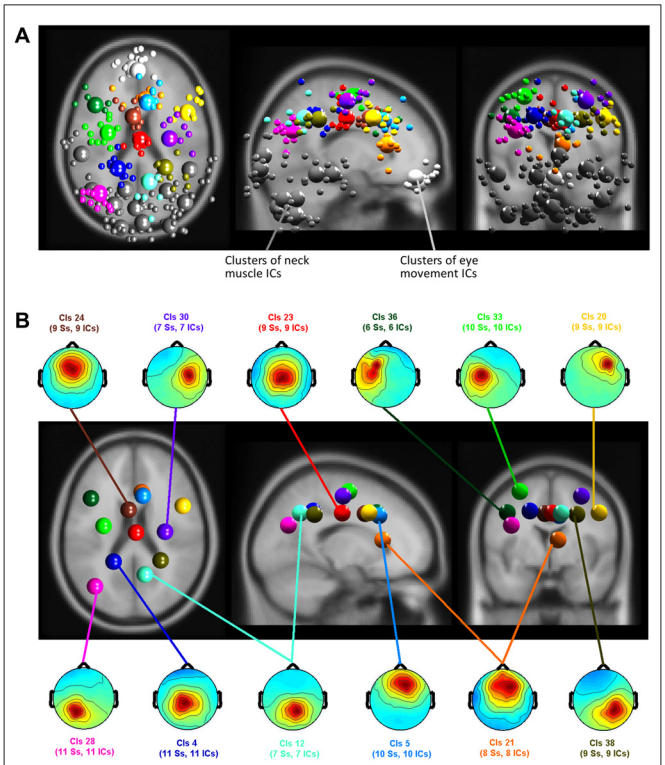


FIGURE 7 | (A) Equivalent-dipole locations of neck muscle (dark gray), eye movement (bright gray) and brain-based independent components (ICs) and their cluster centroids (large spheres, corresponding color) projected to the horizontal, sagittal, and coronal views of the standard Montreal Neurological Institute (MNI) brain. **(B)** Mean projections to the scalp of brain-based IC cluster centroids with index (Cls #), number of participants (# Ss), and number of ICs (# ICs) for each cluster.

Greenhouse-Geisser corrected p-values are reported in case of non-sphericity. The results revealed a significant main effect of stimulus type ($F_{(2,22)} = 8.58$, $p = 0.010$, $\eta^2 = 0.343$) and a tendency for the response condition ($F_{(1,11)} = 3.32$, $p = 0.084$, $\eta^2 = 0.247$) but no interaction effect ($F_{(2,22)} = 2.99$, $p = 0.208$, $\eta^2 = 0.133$). *Post hoc* HSD contrasts (Tukey, 1949) revealed that the P3 amplitude for targets in the pointing condition was significantly higher than for standards in both response conditions (all $ps < 0.009$) as well as distractors in the button press condition ($p = 0.02$). Comparing P3 amplitudes for targets and distractors in the pointing condition revealed only a trend towards significance ($p = 0.09$) and there was no significant difference between targets in the physical pointing and the button press condition ($p = 0.14$). There were no significant differences between any of the stimuli in the button press condition (all $ps > 0.70$).

While both response conditions were associated with increased P3 amplitudes for targets as compared to standard stimuli and distractors, the physical pointing condition demonstrated stronger amplitude increases in the time range of the P3 as compared to the button press condition. The stronger effect in the pointing condition could have been caused by increased processing demands or a generally higher alertness in a condition that required fast responses to a dynamically moving target. However, because the P3 component was located in a time window that also comprised participants pointing responses, increased P3 amplitudes might have been confounded with non-brain related processes. The rather strong jerks of the rapid pointing movements could have added mechanical artifacts induced by the movement. In addition, physical pointing at a moving target required constant coordination of eye, head, and arm movements that, due to volume conduction of the corneo-retinal potential and neck muscle activity, likely contributed to the P3 component at the sensor level. To further investigate to what extent signals from brain and non-brain sources like eye movements or muscle activity contributed to the sensor signal the correspondent independent component processes were analyzed.

Contributions of Brain, Neck Muscle and Eye Movement Activity Related ICs

Clustering of ICs resulted in 26 clusters with cluster centroids located to the gray matter of the brain model or in regions of the model indicating eye movement or neck muscle activity. **Figure 7** displays clusters of IC processes (smaller spheres) and their respective cluster centroids (larger spheres) reflecting brain dynamics, eye movement and muscle activity.

TABLE 2 | Variance in μV^2 in the −200 to 1000 ms time range for the total data and separately for clusters of eye movement, neck muscle, and brain activity.

		Total variance	Eye variance (pvaf)	Neck variance (pvaf)	Brain variance (pvaf)
Standard	Physical pointing	2.39	1.58 (82.4%)	0.03 (3.6%)	0.17 (7.4%)
	Button press	3.45	2.12 (87.9%)	0.01 (−0.2%)	0.07 (3.8%)
Distractor	Physical pointing	4.46	1.26 (38.2%)	0.56 (12.0%)	0.53 (13.8%)
	Button press	3.04	2.71 (83.3%)	0.04 (−2.7%)	0.14 (8.2%)
Target	Physical pointing	19.32	1.98 (15.3%)	6.09 (56.6%)	1.33 (10.6%)
	Button press	3.17	1.62 (63.9%)	0.05 (3.6%)	0.43 (34.3%)

Brackets show the corresponding pvaf in %. Columns are displaying values separately for the cluster combinations and rows display values to standard, distractor and target stimuli in the physical pointing and button press condition.

Back projection of event-related activity originating from different clusters to the sensors allowed for quantifying the contribution of brain and non-brain sources to the sensor P3 component. ERPs of clusters with a centroid located to the gray matter of the brain as well as clusters representing eye movement and neck muscle activity were selected for back projection. The absolute variance and the percent residual variance accounted for (pvaf) with respect to the P3 envelope was computed for all clusters for the time interval between 200 ms before stimulus onset to 1000 ms post stimulus. The pvaf of a specific cluster is defined as $1 - R$ where, R is the quotient of the absolute variance of the remaining clusters (after excluding the considered one) and the absolute variance of all clusters. The pvaf has an upper bound of 100% but can be negative if its projection to the scalp electrode cancels the projected signal of another cluster. This can happen in case ICs are spatially non-orthogonal. Pvaf values were used to estimate the relative share of certain clusters within one condition. Absolute variances, in contrast, allowed for comparing the contributions of one or more clusters to the sensor level in different response conditions where relative values could be misleading due to differences in overall absolute activity.

Table 2 shows the resulting absolute variances and pvafs. Here, the total variance refers to all 36 clusters resulting from the clustering, while neck variance refers to 12 clusters indicating neck muscle activity, eye variance refers to two clusters contributing to horizontal and vertical eye movements, and brain variance to 12 clusters located to the gray matter of the brain. **Figure 8** displays in gray the back-projected summed sensor signal envelope based on all brain, eye, and neck muscle clusters and in red from left to right the contribution of clusters accounting for eye movements, neck muscle activity, and brain activity, respectively.

Relative Contribution of Clusters to the Envelope

Button Press Condition

The relative contributions to the ERP envelope for standard stimuli in the button press condition was high for eye movement activity, only marginal for neck muscle activity, and low for brain processes (eye: 87.9%, neck: −0.2%, brain: 3.8%). Decreasing contribution of eye movement activity and increasing contributions of brain processes was observed for distractor stimuli (eye: 83.3%, neck: −2.7%, brain: 8.2%) and target stimuli (eye: 63.9%, neck: 3.6%, brain: 34.3%).

Physical Pointing Condition

The contributions to the envelope of the ERP for standard stimuli in the pointing condition (eye: 82.4%, neck: 3.6%, brain: 7.4%) were similar to those in the button press condition with slightly stronger contributions of neck muscle activity and brain processes. This trend grew stronger for distractor stimuli (eye: 38.2%, neck: 12.0%, brain: 13.8%) with a pronounced drop in eye movement contribution. For targets neck muscle activity exceeded all other processes considerably (eye: 15.3%, neck: 56.6%, brain: 10.6%).

Absolute Contribution of Clusters to the Envelope

Button Press Condition

In the button press condition the absolute variance of all non-brain and brain processes was relatively stable for standard (3.45 μV^2), distractor (3.04 μV^2) and target stimuli (3.17 μV^2). The absolute contribution of clusters representing eye movements revealed 2.12 μV^2 for standards, 2.71 μV^2 for distractors, and 1.62 μV^2 absolute variance for targets. For clusters with the equivalent dipole model of the cluster centroid located in or near regions of the head model indicative of neck muscles the absolute variance increased from standard (0.01 μV^2) to distractor (0.04 μV^2) and target stimuli (0.05 μV^2). The same trend was observed for clusters representing brain activity contributing 0.07, 0.14, and 0.43 μV^2 absolute variance for standard, distractor, and target stimuli, respectively.

Physical Pointing Condition

In the physical pointing condition the absolute variance of all non-brain and brain processes strongly increased from standard (2.39 μV^2) and distractor (4.46 μV^2) to target stimuli (19.32 μV^2). The absolute contribution of clusters representing eye movements revealed lower values compared to the button press condition explaining 1.58 μV^2 for standards, 1.26 μV^2 for distractors, and 1.98 μV^2, for targets. The absolute variance for clusters representing neck muscle activity increased from standard (0.03 μV^2) to distractor (0.56 μV^2) and target stimuli (6.09 μV^2). A comparable pattern was observed for brain activity with

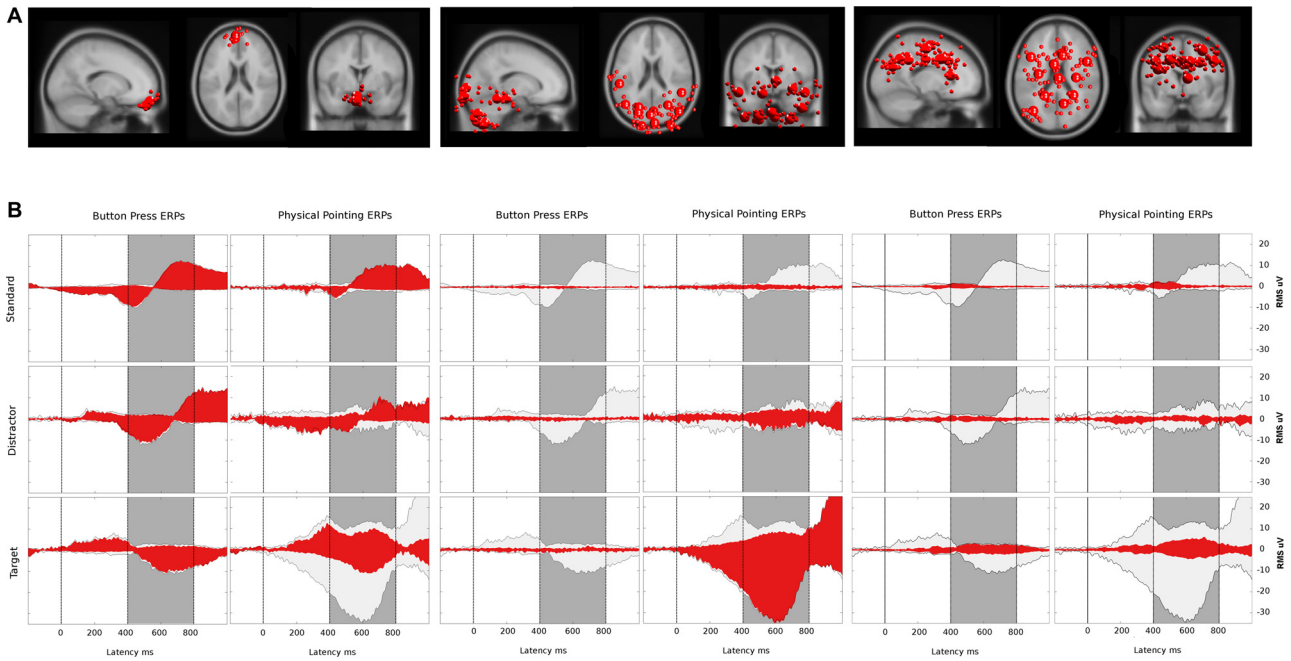


FIGURE 8 | (A) Bigger spheres represent cluster centroids and smaller spheres display individual ICs representing horizontal and vertical eye movement activity (left), neck muscle activity (middle) and brain activity (right). The two clusters representing eye activity in the left column consisted of 12 ICs and 11 ICs, respectively. Neck muscle activity in the middle column is represented by 12 clusters comprising an average of 7.8 ICs (range: 4–10, $\sigma = 1.5$). Brain activity is represented in the right column by 12 clusters comprising on average 8.8 ICs (range: 6–11, $\sigma = 1.5$) from 12 participants. Cluster locations are projected onto the standard MNI brain volume and displayed in sagittal, horizontal, and coronal views. **(B)** Red: ERP contributions of clusters representing eye movement (left), neck muscle (middle), and brain (right) activity. Light gray: ERP envelopes of all 36 back-projected clusters. The dark gray area displays the latency range of the P3 component from 400–800 ms after a color change. The left and right columns display envelopes for the button press and the physical pointing condition, respectively, with rows displaying from top to bottom the different stimuli (standard, distractor and target).

the lowest absolute variance for standard (0.17 μV^2) and distractor stimuli (0.53 μV^2), followed by target stimuli (1.33 μV^2).

In summary, the absolute variance and the increase in absolute variance for clusters representing brain and neck muscle activity were more pronounced in the physical pointing condition, with clusters representing neck muscle activity explaining by far the highest amount of the sensor envelope for target stimuli. In contrast, eye movement contributions were lower for standard and distractor stimuli in the physical pointing condition.

Compared to neck muscle and eye movement activity, the contribution of brain processes to the surface potential was relatively small in both response conditions demonstrating a prominent role of non-brain sources for sensor based ERP analyses during active movements of the head and upper torso.

To further investigate the brain dynamics accompanying target processing in the physical pointing as compared to the button press condition, all non-brain clusters were excluded and only brain-related activity was back projected to the sensor level.

Relative Contributions of Brain Activity to the Sensor Event-Related Potential

Examining the grand average ERP from back-projecting all clusters representing brain activity revealed which clusters contributed most to the sensor level variance in the time window of the P3 component of the ERP. **Table 3** displays the explained absolute and relative (pvaf) variance for the parietal and anterior cingulate cortex (ACC) clusters for each condition in the 400–800 ms time window. For pvafs and absolute variances of all brain clusters, see Supplementary Table 1.

The absolute variance of the sensor ERP explained by brain processes increased from standard (0.14 μV^2) to distractor (0.25 μV^2), and target stimuli (0.87 μV^2) in the button press

TABLE 3 | Variance in μV^2 in the 400–800 ms time range for all clusters contributing to brain activity and separately for the parietal and ACC clusters.

		Total brain variance	Parietal variance (pvaf)	ACC variance (pvaf)
Standard	Physical pointing	0.32	0.004 (6.2%)	0.098 (68.6%)
	Button press	0.14	0.004 (3.6%)	0.055 (72.5%)
Distractor	Physical pointing	0.43	0.020 (11.7%)	0.094 (39.5%)
	Button press	0.25	0.010 (20.1%)	0.076 (58.8%)
Target	Physical pointing	3.11	0.954 (55.4%)	0.486 (0.3%)
	Button press	0.87	0.130 (38.2%)	0.084 (1.5%)

Brackets show the corresponding pvaf in %. Columns are displaying values separately for the cluster combinations and rows display values to standard, distractor and target stimuli in the physical pointing and button press condition.

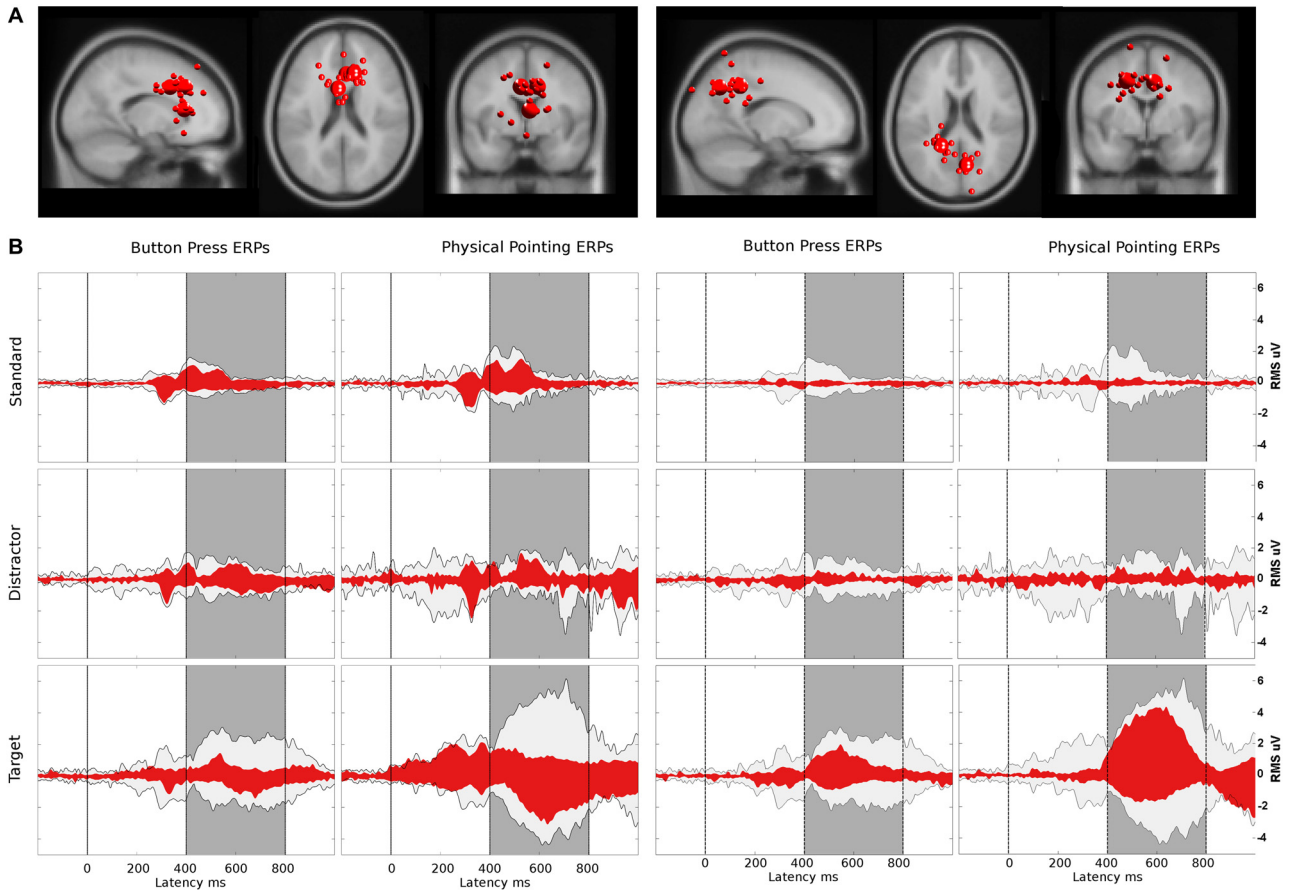


FIGURE 9 | (A) Bigger spheres represent cluster centroids and smaller spheres individual ICs with the cluster centroid located in or near the anterior cingulate cortex (ACC; left) and the dorsal parietal cortex (right). Cluster locations are projected onto the standard MNI brain volume and displayed in sagittal, horizontal, and coronal views. One cluster, located in the left ventral ACC (talairach coordinates: $x = -5$, $y = 9$, $z = 34$, corresponding to BA 24/BA 32) consisted of nine ICs. A second cluster located near the right ventral ACC ($x = 3$, $y = 24$, $z = 6$, near to BA 24) comprised eight ICs and a third located near to the right dorsal ACC ($x = 12$, $y = 26$, $z = 31$, corresponding to BA 9/BA 32) comprised 10 ICs from 12 participants. For the clusters near the parietal cortex, one was located in the left parietal cortex (talairach coordinates: $x = -19$, $y = -42$, $z = 39$, corresponding to BA 31) and consisted of 11 ICs. A second cluster located in the posterior parietal cortex ($x = 11$, $y = -66$, $z = 38$, corresponding to BA 7) comprised seven ICs from 12 participants. **(B)** Red: ERP contributions of the clusters located in or near the ACC and the parietal cortex, respectively; light gray: ERP envelope computed by back-projecting all clusters located in the gray matter of the brain model. The dark gray area displays the latency range of the P3 component from 400–800 ms after a color change which was used for calculating corresponding pvafs. The left and right columns display envelopes for the button press and the physical pointing condition, respectively, with rows displaying from top to bottom the different stimuli (standard, distractor and target).

condition. The same trend was observed for the physical pointing condition with lowest absolute variance for standards (0.32 μV^2) and distractor stimuli (0.43 μV^2), followed by target stimuli (3.11 μV^2). The amount of variance explained and the increase in explained variance was stronger in the physical pointing condition.

The general pattern observed for the contribution of all brain clusters was also observed for the backprojection of a subset of clusters with their centroids located in or near the anterior cingulate and parietal cortex. Three clusters (Cls 5, 21, and 24) representing brain activity in or near the ACC explained lower absolute variance for standard (0.098 μV^2) and distractor stimuli (0.094 μV^2) than for target stimuli (0.486 μV^2) in the physical pointing condition. A different contribution was observed in the button press condition with increasing absolute variance for standards (0.055 μV^2) to distractors (0.076 μV^2), and targets (0.084 μV^2). Because of the general increase in absolute variance in the target condition the relative contribution of the ACC clusters was considerably more pronounced for standard and distractor stimuli than for the target related P3 (see **Figure 9**). The relative contribution of the ACC clusters for standard stimuli was 68.6% and 72.5%, for distractor stimuli 39.5% and 58.8% and for target stimuli 0.3% and 1.5% in the physical pointing and the button press condition, respectively.

Parietal clusters explained increasing variance with the lowest contribution for standard stimuli (0.004 μV^2), followed by distractor (0.010 μV^2) and target stimuli (0.130 μV^2) in the button press condition. This increase from standard to target was also observed for the physical pointing condition with 0.004,

0.020, and 0.954 μV^2 for standard, distractor and target stimuli, respectively. With 55.4% in the physical pointing condition and 38.2% in the button press condition the two parietal clusters contributed the most to the P3 signal for target stimuli. The right panel of **Figure 9** displays two clusters located in or near the parietal lobe and their summed backprojected ERP activity relative to the envelope of all ICs representing brain activity.

Beyond the contribution of the described clusters located in or near the ACC and parietal lobe, other clusters also contributed to the sensor envelope for target stimuli in the P3 time range in the physical pointing condition. These clusters were located in or near the junction of the left parietal and occipital cortex ($x = -40$, $y = -73$, $z = 27$ corresponding to BA 39/BA 19) explaining 38.3%, the right motor and premotor cortex ($x = 40$, $y = -6$, $z = 54$, corresponding to BA 6/BA 4) explaining 14.8%, and the left dorsolateral prefrontal cortex ($x = -43$, $y = 22$, $z = 31$, corresponding to BA 9) explaining 9.8% of variance of the sensor envelope (see Supplementary Table 1 for additional cluster contributions in the button press condition).

The Contribution of Brain Activity to the P3 at Pz'

To analyze the brain dynamic contribution to the maximum of the P3 at the central parietal electrode, only ICs with their equivalent dipole model located to the gray matter of the brain were back projected to Pz' (see **Figure 10**). The resulting summed activity was analyzed with respect to the response condition and stimulus type. To this end mean ERP amplitudes at Pz' were calculated for a time window ranging from 400 to 800 ms after a color change of the sphere and tested for statistical differences using a 2 × 3 repeated measures ANOVA with the factors response condition (physical pointing vs. button press) and stimulus type (standard, distractor, target). Greenhouse–Geisser correction was performed in cases where the assumption of sphericity was violated.

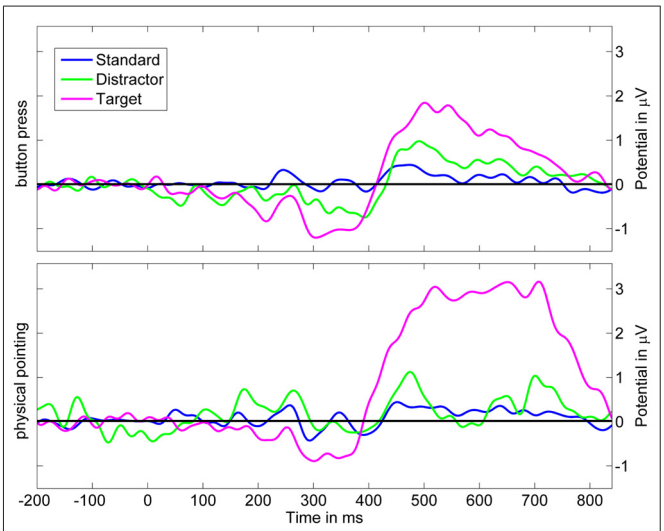


FIGURE 10 | Grand average ERP at Pz' based on backprojection of clusters representing brain activity. Upper row displays scalp potentials for the button press condition, the lower row for the physical pointing condition. Blue: standard stimuli, green: distractor stimuli, magenta: target stimuli.

The analysis revealed a significant main effect of the response condition ($F_{(1,11)} = 11.70$; $p = 0.006$; $\eta^2 = 0.515$) and stimulus type ($F_{(2,22)} = 16.04$; $p = 0.001$; $\eta^2 = 0.593$). The interaction of both factors was also significant ($F_{(2,22)} = 12.47$; $p = 0.003$; $\eta^2 = 0.531$). *Post hoc* HSD contrasts revealed that P3 amplitudes were significantly higher for targets in the pointing condition as compared to standards and distractors in the pointing condition (all $ps < 0.001$) as well as for standards, targets, and distractors in the button press condition (all $ps < 0.001$). In the button press condition the P3 amplitude was significantly higher for targets than for standard stimuli ($p < 0.02$) but did not differ from distractor stimuli ($p > 0.19$).

DISCUSSION

In the present study, a visual oddball paradigm was used to investigate the feasibility of MoBI during volatile rapid movements. The systematic manipulation of response requirements to color changes of a dynamically moving object allowed for a direct comparison of ERPs during simple button presses and active physical pointing. Whereas earlier studies demonstrated that treadmill walking introduces comparatively more eye movements than neck muscle activity (Gramann et al., 2010a) the impact of neck muscle activity was much stronger in the present study with non-stereotyped pointing movements accompanying a wide range of different velocities and movement directions. To react properly in the physical pointing condition participants were requested to move fast and accurately requiring continuous tracking of the stimulus accompanied by eye and head movements and, whenever a target appeared, rapid arm and head movements integrating visual information from the dynamically moving object to intercept the target. This mimicked the fundamental difference between traditional imaging approaches using simple button responses and the MoBI approach allowing for natural interaction with the environment.

The present study revealed important new insights into the brain dynamics accompanying physical interaction with a moving object. Firstly, the study clearly demonstrated that MoBI is feasible for recording and analyzing embodied cognitive processes and the accompanying brain/body dynamics during volatile rapid movements in a realistic 3-D environment. Secondly, applying blind source separation methods to the EEG signals recorded during the visual oddball paradigm allowed for separating and clustering ICs corresponding to neck muscle activity, eye movements or brain processes. This way it was possible to analyze the contribution of different clusters to the scalp signal revealing strong activity of neck muscles during the physical pointing response resulting from head orientation changes and compensation of shoulder and arm movements during pointing. Thirdly, movement onsets and corresponding reaction times in the physical pointing condition demonstrated significantly faster response onsets as compared to the button press condition. Fourthly, analysis of the data in the time range

of the P3 component revealed a clear P3 in both response conditions at the sensor level as well as the level of cluster activity. This manifested in significantly higher mean ERP amplitudes for target stimuli as compared to standard stimuli as well as increasing absolute variance for standard, to distractor, and target stimuli in both response conditions. Finally, back-projecting all brain-related clusters to the centro-parietal sensor showed significantly higher P3 amplitudes for target stimuli in the physical pointing condition compared to the button press condition. This finding indicates different brain dynamics for different behavioral states and has far-reaching implications in the field of Neuroergonomics.

Natural Cognition and the Contribution of Brain and Non-Brain Sources

During physical interaction with a dynamically moving object, non-brain sources stemming mainly from eye movements and neck muscle activity as well as mechanical artifacts strongly diminished the observable fraction of brain activity recorded on the scalp and avoided meaningful analysis of sensor-based potentials without further preprocessing. However removing all ICs not associated with brain activity allowed for analyzing the P3 component and the contribution of different clusters to its time course revealing the following findings.

Clusters contributing to the sensor P3 component were mostly in line with the results of previous studies. As in Makeig et al. (2004), central parietal, motor and occipital processes contributed to the P3 with the largest contribution of parietal clusters to the onset of target stimuli. The contribution of brain processes located near or in the ACC was in line with previous findings using an oddball paradigm during treadmill walking (Gramann et al., 2010a).

The explained variance of brain related sources increasing from standard, to distractor, and target stimuli were found in both response conditions, with a stronger effect in the physical pointing condition. This is consistent with the assumption that a potential physical interaction with the environment requires additional cognitive and motor processes and thus leads to higher computational effort. In the present study it was necessary to track the position and movement direction of the relevant object and body parts required for orienting to and interacting with the stimulus. Physical interaction with target stimuli required action planning, execution, and control. Those processes were not required for frequent standard and rare distractor stimuli reflected in smaller amplitudes and lower variance in clusters reflecting brain processes. However, distractors attracted more attention and potentially triggered an initial response. This response had to be suppressed resulting in additional inhibitory processes and accompanying brain activity as indicated by higher variance for distractor stimuli than for standard stimuli.

Clusters representing neck muscle activity also accounted for increasing variance from standard, to distractor and target stimuli in both conditions. The increase was stronger in the physical pointing condition where a correct response to the target required a pointing movement comprising movement of the head, shoulder and arm. Those movements were accompanied by strong neck muscle activity as observed for target stimuli in the physical pointing condition. Since the readiness to act was very high, as indicated by faster response times and the absence of any misses, it is likely that rare distractor stimuli caused the initiation of response movements. Even in case the response was subsequently inhibited for distractor stimuli, response initiation would be reflected in higher neck muscle contribution to distractor than to standard stimuli.

Finally, clusters representing eye movements explained more variance in the sensor signal in the button press as compared to the physical pointing condition for standard and distractor stimuli. One possible explanation is that in the physical pointing condition head alignment to the stimulus position facilitated physical movements in that direction. As a consequence of increasing head movements during stimulus tracking, less eye movements were required for keeping the moving stimulus in the visual field. However, since the sphere kept moving after the color change, successful pointing movements to targets required an ongoing prediction of its future position. This caused extended coordination of eye and arm movement resulting in an increase of variance explained by eye movements. In the button press condition a simple button press was sufficient to respond to a color change requiring no further coordination of eye movement and physical response. In addition, the visual stimulus stopped moving after response execution rendering stimulus tracking unnecessary. This would have resulted in a decrease of corresponding variance for target trials compared to distractor and standard trials in both conditions.

Limits of MoBI

The present study required continuous head and eye movements during stimulus tracking causing electrical potentials on the surface electrodes superposing the EEG signal. This happened especially for target stimuli in the physical pointing condition where a correct response demanded arm movements accompanied by strong jerks associated with increased neck muscle activity. Subsequently no significant mean ERP difference was found on the scalp electrodes in the P3 time range between physical pointing and button press for target stimuli. Volume conducted non-brain activity in the recorded EEG signal is an inevitable consequence of active movements of the participants. Using ICA for separating brain related from non-brain related activity and back projecting the former to the Pz' revealed the expected differences between those conditions. Thus, MoBI proved feasible for analyzing event-related EEG dynamics of participants performing rapid pointing movements in a realistic 3-D environment.

However some caveats were identified in this study indicating potential constraints of the MoBI approach for investigating natural movements. These included an increase of artifact contaminated trials and channels as well as higher residual variances compared to EEG studies with stationary participants that are not allowed to move their heads.

A relatively high number of trials had to be removed due to inconsistencies between markers written online during the experiment and those derived afterwards from

the velocity profiles. Thus the amount of considered data was decreased impeding statistical analysis especially in the physical pointing condition. Future MoBI experiments will need a setup fully covered with fixed cameras minimizing the risk of LED occlusions and camera movement causing such inconsistencies.

Related to the reduced number of trials due to technical problems, the impact of movement-related mechanical artifacts like cable sway was reflected in a specific distribution of the probabilities for channels to be removed. Jerks and micromovements of the electrodes over the skin surface associated with fast response movements led to impedance changes with strong artifactual activity affecting the outmost neck electrodes in the posterio-inferior regions. The central midline area as well as the two lateral regions over the scalp in contrast were most likely affected by cable pull during head movements due to the cable routing over the scalp to the back. This is one likely explanation for the high rejection rate of those electrodes. The lateral regions near the mastoid processes were predestined for bad contact with high impedance leading to artifactual activity due to the fit of the cap. For further MoBI experiments a redesign of electrode attachment or cable-free electrodes has to be considered to increase the number of channels that can be analyzed.

Finally, in contrast to non-MoBI studies the ICA results revealed many ICs with large residual variances that would not be included into the clustering process applying standard selection criteria (e.g., Gramann et al., 2010b). However muscle activity and eye movements originate in regions of the head that are usually not included in the head model for source reconstructions rendering it difficult to calculate suitable dipole models. Moreover, muscle contraction causes tissue movements which could result in dipole displacement. Thus, although in general higher residual variance is associated with decreased result accuracy, dipoles with relatively large residual variances were included in the analysis. A future improvement would be the introduction of forward models including neck muscles and their contraction profiles as additional parameter for the inverse solution.

Implications on Neuroergonomics Research

The faster response onsets in the physical pointing condition might be the consequence of another brain dynamic state caused by the need of physical interaction with the stimulus. In addition, the physical pointing condition might have led to increased motivation and more fun for this response format as reported by the participants after the experiment. However, using movement and velocity profiles for the purpose of additional brain activity analysis requires the definition of corresponding features that are widely used and accepted. Defining the movement onset as a fixed percentage of the corresponding maximum velocity as in this study was only one possible solution. Other criteria might be useful in different contexts like a fixed absolute velocity or acceleration value which would be independent of individual movement differences. A general definition should be discussed and established to increase comparability of experimental results in the field of Neuroergonomics and for MoBI research in general. Importantly, comparing different onset criteria starting from 5% of the corresponding peak velocity up to 53% of the corresponding peak velocity still indicated faster responses in the physical pointing condition as compared to button presses. Whether this was simply due to the fact that participants enjoyed the physical response format or whether interception of a dynamically moving objects was associated with a generally different behavioral and brain dynamic state will have to be investigated in future experiments. There are clear arguments in favor of state differences in brain dynamics depending on the behavioral state. Introducing a task that requires large volatile movements not only produces muscle activity and eye movements but also requires additional processes that allow for movement planning, control and execution. This additional processes will be reflected in changes in brain dynamics. Moreover, the oddball paradigm required constant attention directed towards a sphere that moved within the borders of a large screen in front of participants. In addition, the physical pointing condition necessitated the prediction of the targets' movement to integrate this information with proprioceptive information about position and orientation of the arm and hand for concurrent dynamic motor planning and execution. Continuous observation and integration of environmental aspects with complex motor programs causes higher computational effort and can be assumed to lead to different brain dynamic states compared to passive observation. This is indicated by the significantly increased amplitude of the P3 and faster response onsets in the physical pointing condition compared to the button press condition. It would be important for future investigations to analyze the impact of stimulus speed on the brain dynamic state since higher speeds increase task difficulty and thus affect head and eye movement velocities.

This has significant implications for Neuroergonomics investigating the brain at work, especially in case the working environment requires physical interaction with a dynamic system. The physical pointing task required the participants to actively interact with their environment using fast, precise movements of the upper torso and the arm and hand. This generalizes to a wide range of working tasks where people have to manipulate objects as, for example, in assembly-line work or construction trade. Future studies might investigate the brain dynamics underlying spatially extended movements with different velocities including team sports or spatial orientation with or without navigation assistance. Studying the brain activity in the described work settings could provide valuable insights into the cognitive processes and the limits of the cognitive system and thus allow for suggestions how to increase system safety. For example, the degree of interaction seems to be one factor improving working environments by influencing motivation, reaction time and task complexity. Another factor to be considered is the body posture since active movement is naturally associated with an upright posture whereas cognitive neuroscientists still investigate sitting or lying participants. Changes in brain dynamics due to different body postures could have an impact on result quality and information processing speed.

Investigating the human brain dynamics accompanying physical interaction with dynamically moving objects for the first time, this study clearly demonstrated that it is possible to record and analyze EEG activity during volatile rapid movements. Thus, future MoBI studies examining the mentioned aspects will have an important impact on Neuroergonomics specifically and cognitive neuroscience in general.

AUTHOR CONTRIBUTIONS

KG and EJ have been actively involved in the experimental design. EJ recruited experimental participants, acquired and analyzed the data. KG participated in the data analysis. All authors contributed to the interpretation of the data, revised the manuscript and approved the final version to be published. EJ and KG agree to be accountable for all aspects of the work and ensure that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fnhum.2016.00306/abstract

REFERENCES

Acar, Z. A., and Makeig, S. (2013). Effects of forward model errors on EEG source localization. *Brain Topogr.* 263, 378–396. doi: 10.1007/s10548-012-0274-6

Bell, A. J., and Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Comput.* 7, 1129–1159. doi: 10.1162/neco.1995.7.6.1129

Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21. doi: 10.1016/j.jneumeth.2003.10.009

Gramann, K. (2013). Embodiment of spatial reference frames and individual differences in reference frame proclivity. *Spat. Cogn. Comput.* 13, 1–25. doi: 10.1080/13875868.2011.589038

Gramann, K., Ferris, D. P., Gwin, J., and Makeig, S. (2014). Imaging natural cognition in action. *Int. J. Psychophysiol.* 91, 22–29. doi: 10.1016/j.ijpsycho.2013.09.003

Gramann, K., Gwin, J. T., Bigdely-Shamlo, N., Ferris, D. P., and Makeig, S. (2010a). Visual evoked responses during standing and walking. *Front. Hum. Neurosci.* 4:202. doi: 10.3389/fnhum.2010.00202

Gramann, K., Onton, J., Riccobon, D., Mueller, H. J., Bardins, S., and Makeig, S. (2010b). Human brain dynamics accompanying use of egocentric and allocentric reference frames during navigation. *J. Cogn. Neurosci.* 22, 2836–2849. doi: 10.1162/jocn.2009.21369

Gramann, K., Gwin, J. T., Ferris, D. P., Oie, K., Jung, T. P., Lin, C. T., et al. (2011). Cognition in action: imaging brain/body dynamics in mobile humans. *Rev. Neurosci.* 22, 593–608. doi: 10.1515/RNS.2011.047

Grillon, C., Courchesne, E., Ameli, R., Elmasian, R., and Braff, D. (1990). Effects of rare non-target stimuli on brain electrophysiological activity and performance. *Int. J. Psychophysiol.* 9, 257–267. doi: 10.1016/0167-8760(90)90058-l

Gwin, J. T., Gramann, K., Makeig, S., and Ferris, D. P. (2010). Removal of movement artifact from high-density EEG recorded during walking and running. *J. Neurophysiol.* 103, 3526–3534. doi: 10.1152/jn.00105.2010

Gwin, J. T., Gramann, K., Makeig, S., and Ferris, D. P. (2011). Electrocortical activity is coupled to gait cycle phase during treadmill walking. *Neuroimage* 54, 1289–1296. doi: 10.1016/j.neuroimage.2010.08.066

Kothe, C. (2014). Lab Streaming Layer (LSL). Available online at: https://github.com/sccn/labstreaminglayer. Accessed on October 26, 2015.

Lee, T. W., Girolami, M., and Sejnowski, T. J. (1999a). Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources. *Neural Comput.* 11, 417–441. doi: 10.1162/089976699300016719

Lee, T. W., Lewicki, M. S., Girolami, M., and Sejnowski, T. J. (1999b). Blind source separation of more sources than mixtures using overcomplete representations. *IEEE Signal Process. Lett.* 6, 87–90. doi: 10.1109/97.752062

Lewicki, M., and Sejnowski, T. (2000). Learning overcomplete representations. *Neural Comput.* 12, 337–365. doi: 10.1162/089976600300015826

Makeig, S., Bell, A. J., Jung, T. P., and Sejnowski, T. J. (1996). Independent component analysis of electroencephalographic data. *Adv. Neural Inf. Process. Syst.* 8, 145–151.

Makeig, S., Delorme, A., Westerfield, M., Jung, T. P., Townsend, J., Courchesne, E., et al. (2004). Electroencephalographic brain dynamics following manually responded visual targets. *PLoS Biol.* 2:e176. doi: 10.1371/journal.pbio.0020176

Makeig, S., Gramann, K., Jung, T. P., Sejnowski, T. J., and Poizner, H. (2009). Linking brain, mind and behavior. *Int. J. Psychophysiol.* 73, 95–100. doi: 10.1016/j.ijpsycho.2008.11.008

Mehta, R. K., and Parasuraman, R. (2013). Neuroergonomics: a review of applications to physical and cognitive work. *Front. Hum. Neurosci.* 7:889. doi: 10.3389/fnhum.2013.00889

Mijovic, P., Kovic, V., De Vos, M., Mačužic, I., Jeremic, B., and Gligorijevic, I. (2016). Benefits of instructed responding in manual assembly tasks: an ERP approach. *Front. Hum. Neurosci.* 10:171. doi: 10.3389/fnhum.2016.00171

Ojeda, A. (2011). MoBILAB. Available online at: http://sccn.ucsd.edu/wiki/MoBILAB. Accessed on October 29, 2015

Oostenveld, R., and Oostendorp, T. F. (2002). Validating the boundary element method for forward and inverse EEG computations in the presence of a hole in the skull. *Hum. Brain Mapp.* 17, 179–192. doi: 10.1002/hbm.10061

Palmer, J. A. (2016). AMICA algorithm. Available online at: http://sccn.ucsd.edu/~jason/amica_web.html. Accessed on May 17, 2016.

Palmer, J. A., Kreutz-Delgado, K., and Makeig, S. (2006). "Super-Gaussian mixture source model for ICA," in *Lecture Notes in Computer Science*, J. Rosca, D. Erdogmus, J. C. Príncipe and S. Haykin (Heidelberg: Springer), 854–861.

Palmer, J. A., Makeig, S., Delgado, K. K., and Rao, B. D. (2008). "Newton method for the ICA mixture model," in *2008 IEEE International Conference on Acoustics, Speech and Signal Processing* (Las Vegas, NV), 1805–1808.

Parasuraman, R. (2003). Neuroergonomics: research and practice. *Theor. Issues Ergon. Sci.* 4, 5–20. doi: 10.1080/14639220210199753

Tukey, J. W. (1949). Comparing individual means in the analysis of variance. *Biometrics* 5, 99–114. doi: 10.2307/3001913

Wascher, E., Heppner, H., Kobald, S. O., Arnau, S., Getzmann, S., and Möckel, T. (2016). Age-sensitive effects of enduring work with alternating cognitive and physical load. A study applying mobile EEG in a real life working scenario. *Front. Hum. Neurosci.* 9:711. doi: 10.3389/fnhum.2015.00711

Wilson, M. (2002). Six views of embodied cognition. *Psychon. Bull. Rev.* 9, 625–636. doi: 10.3758/bf03196322

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Jungnickel and Gramann. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution and reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 13 January 2016
doi: 10.3389/fnhum.2015.00711

Age-Sensitive Effects of Enduring Work with Alternating Cognitive and Physical Load. A Study Applying Mobile EEG in a Real Life Working Scenario

Edmund Wascher, Holger Heppner, Sven O. Kobald, Stefan Arnau, Stephan Getzmann and Tina Möckel*

IfADo-Leibniz Research Centre for Working Environment and Human Factors, Dortmund, Germany

Ergonomic assessment of a workplace requires the evaluation of physical as well as cognitive aspects of a particular working situation. In particular the latter is hardly possible without interfering in the natural setting. Mobile acquisition of neurophysiological measures (such as parameters of the EEG) may close this gap. At a simulated workplace we tracked older and younger participants with mobile EEG during a 4–5 h work shift. They had to perform either a monotonous cognitive task, a self-paced cognitive task or a self-paced physical task in a predefined order. Self assessment, behavioral performance and spectral measures of the EEG (before most alpha power) indicated that younger participants suffered from monotony. Older adults, on the other hand, were overall impaired by inefficient information processing. This was visible in EEG variations time-locked to eye blinks (blink-related synchronizations), a new measure to investigate cognitive processing in real life environments. Thus, we were able to distinguish between active and passive task-related aspects of mental fatigue without impinging on the natural working situation.

Keywords: working environment, mental fatigue, mobile EEG, aging

OPEN ACCESS

Edited by:
Klaus Gramann, Berlin Institute of Technology, Germany

Reviewed by:
Juliana Yordanova, Bulgarian Academy of Sciences, Bulgaria
Lutz Jäncke, University of Zurich, Switzerland

***Correspondence:**
Edmund Wascher
wascher@ifado.de

Received: *09 October 2015*
Accepted: *21 December 2015*
Published: *13 January 2016*

Citation:
Wascher E, Heppner H, Kobald SO, Arnau S, Getzmann S and Möckel T (2016) Age-Sensitive Effects of Enduring Work with Alternating Cognitive and Physical Load. A Study Applying Mobile EEG in a Real Life Working Scenario. Front. Hum. Neurosci. 9:711. doi: 10.3389/fnhum.2015.00711

INTRODUCTION

Evaluation of workplaces may take place on quite different levels. Traditional ergonomics focuses on physiological and physical factors of working environment, applying surveys to different aspects of the workplace in order to support and protect the worker (McAtamney and Corlett, 1993; Hignett and McAtamney, 2000; Halim et al., 2012). More recently, virtual models of humans help to evaluate many of these aspects of work (Lämkull et al., 2007; Bandouch et al., 2008). Cognition, however, and human factors affecting cognition, can be addressed only superficially in real life situations. While experimental psychology may address distinct cognitive aspects that play a role in working situations, this can hardly apply for the complex interaction of cognitive requirements that a worker faces during a work shift. Neurophysiological measures that can be taken while regular work is performed may help to close this gap. In particular, the increasing accessibility of mobile neurophysiological technology that allows for an "online" registration of work-related parameters is a huge challenge and chance for ergonomic evaluation (Wascher et al., 2014a).

In the present study, we focused on a neuroergonomic evaluation of the interaction between mental fatigue and age-related changes in cognitive performance. The proportion of employed older adults is continuously increasing with the general demographic change. While older employees may benefit in many cases from professional experience, they are also facing physiological and cognitive decline, which makes it difficult to keep up with their younger colleagues, in particular when long-term working situations are considered. There is some evidence from laboratory experiments that not only physical but also mental fatigue raises faster with higher age (Wascher and Getzmann, 2014). On the other hand, it has been shown that varying the cognitive task may prevent older adults from accelerated raise of mental fatigue (Falkenstein et al., 2002). However, does this finding hold also for situations in which varying tasks require flexible retrieval of cognitive and physical capabilities continuously, as it may be the case during a working shift?

We addressed this question based on a multi-level model of fatigue and its possible relation of age-related cognitive decline by pinning the observed behavior measures down to neurophysiological mechanisms.

Long lasting activity, independently whether it is physical or cognitive, leads to a decline in capabilities of a worker (Halim et al., 2012; Lerman et al., 2012). Despite this communality and the usage of the common term of "fatigue," physical and cognitive declines have quite different underlying mechanisms and consequences. Physical fatigue goes along with metabolic changes in the muscle, which leads a decline in physiological capabilities. Additionally, an increase in the plasma fatty acid level may lead to an increase in tryptophan. This is assumed to increase the 5-HT concentration in the brain and thereby contributing to central fatigue (Newsholme et al., 1992), a mental contribution to physical fatigue. Although central aspects may play an important role for the accessibility of physical resources (Marcora et al., 2009; Mehta and Parasuraman, 2013), the peripheral exhaustion of resources is the core aspect of physical fatigue.

Mental fatigue, on the other hand, is not known to go along with any physiological resource consuming aspect (Hockey, 2013). Apart from circadian rhythms and sleeping behavior (sleep-related fatigue = SR), mental fatigue may derive either from cognitive overload (active task-related fatigue = aTR) or from mental underload (passive task-related fatigue = pTR) due to, e.g., monotony (May and Baldwin, 2009). Those latter two mechanisms end up in distinct types of fatigue that have different consequences for subsequent activity. While mental aTR was found to persist even after completion of the task and not to depend on the amount of motivation (van der Linden et al., 2003), pTR is strongly coupled with the motivational system (Boksem et al., 2006; Boksem and Tops, 2008; Bonnefond et al., 2011). Both in laboratory basic research studies as well as in applied contexts it was demonstrated that pTR can be efficiently effaced by incentives. Even after long lasting mental tasks, the instruction to put more effort into a given task is efficient to restore cognitive performance almost completely to the level that has been shown at the beginning of the experiment (Boksem et al., 2006). Therefore, pTR has been framed in motivational terms. Boksem and Tops (2008), for example, proposed that pTR reflects an imbalance between resources invested and outcome. Whenever the effort that is needed to perform on an adequate level is too high in relation to the outcome that is generated, motivation declines and, as a consequence, cognitive processing becomes less efficient.

Until now, the different aspects of mental fatigue have been investigated in isolation and under controlled laboratory settings. However, one has to be aware that most working activities that require physical and cognitive effort also contain periods in which monotonous cognitive tasks have to be performed. Such changes in duties may prevent at least from pTR, as has been demonstrated in studies that investigated effects of aging on mental fatigue. When tasks changed during the experimental session no mental fatigue was observed in older adults (Falkenstein et al., 2002), whereas performing the same task for a longer period of time led to a clear age-related decline of attentional performance (Wascher and Getzmann, 2014).

Age-related effects on pTR are insofar of central interest in fatigue research as the well-known decline of structures in the frontal lobe of the human brain (Chao and Knight, 1997) affects primarily those structures that are also involved in motivational processes and thereby in upholding cognitive performance in non-demanding situations (Berridge and Robinson, 1998). Beside this, lower muscle strength with higher age may contribute to central fatigue when physical demanding tasks are part of a working situation.

We designed a simulated workplace that resembled the post room of a German wholesale house (see Funding). The tasks of the participants recreated parts of the real workflow but were nevertheless controlled experimental settings. Participants had to perform a monotonous stimulus-response task, a self-paced cognitive task, and a physical task (moving and sorting boxes of different weights and sizes) in a repetitive sequence for about 4–5 h. Behavioral performance was measured in the cognitive tasks and self estimation of experienced fatigue and motivation were repeatedly taken. One of the core questions of the present study was, to what degree a neuroergonomic approach may help to get objective data from task load, effort, and fatigue-related changes in cognitive processing. To this end, beside the "classical" measures of behavioral performance and self estimation, mobile EEG was recorded continuously, while the participants were freely moving in an office-like room, dealing with the different tasks.

Since there is hardly any literature to such experimental setting, we focused firstly on well-known aspects of the EEG that have been related to mental fatigue. This includes before most oscillatory activity. It has been repeatedly shown that brain oscillations are slowing down with mental fatigue, indicated by increasing power in the alpha and the theta band of the EEG (Lal and Craig, 2001; Akerstedt et al., 2004; Wascher et al., 2014b). Given that most of these studies used longer lasting monotonous tasks, these effects may be related before most to pTR. In particular, the increase in alpha activity may well be a correlate of decreasing motivation and a withdrawal of attention that lead to a kind of idle state in the sensory and attention

related structures of the brain (Hanslmayr et al., 2012) that should be most pronounced in the monotonous stimulus-response task. The increase in theta activity may be rather related to increasing effort that is invested to keep performance high (Sarter et al., 2001, 2006) which should be stronger in self-paced tasks. Most interestingly, a kind of slowing has been also reported with increasing age, as indicated by a reduction in the individual alpha frequency (for a review see Klimesch, 1999), but so far not for mental fatigue.

Besides these rather energetical aspects of brain activity, we asked for specific neuronal processes that can be related to information processing, and how they change with age and fatigue. In laboratory settings, stimuli are presented at distinct time points and cortical activity is measured time-locked to these events (so-called event-related potentials or oscillations). Such events are not accessible in real life situations. Adding additional stimuli to a real life situation may substantially alter the task of participants, leading to, e.g., attentional distraction. Distinct events from the surrounding that may be identified by scene cameras would be not comparable across different tasks, and no sufficient number of repetitions of comparable events is guarantied. Events that occur independently of a particular task, repetitively also in real life situations are so-called eye events. Horizontal eye-movements (e.g., saccades) are the core human behavior related to spatial orientation of attention. More interesting for the temporal segmentation of incoming information are eye blinks that occur primarily at the end of an information processing sequence (e.g., Doughty, 2001; Wascher et al., 2015). Recently, we demonstrated that time locking of EEG activity to eye blinks provide reliable measures for cognitive effort (Wascher et al., 2014a). Because event-related potentials did not reliably show time on task related changes in a previous task (see Wascher and Getzmann, 2014), we applied event-related synchronization/desynchronization (ERS/ERD; Pfurtscheller and Aranibar, 1979) analyses to the eye-blink related data. Due to the lower time resolution, these data might be more robust in complex experimental situations like a workplace simulation. Moreover, phasic changes in brain oscillatory activity (in particular in the Alpha band) appear to be reliable correlates of signal processing (Klimesch et al., 2002; Müller et al., 2009).

Taken together, applying these methods to a working situation that resembles a real workplace should provide information about age and fatigue-related changes in information processing and the underlying neuronal mechanisms. The aim of the present study was to go beyond the description of age-related differences in performance.

METHODS

Participants

Thirteen younger adults (20–29 years old, mean age 25.3) and 12 older adults (55–73 years old, mean age 64.4) took part in the experiment. All participants had normal or corrected to normal vision, were of good physical health, and reported no history of psychiatric or neurological diseases. For the entire procedure (lasting around 5–6 h including preparation) participants received 60 €.

Prior to the experiment participants gave written informed consent. The study was approved by the local ethics committee and according to the Declaration of Helsinki.

Task, Stimuli, and Procedure

The experiment took place in an office room (3.50 × 4.80 m) with partly covered windows. Tables stood along the walls where the boxes for the physical task were placed (see **Figure 1**). On one of the tables, a computer monitor was positioned for the presentation of instructions for the physical task and the cognitive tasks. A research assistant who controlled the experiment sat in another room and monitored the EEG recordings. Participants could reach the assistant via phone at all times. The assistant only entered the room during breaks or to correct any technical failures if necessary.

The experiment consisted of three tasks, which were repeated in a predefined sequence within each block. Blocks started and ended with a computerized version of the d2-task, that was closely oriented on the original paper-pencil version (Brickenkamp, 1962). In between, a block of the Simon task was presented, followed by the physical task and again the Simon task had to be performed. This procedure was repeated three times with short breaks in between and added up to an overall duration of the work shift of about 4–4.5 h. Each subtest was defined for a pre-defined duration, thus the self-paced tasks were stopped no matter how many passes were finished.

In the d2-task, three lines of 57 d's or p's each with one to four marks (in the form of single or double quotes) above and/or below the letter were presented. The participants had to

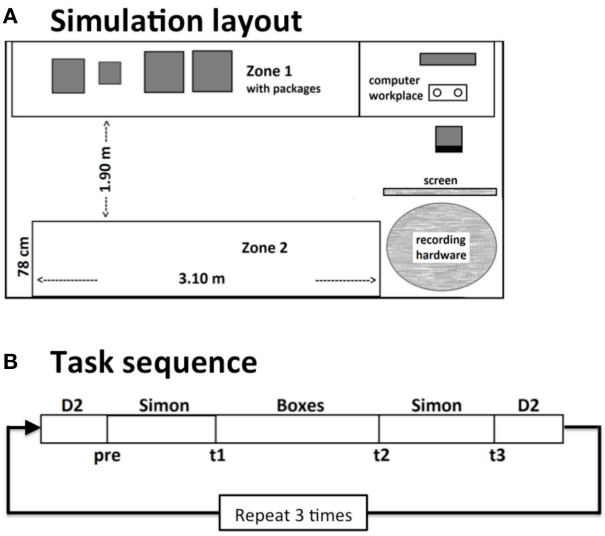


FIGURE 1 | Schematic layout of the workplace simulation (A). In an office-like room, participants either had to perform a monotonous or a self-paced cognitive task at a computer. In between they had to move boxes between two zones. Thereby they had to follow randomly changing sorting instructions. This procedure was repeated 3 times in a 4–4.5 h lasting work shift **(B)**.

mark as much d's with exactly two marks as possible in a given time window using a computer mouse, while simultaneously ignoring all p's and the d's with less or more than two marks. The participants had 20 s per line. Then a sound signaled to proceed with the next line. After the three lines were done, a new screen appeared with three new lines. In each d2-task block, five screens with three lines each were presented. Overall, one d2-task block took about 5 min. The d2-task served as a self-paced cognitive task.

In the Simon task (Simon, 1969) one of two symbols (either a square or a diamond) was presented on the left or the right side of a fixation cross for 150 ms. The participants had to decide which one of the symbols was shown by responding with either the left or the right hand, while ignoring the side on which the stimulus was presented. Thus, a trial could be either corresponding (stimulus presentation and response on the same side) or non-corresponding (stimulus presentation and response on different sides). The inter-stimulus interval was 1800 (±500) ms, and 704 stimuli were presented overall. Each block of the Simon task took about 21 min. The Simon task served as an externally-paced, monotonous cognitive task.

Before and after each Simon task the participants were asked to rate their subjectively experienced amount of mental fatigue and their motivation to continue with the task on a 9-point Likert scale.

In the physical box-sorting task (Boxes), participants had to handle 12 cardboard boxes of three different sizes and three different weights (0.5–15 kg). The boxes were placed on waist high tables, which formed two "zones" on opposite sides of the room. The distance between the two opposite zones was about 190 cm. Participants had to carry the boxes between these zones. They had to sort them according to size, weight, or label, consisting of either a letter (A, B, C) or a number (1–12) both attached to the boxes. Boxes always had to be arranged in three groups of four objects each. In case of sorting by numbers, boxes 1–4, 5–8, and 9–12 had to be put together. In case of sorting by letters, boxes A, B, and C had to be put together. The sorting rules were presented on a computer screen. After finishing one sorting task the participant had to press a button on the keyboard to get new instructions. The order of the sorting rules was randomized. There was no time limit for a single sorting task. Overall, one physical task block was performed for 25 min. A message on the screen signaled the ending of the block. The physical task was not paced at all.

EEG Data Recording and Processing

EEG was recorded from 60 standard electrode sites using an active electrode system (ActiCap; BrainProducts). Vertical eye movements and blinks were measured from two electrodes above and below the right eye (vEOG). Two electrodes at the outer canthi of both eyes were used for the measurement of horizontal eye movements (hEOG). Electrode impedance was kept below 10 kΩ. EEG and EOG were digitized at 1000 Hz and submitted via a WiFi module (MOVE; BrainProducts) to a BrainAmp MR plus EEG amplifier (BrainProducts). Data was recorded with a resolution of 0.1 μV, a Low Cutoff at DC and a High Cutoff at 250 Hz. Transmitter and the ActiCap Control Box were placed in a belt bag at the lower back of the participants. They could move around without any restrictions.

Data were offline re-referenced to averaged mastoids and a bandpass filter (0.5–45 Hz) was applied. Data were set up both for regular and event-related frequency analyses (event-related desynchronizations/synchronizazions = ERD/ERS: see Pfurtscheller and Aranibar, 1979; Klimesch, 1999; Pfurtscheller and Lopes da Silva, 1999). Because of the large structural differences between tasks, no task-inherent temporal markers for event-related analyses were available across tasks. Therefore, we referred to eye blinks as temporal marker for information processing (Wascher et al., 2014a, 2015). Only singular blinks were used for this procedure, which were not followed by another blink within 700 ms. Additionally, blinks were excluded that were accompanied by marked horizontal eye movements (for more detailed information about the blink detection mechanism, see Wascher et al., 2015). Data segments from −1000 to 2000 ms around the maximum of blink-related activity in the bipolar vEOG were extracted. An interval between −450 and −250 ms served as baseline. This interval was selected to avoid any temporal overlap with ongoing vertical eye movements. After statistics-based artifact removal as implemented in EEGLAB (Delorme and Makeig, 2004), an independent component analysis (ICA) was applied (data down-sampled to 250 Hz). Independent components (ICs) reflecting artifacts were identified and rejected using ADJUST (Mognon et al., 2011). The remaining ICs were tested for biological plausibility based on their scalp maps. The goodness of fit for modeling each IC with a single equivalent current dipole was calculated by submitting individual component maps to an automatic source localization algorithm (DIPFIT, contributed to EEGLAB by Oostenvelt et al., 2003), using a standard four-shell spherical head model. Any IC with a residual variance of more than 40% was automatically removed from the data (for a similar procedure see Debener et al., 2005).

For the analyses of frequency spectra, fast-fourier transformations (FFTs) were applied to the extracted segments, using the spectopo function of EEGLAB. In order to provide a sufficient resolution of frequencies, data were padded with zeros to a length of 2048 data points (freqfrac = 4). For ERD/ERS analyses, the matrix of valid ICs was projected back to the continuous data set for band-pass filtering (4–7.5 Hz for Theta activity; 8–12 Hz for Alpha activity). Segments that were marked as artificial in the preprocessing pipeline were removed from those data as well.

Data Analysis

Self Assessment

Analyses of variance for repeated measurements (ANOVAs) were conducted for the subjective measures (i.e., the rated mental fatigue and rated motivation to continue the task) with the between-subject factor Age (younger, older), and the within-subject factors Time on Task (ToT; across the three blocks) and Sequence (order within each block).

Behavioral Data (Simon Task)

ANOVAs were conducted for response times and error rates in the Simon task with the between-subject factor Age (younger, older) and the within-subject factors ToT, Sequence (task run before vs. after physical task), and S-R Correspondence (relates to the spatial relation between stimulus and response location: corresponding vs. non-corresponding).

EEG Data

All EEG analyses were restricted to FCz and POz, two electrodes that are commonly reported in studies investigating mental fatigue.

Since several individuals in the sample showed either multiple peaks in the EEG spectrum or no peak at all, we chose the gravity frequency method in order to determine the individual Alpha frequency (IAF; see Klimesch, 1999). Gravity frequency (GF) is defined as the weighted sum of spectral estimates in the Alpha range divided by the total Alpha power (Goljahani et al., 2012). Extracted power measures were individually adjusted to GFs (for a review see Klimesch, 1999). Lower alpha power was defined as the mean power between GF–2 Hz and GF. Upper Alpha ranged from GF to GF + 2 Hz, Theta from GF − 5 to GF − 3 Hz and Beta from GF + 5 to GF + 18. GF and the mean power in all bands were entered into ANOVAs with the between subject factor age and the within subject factors Task (3; D2, Simon task, Boxes), Time on Task (3), and Electrode (2; FCz, POz).

For ERD/ERS analyses, band-pass filtered data were squared and set into relation to the mean power in the baseline (−1000 to 0 ms relative to the blink maximum). The most impressive effect occurred immediately after the re-opening of the eyes (see also **Figure 8**). Therefore, ERD/ERS were measured in a distinct time windows between 0 and 300 ms after the blink maximum. Mean ERD/ERS were calculated for this time window and entered into the same analysis as power values and GF.

For factors with more than two levels, Greenhouse-Geisser adjusted *p*-values are reported where appropriate. Additionally, effect sizes by means of partial eta squared (η_p^2) are reported for significant results. Post-tests were Bonferroni corrected. Signal analyses were performed on MATLAB®. All statistical analyses were conducted using GNU R (R Core Team, 2012). Plots were drawn using VEUSZ (Jeremy Sanders, 2013; http://home.gna.org/veusz/).

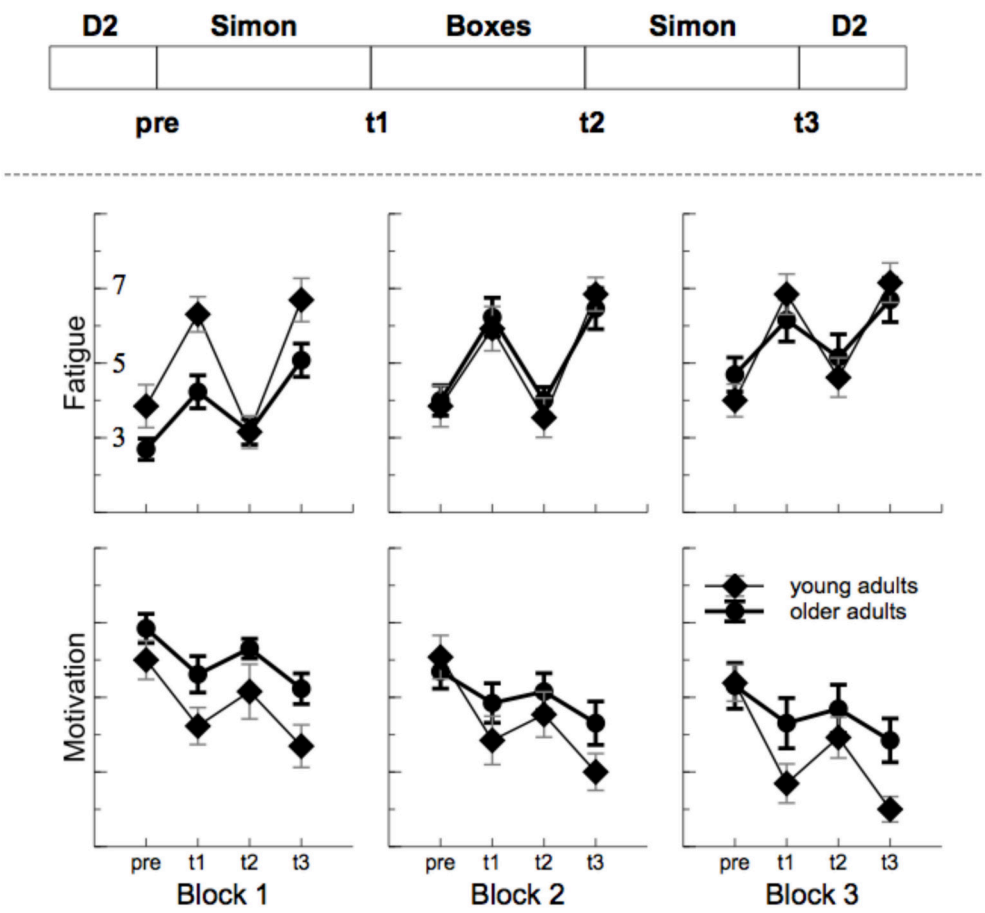


FIGURE 2 | Sequence of task and outcome of the self assessment. Fatigue increased and motivation decreased with time on task. In particular after the monotonous Simon task (t1 and t3).

RESULTS

Self Assessment

Fatigue increased with ToT, $F_{(2, 48)} = 9.70$, $p = 0.001$, $\eta_p^2 = 0.29$, and motivation decreased, $F_{(2, 48)} = 13.92$, $p < 0.001$, $\eta_p^2 = 0.37$. For both scales (see **Figure 2**), a clear modulation was found with the task performed [fatigue: $F_{(3, 72)} = 56.04$, $p < 0.001$, $\eta_p^2 = 0.70$, motivation: $F_{(3, 72)} = 28.94$, $p < 0.001$, $\eta_p^2 = 0.55$]. In particular, after the Simon task self-experienced fatigue was high and motivation was low. The time on task effect was more pronounced in older participants for mental fatigue, interaction ToT by Age: $F_{(2, 48)} = 3.33$, $p = 0.061$, $\eta_p^2 = 0.12$, but not for motivation ratings, $F_{(4, 48)} = 0.069$, $p > 0.2$, indicating that older adults experienced a stronger increase in mental fatigue than younger adults. On the other hand, for both measures some evidence for an interaction of age by task was found, fatigue: $F_{(3, 72)} = 2.44$, $p = 0.108$, $\eta_p^2 = 0.09$, motivation: $F_{(3, 72)} = 3.23$, $p = 0.045$, $\eta_p^2 = 0.12$, indicating more impact of the Simon task in younger compared to older adults.

Behavioral Data (Simon Task)

Older adults responded marginally slower than younger participants in the Simon task (see **Figure 3**), $F_{(1, 23)} = 3.99$, $p = 0.058$, $\eta_p^2 = 0.15$, and responses were faster for S-R corresponding trials, $F_{(1, 23)} = 6.25$, $p < 0.001$, $\eta_p^2 = 0.76$. No overall effect of ToT was found, $F_{(2, 46)} = 2.16$, $p = 0.132$, $\eta_p^2 = 0.09$, however, within blocks (Sequence), response times were faster after the physical task, $F_{(1, 23)} = 7.67$, $p = 0.007$, $\eta_p^2 = 0.25$. This phenomenon was most pronounced at the beginning of the experiment, interaction of ToT by Sequence: $F_{(2, 46)} = 3.99$, $p = 0.007$, $\eta_p^2 = 0.21$. No systematic variation of time on task parameters with age was found.

No age effect was found for error rates, $F_{(1, 23)} = 0.40$, $p = 0.534$, $\eta_p^2 = 0.02$, but there was an increase of error rates with non-corresponding trials, relative to corresponding trials, $F_{(1, 23)} = 39.31$, $p < 0.001$, $\eta_p^2 = 0.63$. Error rates slightly increased with ToT, $F_{(2, 46)} = 5.16$, $p = 0.013$, $\eta_p^2 = 0.18$. The latter effect was more pronounced in younger adults, interaction ToT by age: $F_{(2, 46)} = 2.79$, $p = 0.079$, $\eta_p^2 = 0.18$, indicating that younger participants committed more errors in the last block of the experiment.

EEG Data

Gravity Frequency (GF), Alpha, and Theta power

GF (see **Figure 4**) did not overall vary with age, $F_{(1, 24)} = 1.70$, $p = 0.205$, $\eta_p^2 = 0.07$, but an interactions of age by channel $F_{(1, 24)} = 26.91$, $p < 0.001$, $\eta_p^2 = 0.53$, was found. No age effect was visible at the anterior lead, $F_{(1, 24)} = 0.10$, $p > 0.05$, whereas GF was reduced with higher age at the posterior electrode, $F_{(1, 24)} = 8.12$, $p = 0.018$, $\eta_p^2 = 0.25$. Also the effect of ToT did not reach significance, $F_{(1, 24)} = 2.93$, $p = 0.200$, $\eta_p^2 = 0.11$. However, GF strongly varied with the task performed, $F_{(2, 48)} = 12.14$, $p < 0.001$, $\eta_p^2 = 0.34$, and the effect of task was qualified by electrode position, $F_{(2, 48)} = 13.74$, $p < 0.001$, $\eta_p^2 = 0.36$. At frontal leads, GF was higher in the Simon task compared to the self-paced D2 task (D2), $F_{(1, 24)} = 17.18$, $p < 0.001$, $\eta_p^2 = 0.41$. Also, the physical task showed higher GFs than the D2 at the anterior lead, $F_{(1, 24)} = 10.00$, $p = 0.008$, $\eta_p^2 = 0.29$. At POz, again the Simon task evoked higher GFs compared to the D2, $F_{(1, 24)} = 19.79$, $p < 0.001$, $\eta_p^2 = 0.45$. At this electrode location no difference in GF was found between the two self-paced tasks (D2 and Boxes), $F_{(1, 24)} = 2.31$, $p > 0.2$.

Alpha power (see **Figures 5**, **7**) increased with ToT, $F_{(1, 24)} = 17.76$, $p < 0.001$, $\eta_p^2 = 0.43$ and varied with task, $F_{(2, 48)} = 10.96$,

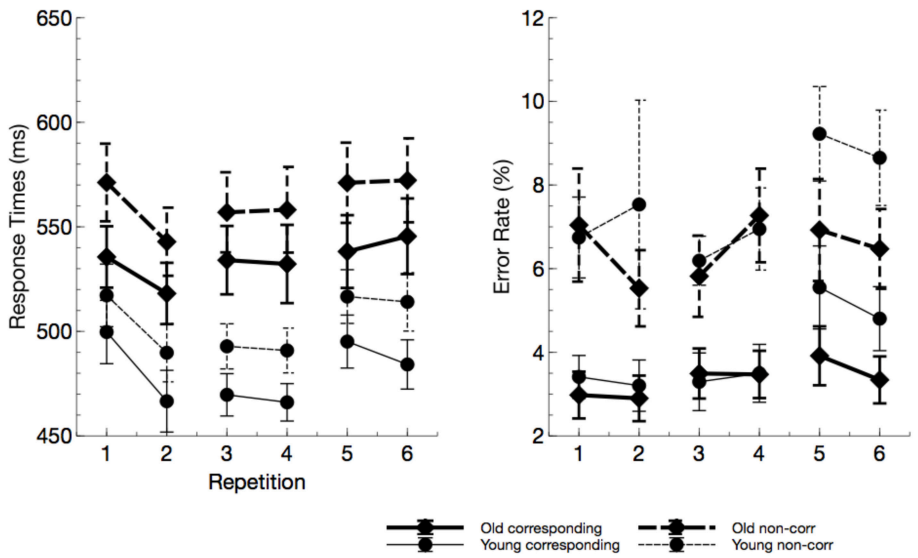


FIGURE 3 | Response times and error rates from the Simon task ("repetitions" reflect the factor sequence). Older adults responded slower but not less accurate. In particular in the last block, error rate in young participants markedly increased due to mental fatigue.

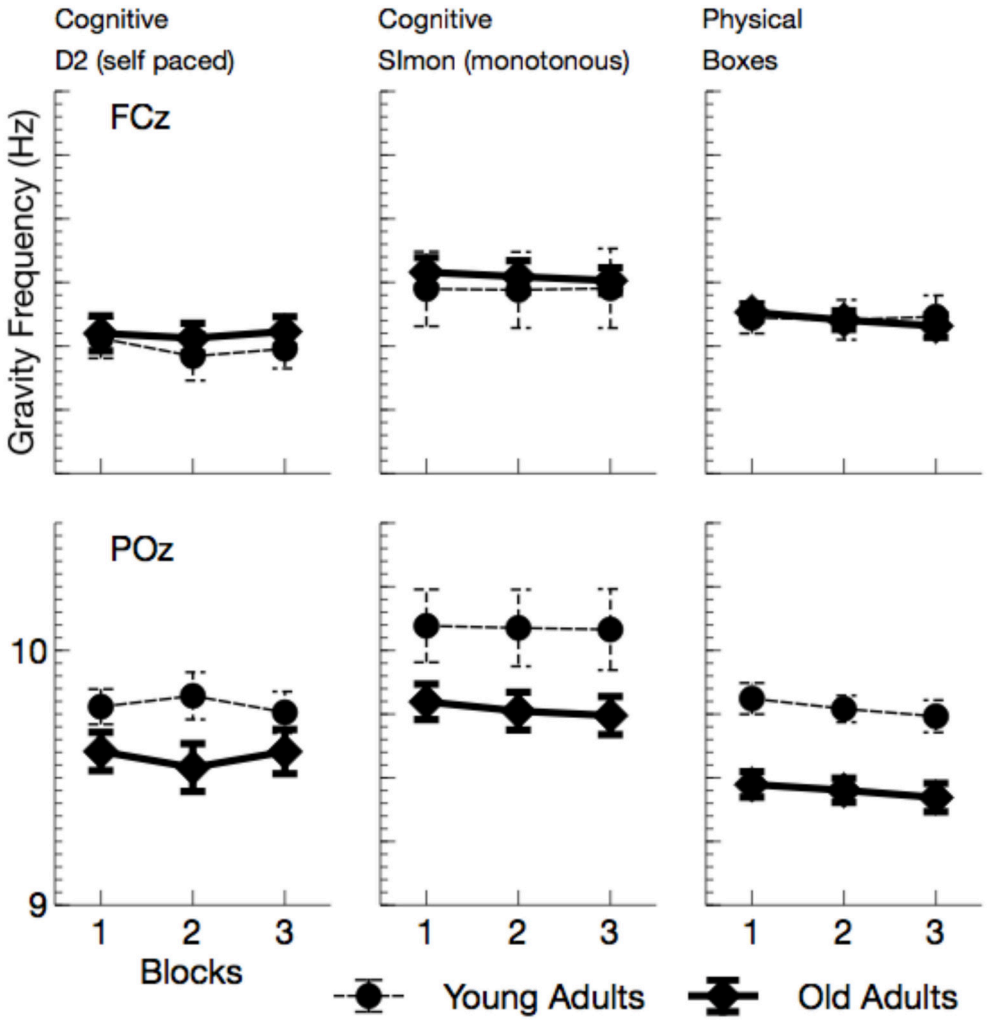


FIGURE 4 | Mean Gravity Frequency (with standard error of mean), separately for tasks and channels. Most evident is the reduction of GF at posterior leads in all tasks.

$p < 0.001$, $\eta_p^2 = 0.31$. Pairwise comparisons revealed that in the cognitive tasks, alpha power was higher in the monotonous Simon task than in the self-paced D2 task, $F_{(1, 24)} = 15.40$, $p = 0.002$, $\eta_p^2 = 0.39$. Comparing the two self-paced tasks (D2 and Boxes), alpha power was higher when participants performed the physical task, $F_{(1, 24)} = 18.18$, $p < 0.001$, $\eta_p^2 = 0.43$. The effect of task was modulated both by age, $F_{(2, 48)} = 5.81$, $p = 0.012$, $\eta_p^2 = 0.19$, and by ToT, $F_{(2, 48)} = 6.97$, $p = 0.004$, $\eta_p^2 = 0.23$.

While ToT effects were obtained for both cognitive tasks [D2: $F_{(1, 24)} = 17.69$, $p < 0.001$, $\eta_p^2 = 0.42$; Simon: $F_{(1, 24)} = 14.72$, $p = 0.003$, $\eta_p^2 = 0.38$; Boxes: $F_{(1, 24)} = 4.44$, $p = 0.138$, $\eta_p^2 = 0.16$], the age effect was restricted to the Simon task. Alpha power was enhanced in younger adults, $F_{(1, 24)} = 4.83$, $p = 0.076$, $\eta_p^2 = 0.17$.

Theta power (see **Figures 6**, **7**) was reduced in older adults, $F_{(1, 24)} = 5.29$, $p = 0.030$, $\eta_p^2 = 0.18$, and varied with the task performed, $F_{(2,48)} = 26.64$, $p < 0.001$, $\eta_p^2 = 0.53$. Theta power did not differ between the two cognitive tasks (D2 and Simon), $F_{(1, 24)} = 0.19$, $p > 0.5$, but was markedly increased in the physical task compared to the self-paced cognitive D2 task, $F_{(1, 24)} = 37.66$, $p < 0.002$, $\eta_p^2 = 0.61$. No effect of ToT was observed, $F_{(1, 24)} = 0.20$, $p > 0.5$. Task effects were more pronounced at frontal leads, $F_{(2, 48)} = 46.81$, $p < 0.001$, $\eta_p^2 = 0.66$, and varied across age groups, $F_{(2, 48)} = 3.67$, $p = 0.033$, $\eta_p^2 = 0.13$. Significant age effects were only observed in the cognitive tasks, D2: $F_{(1, 24)} = 8.96$, $p = 0.018$, $\eta_p^2 = 0.27$, Simon: $F_{(1, 24)} = 10.48$, $p = 0.012$, $\eta_p^2 = 0.30$, but not in the physical task, $F_{(1, 24)} = 0.56$, $p > 0.5$.

Blink-Related Desynchronization/Synchronization (ERD/ERS) of the EEG

As depicted in **Figure 8**, alpha activity synchronized after the eyes were opened. This effect was strongly modulated by experimental factors and differed across age groups. In the following, statistics will be reported for the mean ERD/ERS in the time window

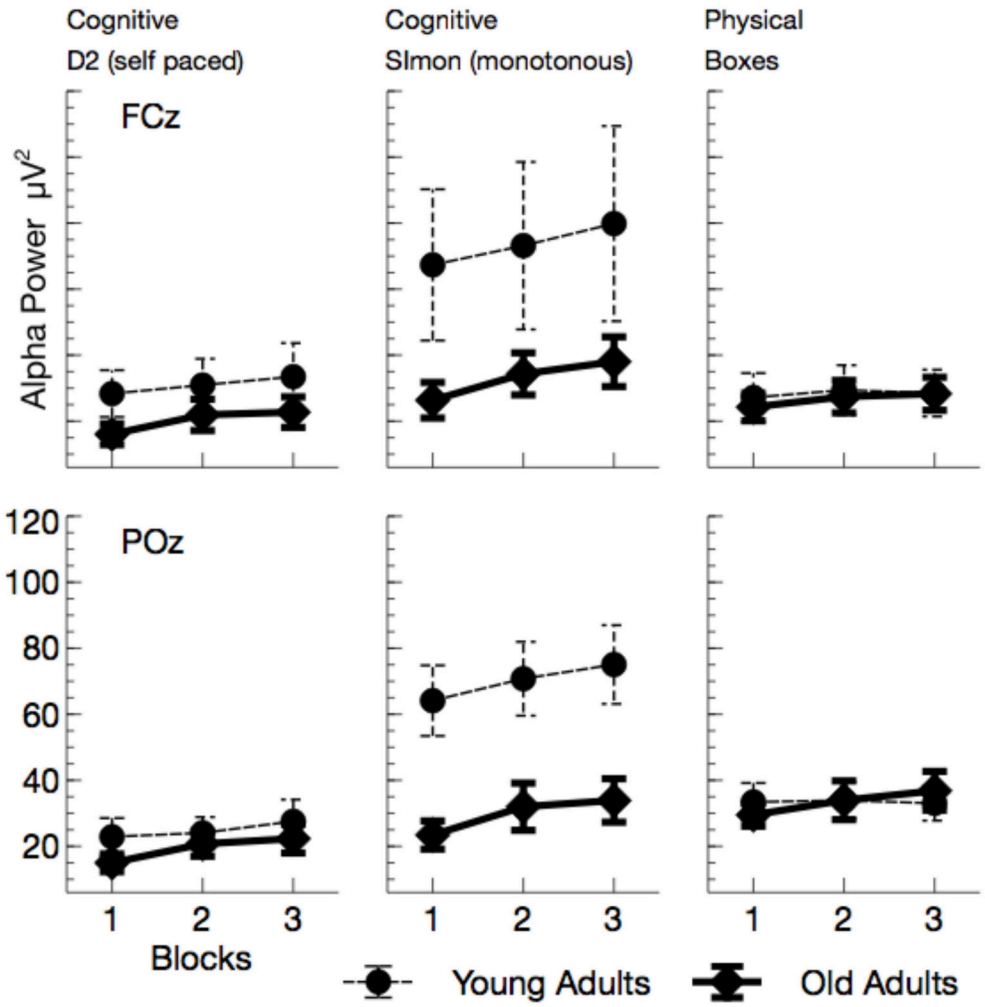


FIGURE 5 | Mean Alpha power (with standard error of mean), separated for tasks and channels. Besides a general effect of ToT, younger adults show a massive increase in alpha power in the monotonous Simon task.

between 0 and 300 ms after the maximum of the blink in the EOG.

Alpha ERD/ERS

Event-related synchronizations in the alpha band (see **Figures 8**, **9**) after the blink were enhanced for older adults, $F_{(1,\ 24)} = 9.81$, $p = 0.005$, $\eta_p^2 = 0.29$, and varied with the task performed, $F_{(2,\ 48)} = 6.97$, $p = 0.002$, $\eta_p^2 = 0.23$. Pairwise comparisons of tasks, however, did not show any significant effects [Simon vs. D2, $F_{(1,\ 24)} = 3.95$, $p = 0.118$, $\eta_p^2 = 0.14$, D2 vs. self-paced physical task, $F_{(1,\ 24)} = 2.21$, $p = 0.300$, $\eta_p^2 = 0.08$]. In the overall analysis also a number of interactions was observed, Age by ToT: $F_{(1,\ 24)} = 5.91$, $p = 0.023$, $\eta_p^2 = 0.20$, Age by Task by Channel: $F_{(2,48)} = 3.93$, $p = 0.026$, $\eta_p^2 = 0.14$, Age by Channel by ToT: $F_{(1,\ 24)} = 4.38$, $p = 0.047$, $\eta_p^2 = 0.15$, Age by Task by channel by ToT: $F_{(2,\ 48)} = 3.52$, $p = 0.038$, $\eta_p^2 = 0.13$, indicating that ERS was modulated by all experimental factors. ERS systematically increased with ToT in younger adults, $F_{(1,\ 12)} = 11.25$, $p = 0.012$, $\eta_p^2 = 0.48$, but not in older ones, $F_{(1,\ 12)} = 1.49$, $p = 0.492$, $\eta_p^2 = 0.11$. Post-tests separate for each task revealed no significant interactions of Age by ToT [D2: $F_{(1,\ 24)} = 4.72$, $p = 0.120$, $\eta_p^2 = 0.16$, Boxes: $F_{(1,\ 24)} = 5.47$, $p = 0.084$, $\eta_p^2 = 0.19$, Simon task, $F_{(1,\ 24)} = 0.09$, $p > 0.5$].

Theta ERD/ERS

Overall, Theta ERS (see **Figure 10**) was enhanced in older adults, $F_{(1,\ 24)} = 7.52$, $p = 0.011$, $\eta_p^2 = 0.24$, and varied with the task performed, $F_{(2,\ 48)} = 3.00$, $p = 0.059$, $\eta_p^2 = 0.11$. Theta ERS was slightly higher in the Simon task compared to the self-paced D2 task, $F_{(1,\ 24)} = 4.31$, $p = 0.049$, $\eta_p^2 = 0.15$, but did not differ between the two self-paced tasks, $F_{(1,\ 24)} = 0.21$, $p = 0.652$, $\eta_p^2 = 0.01$. The effect of task was modulated by a number of other variables, Age by Channel by Task: $F_{(2,\ 48)} = 3.01$, $p = 0.059$, $\eta_p^2 = 0.11$, Age by Task by ToT: $F_{(2,\ 48)} = 5.69$, $p = 0.006$, $\eta_p^2 = 0.19$, Channel by Task by ToT: $F_{(2,\ 48)} = 5.38$, $p = 0.008$,

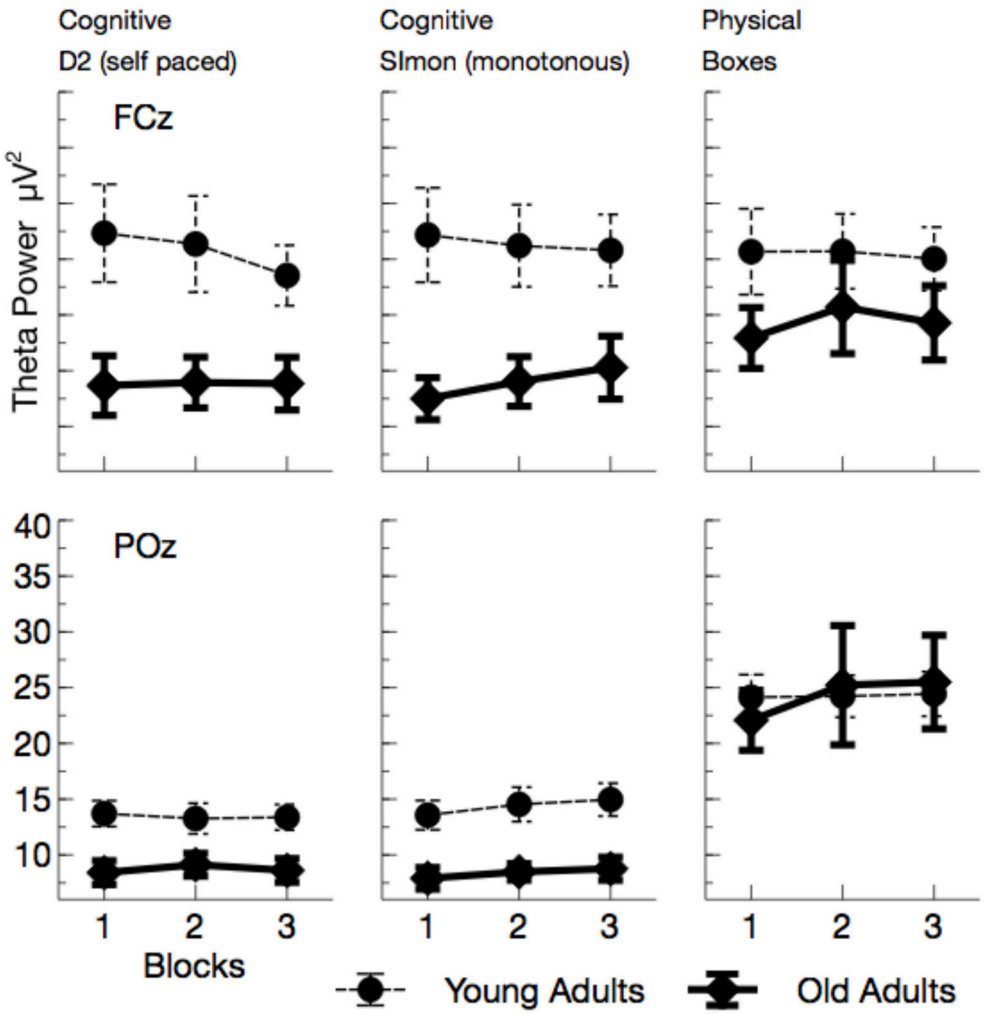


FIGURE 6 | Mean Theta power (with standard error of mean), separated for tasks and channels. Most prominently, frontal Theta was reduced in older adults across all tasks. At POz, reduction of Theta power in older adults was restricted to cognitive tasks.

$\eta_p^2 = 0.18$. All of those interactions reached significance in the young group, but failed to do so in older adults, reflecting the fact that Theta ERS strongly increased in variance in older adults. *Post-hoc* Tests revealed evidence toward an increase in Theta ERS for older adults in all tasks, D2: $F_{(1, 24)} = 5.00$, $p = 0.105$, $\eta_p^2 = 0.17$, Simon: $F_{(1, 24)} = 7.21$, $p = 0.039$, $\eta_p^2 = 0.23$, Boxes: $F_{(1, 24)} = 5.10$, $p = 0.099$, $\eta_p^2 = 0.18$. An interaction of Age by ToT was only observed in the self-paced cognitive task, D2: $F_{(1, 24)} = 7.31$, $p = 0.036$, $\eta_p^2 = 0.23$, Simon: $F_{(1, 24)} = 0.94$, $p > 0.5$, Boxes: $F_{(1, 24)} = 0.00$, $p > 0.5$.

In sum, the EEG data showed a pattern that is well comparable to previous experimental settings. Alpha power increased with ToT (see Wascher et al., 2014b) and showed a marked reduction in older adults. The latter effect, however, was restricted to the monotonous cognitive task that resembled most a regular cognitive experiment. Theta power was reduced in older adults and also systematically varied with the task performed. Age-related effects in this measure were more pronounced in cognitive tasks. Finally, a strongly enhanced synchronization of both frequency bands was observed when ERS/ERD were investigated time-locked to the blink maximum in the EEG.

DISCUSSION

In the present study, participants simulated a short (4–5 h) working shift in the post room of a wholesales house. They moved parcels and interacted with a computer in a repetitive sequence. The cognitive tasks on the computer were either repetitive (and rather monotonous) or self-paced. The design of the study was inspired by the real workflow in this particular working environment. During the entire shift, the EEG of the participants was recorded by mobile EEG equipment that did not restrict free movement and thus allowed natural behavior at any moment.

On average, subjectively experienced fatigue remained rather stable in younger adults during the entire shift, but increased for older adults. For both groups, however, fatigue was highly

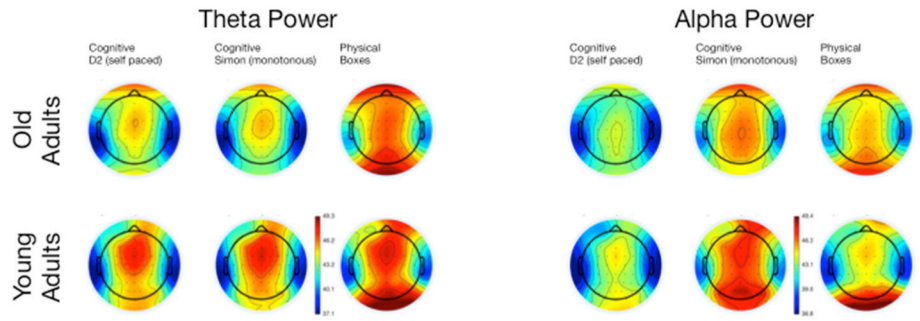


FIGURE 7 | Topographical maps of Theta and Alpha Power for the three tasks, separated for young and old adults.

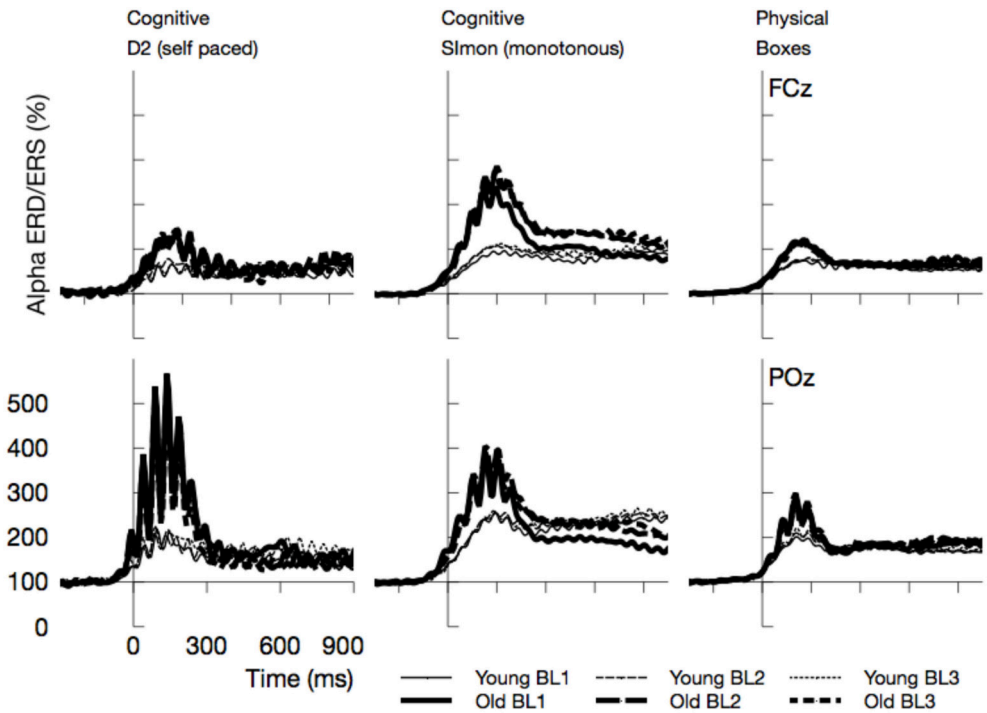


FIGURE 8 | Time course of Alpha synchronization for the three tasks and two channels, superposed for the three Blocks in the experiment (BL1, BL2, BL3). Data are time locked to the maximum of vEOG activation during the blink. Most prominently, Alpha synchronization was more pronounced in older adults.

related to the task performed. After the monotonous computer task, fatigue ratings were substantially increased compared to the physical task. This finding nicely stresses the role of monotony for the experience of mental fatigue, which was more pronounced in younger adults who were obviously subjectively more affected by the monotonous task compared to older participants. These effects go along with a local decline in motivation in the monotonous cognitive task that was found only in the younger participants.

The enhanced impact of monotony upon younger participants is also nicely mirrored in behavioral and neurophysiological data. With increasing time on task, error rates but not response times increased in younger adults. At the end of the working shift, this led to an accuracy in the Simon task that was even lower in the younger, than in the older participants. Considering the EEG, younger participants showed markedly increased alpha activity in this particular task. Referring to the assumption that high alpha power is related to an idle state of the attentional system (Hanslmayr et al., 2012), younger adults might have switched to a state of attentional withdrawal (see Wascher et al., 2014b).

Within the theoretical framework described in the Section Introduction, in which mental fatigue may result either from cognitive overload or from mental underload (May and Baldwin, 2009), the fatiguing factor in this case is a passive one (pTR), namely monotony and the decline of motivation that goes along with that. Older adults appear to deal better with monotony.

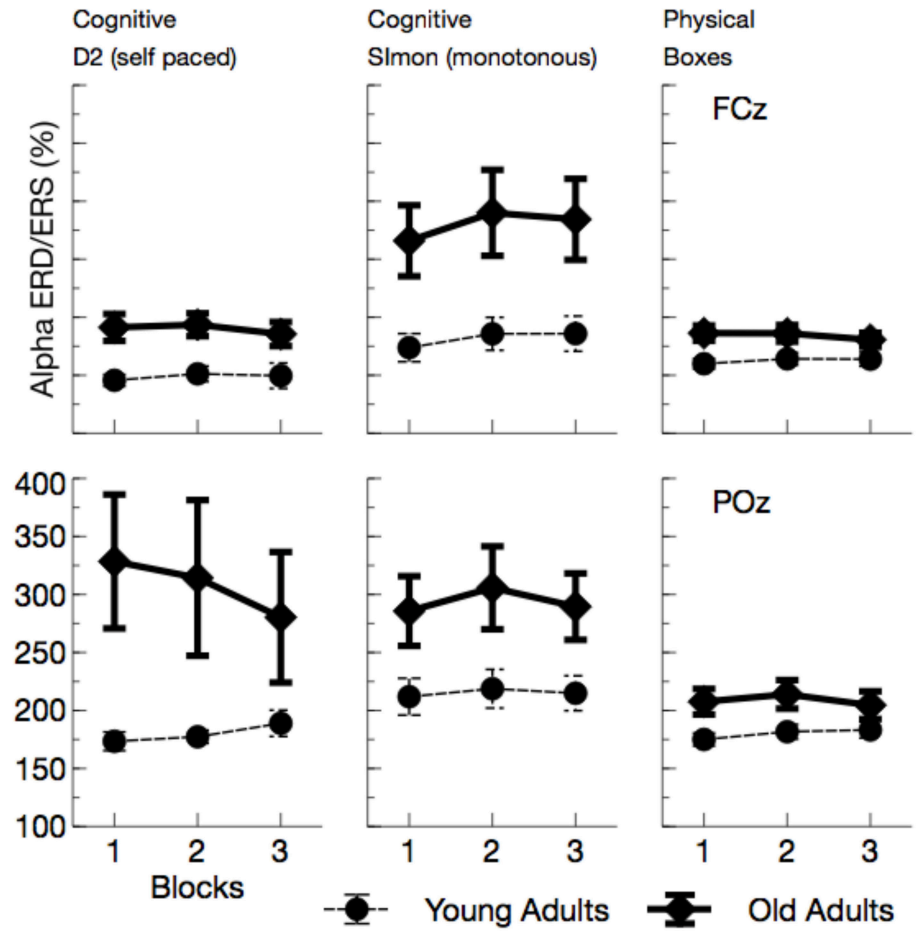


FIGURE 9 | Mean Alpha synchronization in a time window between 0 and 300 ms after the maximum of the eye-blink activity in the vEOG. Enhanced Alpha synchronization in older adults was visible in all tasks. Due to effects of ToT, ERS/ERD measures became more similar between age groups with the duration of the ongoing experiment.

Factors that made them tired were more widespread across tasks. We can't rule out that in particular the physical task was more demanding for older adults. Factors like central fatigue, i.e., a decrease of cognitive capabilities due to muscular strain (e.g., Davis, 1995; Blomstrand, 2001) might have influenced their behavior in terms of an active task-related (aTR) factor. Both, age (Müller et al., 2009) and central fatigue (Hilty et al., 2011) have been reported to go along with larger cortical phase synchronization. In particular early synchronizations in the EEG might indicate that older adults were far more driven by external signals (see Klimesch et al., 2002) compared to young participants (Zacks and Hasher, 1997; Lustig et al., 2007; Wascher et al., 2011, 2012). This is in accordance with a number of laboratory studies that showed amplified early EEG responses both in evoked potentials and in time frequency based analyses (e.g., Müller et al., 2009). This stronger impact of stimulation is assumed to be due to reduced executive cognitive control with increasing age that may affect numerous cognitive functions (Gazzaley et al., 2005, 2008; Grady et al., 2006). As a neurophysiological correlate for this deficit, reduced frontal theta activity has been discussed (Cummins and Finnigan, 2007) which was also found in the present study. This latter effect, however, was restricted to the computer-based cognitive tasks and disappeared when boxes had to be sorted. Thus, we can demonstrate that the decrease in theta power is not a global fact with increasing age, but rather is a task-dependent decline. Thus, cognitive tasks appear to be more demanding for older adults, because of deficient signal handling when information enters the system. Too much irrelevant information might be processed (Wascher et al., 2012) which is resource consuming. Therefore, mental fatigue in older adults is at least in parts related to the exhaustion of cognitive functions.

Finally, regarding effects of time on task, alpha activity showed the well-known pattern of increasing power. In contrast to pure laboratory experiments (Wascher et al., 2014b), no saturation is visible in any task in the present study. This phenomenon might be due to the alternation of tasks that interrupted monotony. In particular, the huge increase of alpha power in younger participants in the Simon task indicate that monotony was an important factor that drove

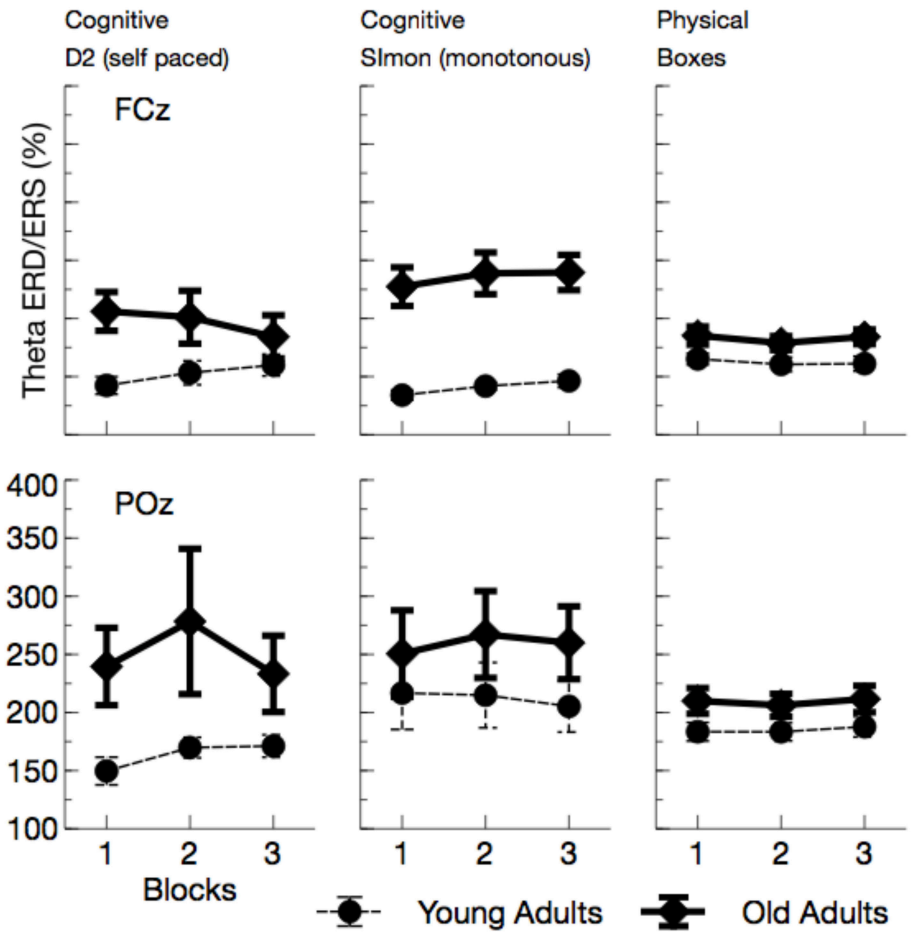


FIGURE 10 | Mean Theta synchronization in a time window between 0 and 300 ms after the maximum of the eye-blink activity in the EOG. Also in this frequency band, more synchronization was observed in older adults. As for Alpha ERS/ERD a kind of convergence across age groups appeared, at least for the self-paced cognitive D2 task.

alpha power. More interestingly, both measures of event-related synchronization/desynchronization showed convergence between age groups with time on task. When younger participants were impaired in particular by the passive task-related factor of monotony, a decline in motivation should go along with that. Reduced motivation is correlated with reduced activity in the frontal dopaminergic motivation system (Berridge and Robinson, 1998). An impairment of executive control functions was the consequence and lead to more stimulus-driven behavior. This transient state resembles the aging brain that lacks frontal activity due to physical decline (Bäckman et al., 2000).

Taken together, these results show that applying neural measures to a real life work situation provided substantial information about mechanisms and causes of mental fatigue in younger and older adults. The core results were highly comparable to laboratory studies and therefore, validity and reliability of data appears to be sufficient. The diversity of tasks additionally provided important insight into the meaning and usefulness of particular neurophysiological measures for neuro-ergonomics. Most importantly, blink-related activity in the EEG (Berg and Davies, 1988) was systematically changing with the task performed and with other experimental factors. As has been shown before (Wascher et al., 2014a), cognitive demands and cognitive strategies are reflected in these measures. Thus, it can be assumed that the re-opening of the eyes, after a blink has been executed, denotes a moment when new information enters the system, very similar to the presentation of a visual stimulus. This fact allows to measure event-related EEG analyses without any external stimulation. In particular in working situations, nothing has to be changed to the natural environment. Nevertheless, aspects of information processing can be specifically addressed.

In summary, the present study demonstrates that mobile EEG provides substantial information about information processing at the workplace and its alteration due to fatigue or age related aspects. The data pattern in self assessments, behavioral data, and neurophysiological measures nicely indicates that younger participants suffered before most from monotony. Passive task-related fatigue led to deficits in information processing with time on task. Older adults, on the other hand, were challenged from the very beginning of the work shift by altered information

processing. Due to declined executive control mechanisms, their information processing was much more stimulus-driven. Thus, the active process of overcoming this deficit appears to play a major role for mental fatigue in older worker in this particular working situation. Addressing this issue when designing working environment and work flow could substantially improve life quality of employees.

AUTHOR CONTRIBUTIONS

EW wrote the MS and analyzed data. HH contributed to the design of the study and recorded data. SK was involved in the development of the theoretical framework. SA helped developing methods of analysis. SG contributed with respect to the theoretical embedding of age related cognitive decline. TM provided theoretical aspects with respect to mental fatigue.

FUNDING

This research was supported by funding from the German Social Accident Insurance Institution for the trade and logistics sector (BGHW), the Federal Labour Office (Bundesagentur für Arbeit) and the "Metro AG."

REFERENCES

Akerstedt, T., Knutsson, A., Westerholm, P., Theorell, T., Alfredsson, L., and Kecklund, G. (2004). Mental fatigue, work and sleep. *J. Psychosom. Res.* 57, 427–433. doi: 10.1016/j.jpsychores.2003.12.001

Bäckman, L., Ginovart, N., Dixon, R. A., Wahlin, T.-B. R., Wahlin, A., Halldin, C., et al. (2000). Age-related cognitive deficits mediated by changes in the striatal dopamine system. *Am. J. Psychiatry* 157, 635–637. doi: 10.1176/ajp.157.4.635

Bandouch, J., Engstler, F., and Beetz, M. (2008). Accurate human motion capture using an ergonomics-based anthropometric human model. *Lect. Notes Comp. Sci.* 5098, 248–258. doi: 10.1007/978-3-540-70517-8_24

Berg, P., and Davies, M. B. (1988). Eyeblink-related potentials. *Electroencephalogr. Clin. Neurophysiol.* 69, 1–5. doi: 10.1016/0013-4694(88)90029-6

Berridge, K. C., and Robinson, T. E. (1998). What is the role of dopamine in reward: hedonic impact, reward learning, or incentive salience? *Brain Res. Rev.* 28, 309–369.

Blomstrand, E. (2001). Amino acids and central fatigue. *Amino Acids* 20, 25–34. doi: 10.1007/s007260170063

Boksem, M. A. S., Meijman, T. F., and Lorist, M. M. (2006). Mental fatigue, motivation and action monitoring. *Biol. Psychol.* 72, 123–132. doi: 10.1016/j.biopsycho.2005.08.007

Boksem, M. A. S., and Tops, M. (2008). Mental fatigue: costs and benefits. *Brain Res. Rev.* 59, 125–139.

Bonnefond, A., Doignon-Camus, N., Hoeft, A., and Dufour, A. (2011). Impact of motivation on cognitive control in the context of vigilance lowering: an ERP study. *Brain Cogn.* 77, 464–471. doi: 10.1016/j.bandc.2011.08.010

Brickenkamp, R. (1962). *Aufmerksamkeits-Belastungs-Test (Test d2)*. Göttingen: Hogrefe.

Chao, L. L., and Knight, R. T. (1997). Prefrontal deficits in attention and inhibitory control with aging. *Cereb. Cortex* 7, 63–69. doi: 10.1093/cercor/7.1.63

Cummins, T. D. R., and Finnigan, S. (2007). Theta power is reduced in healthy cognitive aging. *Int. J. Psychophysiol.* 66, 10–17. doi: 10.1016/j.ijpsycho.2007.05.008

Davis, J. M. (1995). Central and peripheral factors in fatigue. *J. Sports Sci.* 13, 49–53. doi: 10.1080/02640419508732277

Debener, S., Makeig, S., Delorme, A., and Engel, A. K. (2005). What is novel in the novelty oddball paradigm? Functional significance of the novelty P3 event-related potential as revealed by independent component analysis. *Cogn. Brain Res.* 22, 309–321. doi: 10.1016/j.cogbrainres.2004.09.006

Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21. doi: 10.1016/j.jneumeth.2003.10.009

Doughty, M. J. (2001). Consideration of three types of spontaneous eyeblink activity in normal humans: during reading and video display terminal use, in primary gaze, and while in conversation. *Optom. Vis. Sci.* 78, 712–725. doi: 10.1097/00006324-200110000-00011

Falkenstein, M., Hoormann, J., and Hohnsbein, J. (2002). Inhibition-related ERP components: variation with modality, age, and time-on-task. *J. Psychophysiol.* 16, 167–175. doi: 10.1027//0269-8803.16.3.167

Gazzaley, A., Clapp, W., Kelley, J., McEvoy, K., Knight, R. T., and D'Esposito, M. (2008). Age-related top-down suppression deficit in the early stages of cortical visual memory processing. *Proc. Natl. Acad. Sci. U.S.A.* 105, 13122–13126. doi: 10.1073/pnas.0806074105

Gazzaley, A., Cooney, J. W., Rissman, J., and D'Esposito, M. (2005). Top-down suppression deficit underlies working memory impairment in normal aging. *Nat. Neurosci.* 8, 1298–1300. doi: 10.1038/nn1543

Goljahani, A., D'Avanzo, C., Schiff, S., Amodio, P., Bisiacchi, P., and Sparacino, G. (2012). A novel method for the determination of the EEG individual alpha frequency. *Neuroimage* 60, 774–786. doi: 10.1016/j.neuroimage.2011.12.001

Grady, C. L., Springer, M. V., Hongwanishkul, D., McIntosh, A. R., and Winocur, G. (2006). Age-related changes in brain activity across the adult lifespan. *J. Cogn. Neurosci.* 18, 227–241. doi: 10.1162/jocn.2006.18.2.227

Halim, I., Omar, A. R., Saman, A. M., and Othman, I. (2012). Assessment of muscle fatigue associated with prolonged standing in the workplace. *Saf. Health Work* 3, 31–42. doi: 10.5491/SHAW.2012.3.1.31

Hanslmayr, S., Staudigl, T., and Fellner, M.-C. (2012). Oscillatory power decreases and long-term memory: the information via desynchronization hypothesis. *Front. Hum. Neurosci.* 6:74. doi: 10.3389/fnhum.2012.00074

Hignett, S., and McAtamney, L. (2000). Rapid entire body assessment (REBA). *Appl. Ergon.* 31, 201–205. doi: 10.1016/S0003-6870(99)00039-3

Hilty, L., Langer, N., Pascual-Marqui, R., Boutellier, U., and Lutz, K. (2011). Fatigue-induced increase in intracortical communication between mid/anterior insular and motor cortex during cycling exercise. *Eur. J. Neurosci.* 34, 2035–2042. doi: 10.1111/j.1460-9568.2011.07909.x

Hockey, R. (2013). *The Psychology of Fatigue: Work, Effort and Control*. Cambridge: Cambridge University Press.

Klimesch, W. (1999). EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain Res. Rev.* 29, 169–195. doi: 10.1016/S0165-0173(98)00056-3

Klimesch, W., Schabus, M., Doppelmayr, M., Gruber, W., and Sauseng, P. (2002). Evoked oscillations and early components of event-related potentials: an analysis. *Int. J. Bifurcat. Chaos* 14, 705–718. doi: 10.1142/S0218127404009417

Lal, S. K. L., and Craig, A. (2001). A critical review of the psychophysiology of driver fatigue. *Biol. Psychol.* 55, 173–194. doi: 10.1016/S0301-0511(00)00085-5

Lämkull, D., Hanson, L., and Örtengren, R. (2007). The influence of virtual human model appearance on visual ergonomics posture evaluation. *Appl. Ergon.* 38, 713–722. doi: 10.1016/j.apergo.2006.12.007

Lerman, S. E., Eskin, E., Flower, D. J., George, E. C., Gerson, B., Hartenbaum, N., et al. (2012). Fatigue risk management in the workplace. *J. Occup. Environ. Med.* 54, 231–258. doi: 10.1097/JOM.0b013e318247a3b0

Lustig, C., Hasher, L., and Zacks, R. T. (2007). "Inhibitory deficit theory: recent developments in a "new view,"" in *The Place of Inhibition in Cognition,* eds D. S. Gorfein and C. M. MacLeod (Washington, DC: American Psychological Association), 145–162. doi: 10.1037/11587-008

Marcora, S. M., Staiano, W., and Manning, V. (2009). Mental fatigue impairs physical performance in humans. *J. Appl. Physiol.* 106, 857–864. doi: 10.1152/japplphysiol.91324.2008

May, J. F., and Baldwin, C. L. (2009). Driver fatigue: the importance of identifying causal factors of fatigue when considering detection and

countermeasure technologies. *Transport. Res. F Psychol. Behav.* 12, 218–224. doi: 10.1016/j.trf.2008.11.005

McAtamney, L., and Corlett, E. N. (1993). RULA: a survey method for the. *Appl. Ergon.* 24, 91–99. doi: 10.1016/0003-6870(93)90080-S

Mehta, R. K., and Parasuraman, R. (2013). Neuroergonomics: a review of applications to physical and cognitive work. *Front. Hum. Neurosci.* 7:889. doi: 10.3389/fnhum.2013.00889

Mognon, A., Jovicich, J., Bruzzone, L., and Buiatti, M. (2011). ADJUST: An automatic EEG artifact detector based on the joint use of spatial and temporal features. *Psychophysiology* 48, 229–240. doi: 10.1111/j.1469-8986.2010.01061.x

Müller, V., Gruber, W., Klimesch, W., and Lindenberger, U. (2009). Lifespan differences in cortical dynamics of auditory perception. *Dev. Sci.* 12, 839–853. doi: 10.1111/j.1467-7687.2009.00834.x

Newsholme, E. A., Blomstrand, E., and Ekblom, B. (1992). Physical and mental fatigue: metabolic mechanisms and importance of plasma amino acids. *Br. Med. Bull.* 48, 477–495.

Oostenvelt, R., Delorme, A., and Makeig, S. (2003). *DIPFIT: Equivalent Dipole Source Localization of Independent Components.* Available online at: http://sccn.ucsd.edu/eeglab/dipfittut/dipfit.html

Pfurtscheller, G., and Aranibar, A. (1979). Evaluation of event-related desynchronization (ERD) preceding and following voluntary self-paced movement. *Electroencephalogr. Clin. Neurophysiol.* 46, 138–146. doi: 10.1016/0013-4694(79)90063-4

Pfurtscheller, G., and Lopes da Silva, F. H. (1999). Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clin. Neurophysiol.* 110, 1842–1857. doi: 10.1016/S1388-2457(99)00141-8

R Core Team (2012). *R: A Language and Environment for Statistical Computing.* Vienna: R Foundation for Statistical Computing.

Sarter, M., Gehring, W. J., and Kozak, R. (2006). More attention must be paid: the neurobiology of attentional effort. *Brain Res. Rev.* 51, 145–160. doi: 10.1016/j.brainresrev.2005.11.002

Sarter, M., Givens, B., and Bruno, J. P. (2001). The cognitive neuroscience of sustained attention: where top-down meets bottom-up. *Brain Res. Rev.* 35, 146–160. doi: 10.1016/S0165-0173(01)00044-3

Simon, J. R. (1969). Reactions toward the source of stimulation. *J. Exp. Psychol.* 81, 174–176. doi: 10.1037/h0027448

van der Linden, D., Frese, M., and Meijman, T. F. (2003). Mental fatigue and the control of cognitive processes: effects on perseveration and planning. *Acta Psychol.* 113, 45–65. doi: 10.1016/S0001-6918(02)00150-6

Wascher, E., and Getzmann, S. (2014). Rapid mental fatigue amplifies age-related attentional deficits. *J. Psychophysiol.* 28, 215–224. doi: 10.1027/0269-8803/a000127

Wascher, E., Falkenstein, M., and Wild-Wall, N. (2011). Age related strategic differences in processing irrelevant information. *Neurosci. Lett.* 487, 66–69. doi: 10.1016/j.neulet.2010.09.075

Wascher, E., Heppner, H., and Hoffmann, S. (2014a). Towards the measurement of event-related EEG activity in real-life working environments. *Int. J. Psychophysiol.* 91, 3–9. doi: 10.1016/j.ijpsycho.2013.10.006

Wascher, E., Heppner, H., Möckel, T., Kobald, S. O., and Getzmann, S. (2015). Eye-blinks in choice response tasks uncover hidden aspects of information processing. *EXCLI J.* 15, 1207–1218. doi: 10.17179/excli2015-696

Wascher, E., Rasch, B., Sänger, J., Hoffmann, S., Schneider, D., Rinkenauer, G., et al. (2014b). Frontal theta activity reflects distinct aspects of mental fatigue. *Biol. Psychol.* 96, 57–65. doi: 10.1016/j.biopsycho.2013.11.010

Wascher, E., Schneider, D., Hoffmann, S., Beste, C., and Sänger, J. (2012). When compensation fails: attentional deficits in healthy ageing caused by visual distraction. *Neuropsychologia* 50, 3185–3192. doi: 10.1016/j.neuropsychologia.2012.09.033

Zacks, R., and Hasher, L. (1997). Cognitive gerontology and attentional inhibition: a reply to Burke and McDowd. *J. Gerontol. B Psychol. Sci. Soc. Sci.* 52B, P274–P283.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Wascher, Heppner, Kobald, Arnau, Getzmann and Möckel. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 20 April 2016
doi: 10.3389/fnhum.2016.00171

Benefits of Instructed Responding in Manual Assembly Tasks: An ERP Approach

Pavle Mijović[1], Vanja Ković[2], Maarten De Vos[3], Ivan Mačužić[1], Branislav Jeremić[1] and Ivan Gligorijević[1]*

[1] Department for Production Engineering, Faculty of Engineering, University of Kragujevac, Kragujevac, Serbia, [2] Laboratory for Neurocognition and Applied Cognition, Department for Psychology, Faculty of Philosophy, University of Belgrade, Belgrade, Serbia, [3] Department of Engineering, The Institute of Biomedical Engineering, University of Oxford, Oxford, UK

The majority of neuroergonomics studies are focused mainly on investigating the interaction between operators and automated systems. Far less attention has been dedicated to the investigation of brain processes in more traditional workplaces, such as manual assembly, which are still ubiquitous in industry. The present study investigates whether assembly workers' attention can be enhanced if they are instructed with which hand to initiate the assembly operation, as opposed to the case when they can commence the operation with whichever hand they prefer. For this aim, we replicated a specific workplace, where 17 participants in the study simulated a manual assembly operation of the rubber hoses that are used in vehicle hydraulic brake systems, while wearing wireless electroencephalography (EEG). The specific EEG feature of interest for this study was the P300 components' amplitude of the event-related potential (ERP), as it has previously been shown that it is positively related to human attention. The behavioral attention-related modality of reaction times (RTs) was also recorded. Participants were presented with two distinct tasks during the simulated operation, which were counterbalanced across participants. In the first task, digits were used as indicators for the operation initiation (Numbers task), where participants could freely choose with which hand they would commence the action upon seeing the digit. In the second task, participants were presented with arrows, which served as instructed operation initiators (Arrows task), and they were instructed to start each operation with the hand that corresponded to the arrow direction. The results of this study showed that the P300 amplitude was significantly higher in the instructed condition. Interestingly, the RTs did not differ across any task conditions. This, together with the other findings of this study, suggests that attention levels can be increased using instructed responses without compromising work performance or operators' well-being, paving the way for future applications in manual assembly task design.

Keywords: neuroergonomics, wireless electroencepholagraphy, event-related potentials, P300, attention, manual assembly

OPEN ACCESS

Edited by:
Klaus Gramann, Berlin Institute of Technology, Germany

Reviewed by:
Rolf Verleger, Universität zu Lübeck, Germany
Edmund Wascher, Leibniz Research Centre for Working Environment and Human Factors, Germany

****Correspondence:***
Pavle Mijović
p.mijovic@kg.ac.rs

Received: *01 October 2015*
Accepted: *04 April 2016*
Published: *20 April 2016*

Citation:
Mijović P, Ković V, De Vos M, Mačužić I, Jeremić B and Gligorijević I (2016) Benefits of Instructed Responding in Manual Assembly Tasks: An ERP Approach. Front. Hum. Neurosci. 10:171. doi: 10.3389/fnhum.2016.00171

INTRODUCTION

The importance of studying the human brain processes while executing everyday complex tasks in naturalistic environments was pinpointed by Parasuraman (2003), through a new direction in human factors and ergonomics (HF/E) research. This novel direction was tentatively named neuroergonomics (Parasuraman, 2003; Parasuraman and Rizzo, 2006; Parasuraman, 2011). Although Parasuraman and Wilson (2008) modestly stated that neuroergonomics should not be thought of as revolutionary, but rather as another step in HFE research, the growing body of neuroergonomics research refutes this statement. In fact, ever advancing technology has facilitated neuroergonomics research and only 12 years from its inception it has become one of the principal directions in HFE research. Ultimately, understanding brain processes in naturalistic environments can lead to improvement of existing industrial processes design and to creation of safer and more efficient working conditions (Parasuraman, 2003), consequently improving the operators' overall well-being.

Neuroergonomics has had significant success in evaluating brain activity in its interaction with automated systems, through the studies of mental workload, dual-task performance (Ayaz et al., 2013) and operators' vigilance (Warm and Parasuraman, 2006; Warm et al., 2009). Additionally, it has gone a step further with the development of state-of-the-art neuroadaptive systems facilitating the mutual interaction between an automated system and operators, in the sense that both human and the system can initiate a change in the level of automation when needed (Scerbo, 2006; Mehta and Parasuraman, 2013). On the one hand, this trend is understandable as industry, over several decades, has tried to reach the "lights-out manufacturing" concept (Tompkins et al., 2010), i.e., completely automated factories which can operate without the direct presence of human operators in the production processes. In that case, human supervisory control of automated systems becomes essential (Sheridan and Parasuraman, 2005), as human operators would be solely responsible for controlling the automated production systems (Warm et al., 2008). On the other hand, although automation is becoming ubiquitous in industry and everyday life (Parasuraman and Wilson, 2008), the "lights-out" concept is still rather futuristic and there is still a need for human manual operations in production processes. This is especially notable in assembly tasks and processes where costs related to process automation are generally not justifiable (Tang et al., 2003).

For these reasons, it is evident that neuroergonomics studies should pay additional attention to more traditional workplaces, through investigation of concurrent physical and cognitive work. This approach has received far less attention in neuroergonomic studies (for review see Mehta and Parasuraman, 2013). For example, in the car manufacturing industries the majority of processes are automated, however human operators play a crucial role in the final car cockpit and interior assembly, i.e., final assembly (Michalos et al., 2010). Typically, manual assembly tasks require a large number of repetitions and are monotonous in nature, thus leading to hypo-vigilance of operators (Spath et al., 2012). In turn, operators have difficulty in sustaining the desired level of attention during the task, and therefore, the risk of work-related injuries, material damage or even accidents is increased (Kletz, 2001). Therefore, employing existing neuroimaging techniques to understand the way the brain processes various stimuli in this class of tasks could be beneficial, as the task design could be optimized in such a way as to obtain and maintain sufficient operator attention, thereby avoiding possibly hazardous situations.

An extensive review of neuroimaging techniques applicable to neuroergonomics research has been recently published by Mehta and Parasuraman (2013). Although functional near infrared spectroscopy (fNIRS) presents a convenient technique for the neuroergonomics research in naturalistic setting due to its light weight and portability (Ayaz et al., 2010, 2012; Mehta and Parasuraman, 2013), it suffers from low temporal resolution and its use in dynamic everyday environments is still somewhat limited (Gramann et al., 2011). On the other hand, Electroencephalography (EEG) and therefrom derived event related potentials (ERPs) belong to the neuroimaging techniques that directly measure brain activity (Gramann et al., 2011; Mehta and Parasuraman, 2013) and both EEG and ERPs possesses high temporal resolution (down to the order of milliseconds) making them suitable for real-time investigation of brain dynamics in complex environments (Gramann et al., 2011). Even though Parasuraman (1990) proposed the introduction of ERPs in ergonomics research, until recently the traditional EEG recording suffered from long wiring between the electrode cap and amplifier unit, which engenders the artifacts that degrades signal quality (Debener et al., 2012). Additionally, EEG recordings usually required shielded, dimly lit and sound attenuated rooms, which was one of the main precondition for its recording, thus limiting its use in naturalistic environments (Gramann et al., 2011). However, these problems were recently overcome by the development of wearable EEG systems, empowering its use in everyday and applied settings (Debener et al., 2012; De Vos et al., 2014; Wascher et al., 2014; Mijović et al., 2016). Consequently, operators' brain dynamics can nowadays be successfully investigated with wearable EEG in faithfully replicated workplaces, by simulating the work activity (Wascher et al., 2014; Mijović et al., 2016). This can provide insight in how the brain responds to complex industrial tasks and these findings can contribute to more efficient task designs.

The aim of this article is the investigation of assemblers' mental states, by utilizing ERPs in a realistically replicated workplace. Neuroergonomics implies that overt performance measurements are unreliable (Parasuraman, 2003), since they do not provide the possibility for timely investigation of the underlying covert cognitive processes during everyday tasks. To get better insights into the temporal course of the underlying attention processes engaged in manual assembly operation, we selected two tasks in which we triggered goal-directed actions of workers by presenting them with either digits (in one) or arrows (in the other task) prior to initiating the operation. In this way we wanted to elicit the P300 ERP component (also

called P3 or P3b), which is represented by the positive ERP voltage deflection that usually appears between 300 and 500 ms after appearance of the task-relevant stimuli (Polich and Kok, 1995; Verleger et al., 2005). The P300 component is often used to identify the depth of cognitive information processing and its amplitude and latency are considered to be related to the human attention level (Johnson, 1988; Polich, 2007; De Vos et al., 2014).

The P300 complex is the most prominent over the midline scalp sites (Polich, 2007) and it is among the most prominent ERP components (Verleger et al., 2014), making it one of the most studied components of human ERP. However, there is still a lack of consensus regarding what brain functions the P300 component represents (the arguments are briefly summarized in Verleger et al., 2014). One influential view is that the P300 component can be explained through the context updating hypothesis that was proposed by Donchin (1981) and which governs that the P3 reflects the updating of working memory that is related to task-relevant and unexpected events. The context updating theory assumes that the mental process that elicit the P3 component reflects a revision of the model of the environment rather than serving to organize a response to the eliciting stimulus (Verleger et al., 2005). In other words, it is assumed that following an initial sensory processing, an attention-related process evaluates the presentation of the previous event in working memory and if a new stimulus in a train of standard stimuli is detected, the attention-related process updates, which is followed by production of the P300 component (Polich, 2007). However, we have also witnessed arguments against the context updating theory (Verleger et al., 2005, 2014). In fact, Verleger et al. (2005) proposed a new hypothesis in which they argued that the P300 component is related both to stimuli processing and organizing the response. In order to prove this hypothesis, Verleger et al. (2005) compared the P3 amplitude in stimulus- and response-locked ERPs and they found that both P3 amplitudes were comparable. Therefore, it was confirmed that P300 amplitude does not reflect just the simple reaction to stimulus change. Rather, P300 reflects a process that mediates between perceptual analysis and response (Verleger et al., 2005), i.e., it is related to the organization of the response and it depends on the stimulus-response links (Verleger et al., 2014).

Based on these findings, the present study investigated whether and how the neural correlates of goal-directed actions would differ if the operators were requested to initiate the simulated assembly operation spontaneously (upon seeing a digit), as opposed to the condition where participants were instructed with which hand to commence the operation (upon seeing an arrow). In the spontaneous condition (the Numbers task), we adopted the stimuli from the original SART paradigm that is a simple "go/no-go" task, which consists of consecutively presenting digits from "1" to "9" and participants are required to give a speedy response on all stimuli, with the exception of digit "3" (Robertson et al., 1997). The main difference between the original SART and the Numbers paradigm (used in our study) is that the digits in Numbers are randomized. Further, in the original SART paradigm it is requested that participants provide the speedy response with the index finger upon the stimulus presentation. However, this would impede the simulation of the real working operation, since it would require an additional, task-unrelated operation from participants. Instead, in the Numbers paradigm, participants were instructed to initiate the assembly operation as soon as the visual (target) stimulus appeared on the screen, with whichever hand they felt more comfortable (the assembly operation is explained in detail in Section "Simulated Assembly Operation"). For the instructed responding (the Arrows) task, we adopted the stimuli and procedures from Donkers and van Boxtel (2004). The Arrows task is essentially a choice reaction task, where the arrows pointing to the left and right appear on the screen; white arrows represent the target ("go") condition, while red arrows represent the "no-go" stimulus. The main difference between the Numbers and Arrows tasks was that in the Numbers task participants could freely choose the hand with which they would initiate the assembly operation, while in the Arrows task, participants were instructed to commence each operation with the hand that corresponds to the direction in which the white arrow on the screen was pointing. An important notion is that not only the simple stimulus difference between the tasks was varied (digit vs. arrow), but also the informational value of those stimuli: the Arrows task arguably requires stimulus-response mapping, which in turn requires more cognitive evaluation, consequently inducing higher-level attentional processing than in the simple "go/no-go" task. In both the task specific and spontaneous condition, the visual stimuli (digits and arrows) appeared in the center of a screen that was placed in front of the participants.

We expected attention, when assessed through the P300 amplitude, to be more enhanced in the instructed responding (Arrows) task, compared to the one where participants could initiate the assembly operation upon seeing the task unspecific cue (Numbers task). Further, we wanted to investigate whether the difference in the task condition would also influence the reaction times (RTs), as the performance of the participants is also important, since this study simulates the naturalistic assembly task replicated from the industry. In other words, we wanted to investigate whether the participants would be slower in the case when they are instructed with which hand they should start the assembly operation, as compared to the condition when they can spontaneously initiate the assembly operation with whichever hand they prefer.

MATERIALS AND METHODS

Participants

Seventeen healthy subjects, from which one was left-handed, aged between 19 and 21 years volunteered as participants in the study. Due to abnormalities in the recording three subjects were excluded from further analysis, leaving a total of 14 participants. The study was restricted to male participants, both to exclude possible inter-gender differences and to replicate the selected job task more faithfully, since in the company that supported our research only males occupy the specific workplace under study.

Participants did not report any past or present neurological or psychiatric conditions and were free of medication and psychoactive substances. They were instructed not to take any alcoholic drinks prior to, nor on the day of, participation in the study. All participants had normal or corrected-to-normal vision. They agreed to participate in the study and signed informed consent after reading the experiment summary in accordance with the *Declaration of Helsinki*. The Ethical Committee of the University of Kragujevac approved the study and procedures for the participants.

Replication of the Workplace

As we stated in the introduction, reliable EEG recording still relies on wet electrodes, limiting on-site industrial EEG recording. For that reason, we simulated the production process of the rubber hoses, which are used in hydraulic brake systems in the automotive industry, in a faithfully replicated workplace (**Figure 1**). Full-scale replica of the specific workplace was created at the laboratory of University of Kragujevac, in consultation with the car sub-component manufacturing company. In order to create a naturalistic environment, all major elements from the real factory settings have been included while preserving respective spatial ratios and replicating ambient conditions.

The laboratory was air-conditioned and microclimate conditions controlled, keeping the ambient temperature at 24 ± 1°C while the measured relative air humidity value was between 40% and 60%. The luminance at the real workplace was also replicated from the industrial settings, using the same lighting and maintaining the luminance value at 810 lx. Finally, the noise trace was obtained by recording sounds in the vicinity of the original production facility, using cardiodid condenser microphone AT2020USB (Audio-technica, Japan), and this was replayed during the experiments with an SW-HF 5.1 6000 surround multimedia speaker (Genius, Taiwan). The ambient (light, noise) and microclimate (temperature, humidity) condition values were obtained using multifunctional environmental meter device PCE-EM882 (PCE instruments, UK).

The experimental setup used in this study was similar to previously reported studies (Mijović et al., 2015a,b, 2016), while the experimental task and procedure were modified. For clarity, we will repeat the detailed experimental setup here.

Simulated Assembly Operation

In the production process, an operator carries out a crimping operation in order to join a metal extension to a rubber hose. This



FIGURE 1 | Left image—Real workplace (replicated from our industrial partner); Right image—Replicated workplace.

single operation, carried out in a sitting position, consists of eight simple steps (actions). Step-by-step simulated operation, carried out by participants in the replicated working environment, is graphically presented in **Figure 2A** and explained in detail further in the text.

The major production steps can be summarized as follows (**Figure 2A**): first, the information to initiate the simulated assembly operation is presented to the participant, in the form of visual stimulus (step 1, explained in detail in Section "Experimental Procedure"), upon which he is instructed to instantly initiate the operation by taking the metal part (step 2) and the rubber hose (step 3). Following this, participants should place the metal part on the hose (step 4) and place both inside the crimping machine (step 5). Once the rubber hose and metal part are correctly placed inside the opening, the industrial green lamp lights and presents a visual cue to the participant, informing him that the part has been correctly placed. Participant then proceed by promptly pressing the pedal, which initiates the improvised machine and replicates the real machines' crimping sound with a duration of 3500 ms (step 6). The real crimping operation that would happen upon pressing the pedal was avoided, preserving its major aspects from operator's perspective—the sound it produces and the cessation of which indicates the end of machine operation, analogously to the real case. Upon completion of the simulated crimping process, the participant removes the component and places it in the box with completed parts (step 7). Finally, following these steps, the participant sits still, waiting for the subsequent stimulus (step 8) indicating the next-in-line operation.

Although the assembly task consists of eight sub-actions, the whole operation lasts less than 10 s and a single operator

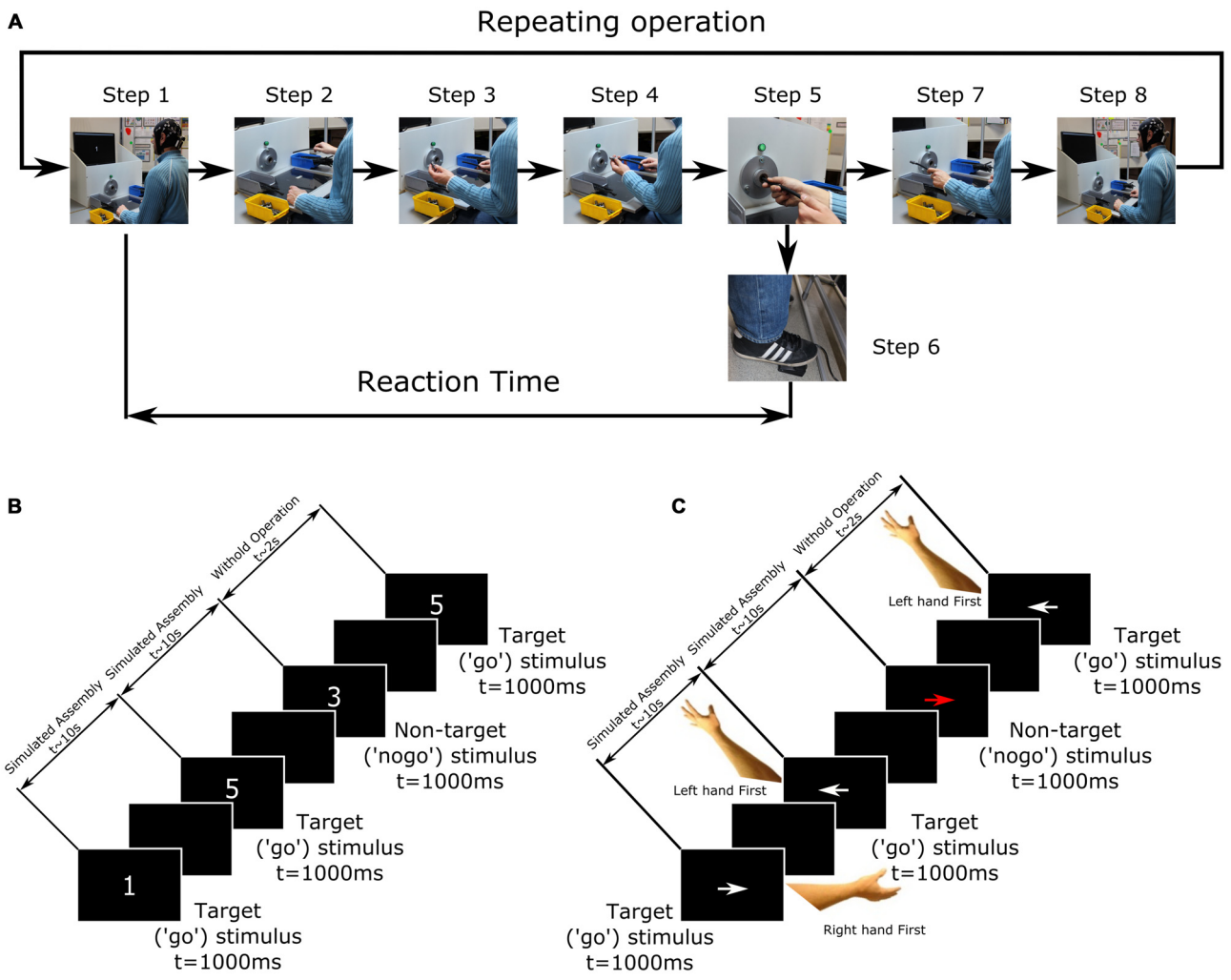


FIGURE 2 | (A) Step by step representation of the simulated working process. Step 1—Stimulus presentation; Step 2—Taking the rubber hose; Step 3—Taking the metal part; Step 4—Placing metal part on the rubber hose; Step 5—Insertion of the uncompleted part inside the improvised machine opening; Step 6—Pressing the pedal in order to initiate the simulated crimping operation; Step 7—Placing the completed into the box with completed parts; Step 8—Waiting for the successive stimulus presentation. **(B)** Graphical representation of the Numbers Task. **(C)** Graphical Representation of the Arrows task.

completes between 2500–3000 elements during a work shift. Hence, this workplace represents a typical example of a repetitive, monotonous operational task in industrial assembly settings.

Preparation

Each participant arrived to the laboratory at 9:00 a.m. Upon carefully reading the experiment summary and signing the informed consent for participation in the study, participants started the training session in order to gain familiarity with the task. Due to its simplicity, they were given 15 min for practicing, following which they confirmed their readiness to start the experiment. Finally, an EEG cap and amplifier were mounted on the participant's head (as explained in the Section "EEG Recording") and the recording started around 9:30 a.m.

Experimental Procedure

During the experiment, at least two experimenters were constantly present in the laboratory in order to assure that experimental procedures were strictly followed. The experimenters were seated behind an opaque board (so that participants could not see them during the task) and they observed the participants through a red-blue-green (RGB) camera that recorded the entire experiment.

Participants were seated in a comfortable chair in front of an improvised workplace including the improvised machine (**Figure 1**). In order to extract the ERP component from continuous EEG recording, a single functional modification in the simulated assembly task was made. Simultaneously with the simulated assembly process, the participants were subjected to either the Numbers (**Figure 2B**) or Arrows (**Figure 2C**) task to prompt initiation of the assembly operation. Both tasks were presented on the 24" screen from a distance of approximately 100 cm in a balanced order across participants (with a 15 min break between the tasks). The screen was height adjustable and the center of the screen was set to be level with participants' eyes. Upon presentation of the stimuli on the screen, the participants were instructed to complete the previously explained assembly operation (also graphically presented in **Figure 2A**).

All the stimuli were presented for 1000 ms on a black screen background. In both tasks the appearance of the stimuli was randomized, with the condition that forbade the two consecutive appearance of the "no-go" stimuli (digit "3" in Numbers, and red arrow in Arrows task). Additionally, in the Numbers tasks, five randomly allocated digit sizes were presented to increase the demands for processing the numerical value and to minimize the possibility that subjects would set a search template for some perceptual feature of the "no-go" trial (the digit "3"). Digit font sizes were 60, 80, 100, 120 and 140 in Arial text font (similar to Dockree et al., 2005). The main difference between the tasks is that in the Arrows tasks the participants were instructed to initiate the simulated operation with the right hand (step 2) if the white arrow was pointing to the right, or with the left hand (step 3) if pointing left (as depicted on **Figure 2C**), while in the Numbers task, the participants could freely choose between step 2 or step 3 (from the **Figure 2A**) upon seeing the digit. Each task consisted of 500 stimuli, where the probability of appearance of the "no-go" stimuli was set at 10% (50 in total), while the "go" stimuli were presented 450 times. The inter-stimulus interval (ISI) between two consecutive "go" stimuli was on average 11,240 ms (STD = 410 ms), while between "no-go" and following "go" stimuli the average ISI was 3210 ms (STD = 120 ms). The duration of the each task was around one and a half hours, upon which participants had a 15 min break, before starting the second task. Thus, the whole experiment lasted around 3 h and 15 min.

The task specifications were programmed in Simulation and Neuroscience Application Platform (SNAP)[1], developed by the Swartz Center for Computational Neuroscience (SCCN). As explained in Bigdely-Shamlo et al. (2013) and Gramann et al. (2014), SNAP is a python-based experiment control framework that is able to send markers as strings to Lab Streaming Layer (LSL)[2]. LSL is a real-time data collection and distribution system that allows multiple continuous data streams as well as discrete marker timestamps to be acquired simultaneously in an eXtensible Data Format (XDF)[3]. This data collection method provides synchronous, precise recording of multi-channel, multi-stream data that is heterogeneous in both type and sampling rate (Bigdely-Shamlo et al., 2013; Gramann et al., 2014), and is obtained via a local area network (LAN).

EEG Recording

EEG data acquisition was performed using the SMARTING (mBrainTrain, Serbia) wireless EEG system, with a sampling frequency of 500 Hz and 24-bit data resolution. The small and lightweight EEG amplifier (85 × 51 × 12 mm, 60gr) is tightly connected to a 24-channel electrode cap (Easycap, Germany) at the occipital site of the participant's head, using an elastic band. The connection between the EEG amplifier and recording computer was obtained using a Bluetooth connection, and the data were streamed to the described LSL recorder. The design of the cap-amplifier unit ensured minimal isolated movement of individual electrodes, cables, or the amplifier, which strongly reduced electromagnetic interference and movement artifacts. Further, the small dimensions of the recording system provided full mobility and comfort to the participants, as movement constraints were not imposed. The electrode cap contained sintered Ag/AgCl electrodes that were placed based on the international 10–20 System: Fp1, Fp2, Fz, F7, F8, FC1, FC2, Cz, C3, C4, T7, T8, CPz, CP1, CP2, CP5, CP6, TP9, TP10, Pz, P3, P4, O1 and O2. The electrodes were referenced to FCz and the ground electrode was AFz. Before initiation of the experiments, the experimental procedure imposed that the electrode impedances must be below the 5 kΩ value, which was confirmed by the device acquisition software.

ERP Processing

EEG signal processing was performed offline using EEGLAB (Delorme and Makeig, 2004) and MATLAB (Mathworks Inc.,

[1]https://github.com/sccn/SNAP
[2]https://code.google.com/p/labstreaminglayer/
[3]https://code.google.com/p/xdf/

Natick, MA, USA). EEG data were first bandpass filtered in the 1–35 Hz range, following which the signals were re-referenced to the average of the mastoid channels (Tp9 and Tp10). Further, an extended infomax Independent Component Analysis (ICA) was used to semi-automatically attenuate contributions from eye blink and (sometimes) muscle artifacts (as explained in Viola et al., 2009; De Vos et al., 2010, 2011). After this data preprocessing, ERP epochs were extracted from −200 to 800 ms with respect to timestamp values of "go" and "no-go" stimuli indicated by the SNAP software. Baseline values were corrected by subtracting mean values for the period from −200 to 0 ms from the stimuli. The identified electrode sites of interest for the ERP analysis in this study were Fz, Cz, CPz and Pz, as the P300 component is most prominent over the central and parieto-central scalp locations (Picton, 1992).

For the "no-go" condition we extracted and averaged the ERPs across the trials. For the "go" condition, the ERPs that preceded the "no-go" condition were calculated. Following these steps, the grand average (GA) ERPs across participants were formed. Further, the P300 amplitude was calculated for both "go" and "no-go" conditions and for each experimental condition, using mean amplitude measure (Luck, 2005) in the time window from 350 to 450 ms, with regard to the time stamps of the stimuli. Finally, the statistical analysis on the obtained results was carried out.

Reaction Times

As already stated in Section "Experimental Procedure", our experimental design did not allow subjects to react with the button press upon seeing the visual "go" stimulus. Therefore, the RT could not be measured in the traditional fashion, as the time elapsed between the stimulus presentation and the response by the participants (usually executed with the right index finger). Instead, the RTs here were measured as the time elapsed between the stimulus presentation (step 1) and the pedal press (step 6 from the Section "Preparation", also depicted on the **Figure 2A**). The pedal used in our study was actually a modified mouse button and it was connected to the recording computer via USB connection. As LSL is capable of real-time recording of the timestamps of the mouse button press, it enabled us to gather precise information regarding the time when pedal was pressed. This allows the calculation of RTs, as the difference between timestamps from stimulus presentation (operation initiation) and the beginning of the machine simulated crimping process.

Error Processing

Errors of omission were classified as the errors occurring when participants did not respond to the appearance of the "go" stimuli. The commission errors processing was challenging, since our task did not require a speeded button press and therefore, the errors of commission were difficult to interpret. In fact, the most obvious classification of commission errors would be when participants completely execute the simulated operation upon appearance of the "no-go" stimuli. However, it is important to note that participants sometimes made slight movements upon appearance of the "no-go" stimuli (in sense that they showed intention to initiate the action) and then they inhibited the response upon realization that it was a "no-go" stimulus. This kind of error we classified as near-misses. The identification of the near misses and commission errors was conducted initially by the experimenters in the room and subsequently confirmed in an off-line analysis, by replaying the videos recorded with the RGB camera during the experiment.

Statistical Analysis

The statistical analysis was performed using IBM SPSS software. The ERPs used for statistical analysis included all ERPs related to the "no-go" condition and 50 ERPs related to "go" preceding the "no-go" condition. The 4 × 2 × 2 × 2 repeated measures analysis of variance (ANOVA) was conducted with Site (Fz, Cz, CPz and Pz), Task (Arrows vs. Numbers) and Condition ("go/no-go") as within subject factors and Order of presentation (first vs. second) as between-subject factor. Additionally, a 2 × 2 ANOVA comparing RTs across Task (Arrows vs. Numbers) as within subject factors and Order of presentation (first vs. second) as between subject factor was conducted. Finally, we carried out a 2 × 2 ANOVA comparing near misses across Task (Arrows vs. Numbers) as within subject factors and Order of presentation (first vs. second) as between subject factor. Greenhouse-Geissser corrections (FG) were applied where necessary. Since the participants did not make any omission errors and only seven commission errors occurred across the participants they were exempted from further statistical analysis.

RESULTS

Behavioral Results

Reaction Times

The 2 × 2 ANOVA comparing RTs across Task (Arrow vs. SART) condition as within subject factor and Order of presentation (first vs. second) as between subject factor revealed neither significant main effects, nor interaction effects.

Errors

As stated in the "Materials and Methods" Section (Section "Statistical Analysis"), the participants did not make any omission errors and the low number of omission errors were not statistically analyzed. However, regarding near-misses, the ANOVA revealed only a significant effect of task ($F_{(1,8)} = 11.9$, $p < 0.01$, $\eta = 0.60$) with more near-misses occurring in the Numbers compared to the Arrows task.

ERP Results

The GA ERPs for each task (Arrows and Numbers), each condition ("go/no-go") and each electrode site under study (Fz, Cz, CPz and Pz) are depicted in **Figure 3**.

The 4 × 2 × 2 × 2 ANOVA revealed that the ERPs differed depending on the condition (Go/No-Go: $F_{(1,12)} = 5.99$, $p < 0.05$, $\eta = 0.33$), the task (Task: $F_{(1,12)} = 17.06$, $p < 0.001$, $\eta = 0.59$), the order of presentation (Order of presentation: $F_{(1,12)} = 15.635$, $p < 0.01$, $\eta = 0.57$) and across the scalp (Site: $F_{(1.48,17.75)} = 5.352$,

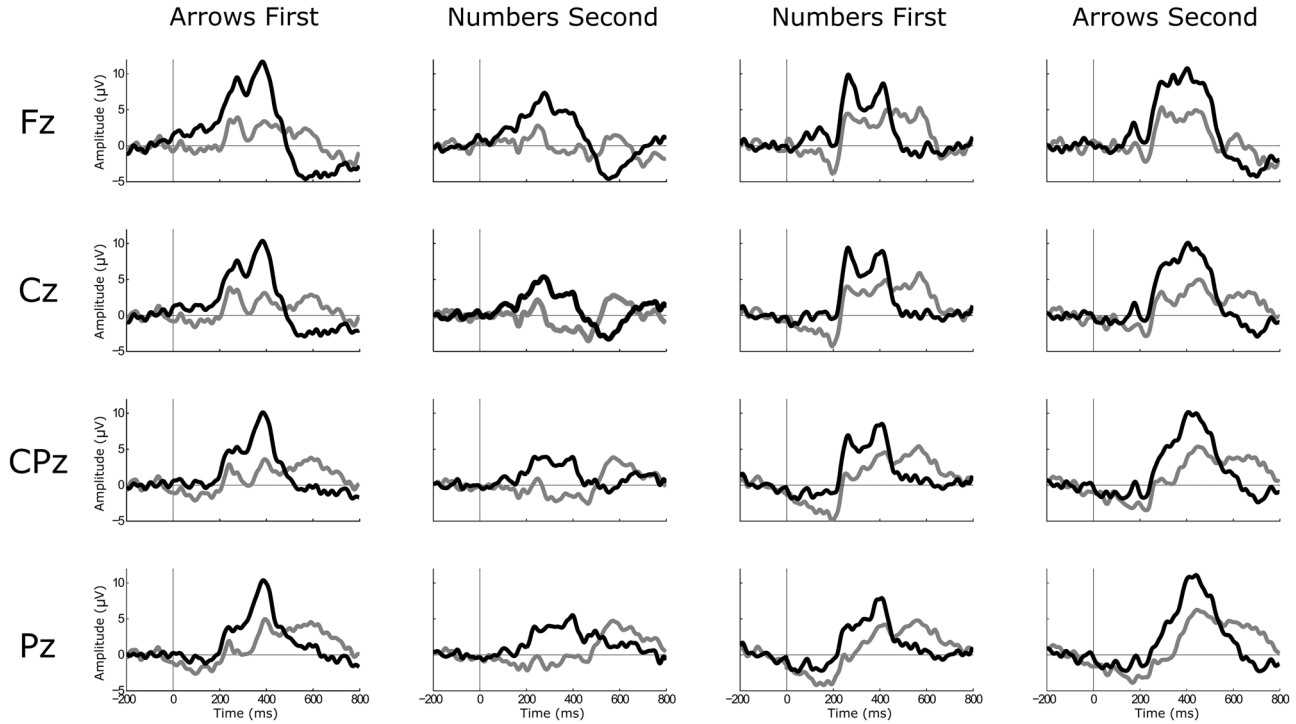


FIGURE 3 | Graphical representation of the grand average (GA) event-related potentials (ERPs) for each task and each electrode location under study. The black line represents the "go" condition, while the gray line depicts the GA ERPs for the "no-go" condition.

$p < 0.05$, $\eta = 0.31$). Namely, the P300 amplitudes elicited for "go" trials were higher than for "no-go" trials ($M = 5.73$, $SD = 1.47$; $M = 2.25$, $SD = 1.41$, respectively). Further, the Arrow task produced higher amplitudes in comparison to Numbers ($M = 5.24$, $SD = 1.11$; $M = 2.73$, $SD = 1.46$, respectively). The P300 amplitudes elicited with regard to the Order of presentation demonstrated higher amplitudes for whichever task was presented first in comparison to second task ($M = 5.11$, $SD = 1.31$; $M = 2.86$, $SD = 1.54$, respectively). Finally, amplitudes elicited at Pz were significantly higher than the amplitudes at the other three sites and amplitudes at CPz site were higher than at Cz and Fz sites at the $p < 0.05$ level. All the other comparisons were significant in the same direction apart from the Fz-Cz difference.

Figure 4 depicts the GA ERPs elicited over all four electrode sites under study for the "go" condition.

The P300 amplitude differences for all four sites and depending on the task representation order are presented in **Figure 5**.

DISCUSSION

The present study investigated whether operators' attention is enhanced when they are instructed with which hand to initiate the manual assembly operation, as compared to spontaneous and free choice of preferred hand. The attention was assessed through the P300 amplitude, as it is widely accepted that the P300 amplitude is positively related to the human level of attention (Ford et al., 1994; Polich, 2007; De Vos et al., 2014). For this aim we simulated a manual assembly operation, where we provided the participants with two distinct psychological tasks (Numbers and Arrows) simultaneously with the simulated operation.

The P300 components' amplitude was significantly higher in magnitude for the frequent "go" (target), than for the infrequent "no-go" condition (as presented on the **Figure 3**). This finding is in contrast to the majority of previously reported studies where an infrequent target condition elicits a higher magnitude of the P300 amplitude, since the participants are usually required to note the occurrence of infrequent targets by button press or by silent counting (Strüber and Polich, 2002). On the other hand, in our task target stimuli were the frequent ones, as the continuity of operation in manual assembly is essential, while the participants were instructed just to sit still and with no actions during the infrequent "no-go" condition. As such, it is not surprising that the lower magnitude of the P300 amplitude were elicited in infrequent non-target condition, as passive stimulus processing induces smaller P300 amplitudes than active tasks (Polich, 2007). This was also supported by the results from the study of Potts et al. (2004), where they reported that the P300 amplitude was larger in frequent "go" condition as compared to rare non-target condition in the task where the ratio between "go" and "no-go" condition was 80/20. Moreover, it was found that the ISI between target stimuli influences the P300 amplitude, in the sense that a short ISI leads to decreased amplitude, while relatively long ISIs elicit the

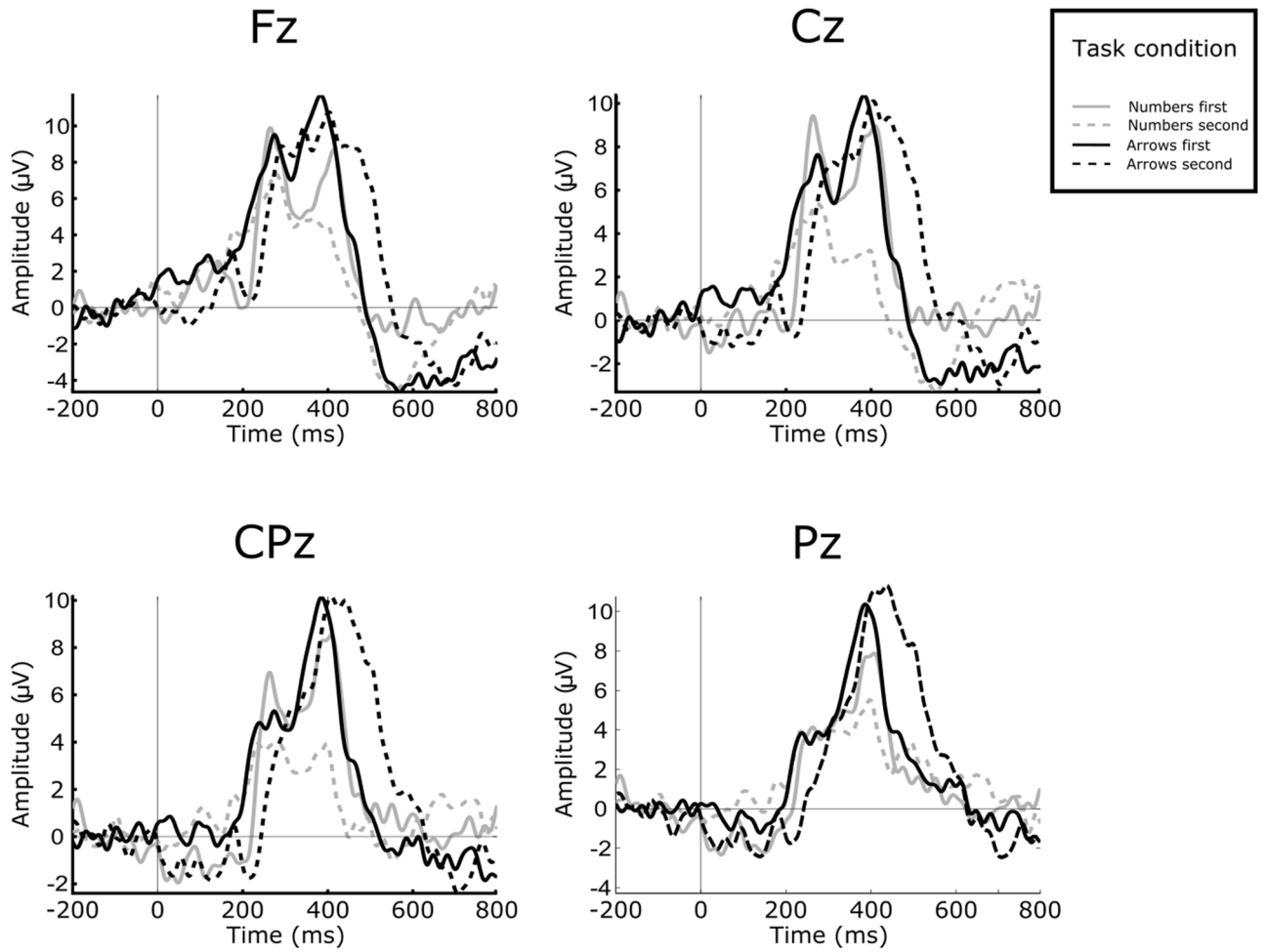


FIGURE 4 | The GA ERPs elicited for "go" condition in all four experimental conditions. ERPs elicited for The Numbers task are represented with the gray color, while the ERPs elicited in the Arrows task are depicted with the black color. The full line represents that the task was presented as a first task and the dashed line if the task was presented as second task.

higher P300 amplitude, which is the case even in the single-stimulus paradigm (Strüber and Polich, 2002; Polich, 2007). This was the case also in our study, since the ISI was relatively long (approximately 11 s) and we believe that it was suitable for eliciting the P300 amplitude even in the frequent target condition.

The main finding of the present study is that the P300 amplitude was considerably higher in magnitude when participants were instructed with which hand to initiate the simulated assembly operation, as compared to the case when participants could freely choose the preferred hand for the operation initiation. This may not be surprising, since in the choice reaction task (Arrows) participants were subjected to slightly higher demands of the incoming stimuli evaluation, as they were un-aware of the direction in which the white arrow stimuli would point. On the other hand, the digit stimulus carries considerably lower information, as participants are required just to make distinction whether it is a "go" or "no-go" stimulus and to perform their action accordingly, i.e., the participants may stop evaluating the content of the stimuli after some time. Therefore, the response selection requirements during the Arrows task are substantially higher than in Numbers task, which may lead to increased P300 amplitude in the condition which required instructed responding from the participants (Verleger et al., 2005, 2014). Following this finding, it may be proposed that the workers on repetitive and monotonous assembly tasks should not receive information solely on whether they should initiate the operation or not, but it should be beneficial if they receive information that carries slightly higher cognitive demands. In fact, the task that consisted of the stimuli with the higher cognitive demands induced the higher P300 amplitude, which may be related to the attention of the worker for the task in hand. An important notion, however, is that there is possibility that the P300 amplitude in this study does not reflect solely the attention level of a worker, but it also may be influenced by the different cognitive demands of the tasks. For that reason, it is important to further investigate whether the P300 amplitude was influenced by the presented

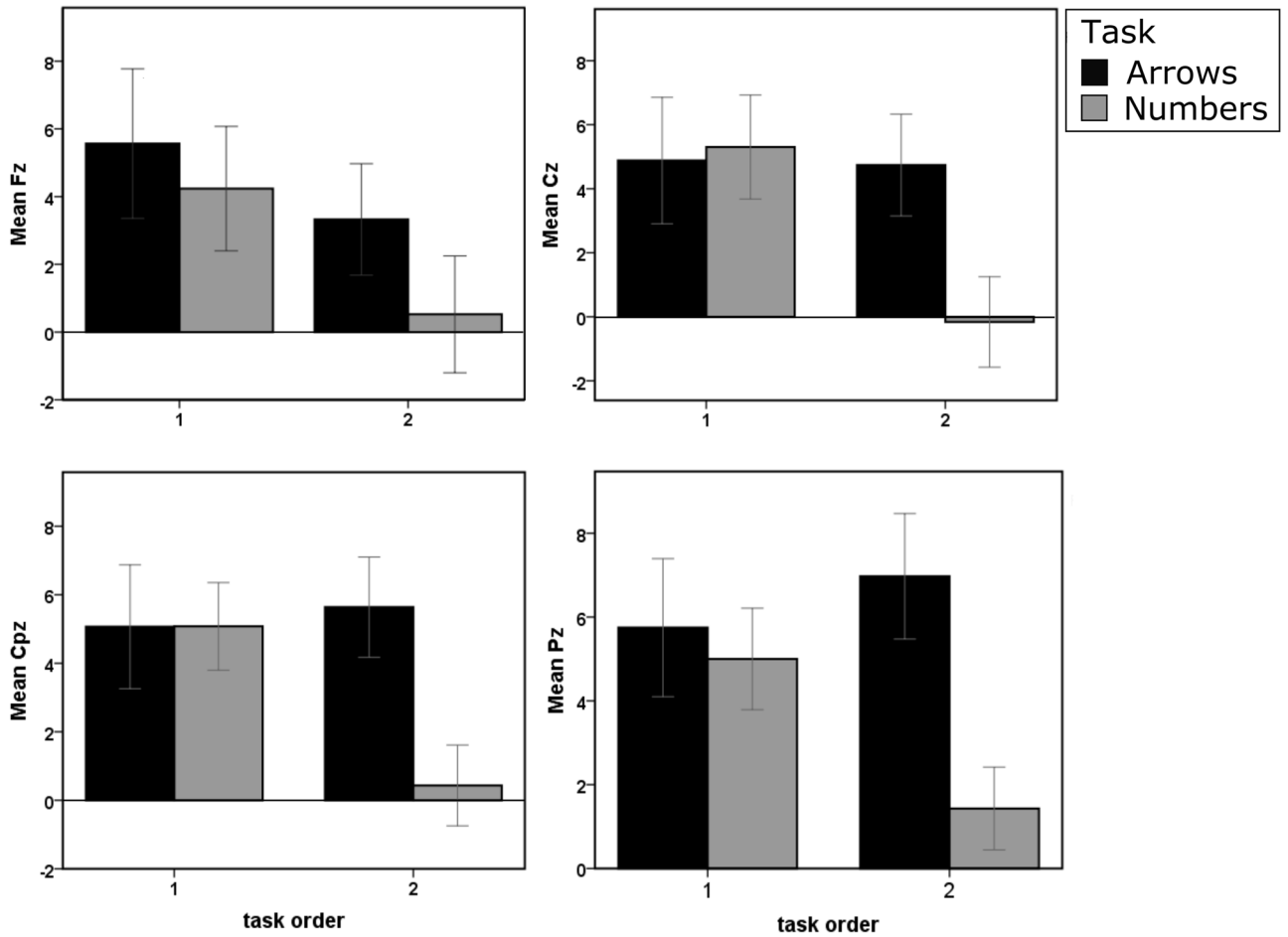


FIGURE 5 | The amplitude values for all four electrode sites and for all experimental conditions. The black color depicts the Arrows task, while the Numbers task is represented with the gray color. The error bars represents ±2 SE.

task demands, or it was solely related to the attention of the workers.

Interestingly, although it was expected that the RTs could differ between the two tasks, this was not the case in our study. One of the possible reasons for the absence of the response time effect could be the methodology used for the RTs calculation. In fact, the time period for RT calculation is much longer than in the conventional studies, where a speeded response from the participants is expected. Apart from that, the RT calculation includes several coordinated hand movements before the foot switch is pressed. All of these could induce a large variation within and between subject conditions, which may induce inaccuracy of the RT methodology used in this study. Further, with regard to behavioral measurements, the number of commission errors was relatively low and did not differ between the tasks. However, there was significantly higher amount of near-misses in the Numbers than in the Arrows task. The fact that there was larger number of near-misses in the Numbers task may be expected, as the Arrows task imposes a higher workload on the participants, due to the higher response selection requirements, and as it was previously reported, the errors and mental workload are related according to a U-shaped curve (Desmond and Hoyes, 1996).

Although we showed that the Arrows task produced a higher P300 amplitude than the Numbers task, one could argue about the selection of the tasks, as the stimulus type between task conditions significantly differed (digits vs. arrows). The main reason for not investigating the difference between instructed and non-instructed condition with the same type of stimuli was the avoidance of the interference effect (Pashler, 1994). In fact, if only stimuli from Numbers task were used and dedicated the directions to specific digits in the hand instructing task (e.g., odd numbers means left and even numbers right hand first), it would be highly likely that the memory would strongly influence the attention processing. On the other hand, if we only used the Arrows stimuli type, an undesired bias would be included in the condition when participants could initiate the operation with their preferred hand. An additional concern is whether the two distinct psychological tests trigger different

attentional resources, given that they are composed of different stimulus types and that the Arrows task alternates the response hand, while in the Numbers task participants could respond with whichever hand they preferred. The answer to this doubt could be found in premotor theory of attention (Rizzolatti et al., 1994), which states that attention orienting processes are triggered during uni-manual response preparation and that the orienting processes are assumed to be equivalent to the processes elicited during instructed endogenous shifts of spatial attention (Eimer et al., 2005). Moreover, Ranzini et al. (2009) also used the tasks with Arabic digits and Arrows and they demonstrated that processes evoked by these cues are alike and that the volitional and non-volitional attentional shifts rely on the same fronto-parietal brain networks. Thus, both Numbers and Arrows tasks should evoke the same cognitive resources of attention, which gives legitimacy to the choice of the tasks used in this study.

One of the limitations of the present study is that it was conducted in a simulated working environment, instead of a real factory setting. The main reason for this was usage of the wet-electrode EEG recording system, which is still uncomfortable for application in actual industrial environments. Nevertheless, we replicated both the spatial dimensions and ambient conditions and performed the wearable EEG study, demonstrating its applicability for the investigation of covert cognitive processes in naturalistic environments for HF/E studies. Another limitation is that, simultaneously with the simulated operation, we used two distinct psychological tests, with the aim of eliciting the P300 ERP component. Although it could be argued that psychological tests could interfere with the simulated operation, an important notion is that the assembly workers should be provided with timely information regarding the performed operation (Stork and Schubö, 2010). Therefore, we believe that this modification did not significantly differ from the actual assembly operation in industrial environments. Moreover, in naturalistic settings it is usually hard to isolate and analyze the specific cognitive process, since they should first be evoked and co-occurring cognitive factors should be isolated (Bulling and Zander, 2014). Thus, this modification in the information presentation to the participants was necessary in order to elicit the anticipated P300 ERP component during the simulated assembly operation. Unfortunately, the present study is unable to compare brain responses between self-paced (as in this specific workplace) and externally paced work routines that we used in our study. This issue should be addressed in future studies.

The present study demonstrated that wearable EEG recording could be beneficial for task design in HF/E studies. Future studies should investigate whether the reported findings also hold for similar job positions, which are monotonous and repetitive in nature but require continuous focus of the worker on the industrial task (e.g., quality control tasks). Although the present study utilized wearable EEG in a faithfully replicated workplace environment, it seems that it is just a matter of time until EEG systems will be willingly accepted for everyday use (van Erp et al., 2012; Mihajlovic et al., 2015). This could even lead to the application of passive brain-computer interfaces, which could be used for real-time assessment of the cognitive user states in industrial environments (Zander and Kothe, 2011). Nevertheless, the fact that it is nowadays possible to investigate brain dynamics during natural movements (without imposing movements constraints) of the recorded individual brings us a step closer to the guiding principle of the neuroergonomics, that is, to investigate how the brain carries out the complex tasks of everyday life and not just simplified and artificial tasks in the laboratory settings (Parasuraman and Rizzo, 2006).

CONCLUSION

Comparing monotonous ("go/no-go") Numbers task to the choice-reaction (Arrows) task, which instructs the participants with which hand to commence the assembly operation, the present study indicates that the latter is more suitable to preserve participants' attention during the initiation of externally-paced assembly task. This finding was achieved through investigation of the ERP waveform, where it was found that the P300 amplitude, which is related to the level of attention, was enhanced in the task that instructed the participants with which hand to initiate the simulated assembly operation. This study demonstrated the potential benefits of introducing the EEG measurements in the industrial task design, as from the presented results it may be concluded that in in monotonous assembly tasks, instructed responding, or a similar method of engagement, should be imposed on operators, since it is indicated that additional engagement enhances the worker's attention.

AUTHOR CONTRIBUTIONS

Study design and protocols were created by PM, VK, MDeV, BJ, IM and IG. Data acquisition was performed by PM and IG. Data analysis was performed by PM, VK, MDeV, IM and IG. The data interpretation was performed by PM, VK, BJ, IM and IG. The manuscript was written by PM and VK and critical editing was performed by MDeV, BJ, IM and IG. Final approval of the version was obtained from all co-authors. All co-authors agree on all aspects of the work and ensure that questions related to the accuracy and integrity of any part of the submitted work are appropriately investigated and resolved.

ACKNOWLEDGMENTS

This research is financed under EU—FP7 Marie Curie Actions Initial Training Networks—FP7-PEOPLE-2011-ITN, project name "Innovation Through Human Factors in Risk Analysis and Management (InnHF)", project number: 289837. We would further like to acknowledge company "Gomma Line" (Serbia), for their assistance and advisory during the experimental setup phase. We would like to acknowledge Nora Balfe from Center for Innovative Human Systems (CIHS), Trinity College Dublin, Department for Psychology, for the critical comments of the manuscript.

REFERENCES

Ayaz, H., Onaral, B., Izzetoglu, K., Shewokis, P. A., McKendrick, R., and Parasuraman, R. (2013). Continuous monitoring of brain dynamics with functional near infrared spectroscopy as a tool for neuroergonomic research: empirical examples and a technological development. *Front. Hum. Neurosci.* 7:871. doi: 10.3389/fnhum.2013.00871

Ayaz, H., Shewokis, P. A., Bunce, S., Izzetoglu, K., Willems, B., and Onaral, B. (2012). Optical brain monitoring for operator training and mental workload assessment. *Neuroimage* 59, 36–47. doi: 10.1016/j.neuroimage.2011.06.023

Ayaz, H., Willems, B., Bunce, B., Shewokis, P. A., Izzetoglu, K., Hah, S., et al. (2010). "Cognitive workload assessment of air traffic controllers using optical brain imaging sensors," in *Advances in Understanding Human Performance: Neuroergonomics, Human Factors Design and Special Populations*, eds T. Marek, W. Karwowski, and V. Rice (Boca Raton, FL: CRC press), 21–31.

Bigdely-Shamlo, N., Kreutz-Delgado, K., Robbins, K., Miyakoshi, M., Westerfield, M., Bel-Bahar, T., et al. (2013). "Hierarchical event descriptor (HED) tags for analysis of event-related EEG studies," in *Global Conference on Signal and Information Processing (GlobalSIP)*, Austin, TX, 1–4.

Bulling, A., and Zander, T. O. (2014). Cognition-aware computing. *IEEE Perv. Comp.* 13, 80–83. doi: 10.1109/mprv.2014.42

De Vos, M., De Lathauwer, L., and Van Huffel, S. (2011). Spatially constrained ICA algorithm with an application in EEG processing. *Sign. Process.* 91, 1963–1972. doi: 10.1016/j.sigpro.2011.02.019

De Vos, M., Gandras, K., and Debener, S. (2014). Towards a truly mobile auditory brain-computer interface: exploring the P300 to take away. *Int. J. Psychophysiol.* 91, 46–53. doi: 10.1016/j.ijpsycho.2013.08.010

De Vos, M., Riès, S., Vanderperren, K., Vanrumste, B., Alario, F. X., van Huffel, S., et al. (2010). Removal of muscle artifacts from EEG recordings of spoken language production. *Neuroinformatics* 8, 135–150. doi: 10.1007/s12021-010-9071-0

Debener, S., Minow, F., Emkes, R., Gandras, K., and de Vos, M. (2012). How about taking a low-cost, small and wireless EEG for a walk? *Psychophysiology* 49, 1617–1621. doi: 10.1111/j.1469-8986.2012.01471.x

Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21. doi: 10.1016/j.jneumeth.2003.10.009

Desmond, P. A., and Hoyes, T. W. (1996). Workload variation, intrinsic risk and utility in a simulated air traffic control task: evidence for compensatory effects. *Saf. Sci.* 22, 87–101. doi: 10.1016/0925-7535(96)00008-2

Dockree, P. M., Kelly, S. P., Robertson, I. H., Reilly, R. B., and Foxe, J. J. (2005). Neurophysiological markers of alert responding during goal-directed behavior: a high-density electrical mapping study. *Neuroimage* 27, 587–601. doi: 10.1016/j.neuroimage.2005.05.044

Donchin, E. (1981). Surprise!. . . surprise? *Psychophysiology* 18, 493–513. doi: 10.1111/j.1469-8986.1981.tb01815.x

Donkers, F. C., and van Boxtel, G. J. (2004). The N2 in go/no-go tasks reflects conflict monitoring not response inhibition. *Brain Cogn.* 56, 165–176. doi: 10.1016/j.bandc.2004.04.005

Eimer, M., Forster, B., Van Velzen, J., and Prabhu, G. (2005). Covert manual response preparation triggers attentional shifts: ERP evidence for the premotor theory of attention. *Neuropsychologia* 43, 957–966. doi: 10.1016/j.neuropsychologia.2004.08.011

Ford, J. M., Sullivan, E. V., Marsh, L., White, P. M., Lim, K. O., and Pfefferbaum, A. (1994). The relationship between P300 amplitude and regional gray matter volumes depends upon the attentional system engaged. *Electroencephalogr. Clin. Neurophysiol.* 90, 214–228. doi: 10.1016/0013-4694(94)90093-0

Gramann, K., Ferris, D. P., Gwin, J., and Makeig, S. (2014). Imaging natural cognition in action. *Int. J. Psychophysiol.* 91, 22–29. doi: 10.1016/j.ijpsycho.2013.09.003

Gramann, K., Gwin, J. T., Ferris, D. P., Oie, K., Jung, T. P., Lin, C. T., et al. (2011). Cognition in action: imaging brain/body dynamics in mobile humans. *Rev. Neurosci.* 22, 593–608. doi: 10.1515/RNS.2011.047

Johnson, R. (1988). "The amplitude of the P300 component of the event-related potential: review and synthesis," in *Advances in Psychophysiology: A Research Annual*, (Vol. 3), eds P. Ackles, J. R. Jennings and M. G. H. Coles (Greenwich, CT.: JAI press), 69–137.

Kletz, T. A. (2001). *An Engineer's View of Human Error.* (Vol. 3), Rugby: Institution of Chemical Engineers.

Luck, S. J. (2005). *An Introduction to the Event-Related Potential Technique.* Cambridge, MA: MIT Press.

Mehta, R. K., and Parasuraman, R. (2013). Neuroergonomics: a review of applications to physical and cognitive work. *Front. Hum. Neurosci.* 7:889. doi: 10.3389/fnhum.2013.00889

Michalos, G., Makris, S., Papakostas, N., Mourtzis, D., and Chryssolouris, G. (2010). Automotive assembly technologies review: challenges and outlook for a flexible and adaptive approach. *CIRP J. Manuf. Sci. Tech.* 2, 81–91. doi: 10.1016/j.cirpj.2009.12.001

Mihajlovic, V., Grundlehner, B., Vullers, R., and Penders, J. (2015). Wearable, wireless EEG Solutions in daily life applications: what are we missing? *IEEE J. Biomed. Health Inform.* 19, 6–21. doi: 10.1109/jbhi.2014.2328317

Mijović, P., Ković, V., De Vos, M., Ma čužić, I., Todorović, P., Jeremić, B., et al. (2016). Towards continuous and real-time attention monitoring at work: reaction times versus brain response. *Ergonomics* doi: 10.1080/00140139.2016.1142121 [Epub ahead of print].

Mijović, P., Ković, V., Ma čužić, I., Jeremić, B., Todorović, P., Milovanović, M., et al. (2015a). "Do micro-breaks increase the attention level of an assembly worker? An ERP study," in *Proceedings of 6th International Conference of Applied Human Factors and Ergonomics (AHFE 2015)*, Las Vegas, NV.

Mijović, P., Milovanović, M., Minović, M., Ma čužić, I., Ković, V., and Gligorijević, I. (2015b). "Towards creation of implicit HCI model for prediction and prevention of operators' error," in *Human-Computer Interaction: Interaction Technologies*, ed. M. Kurosu (Cham: Springer International Publishing), 341–352.

Parasuraman, R. (1990). "Event-related brain potentials and human factors research," in *Event-Related Brain Potentials: Basic Issues and Applications*, eds J. Rohbaugh, R. Parasuraman, and R. Johnson (New York, NY: Oxford University Press), 279–300.

Parasuraman, R. (2003). Neuroergonomics: research and practice. *Theor. Issues Ergon. Sci.* 4, 5–20. doi: 10.1080/14639220210199753

Parasuraman, R. (2011). Neuroergonomics brain, cognition and performance at work. *Curr. Dir. Psychol. Sci.* 20, 181–186. doi: 10.1177/0963721411409176

Parasuraman, R., and Rizzo, M. (Eds.) (2006). *Neuroergonomics: The Brain at Work.* New York, NY: Oxford University Press.

Parasuraman, R., and Wilson, G. F. (2008). Putting the brain to work: neuroergonomics past, present and future. *Hum. Factors* 50, 468–474. doi: 10.1518/001872008x288349

Pashler, H. (1994). Dual-task interference in simple tasks: data and theory. *Psych. Bull.* 116, 220–244. doi: 10.1037/0033-2909.116.2.220

Picton, T. W. (1992). The P300 wave of the human event-related potential. *J. Clin. Neurophysiol.* 9, 456–479. doi: 10.1097/00004691-199210000-00002

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clin. Neurophysiol.* 118, 2128–2148. doi: 10.1016/j.clinph.2007.04.019

Polich, J., and Kok, A. (1995). Cognitive and biological determinants of P300: an integrative review. *Biol. Psychol.* 41, 103–146. doi: 10.1016/0301-0511(95)05130-9

Potts, G. F., Patel, S. H., and Azzam, P. N. (2004). Impact of instructed relevance on the visual ERP. *Int. J. Psychophysiol.* 52, 197–209. doi: 10.1016/j.ijpsycho.2003.10.005

Ranzini, M., Dehaene, S., Piazza, M., and Hubbard, E. M. (2009). Neural mechanisms of attentional shifts due to irrelevant spatial and numerical cues. *Neuropsychologia* 47, 2615–2624. doi: 10.1016/j.neuropsychologia.2009.05.011

Rizzolatti, G., Riggio, L., and Sheliga, B. M. (1994). Space and selective attention. *Atten. Perform. XV* 15, 231–265.

Robertson, I. H., Manly, T., Andrade, J., Baddeley, B. T., and Yiend, J. (1997). Oops!': performance correlates of everyday attentional failures in traumatic brain injured and normal subjects. *Neuropsychologia* 35, 747–758. doi: 10.1016/s0028-3932(97)00015-8

Scerbo, M. (2006). "Adaptive automation," in *Neuroergonomics: The Brain at Work*, eds R. Parasuraman and M. Rizzo (New York, NY: Oxford University Press), 239–252.

Sheridan, T. B., and Parasuraman, R. (2005). Human-automation interaction. *Rev. Hum. Fact. Ergon.* 1, 89–129. doi: 10.1518/155723405783703082

Spath, D., Braun, M., and Meinken, K. (2012). "Human factors in manufacturing," in *Handbook of Human Factors and Ergonomics*, ed. G. Salvendy (New Jersey, NJ: John Wiley & Sons), 1643–1666.

Stork, S., and Schubö, A. (2010). Human cognition in manual assembly: theories and applications. *Adv. Eng. Inform.* 24, 320–328. doi: 10.1016/j.aei.2010.05.010

Strüber, D., and Polich, J. (2002). P300 and slow wave from oddball and single-stimulus visual tasks: inter-stimulus interval effects. *Int. J. Psychophysiol.* 45, 187–196. doi: 10.1016/s0167-8760(02)00071-5

Tang, A., Owen, C., Biocca, F., and Mou, W. (2003). "Comparative effectiveness of augmented reality in object assembly," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. (New York, NY: ACM), 73–80.

Tompkins, J. A., White, J. A., Bozer, Y. A., and Tanchoco, J. M. A. (2010). *Facilities Planning.* New Jersey, NJ: John Wiley & Sons.

van Erp, J. B., Lotte, F., and Tangermann, M. (2012). Brain-computer interfaces: beyond medical applications. *IEEE Comput.* 45, 26–34. doi: 10.1109/mc.2012.107

Verleger, R., Baur, N., Metzner, M. F., and Śmigasiewicz, K. (2014). The hard oddball: effects of difficult response selection on stimulus-related P3 and on response-related negative potentials. *Psychophysiology* 51, 1089–1100. doi: 10.1111/psyp.12262

Verleger, R., Ja śkowski, P., and Wascher, E. (2005). Evidence for an integrative role of P3b in linking reaction to perception. *J. Psychophysiol.* 19, 165–181. doi: 10.1027/0269-8803.19.3.165

Viola, F. C., Thorne, J., Edmonds, B., Schneider, T., Eichele, T., and Debener, S. (2009). Semi-automatic identification of independent components representing EEG artifact. *Clin. Neurophysiol.* 120, 868–877. doi: 10.1016/j.clinph.2009.01.015

Warm, J. S., Matthews, G., and Parasuraman, R. (2009). Cerebral hemodynamics and vigilance performance. *Mil. Psych.* 21, S75–S100. doi: 10.1080/08995600802554706

Warm, J. S., and Parasuraman, R. (2006). "Cerebral hemodynamics and vigilance," in *Neuroergonomics: The Brain at Work*, eds R. Parasuraman and M. Rizzo (New York, NY: Oxford University Press), 146–158.

Warm, J. S., Parasuraman, R., and Matthews, G. (2008). Vigilance requires hard mental work and is stressful. *Hum. Factors* 50, 433–441. doi: 10.1518/001872008x312152

Wascher, E., Heppner, H., and Hoffmann, S. (2014). Towards the measurement of event-related EEG activity in real-life working environments. *Int. J. Psychophysiol.* 91, 3–9. doi: 10.1016/j.ijpsycho.2013.10.006

Zander, T. O., and Kothe, C. (2011). Towards passive brain-computer interfaces: applying brain-computer interface technology to human-machine systems in general. *J. Neural Eng.* 8:025005. doi: 10.1088/1741-2560/8/2/025005

Conflict of Interest Statement: IG is associated with the mBrainTrain Company, supplier of wireless EEG system "SMARTING" used in this study. However, no financial or other conflicting interests arise from this fact. The other authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Mijović, Ković, De Vos, Mačužić, Jeremić and Gligorijević. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution and reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Human Neuroscience

METHODS
published: 25 April 2016
doi: 10.3389/fnhum.2016.00170

Pre-Trial EEG-Based Single-Trial Motor Performance Prediction to Enhance Neuroergonomics for a Hand Force Task

*Andreas Meinel[1], Sebastián Castaño-Candamil[1], Janine Reis[2] and Michael Tangermann[1]**

[1] Brain State Decoding Lab, Cluster of Excellence BrainLinks-BrainTools, Department of Computer Science, Albert-Ludwigs-University, Freiburg, Germany, [2] Department of Neurology, Albert-Ludwigs-University, Freiburg, Germany

We propose a framework for building electrophysiological predictors of single-trial motor performance variations, exemplified for SVIPT, a sequential isometric force control task suitable for hand motor rehabilitation after stroke. Electroencephalogram (EEG) data of 20 subjects with mean age of 53 years was recorded prior to and during 400 trials of SVIPT. They were executed within a single session with the non-dominant left hand, while receiving continuous visual feedback of the produced force trajectories. The behavioral data showed strong trial-by-trial performance variations for five clinically relevant metrics, which accounted for reaction time as well as for the smoothness and precision of the produced force trajectory. 18 out of 20 tested subjects remained after preprocessing and entered offline analysis. Source Power Comodulation (SPoC) was applied on EEG data of a short time interval prior to the start of each SVIPT trial. For 11 subjects, SPoC revealed robust oscillatory EEG subspace components, whose bandpower activity are predictive for the performance of the upcoming trial. Since SPoC may overfit to non-informative subspaces, we propose to apply three selection criteria accounting for the meaningfulness of the features. Across all subjects, the obtained components were spread along the frequency spectrum and showed a variety of spatial activity patterns. Those containing the highest level of predictive information resided in and close to the alpha band. Their spatial patterns resemble topologies reported for visual attention processes as well as those of imagined or executed hand motor tasks. In summary, we identified subject-specific single predictors that explain up to 36% of the performance fluctuations and may serve for enhancing neuroergonomics of motor rehabilitation scenarios.

Keywords: single-trial performance prediction, trial-by-trial variability, isometric force modulation, hand motor rehabilitation, visuomotor integration, EEG, oscillatory subspace, spatial filtering

OPEN ACCESS

Edited by:
Klaus Gramann, Berlin Institute of Technology, Germany

Reviewed by:
Sara L. Gonzalez Andino, Hôpitaux Universitaires de Genève, Switzerland
Floriana Pichiorri, Fondazione Santa Lucia, IRCCS, Italy
Jörn M. Horschig, Artinis Medical Systems BV, Netherlands

****Correspondence:***
Michael Tangermann
michael.tangermann@blbt.uni-freiburg.de

Received: *15 December 2015*
Accepted: *04 April 2016*
Published: *25 April 2016*

Citation:
Meinel A, Castaño-Candamil S, Reis J and Tangermann M (2016) Pre-Trial EEG-Based Single-Trial Motor Performance Prediction to Enhance Neuroergonomics for a Hand Force Task. Front. Hum. Neurosci. 10:170. doi: 10.3389/fnhum.2016.00170

1. INTRODUCTION

Motor training is utilized in rehabilitation scenarios to accelerate the re-gain of lost motor function after brain injury. State-of-the-art rehabilitation concepts are based on repetitive training tasks with the aim to reach a functional gain (Dobkin, 2004; Timmermans et al., 2009; Langhorne et al., 2011). Most prominent training paradigms comprise mirror training (French et al., 2007),

constraint-induced movement therapy (Wolf et al., 2002), simultaneous bilateral training (Coupar et al., 2010), BCI-supported training (Ang and Guan, 2013) and robot-assisted techniques (Kwakkel et al., 2008). Recent rehabilitation approaches include the training of novel, unfamiliar motor skills instead of training well-known habitual motor tasks, attempting to optimize functional cortical reorganization.

Repetitive paradigms allow for the assessment of motor performance on a very fine-granular time scale. The performance of each single trial can be monitored by metrics such as the length, speed or smoothness of the produced movement trajectory. The distributions and temporal characteristics of trial-wise motor performance variations have been studied by different groups (Abe and Sternad, 2013; van Beers et al., 2013; Wu et al., 2014; Hadjiosif and Smith, 2015). While practicing a motor task over several sessions enables a user for skill acquisition (Lage et al., 2015), trial-by-trial variability of motor performance is a prominent feature which does not fully vanish with training (Cohen and Sternad, 2009; Osu et al., 2015). The underlying neural mechanisms of motor performance fluctuations on short time scales is subject of controversial discussion in literature and is not fully resolved yet (Faisal et al., 2008; Hadjiosif and Smith, 2015; Osu et al., 2015).

In the present work, we aim toward closing this gap. Therefore, trial-wise performance fluctuations of a sequential visuo-motor task (SVIPT; Camus et al., 2009; Reis et al., 2009; Fritsch et al., 2010 are investigated while registering a user's brain activity by EEG. In SVIPT trials, the quality of a movement changes within seconds and from repetition to repetition.

Our *hypothesis* is that subject-specific pre-trial brain signals can partially explain and temporally predict the trial-by-trial fluctuations of the upcoming motor performance. Given that such informative neural markers exist, then the SVIPT paradigm could be altered in order to meet the cognitive ergonomic requirements of each single user. Practically, the starting time point of the upcoming trial can be determined based on the information contained in this pre-trial neural marker. Ideally, such a neuroergonomic closed-loop gating strategy could provide control over the level of difficulty. This should allow to causally influence user performance and ultimately support SVIPT motor learning on the long run.

Paradigms which include *brain-state-dependent experimenting* (see Jensen et al., 2011; Horschig et al., 2014c) require that an informative neural marker can be extracted robustly from brain signals. Given the high dimensionality and noisy characteristic of most types of brain signals, the extraction and decoding of such individual neural markers is a challenging task.

Screening literature on relevant neural markers of visual and motor performance, it is important to make a distinction between the use of single-trial decoding in contrast to the extraction of statistical differences, which may even be reported as group averages. Neural features which correlate with the task performance on the grand average (GA) of a set of subjects have limited usefulness for closed-loop experimenting with a given individual. As inter-subject differences get lost during the averaging, GA features may have low predictive power when tested with data of a novel subject. Research in the field of brain-computer interfaces (BCI) has pushed forward methods for single-trial decoding of individual brain activity (mostly EEG signals) (Millán et al., 2010; Makeig et al., 2012). Results from this field affirm that brain signals and informative features vary strongly between individuals (Müller et al., 2008). To obtain optimal decoding results, BCI data processing pipelines thrive to identify subject-specific informative features. Technically, these are gained either from a calibration recording prior to the online use of the BCI (Blankertz et al., 2007), or by transfer learning methods (Kindermans et al., 2014) which exploit features from pre-trained machine learning models of earlier sessions or previous users. Furthermore, attention needs to be paid to temporal dependencies: brain features may correlate with previous behavior, with simultaneous behavior or may even be predictive for future behavior. Only the latter brain features can serve as a tool for brain-state-dependent experimenting.

Statistical correlates of visual perception performance are reported by several groups. For stimuli near the perception threshold, the pre-stimulus occipital alpha bandpower correlates with the detection performance (van Dijk et al., 2008), even on a single-trial basis using predictive features (Hanslmayr et al., 2007). In addition to bandpower, the pre-stimulus alpha phase was reported to correlate with the detection performance (Busch et al., 2009). Single-trial decoding methods were not applied in those auditory studies, but the reported correlates precede the perception, which may open the possibility for closed-loop experimenting. Based on the findings of Hanslmayr et al. (2007) and van Dijk et al. (2008), there are two examples that set up an online experiment based on occipital alpha bandpower features. Tonin et al. (2013) using EEG data and Horschig et al. (2014b), who employed MEG signals, both decoded covert visual attention in a closed-loop experiment by utilizing single-trial feedback on the detected attention shift. However, both groups did not fully close the loop e.g., by manipulating the perception performance, which may have been possible by selecting suitable brain states for stimulation. Gonzalez Andino et al. (2005) studied a cued reaction time task and identified that gamma band oscillatory activity observed in fronto-parietal regions prior to the stimulus onset correlates with reaction time. Similarly, Hoogenboom et al. (2010) stated that the strength of visually induced gamma band activity is predictive for the detection of stimulus motion. Somatosensory stimuli of low-intensity, but above threshold were delivered and combined with a distracting masker stimulus by Schubert et al. (2009). Investigating perceived vs. missed stimuli in an offline analysis, pre-stimulus beta bandpower over the left frontal cortex was found predictive for the perception performance on the grand average, as well as mu and beta bandpower over the pericentral sensorimotor areas.

In the motor domain, several groups have successfully decoded hand kinematics, using the center-out task as the dominating experimental approach. In their own work, Jerbi et al. (2011) provide a review over the decoding of hand movement parameters such as direction, position and velocity based on brain signals. ECoG signals were used by Pistohl et al. (2008) to decode two-dimensional hand movement trajectories using an autoregressive filtering approach. Considering non-invasive

techniques, Waldert et al. (2008) have decoded (but not temporally predicted) the hand movement direction based on MEG and EEG. Neural correlates which encode the velocity of a movement have been investigated by Bradberry et al. (2009). The decoding of produced grip force based on a phase feature extracted from the beta range has been reported on data of three subjects by Logar et al. (2008). Zaepffel et al. (2013) reported an increased centro-parietal beta power during the planning period of grasping movements, but it was not investigated, if decoding may work on the basis of single trials. Focusing on single-trial methods, Lew et al. (2014) used slow cortical potentials of the EEG from fronto-parietal areas to predict self-paced movement directions a few hundred milliseconds prior to movement onset. Similarly, Hammon et al. (2008) inspected predictive EEG features for planning target directions using a cue-based paradigm.

In the field of BCI research, Maeder et al. (2012) studied a motor imagery paradigm. The single-trial decoding performance of left vs. right hand movement imagery tasks could be correlated to the level of pre-trial alpha bandpower over the sensorimotor cortices. Despite used offline, this neural marker would allow for a predictive intervention in a closed loop. In their statistical analysis, Yang et al. (2014) identified frontal alpha and beta bandpower features which correlate with performance metrics of a reaching task. Proceeding to single-trial methods, Meyer et al. (2014) reported on data of six subjects, who performed a hand positioning task. Their offline analysis revealed that the normalized time-to-target could be predicted based on pre-cue alpha-band activity of the EEG.

The state-of-the-art can be summed up as follows: In the perception domain, several studies have established single-trial performance prediction, partially even in closed-loop applications. The situation is different for the motor domain since only very few studies have investigated subject-specific motor performance prediction in single-trial upon a sufficiently large subject group. Closest to all of these requirements is the study by Meyer et al. (2014). Our research hypothesis builds exactly upon this point. In the context of a hand force task, we propose a generalized *workflow* which identifies subject-specific *predictive* oscillatory EEG features evaluated on a single trial basis.

First, by means of a simulated online analysis, an approach to extract robust and meaningful EEG components is developed. We evaluate, if the information contained in selected components is able to partially explain the trial-by-trial variation of SVIPT performance in a predictive fashion, i.e., the pre-trial component is required to predict the outcome of the upcoming trial. Second, the characteristics of the best performance predictors are investigated by a group-level analysis.

2. MATERIALS AND METHODS

2.1. Hand Motor Paradigm

In the context of hand motor skill learning, Reis et al. (2009) investigated the *Sequential Visual Isometric Pinch Task* (SVIPT), which demands an isometric force control of thumb and index finger. Interestingly, training-induced improvement of the SVIPT generalizes well to other hand motor control tasks, even though pinch grasp activities are rarely displayed during natural behavior patterns. Compared to the original SVIPT setup, brain activity is recorded using electroencephalogram (EEG) during a training session for *post-hoc* offline analysis. The resulting SVIPT setup follows the proposal in Meinel et al. (2015) and is sketched in **Figure 1**.

Each SVIPT trial consists of three phases: a light blue (inactive) cursor appears on the leftmost edge of the T0 field (corresponding to zero force), while the user is touching the sensor only slightly with his non-dominant left hand. The appearance of the cursor indicates the start of the *get-ready* phase, which corresponds to a waiting period with enhanced attention level. Its duration is varied randomly between 2 and 3 s. The transduction of force into cursor movements is deactivated during the get-ready phase. Fixating the cursor, the user will observe a distinct color change of the cursor from light to dark blue. This *go-cue* indicates the beginning of the *running* phase, in which the cursor position can be controlled by applying force to the sensor. As force is transduced into horizontal cursor position, increasing force will move the dark blue cursor to the rightmost position, which is pre-calibrated at session start to represent 30% of the user's maximum force. The user has been instructed to navigate the cursor as quickly and accurately as possible, in order to visit a sequence of target fields (T0, T1, and T2). Overshoots of the cursor are to be avoided. The current target field is visually indicated to the subject by a green shading (see **Figure 1**). Reaching a target field, a dwell time of 200 ms must be fulfilled in order to achieve a successful hit of this target field. Hit events are indicated visually by a switch of the target field (another field is shaded in green), or by the end of the trial. Trials were chosen randomly from two conditions, each with a specific required target field sequence (T1-T0-T2-T0 or T2-T0-T1-T0). A trial was finished by fulfilling the complete sequence – skipping a target was not allowed. Trial duration is presented visually as an immediate performance feedback during the *pause* phase between trials.

2.2. Subjects and Ethics

Overall, 20 right-handed *normally aged* subjects (8 female, average age: 53 years, std: 6 years) were recruited. The subject group resembles the target group of first-stroke patients with respect to age and gender (Ovbiagele and Nguyen-Huynh, 2011). The term *normally aged* was chosen to indicate our selection criteria: the participants did not have any known neurological or psychological history and were probably healthy—even though we can not exclude the possibility that some participants had a history of unrecognized micro stroke events.

The offline study was approved by the Ethics Committee of the University Medical Center Freiburg. Following the principles of the Declaration of Helsinki, written informed consent was given by subjects prior to participation. In one session of about 3–4 h (including EEG setup and washing the hair), every participant controlled the cursor with their left hand for 20 blocks of 20 trials each.

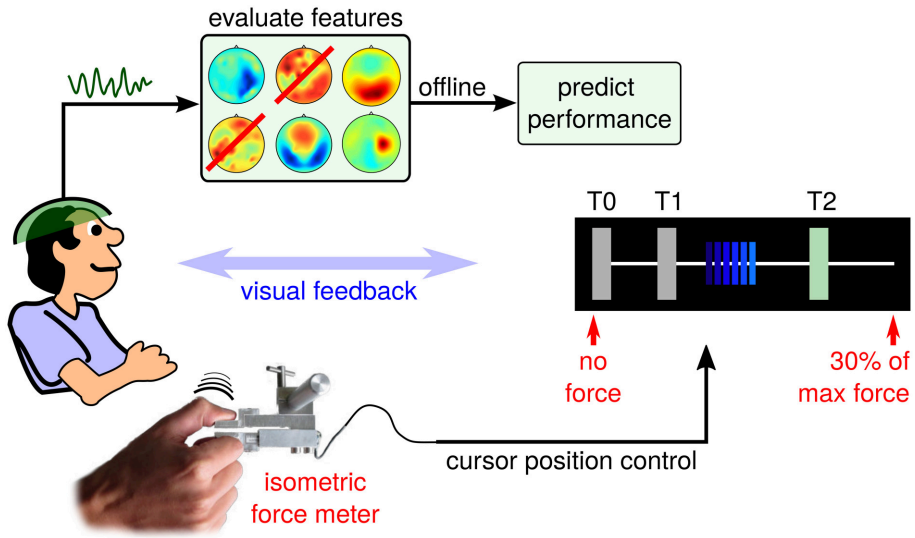


FIGURE 1 | Schematic setup of the EEG-tracked SVIPT. The subject applies force to a sensor using a pinch grasp. Force is transduced into positions of a horizontally moving cursor. EEG activity is recorded before, during and after repeated trials.

2.3. SVIPT Performance Scores

SVIPT enables to capture single trial motor performance. Given a high order motor control, the force profile $F(t)$ of a single-trial is characterized by a quick force ramp up upon the presentation of the go-cue and the avoidance of overshoots. The requested speed-accuracy trade-off can be translated into various performance scores of the SVIPT task. In Tangermann et al. (2015), the authors selected a set of scores, which describe the single-trial performance:

- **Reaction Time/RT**: A quick response upon the *go-cue* is a good start for a successful trial. The time interval between the *go-cue* at time t_{go} and the time point $t_{T0,exit}$, which indicates the cursor leaving the starting field T0, is defined as reaction time.
- **Duration/DUR**: Comparable to RT, a short time duration from the *go-cue* at time t_{go} until the hit of the first target field at time point t_{hit} characterizes a successful trial.
- **Cursor Path Length/CPL**: The total path length the cursor is moved from the *go-cue* to the hit of the first target field is described by the integral over the first temporal derivative of the force profile $F(t)$:

$$CPL \equiv \int_{t_{go}}^{t_{hit}} |\dot{F}(t)|\, dt'$$

- **Integrated Squared Jerk/ISJ**: The level of fine-granular motor control is reflected in variations of the trajectory smoothness. Therefore, jerk—defined as the third derivative of the force profile—is expressed by the ISJ metric, which is defined as:

$$ISJ \equiv \int_{t_{go}}^{t_{hit}} |\frac{d^3F(t)}{dt^3}|^2\, dt'$$

- **Normalized Jerk/NJ**: A unit-free variant of ISJ captures smoothness variations. It is given by the *normalized* jerk:

$$NJ \equiv \sqrt{\frac{ISJ \cdot DUR^5}{2 \cdot CPL^2}}$$

Since there are two conditions of target field sequences, a standardization of the performance scores (except for RT) is the prerequisite for pooling trials of both conditions. Therefore, the extracted metrics of each condition were standardized (zero mean and standard deviation one) prior to pooling. Except RT, the metrics are defined with respect to some end point (e.g., t_{hit}). Choosing this boundary represents a trade-off between (a) harvesting a metric which is temporally close and thus related to the get-ready interval (the interval before the *go-cue*), and (b) including thorough information about the force trajectory of the current trial. To balance the two conflicting goals, we chose the hit of the first target field.

2.4. Data Acquisition and Preprocessing

During a single session, subjects were placed in a chair at 80 cm distance from a 24-inch flat screen. EEG signals from 63 passive Ag/AgCl electrodes (EasyCap) were recorded, which were placed according to the extended 10–20 system. Impedances were kept below 20 $k\Omega$. All channels were referenced against the nose. The signals were registered by multichannel EEG amplifiers (BrainAmp DC, Brain Products) at a sampling rate of 1 kHz. An analog lowpass filter of 250 Hz was applied before digitization. The signal of the force sensor was recorded by an additional amplifier system (BrainAmp ExG, Brain Products).

For outlier identification, the offline preprocessing consisted of high-pass filtering the raw EEG signals at 0.2 Hz, low-pass filtering at 48 Hz and sub-sampling to 500 Hz sample frequency. Therefore, linear butterworth filters of 5th order were applied.

For each trial and all 63 channels, an epoch of 2000 ms duration prior to the *go-cue* was extracted. In order to identify outlier epochs, three rejection methods were applied. First, EEG epochs violating a min-max threshold of 60 μV on frontal channels were excluded from further analysis. Second, a variance threshold on single epochs and channels was applied to remove high-frequent muscular artifacts. Therefore, the variance of single epochs needs to be within P_{up} = 90th percentile and is not allowed to exceed $2 \cdot (P_{up} - P_{low})$ with P_{low} = 10th percentile. Third, epochs belonging to extreme trials, represented by outliers of the motor performance metric, were removed. For this purpose, the following min-max thresholds were defined based on earlier pilot recordings (Meinel et al., 2015). The thresholds were [150, 900] ms for RT, [−1.5, 1.5] for ISJ, [−0.6, 0.6] for CPL, [−1.5, 2] for DUR, and [0, 1300] for NJ. They were applied prior to further data analysis. The total number of trials N_e entering the following offline analysis procedures varied across subjects and performance metrics. Only for 2 out of 20 subjects less than 150 out of the original 400 epochs were remaining after the EEG preprocessing. We discarded data of these subjects from the following analysis. The frequency filtering for our main analysis will subsequently be described in Section 2.6.

2.5. Single-Trial Performance Prediction

In the following, the multivariate variable $\mathbf{x}(t) \in \mathbb{R}^{N_c}$ characterizes the EEG signal recorded from N_c sensors. In addition, $\mathbf{s}(t)$ defines the time course of a neural source. The physics of volume conduction assumes a linear mapping of the source space to the sensor space. The forward (or generative) model thus reads :

$$\mathbf{x}(t) = \mathbf{A}\,\mathbf{s}(t) \tag{1}$$

where the matrix $\mathbf{A} \in \mathbb{R}^{N_c \times M}$ describes the projection of the M sources to the EEG sensor space.

The main goal is to approximate the true neural source $\mathbf{s}(t)$ by $\hat{\mathbf{s}}(t)$, whose power achieves the highest correlation with a predefined external variable $z(t)$, called *target variable* from this point onwards. Several methods can be used to estimate such a source, among them Blind Source Separation (BSS) and source reconstruction techniques. BSS techniques rely solely on an unsupervised framework, which may be a suboptimal approach given the availability of labels in the form of the target variable z. Source reconstruction techniques may provide a high level of interpretability for the results directly in the source space, and may describe non-stationarities in the data and other complex dynamics (Castaño-Candamil et al., 2015b). However, there are three potential drawbacks of source reconstruction approaches (Grech et al., 2008): First, the estimation of $\hat{\mathbf{s}}(t)$ usually creates a rather high computational burden. Second, the methods require a forward model $\mathbf{A}$ for each individual subject, which may not be available in most situations since it corresponds to the exact anatomical description of a subject's brain. Third, as source reconstruction problems are intrinsically ill-posed, the quality of an estimated source depends on additional assumptions, such as the density of sources or their location within the brain.

An alternative to both, BSS and source localization approaches is the family of the so-called supervised spatial-filtering methods. One widely known approach is the Common Spatial Patterns algorithm (CSP; Ramoser et al., 2000), which searches for spatial filters that enhance the contrast between two classes. Consequently, it is well suited for supervised classification problems. A more recent approach, the Source Power Comodulation algorithm (SPoC; Dähne et al., 2014) is adequate for regression problems. As the five SVIPT metrics are continuous variables, we preferred to include the SPoC algorithm into our data analysis framework.

SPoC learns an optimal spatial filter $\mathbf{w_{opt}} \in \mathbb{R}^{N_c \times 1}$ that allows to estimate the source as $\hat{\mathbf{s}}(t)$ by projecting $\mathbf{x}(t)$ into a subspace, which maximizes the correlation between the band power of $\hat{\mathbf{s}}(t)$ and the target variable $z(t)$:

$$\hat{\mathbf{s}}(t) = \mathbf{w_{opt}}^{\top}\mathbf{x}(t) \tag{2}$$

Without loss of generality, the objective function for SPoC may be defined in terms of the epoched data $\mathbf{x}(e)$, where e refers to the e-th epoch. Assuming that the EEG sensor signal has been bandpass-filtered to a narrow frequency band and that the norm of the spatial filter is constrained, e.g., $\mathrm{Var}[\mathbf{w}^{\top}\mathbf{x}(t)] \stackrel{!}{=} 1$, the optimization problem can be solved by maximizing the covariance:

$$\begin{aligned} &\mathbf{w}_{opt} = \underset{\mathbf{w}}{\mathrm{argmax}}\{\mathrm{Cov}[\Phi_x(e), z(e)]\}\ \forall\ e \\ &\text{s.t. } \mathrm{Var}[\mathbf{w}^{\top}\mathbf{x}(t)] \stackrel{!}{=} 1 \end{aligned} \tag{3}$$

where $\Phi_x(e) = \mathrm{Var}[\hat{\mathbf{s}}](e)$. This formulation of the algorithm – called SPoC$_\lambda$ – can be transformed into a generalized eigenvalue problem and thus delivers a set of N_c spatial filters $\mathbf{W} \in \mathbb{R}^{N_c \times N_c}$. In this paper, the SPoC$_\lambda$ algorithm is utilized which subsequently will be abbreviated by the term SPoC.

Applying a SPoC filter $\mathbf{w_{tr}}$ learned from training data $\mathbf{x_{tr}}(t)$, the method allows to estimate the target variable z_{est} on novel, unseen test data $\mathbf{x_{te}}(t)$ on a single-trial basis by calculating the bandpower of the narrowband subspace signal:

$$z_{est} = \mathrm{Var}[\mathbf{w_{tr}}^{\top}\mathbf{x_{te}}(t)](e) \tag{4}$$

Using this relation, we will focus on the prediction of single-trial SVIPT performance using EEG activity within the *get-ready* phase of the trial.

As Haufe et al. (2014) have been pointing out, there is an existing forward model of the form of Equation (1) for every backward model as in Equation (2). Thus, the corresponding spatial activation patterns can be obtained from the spatial filters $\mathbf{W}$ via the covariance matrix $\mathbf{C_x}$ of the data $\mathbf{x}(t)$ via:

$$\mathbf{A} = \mathbf{C_x}\mathbf{W} \tag{5}$$

Note, that any subspace components resulting from the SPoC analysis depend mainly on four hyperparameters. In the temporal domain, two of them define the epoching interval $[t_0, t_0 + \Delta t]$ where t_0 is the starting time relative to the *go-cue* and a duration Δt. In the frequency domain, the lower frequency f_0 and the band width Δf are the hyperparameters describing the band $[f_0, f_0 + \Delta f]$ in which $\mathbf{x}(t)$ is bandpass-filtered.

Even though simple regression of bandpower features on the channel level does not fulfill the requirements of the assumed forward model, we added this simple method for comparison with SPoC. Therefore, channel-wise bandpower features of the training and test set were calculated.

2.6. Selection Criteria for Informative Oscillatory Components

Performing a grid search across subjects and SPoC parameters, we restrict the evaluation to a fixed predictive time interval given by $t_0 = -800$ ms prior to the *go-cue* and a window size of $\Delta t = 750$ ms.

As sketched in **Figure 2**, logarithmically increasing and overlapping frequency bands ranging from $\approx$ 1–100 Hz (55 configurations in total) were evaluated from the original non-filtered signals. For bandpass filtering, linear butterworth filters of 5th order were utilized. As a trial-wise target variable z, the five different performance metrics introduced in Section 2.3 were considered. Evaluating SPoC across the complete study group of 18 subjects, using five different motor performance metrics, sweeping through 55 discrete frequency bands and selecting the highest-ranked components (see details below) per configuration, results in more than 12,000 oscillatory components. In this section, we will describe an offline selection strategy in order to identify a subset of the most robust and informative oscillatory components which qualify to predict single-trial motor performance.

Upon each parameter configuration, a K = 5-fold chronological cross-validation procedure was employed upon the calculation of SPoC (Lemm et al., 2011). Only trials were considered, which survived the data preprocessing (see Section 2.4). From these, N_e EEG pre-trial epochs and their corresponding target variable values z were extracted in chronological order and split into 5 equally-sized folds. Thus, 4-folds served as training data while the remaining one was used for validating the SPoC filter as described in Equation (4). Since each fold served as test fold once, the estimated target variable $z_{est,j}$ of fold j can be concatenated for all N_e epochs, resulting in $z_{est} = [z_{est,j}]_{j\in[1,K]}$. According to Equation (5), on each fold j the corresponding test pattern is given as $\mathbf{a_j} = \mathbf{C_{te,j}w_{tr}}$ utilizing the covariance matrix $\mathbf{C_{te}}$ of the test data $\mathbf{x_{te}}(t)$.

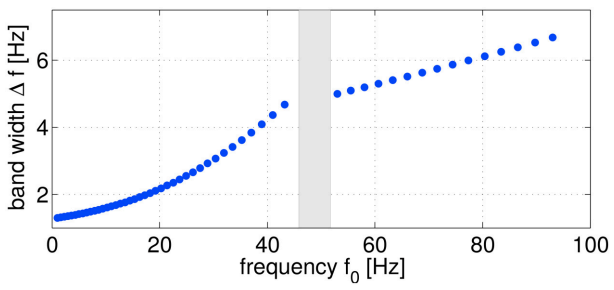


FIGURE 2 | Frequency parameter configurations characterized by the frequency f_0 and the corresponding band width Δf. In total, 55 configurations were used for computing SPoC filters. The omitted points (gray area) correspond to the power line frequency range.

The same cross-validation scheme was applied upon the linear regression model. The whole parameter space of 3600 configurations was screened. Note, that this number is smaller than the number of components delivered by SPoC analysis, since the latter may deliver more than one component per parameter configuration. The regression, which delivers a single component per configuration only, was trained on the training data and finally applied on test data such that an estimate z_{est} was gained on all $N(e)$ trials which had survived the data preprocessing step.

For a given parameter set, SPoC$_\lambda$ returns a set of N_c filters. As described in Tangermann et al. (2015), it is sufficient to take only the first-ranked components into consideration[1]. For this purpose, we applied a rank-based criterion. First, removed the linear trend from the ordered set of N_c eigenvalues. A threshold of $1.5\cdot\sigma(r)$ relative to the standard deviation σ of the resulting N_c residuals r was defined. We restricted the investigation to positive eigenvalues.

Given a single component $\mathbf{w}$, the following set of scores enable to characterize its predictive strength and stability:

- **Correlation characteristics:** As a measure to verify the quality of the predictive strength of a SPoC configuration, the overall correlation of the N_e-many measured performance labels z_{true} with the corresponding predictions z_{est} can be considered:

$$R_{all} = \mathrm{Corr}[z_{true}, z_{est}] \tag{6}$$

  Similarly, the predictive strength in terms of single-trial performance can also be verified by checking the mean of the fold-wise correlations $R_j = \mathrm{Corr}[z_{true,j}, z_{est,j}]$, which rewards temporally stable components:

$$R_{folds} = \frac{1}{K}\sum_{j=1}^{K}\mathrm{Corr}[z_{true,j}, z_{est,j}] \tag{7}$$

  The correlation based metrics R_{all} and R_{folds} come closest to the original optimization objective of the SPoC algorithm. If the trained spatial filters model trial-to-trial fluctuations well, R_{all} and R_{folds} will report a large value, but only R_{folds} allows to discriminate between single-trial predictors and session-trend models. Furthermore, a stable component requires that the correlation of each fold j shares the same sign with R_{all}. Thus, it is reasonable to require a high homogeneity H_{folds}:

$$H_{folds} = \sum_{j=1}^{K}\Theta(\mathrm{sign}(R_{all})\cdot\mathrm{sign}(R_j)), \tag{8}$$

  with $\Theta(x) = 1$ for $x \geq 0$ and $\Theta(x) = 0$ for $x < 0$ representing the unit-step function.
- **Separability of estimated performance:** Simulating the trial-wise online application, the continuous prediction is transferred into a two-class problem. Therefore, we split

[1]The components are sorted according to their eigenvalues. In case of SPoC$_\lambda$, they equal to the covariance between the bandpower features and the target variable.

the N_e prediction values z_{est} into two distributions ($z_{est,h}$ and $z_{est,l}$) based on the 50th percentiles of the true target variable distribution from z_{true}, thus representing high and low performance. The separability of $z_{est,h}$ and $z_{est,l}$ can be quantified by a statistical test. Here, the area under the receiver-operating-characteristic curve (Fawcett, 2006) is reported. It is denoted as z-AUC and has a chance level of 0.5.

- **Stressing the stability:** SPoC is a supervised method, which uses label information to guide the spatial filter calculation. Thus, the robustness of a resulting component can be stressed by introducing label noise. The concept of a step-wise reduction of the SNR of the labels has been introduced by Castaño-Candamil et al. (2015a). Here, SNR levels were varied from −20 dB to 10 dB by adding white noise. Applying SPoC, we estimated the target variable z_{est} for all N_e epochs using 5-fold cross-validation. At each SNR level, three sets of noisy labels z were calculated. For each SNR level, the separability of the resulting z_{est} distribution is verified by the z-AUC value. Regarding the z-AUC values as a function of the SNR, the area under this curve—referred to as AAUC_{SNR}—describes the stability of the component.

To finally identify and select robust and predictive components, we propose to apply three out of these five criteria in parallel. As a prerequisite, the data set needs to consist of at least $N_e = 150$ trials in order to ensure the convergence of the SPoC algorithm (see Dähne et al., 2014; Castaño-Candamil et al., 2015a):

1. The separability of the predicted performance z_{est} can be verified by the resulting z-AUC value. A corresponding threshold $\text{z-AUC}_{min} = 0.59$ was determined according to the 85th percentile across all configurations.
2. The stability of the component is assessed by the AAUC_{SNR}. Here, a threshold $\text{AAUC}_{min} = 0.18$ was determined from the 85th percentile.
3. As an additional stability criterion, we require all fold-wise correlations R_j to share equal sign as R_{all} such that $\text{H}_{folds,min} = 5$.

3. RESULTS

3.1. SVIPT Performance Metrics

Single-trial based SVIPT performance can be assessed by different metrics, as described in Section 2.3. In **Figure 3** examples of the trial-to-trial fluctuations of different metrics are visualized for two subjects. The visualization covers the full sessions, but omits trials removed during the preprocessing. **Figures 3A,D** show the metric reaction time (RT) for two subjects. It is not affected by a session trend. Its distribution is slight asymmetric, which is caused by a physiological limit for the minimal RT. The normalized jerk (NJ) in **Figures 3C,F** behaves in a similar manner. It is affected only slightly by a global trend, but shows a more skewed distribution compared to RT. In contrast, integrated squared jerk (ISJ) depicted in **Figure 3B**, and cursor path length (CPL) in **Figure 3E** both show a strong session trend, which can be explained by the user learning (data not shown here). A comparably strong session trend is present also in the duration metric DUR (data not shown).

The cross-correlations between all five metrics and the shape of their distributions were reported in Tangermann et al. (2015). Metrics ISJ, CPL and DUR showed strong correlations to each other, while RT as well as NJ both are rather independent from the four other metrics.

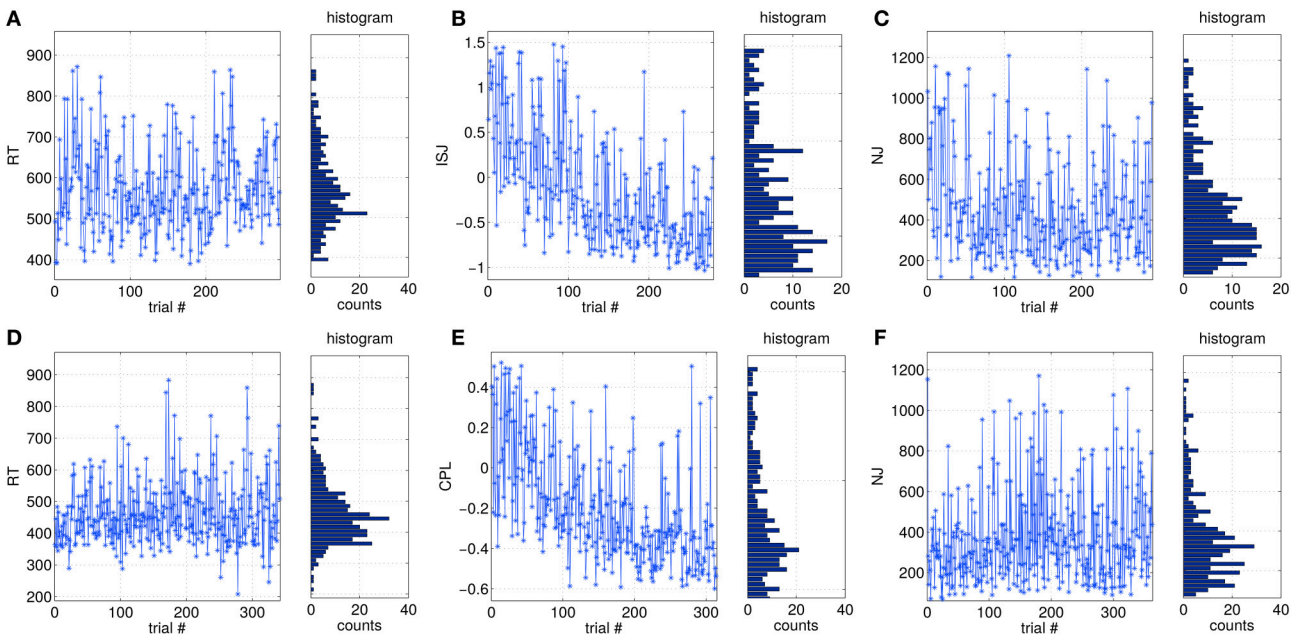


FIGURE 3 | Examples of trial-wise variations of different performance metrics over a full session, and their distributions. (A–C) are taken from data of subject S9, while **(D–F)** are from S13.

3.2. Contrasting SPoC with Linear Regression on Sensor Level

As a baseline comparison for the predictive power of SPoC components, a linear regression model employing channel-wise bandpower features was evaluated as described in Section 2.6. The resulting distributions of the overall correlation R_{all} and the performance separability z-AUC are reported in **Figure 4**. Across all configurations, SPoC delivers a median correlation $R_{all,med} = 0.07$ and a separability of z-AUC$_{med} = 0.54$, while on average the regression performs on chance level. While both methods come up with components revealing z-AUC values above chance level, those with the strongest predictive information are generated by the SPoC method.

3.3. Single-Trial Motor Performance Predictors

In **Figure 5**, five exemplary predictive and robust SPoC components, gained from five different subjects are characterized. Although SPoC components are computed from band-pass filtered data, the resulting filter **w** (gained on all available N_e trials) of a component can be re-applied to non-frequency-filtered epoched data. This spectral content of a SPoC component is shown in **Figure 5A**. The frequency band in which the component was extracted from is indicated by the dashed gray area. Using all available epochs, **Figure 5B** shows the spatial activity pattern gained via Equation (5). In **Figure 5C**, the SPoC filter weights on the 2D-scalp projection are shown. The scatter plot in **Figure 5D** reports on the measured performance metric z_{true} as a function of the predicted performance z_{est} according to the CV scheme described in Equation (4). The data points are colored by the fold index (1–5), which corresponds to the temporal order of the session. Fold 1 represents the beginning of the session, fold 5 its end. In addition, the overall correlation R_{all} reports on the predictive strength of the component. The distributions shown in **Figure 5E** illustrate the separability of the single-trial performance values z_{est}. For this purpose, the estimated labels z_{est} have been reduced to the lower and upper quartile. The corresponding true labels z_{true} were used to compute the quartiles $Q_{low,est}$ and $Q_{high,est}$ and were fitted by a kernel distribution (solid lines). In an ideal case, those quartiles would converge toward the extreme quartiles (Q_{low} and Q_{high}) of z_{true}, which are indicated by dashed lines. As a score of their separability, the score z-AUC as described in Section 2.6 is reported based on the 50th percentile.

The exemplary components in **Figure 5** are selected across the investigated frequency range depicted in **Figure 2**. The predictor of S7 can be assigned to the theta band, those of S9 and S13 correspond to the alpha range, the component for S5 originates from the beta range and the one of S8 was found in the gamma range. Regarding the scatter plots, there are two different types of patterns recognizable: single-trial predictors showing a confined point cloud without a clear trend over time (all examples except for S13), whereas the scatter plot of subject S13 shows a clear trend over the course of the session. The separability plots indicate that the predictive power of a single component nicely matches with the z-AUC value.

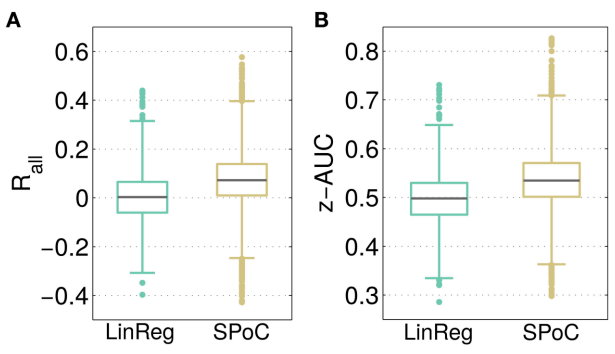


FIGURE 4 | Contrasting the predictive outcome of all 3600 tested parameter configurations for linear regression (LinReg) and over 12,000 configurations for SPoC. In **(A)**, the overall correlation R_{all} between predicted and estimated target variable values is depicted. **(B)** Shows the performance separability z-AUC. Gray lines indicate the median, boxes enclose the 25th to 75th percentile. The whisker length is set to two inter-quartile ranges.

3.4. Testing the Stability of SPoC Components

The stability of an oscillatory component can be challenged by varying the signal-to-noise ratio (SNR) of the target variable z. In **Figure 6**, the z-AUC score is investigated as a function of the SNR for two parameter configurations. **Figure 6A** shows a stable component, where z-AUC is expected to decrease, while for a non-informative component in **Figure 6B** the z-AUC can be expected around the noise floor. Thus, the resulting *area under the* z-AUC *curve* can be assessed as a tool for mapping the stability of the subspace component under challenging noise conditions. In **Figure 6C**, the distribution of this so-called $AAUC_{SNR}$ is reported for all evaluated SPoC components across all 18 subjects. The distribution of $AAUC_{SNR}$ values has its median at 0.07 and is slightly skewed.

3.5. Identification of Robust and Predictive Components

As described in Section 2.6, the first highest ranked components of each parameter configuration have been evaluated, resulting in about 12,000 different subspace components. In **Figure 7**, the configurations are characterized by their stability under noise (AAUC$_{SNR}$), which is plotted in **Figure 7A** as a function of the separability measure z-AUC, in **Figure 7B** as a function homogeneity of the fold-wise sign of the correlation H_{folds} and in **Figure 7C** as a function of the overall correlation R_{all}. A few observations can be made: First, the metrics are not centered at zero. Second, based on all initial configurations (blue data points), AAUC$_{SNR}$ correlates with the z-AUC as well as with R_{all}. The largest AAUC$_{SNR}$ values are evoked by the most homogeneous fold-wise correlation signs with $H_{folds} \geq 3$.

The threshold criteria applied to select the best of the 12,000 subspace components are indicated by red dashed lines, and red dots indicate the components finally selected. As shown in

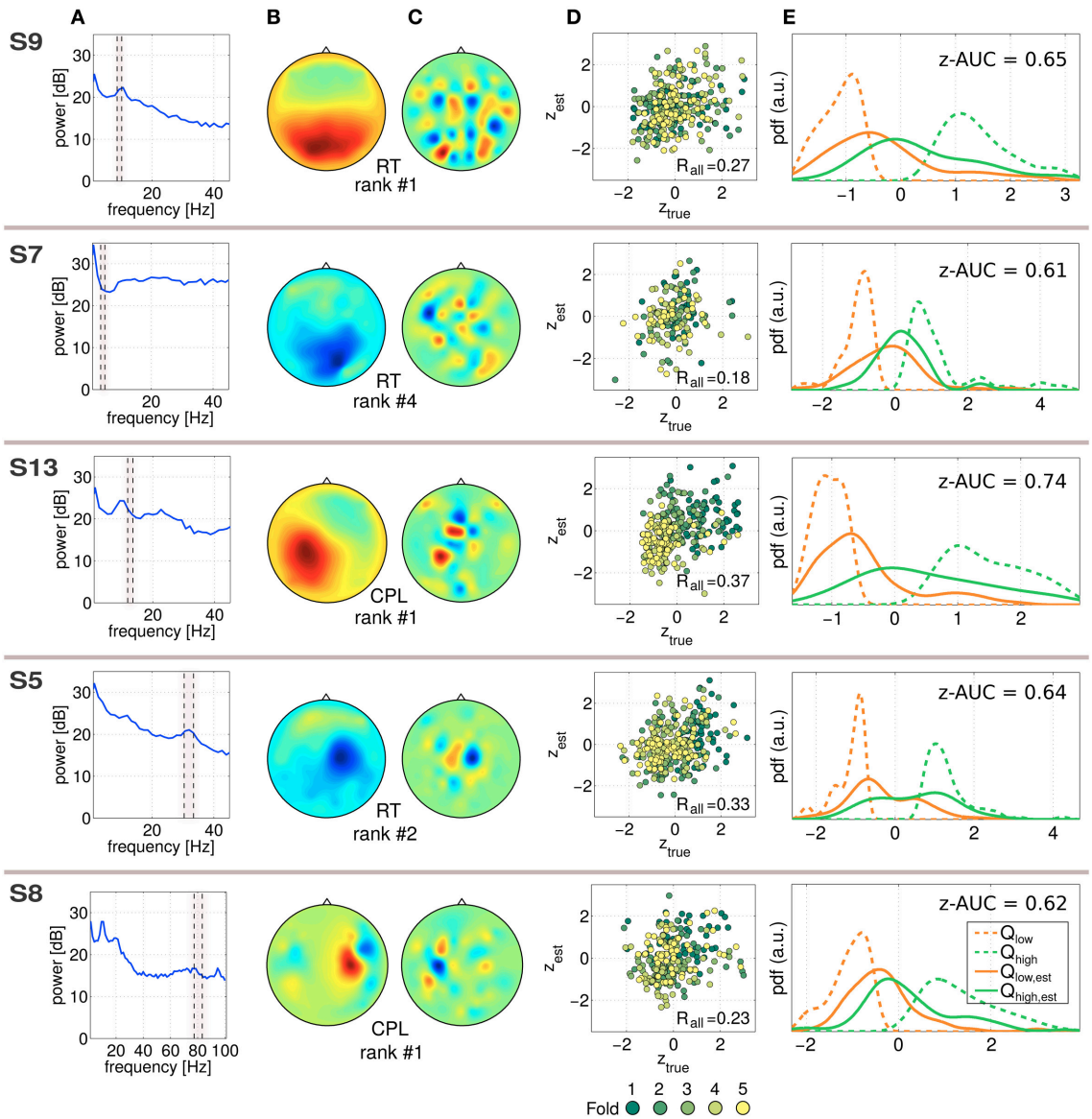


FIGURE 5 | Characterization of *exemplary* predictive SPoC features. Each component is characterized line-wise labeled by the used performance metric and the rank according to the full-session filters. **(A)** Power spectrum of the component applied on non-bandpass filtered full data. The frequency band where the component has been trained is marked by the dashed lines. Note that for the component of S8 a broader frequency range is visualized compared to the other examples. **(B)** Spatial activity pattern. **(C)** Filter weights visualized. **(D)** Scatter plot between true labels z_{true} and the predicted ones z_{est}, color coded by the fold of the chronological cross-validation. **(E)** To illustrate the separability of the prediction, the distribution of z_{true} values has been split using the corresponding trials of the upper and lower quartiles of z_{est}, which resulted in $Q_{low,est}$ and $Q_{high,est}$. As a reference, the extreme quartiles Q_{low} and Q_{high} of z_{true} are also given (dashed curves). In addition, the z-AUC value based on the 50th percentile split is reported.

Figure 8, the overall correlation R_{all} is strongly correlated with the z-AUC metric, such that an additional threshold criterion on R_{all} was not necessary. The most predictive components achieve a correlation value of up to 0.6, corresponding to ${R_{all}}^2 = 0.36$. Assuming a linear relationship between z_{true} and z_{est} as well as normally distributed data, this means that z_{est} can explain up to 36% of the performance variance contained in z_{true}.

In **Figure 9**, all 361 selected components are characterized by histograms in terms of their input parameters. **Figure 9A** displays the subject-wise grouping. In total, 11 out of 18 subjects contribute at least one component, for three subjects more than 50 configurations survive the selection procedure. **Figure 9B** shows the histogram over the number of trials available for the offline analysis. Note, that this histogram is dominated by the best three subjects reported in **Figure 9A**, who contributed a large number of the selected 361 configurations. **Figure 9C** characterizes the selected components assigned to their underlying frequency band $[f_0, f_0+\Delta f]$ (see **Figure 2**). Most components are gained from the alpha- and beta-band range. Interestingly, robust features detected in the gamma band were

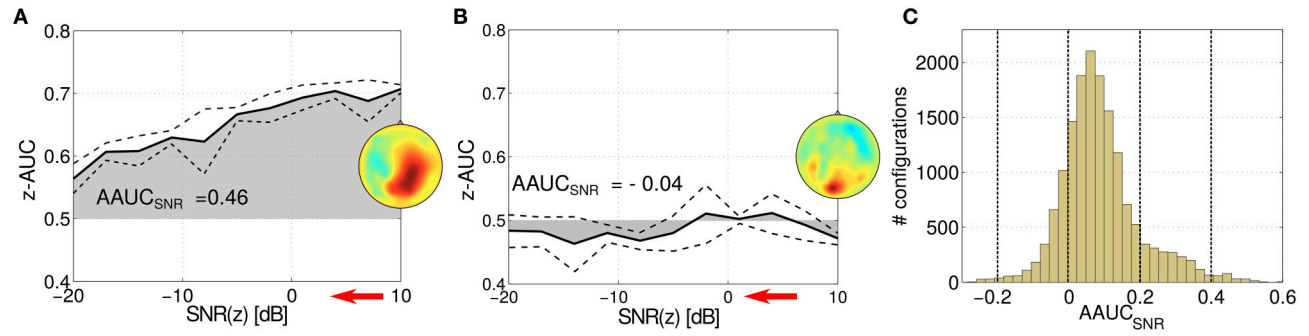


FIGURE 6 | Stressing the stability of two exemplary SPoC components for two different parameter configurations (A,B). While stepwise decreasing the SNR ratio (indicated by the red arrow), z-AUC-values (solid lines) describing the separability of the prediction are plotted together with standard deviations (dashed lines). The area under the z-AUC curve—further on called AAUC_{SNR}—describes the stability of the component under the challenge of added noise. **(C)** Shows the histogram of all $AAUC_{SNR}$ scores evaluated for the considered parameter configurations.

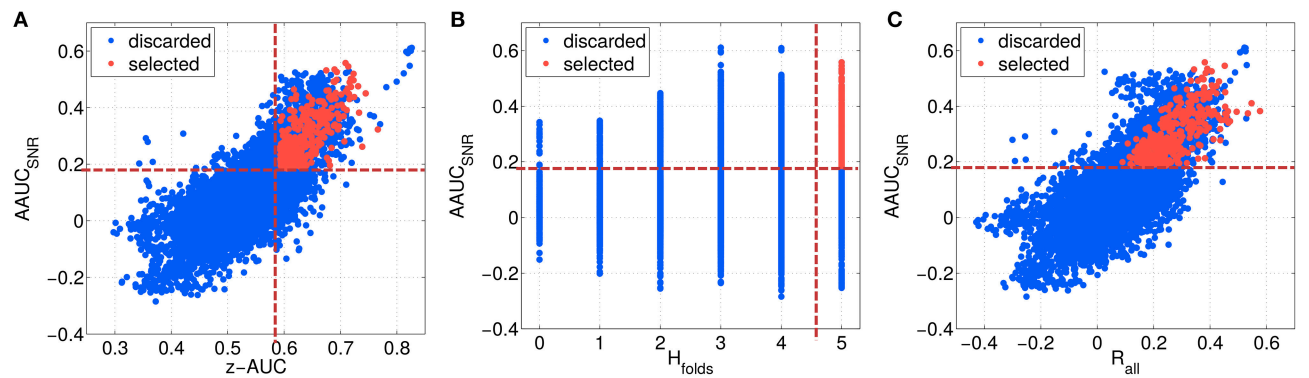


FIGURE 7 | Characterizing the space of SPoC components by several metrics, which describe their stability and predictive information. The SNR-challenged AAUC_{SNR} is given as a function of the performance separability z-AUC **(A)**, in relation to the homogeneity of the correlation sign H_{folds} **(B)**, and dependent on the overall correlation R_{all} **(C)**. Red data points describe the selected SPoC components after applying thresholds (dashed red lines).

dominantly selected for their ability to predict CPL. The slow frequency (<4 Hz) components are dominated by artifactual subspaces. **Figure 9D** reports on the occurrences of the different performance metrics among the selected components. Most components could be extracted for RT (61%), followed by CPL (16%) while all other metrics contribute almost equally well with 6–8% of the selected components. **Figure 9E** provides an overview over the SPoC ranks of the surviving components. The rank ordering corresponds to the eigenvalue ordering of the complete data set. As the number of selected components drop with increased rank, the ranking is associated with the information content of the subspace component.

SPoC as a linear filtering method allows for a limited neurophysiological interpretation of spatial activity patterns. A representative subset of typical scalp topographies from the selected stable and informative subspaces are plotted in **Figure 10**. The components were assigned to three groups. About 70% of components fall into group G1, which comprises patterns ranging from activations in occipital, to central or frontal areas. The maximum activity of those components often is found over one of the hemispheres. About 10% of the components fall in group G2. They show patterns of probable non-neural sources and may represent e.g., eye artifacts, muscular activity or single noisy channels. Group G3 comprises noisy topographies. As indicated by patterns in the intersection area of the three groups, mixed components were observed as well. The detailed parameter configurations of each of the plotted components is listed in **Table 1**.

4. DISCUSSION

We hypothesized that subject-specific pre-trial brain signals contain information which allows to partially explain and temporally predict the trial-by-trial variability of the upcoming motor performance in SVIPT. To test the hypothesis, we developed a workflow which is capable to extract informative oscillatory EEG subspace components and to identify the most robust ones. Simulating an online application, our analysis revealed strong evidence that the band power of the selected components is predictive for the single-trial SVIPT performance. Major findings were that these components indeed exist, but need

to be optimized for individual users. With 11 out of 18, not all, but a majority of the subjects revealed the desired informative features.

In the following we will first discuss the decision to utilize SPoC instead of other alternative analysis methods. In this context, the proposed selection procedure and the stability of SPoC components over time is discussed, with a special focus onto the role of SNR, frequency and the illiteracy phenomenon. In addition, the detected components will be related to existing literature and characterized on a group-level with respect to the covered frequency bands, sub-processes reflected by the components and the time courses revealed. Before concluding, we will describe a neuroergonomically enhanced rehabilitation paradigm as a possible use case of our contribution.

4.1. SPoC and its Alternatives

Designing the data analysis workflow, we built upon our background in BCI. Accordingly, we carefully selected algorithmic building blocks only, if they can be applied in single-trial analysis [e.g., the application of the spatial SPoC filter according to Equation (4)]. This decision should simplify the translation of the presented workflow to closed-loop experiments in the future. The choice of the supervised SPoC algorithm for extracting informative components is supported by its good performance compared to a supervised linear regression of bandpower features on the sensor level (see Section 3.2). This is in accordance with findings of Dähne et al. (2014). On data from an auditory steady-state evoked potential paradigm, these authors reported better results for SPoC compared to both, linear regression and an unsupervised subspace decomposition using independent component analysis (ICA). SPoC does not reconstruct sources of the brain, but instead performs a supervised subspace decomposition. Thus, a SPoC subspace component can not be expected to correspond to a single physical source or even a dipole source (even though such SPoC components are possible). Theoretically even several spread-out brain areas may contribute to a single SPoC component, if they share oscillatory activity which co-varies over time with the labels. The choice between SPoC and source reconstruction approaches (Gonzalez Andino et al., 2005) represents a trade-off—while the latter may facilitate the interpretation of results, SPoC components avoid several of the drawbacks mentioned in Section 2.5. As our workflow was aligned in terms of applicability for single-trial online paradigms, our decision was biased toward SPoC.

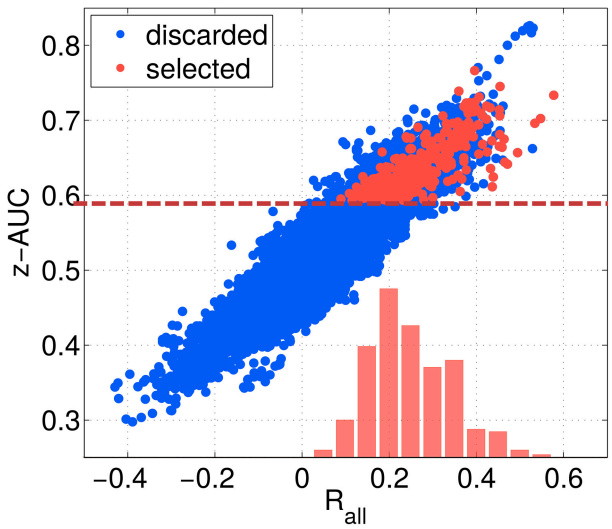


FIGURE 8 | Relation between separability metric z-AUC and the overall correlation R_{all} for all SPoC configurations (blue dots) and the selected ones (red dots). The dashed red line indicates the threshold z-AUC_{min} applied to select most informative components. The red bars indicate the distribution of R_{all} values for the selected components.

4.2. Selection Criteria for Robust and Predictive Components

Over-fitting is a general issue for supervised methods and for SPoC in particular, as no form of regularization was applied. This requires some form of *post-hoc* selection of SPoC components. The situation is aggravated, as SPoC returns full rank filter matrices, which result in a very large numbers of subspaces. However, only a fraction of these can be expected to be informative about the labels. As robustness over time as well as with respect to label noise are important criteria for the potential closed-loop applicability of a component, a single selection criterion (e.g., a threshold on the correlation value) is not sufficient. By that, we selected three criteria (see Section 2.6), which suited best these requirements. Out of the initial five selection criteria, the two scores R_{all} and R_{folds} turned out to

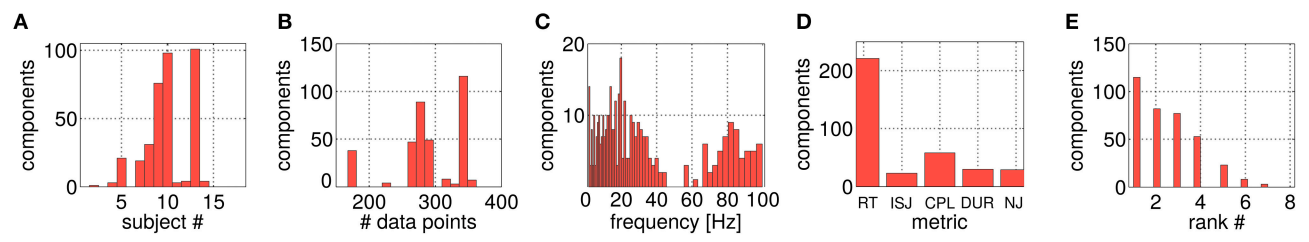


FIGURE 9 | Histograms of different parameters solely for the selected SPoC components. (A) Shows the assignment over the 18 subjects. **(B)** Gives the allocation over data set sizes N_e (with a lower limit of 150 trials). **(C)** Visualizes the distribution across frequency bands. **(D)** Depicts the spread of components over the five utilized motor performance metrics, while **(E)** shows the split according to their SPoC rank positions.

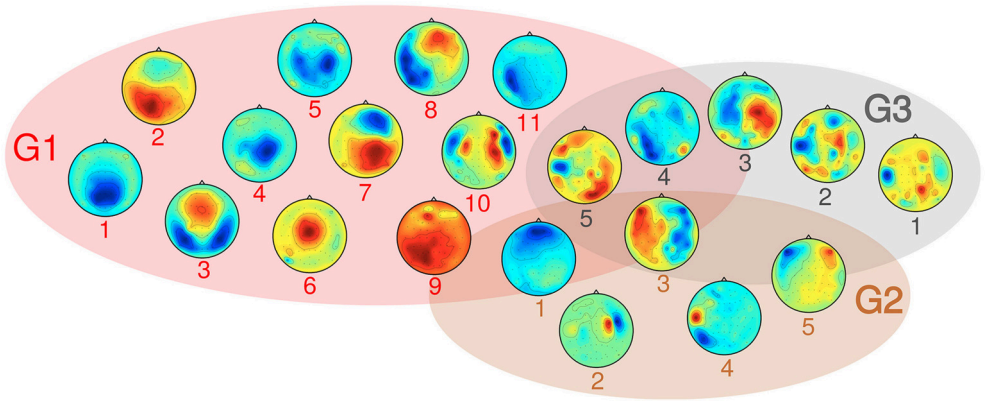


FIGURE 10 | Overview over typical activity patterns resulting from the selected components, grouped in three categories: G1 consists of components with neural origin, G2 comprises artifact-related subspaces and G3 captures non-informative components. Details on their parameter configurations are given in **Table 1**.

TABLE 1 | Parameter configurations for components of groups G1, G2 and G3 as visualized in Figure 10.

	G1				G2				G3			
Comp. no.	**Subject**	**f_0 [Hz]**	**Metric**	**Rank**	**Subject**	**f_0 [Hz]**	**Metric**	**Rank**	**Subject**	**f_0 [Hz]**	**Metric**	**Rank**
1	S13	7.9	NJ	2	S14	1.0	RT	2	S12	27.5	ISJ	1
2	S9	8.7	ISJ	3	S11	27.5	ISJ	1	S10	20.3	RT	1
3	S13	7.9	RT	3	S10	5.0	DUR	1	S9	8.7	ISJ	1
4	S13	8.7	NJ	2	S9	39.0	DUR	2	S15	19.2	CPL	1
5	S10	15.3	RT	6	S10	3.7	RT	3	S10	74.4	NJ	2
6	S5	30.4	RT	3								
7	S9	14.4	RT	5								
8	S9	10.2	ISJ	4								
9	S9	18.2	DUR	1								
10	S8	77.4	CPL	4								
11	S13	74.4	CPL	1								

be beneficial for characterizing the extracted components. Thus, they were omitted for the selection process, since a strong correlation between z-AUC and R_{all} was observed (see **Figure 8**). The same holds for the correlation between z-AUC and R_{folds} (not shown).

An alternative to the current selection procedure would be to relax the thresholds and combine it with additional methods to judge the plausibility of the remaining components *post-hoc*. For ICA components, workflows have been proposed, such as MARA, an automatic classification of artifactual components by Winkler et al. (2014). MARA uses features based on topology, time-frequency analysis and source reconstruction. Similar approaches have been proposed by Daly et al. (2015) and Grosse-Wentrup et al. (2013).

4.3. Influence of SNR on SPoC Components

By applying rather strict selection criteria, weaker but still informative components may have been removed. As a result, the data of some subjects did not reveal informative pre-go oscillatory components. Reasons may be a lower SNR of their data, which hides potential informative content from the SPoC analysis, especially in combination with the limited number of trials used. The work of Castaño-Candamil et al. (2015a) on robustness testing of SPoC components backs this interpretation. In this case, future improvements may be expected by regularization techniques introduced to SPoC, a reduction of the dimensionality prior to applying SPoC, using more data or from transfer learning approaches. However, we can not exclude that informative oscillatory components may not be visible to the EEG or may be absent in some subjects. This problem has been described as BCI "illiteracy." It has predominantly been studied in the context of motor imagery paradigms for the control of BCI applications (Hammer et al., 2012), where decoding the imagery class usually is not possible for a subset of subjects. The BCI illiteracy problem was tackled by novel experimental setups like hybrid BCI paradigms (Allison et al., 2012; Müller-Putz et al., 2015), but could also be alleviated by more advanced decoding methods (Sannelli et al., 2010).

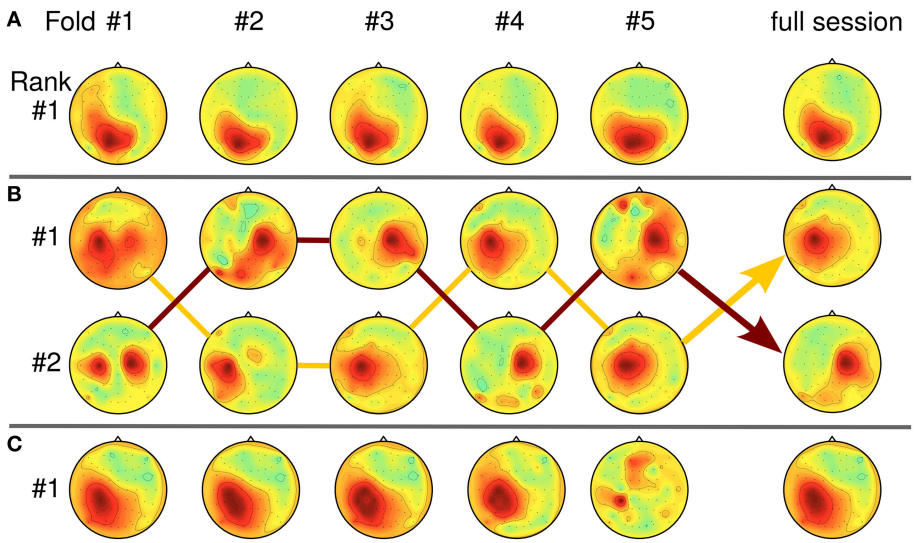


FIGURE 11 | Relation between SPoC rank stability and pattern homogeneity over five cross-validation folds (chronological order). (A) Stationary case: component is first-ranked across all five folds (data of subject S9, $f = [9.4, 11]$ Hz, RT). **(B)** Rank switching: Two almost stable components switch rank positions between folds (S5, $f = [27.5, 30.3]$ Hz, RT). Lines connect the corresponding topologies. **(C)** Intensity variation: intensity of first-ranked component decreases over time folds (S13, $f = [13.6, 15.3]$ Hz, CPL).

4.4. Rank Stability of SPoC Components over Time

In Section 4.3, the relation between SPoC solutions and the SNR of the data has been touched. As SPoC ranks the detected components according to their covariance values, solutions may seem unstable when only the first-ranked component is considered. In real-life data sets, variation of the SNR over time can induce rank switches or mixed components. Tracking a component over multiple runs of the subspace decomposition method is a challenging task, especially as mixtures theoretically can not be distinguished from a single source. However, as similar problems arise for online learning of blind source separation methods like ICA, practical solutions are available (Hsu et al., 2015). **Figure 11** gives examples of stable, stationary components (**Figure 11A**) and of unstable SPoC components (**Figures 11B,C**), both observed over the five chronological cross-validation folds. Instable components may be evoked if the stationarity assumption of SPoC is violated e.g., by slow temporal intensity variations due to user learning. For **Figure 11B**, arrows indicate a possible path through rank positions across folds by connecting corresponding components. Please observe, that SPoC generates cases with even severe variation between folds as those depicted in **Figures 11B,C**. However, such components typically have been removed during the selection process. While mixed, yet stable components may be hard to interpret, they can still be useful for predicting the task performance.

We have observed a high sensitivity of SPoC for small differences in the frequency parameters. Seemingly unstable components which display rank switching behavior (see **Figure 12** at 8.7 Hz) can sometimes be stabilized by slightly changing the frequency, e.g., to 9.4 Hz in this example. Further increase of the frequency to 10.2 Hz again induces instability in this example.

4.5. Characterization of Informative SPoC Components and Sub-Processes

The proposed SPoC workflow delivers a diverse set of oscillatory components, which vary in their topological patterns as well as in their underlying frequency band. This is not surprising, since SVIPT requires the interaction of several cognitive sub-processes in order to reach a good overall performance. For each sub-process, one or more specific neural features may exist, with all of them being informative about the overall outcome of the complex task.

The best components differ between subjects and predominantly occur in the alpha band, followed by beta and gamma band. Our findings are supported by informative features in the alpha and beta-range observed during pre-movement intervals of a hand grasping task (Zaepffel et al., 2013; Meyer et al., 2014; Yang et al., 2014). Furthermore, the informative frequency ranges for SVIPT are comparable to those reported for attention related tasks (Gonzalez Andino et al., 2005; Hoogenboom et al., 2010; van Ede et al., 2012). We obtained best results when using RT as a performance metric, which supports our earlier findings on disjunct data from younger subjects (Castaño-Candamil et al., 2015c; Meinel et al., 2015). RT of course does not automatically lead to a successful trial, but it can be seen as an indicator for a quick ramp-up phase and alertness. For fewer users, presumably those with highest SNR characteristics, informative oscillatory features could be identified for other performance metrics of the force task, too.

Comparing the topological plots of group G1 in **Figure 10** with those reported in literature, it can be observed that many

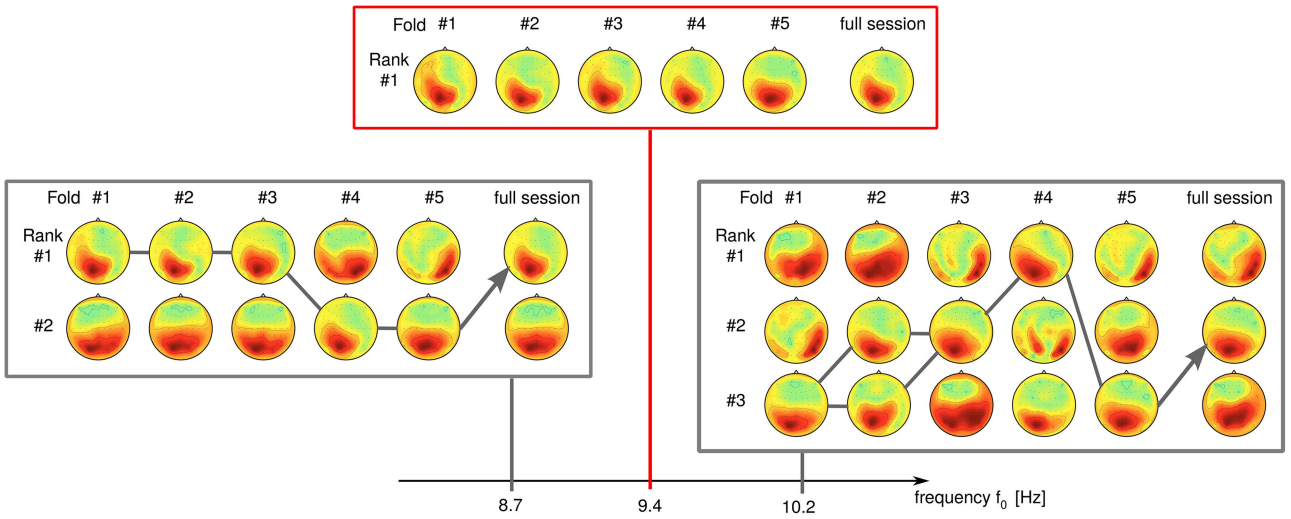


FIGURE 12 | Influence of the frequency band upon the rank stability. While stable at $f_0 = 9.4$ Hz, a component develops rank instability with slight increase/decrease of its frequency band (data of subject S9).

of them resemble patterns emerging for motor imagery tasks in BCI (Krauledat et al., 2008). These often display a clear maximum of activity in channels located over one of the sensorimotor areas (cp. pattern 5 of G1 in **Figure 10** and the pattern of S5 in **Figure 5**) or are located centrally over both hemispheres. While similarity of patterns are by no way a proof for an origin of these oscillatory components in the sensorimotor cortices, the hand force action required to succeed in the SVIPT task would allow for such components.

Other components show a maximum intensity over parietal and occipital areas and may reflect the involvement of the visual system in the SVIPT task. Pattern 2 of **Figure 10** and patterns in **Figure 11A** display a lateralization similar to patterns reported for *directed* and covert visual attention processes (Hanslmayr et al., 2007; Horschig et al., 2014a). Components with a centrally located maximum (cp. pattern 1 of **Figure 10** or the pattern of S9 in **Figure 5**), or with double wing shapes (e.g., pattern 3 of **Figure 10**) resemble components reported for generalized visual attention processes (van Dijk et al., 2008; Meyer et al., 2014). Again, most of these rather clear patterns originate from the alpha frequency band.

While the relevance of several of the selected components cannot be fully interpreted, we do consider these features as added value for neurologists, e.g., by tracking the power time course over sessions for a subject-specific component. Further insight into underlying sub-processes and participating brain areas may be obtained from a *post-hoc* source reconstruction applied upon single SPoC subspaces.

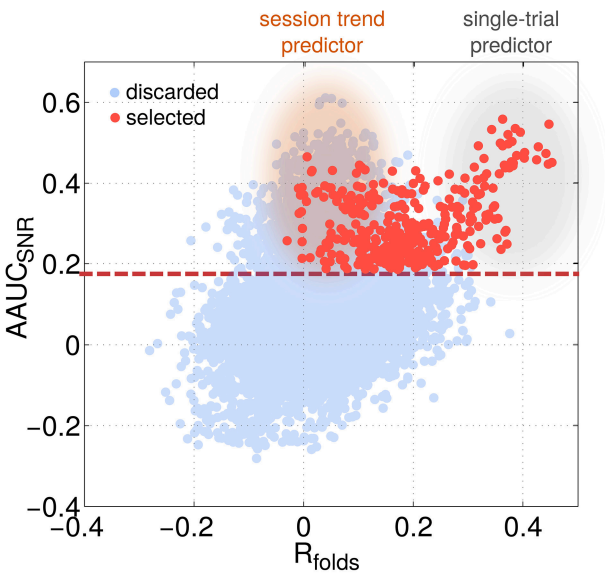


FIGURE 13 | The identification of session trends vs. single-trial variations of performance is possible by localizing a predictor's characteristic with respect to two selection criteria. The scatter plot visualizes $AAUC_{SNR}$ as a function of the mean correlation value across folds R_{folds} for all configurations (light blue) and the selected ones only (red). Two classes of predictors can be identified: single-trial predictors showing a high R_{folds} value while session-trend predictors show a very low R_{folds} value.

4.6. Behavioral Variability on Different Time Scales

Independent of the choice of the exact motor task, subjects generally display two types of performance variations (Chaisanguanthum et al., 2014). First, a large trial-to-trial performance variability is observed from behavioral data. Second, slow performance drifts can occur over the course of a session. Accordingly, SPoC can deliver components, which reflect either one of the two types of performance variations. To tell them apart, a comparison between R_{all} and R_{folds} is helpful. High values for R_{all}, but low for R_{folds} indicate a session trend. If both are high, then the component is informative for

trial-by-trial variation (see *single-trial predictors* and *session trend predictors* in **Figure 13** as well as the examples given in **Figure 5**).

For the purpose of brain-state-informed closed-loop experimenting, single-trial predictors may be more suitable. Session trend predictors, however, may still be useful for pre-cleaning the performance labels. While session trend predictors may reflect an increasing fatigue or a learning effect, it is much harder to determine underlying mechanisms, which cause the rapidly changing trial-to-trial performance of the single-trial predictors (Wu et al., 2014; Hadjiosif and Smith, 2015; Osu et al., 2015). However, our identified components reveal strong evidence that the pre-trial brain activity is partially informative about trail-by-trial variability of motor performance. This in accordance with Churchland et al. (2006) who reported on monkey experiments that at least 30% of behavioral variability could be explained by the fluctuations of preparatory neural activity in the dorsal premotor cortex. However, Chaisanguanthum et al. (2014) stated only a weak relationship between motor cortex activity (PMd/M1) in monkeys and trial-wise fluctuations of behavior.

4.7. Closed-Loop Experimenting as Neuroergonomical Application

The predictive EEG features are extracted from a pre-go interval of each trial. Our pipeline carefully simulated an online scenario, but this approximation of course can not replace the evaluation within a future online study. However, the informative trial-by-trial performance predictors may serve to enhance the neureorgonomical needs of motor rehabilitation scenarios. Since motor performance variability was reported to become larger for stroke patients (Lodha et al., 2010), applying identified patient-specific components within brain-state dependent closed-loop experimenting may enable to causally influence their performance e.g., by manipulating difficulty levels in motor rehabilitation paradigms. So far, BCI methods in stroke rehabilitation (Ang and Guan, 2013) have been used to detect the attempted movement of the affected hand by analyzing informative ERD/ERS features of the EEG and subsequently close the feedback loop for the patient either by triggering a simulated hand movement on a screen (Pichiorri et al., 2015) or by triggering a passive movement of the affected hand, e.g., via an external robotic device or an active orthesis (Ramos-Murguialday et al., 2013).

When implemented in future closed-loop applications, it may be worth to combine SPoC features across multiple frequency bands e.g., by a regression approach. This might allow for enhancing the trial-wise performance prediction, in case the information contained in different frequency bands is independent. Similarly, the combination of predictors based on different performance metrics might serve to gain an enhanced performance estimate.

5. CONCLUSION

In summary, we have shown that the proposed workflow is a suitable basis to identify subject-specific single-trial based neural markers which are predictive for the performance of an upcoming motor task. Those predictors may be valuable building blocks for neuroergonomic applications since they are informative about the status of the visual subsystem as well as the sub-processes involved in hand motor control. Moreover, exploiting those features in future closed-loop experimenting, e.g., by temporal gating of upcoming trials, they will allow for brain-state-informed rehabilitation paradigms. Furthermore, the group-level analysis motivated to utilize our workflow to gain a better understanding of trial-to-trial variations of cognitive sub-processes, which are relevant for a successful rehabilitation outcome.

AUTHOR CONTRIBUTIONS

Conceived and designed the experiments: AM, SC, JR, and MT. Performed the experiments: AM, SC, and MT. Analyzed the data: AM and MT. Contributed reagents/materials/analysis tools: AM, SC, JR, and MT. Wrote the paper: AM and MT.

ACKNOWLEDGMENTS

This work was fully supported by BrainLinks-BrainTools Cluster of Excellence funded by the German Research Foundation (DFG), grant number EXC 1086. The analysis was partly performed on the computational resource bwUniCluster funded by the Ministry of Science, Research and the Arts Baden-Württemberg and the Universities of the State of Baden-Württemberg, Germany, within the framework program bwHPC. The article processing charge was funded by the German Research Foundation (DFG) and the Albert-Ludwigs-University Freiburg in the funding programme Open Access Publishing. The authors would like to thank to Eva-Maria Schlichtmann for her support with the data collection.

REFERENCES

Abe, M. O., and Sternad, D. (2013). Directionality in distribution and temporal structure of variability in skill acquisition. *Front. Hum. Neurosci.* 7:225. doi: 10.3389/fnhum.2013.00225

Allison, B. Z., Brunner, C., Altstätter, C., Wagner, I. C., Grissmann, S., and Neuper, C. (2012). A hybrid ERD/SSVEP BCI for continuous simultaneous two dimensional cursor control. *J. Neurosci. Methods* 209, 299–307. doi: 10.1016/j.jneumeth.2012.06.022

Ang, K. K., and Guan, C. (2013). Brain-computer interface in stroke rehabilitation. *J. Comput. Sci. Eng.* 7, 139–146. doi: 10.5626/JCSE.2013.7.2.139

Blankertz, B., Dornhege, G., Krauledat, M., Müller, K.-R., and Curio, G. (2007). The non-invasive berlin brain–computer interface: fast acquisition of effective performance in untrained subjects. *Neuroimage* 37, 539–550. doi: 10.1016/j.neuroimage.2007.01.051

Bradberry, T. J., Rong, F., and Contreras-Vidal, J. L. (2009). Decoding center-out hand velocity from MEG signals during visuomotor adaptation. *Neuroimage* 47, 1691–1700. doi: 10.1016/j.neuroimage.2009.06.023

Busch, N. A., Dubois, J., and VanRullen, R. (2009). The phase of ongoing EEG oscillations predicts visual perception. *J. Neurosci.* 29, 7869–7876. doi: 10.1523/JNEUROSCI.0113-09.2009

Camus, M., Ragert, P., Vandermeeren, Y., and Cohen, L. G. (2009). Mechanisms controlling motor output to a transfer hand after learning a sequential pinch force skill with the opposite hand. *Clin. Neurophysiol.* 120, 1859–1865. doi: 10.1016/j.clinph.2009.08.013

Castaño-Candamil, S., Höhne, J., Martínez-Vargas, J.-D., An, X.-W., Castellanos-Domínguez, G., et al. (2015b). Solving the EEG inverse problem based on space-time-frequency structured sparsity constraints. *Neuroimage* 118, 598–612. doi: 10.1016/j.neuroimage.2015.05.052

Castaño-Candamil, S., Meinel, A., Dähne, S., and Tangermann, M. (2015a). "Probing meaningfulness of oscillatory EEG components with bootstrapping, label noise and reduced training sets," in *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, (Milan: IEEE), 5159–5162.

Castaño-Candamil, S., Meinel, A., Reis, J., and Tangermann, M. (2015c). P186. correlates to influence user performance in a hand motor rehabilitation task. *Clin. Neurophysiol.* 126, e166–e167. doi: 10.1016/j.clinph.2015.04.282

Chaisanguanthum, K. S., Shen, H. H., and Sabes, P. N. (2014). Motor variability arises from a slow random walk in neural state. *J. Neurosci.* 34, 12071–12080. doi: 10.1523/JNEUROSCI.3001-13.2014

Churchland, M. M., Byron, M. Y., Ryu, S. I., Santhanam, G., and Shenoy, K. V. (2006). Neural variability in premotor cortex provides a signature of motor preparation. *J. Neurosci.* 26, 3697–3712. doi: 10.1523/JNEUROSCI.3762-05.2006

Cohen, R., and Sternad, D. (2009). Variability in motor learning: relocating, channeling and reducing noise. *Exp. Brain Res.* 193, 69–83. doi: 10.1007/s00221-008-1596-1

Coupar, F., Pollock, A., Van Wijck, F., Morris, J., and Langhorne, P. (2010). Simultaneous bilateral training for improving arm function after stroke. *Cochrane Database Syst. Rev.* 4:CD006432. doi: 10.1002/14651858.CD006432.pub2

Dähne, S., Meinecke, F. C., Haufe, S., Höhne, J., Tangermann, M., Müller, K.-R., et al. (2014). SPoC: a novel framework for relating the amplitude of neuronal oscillations to behaviorally relevant parameters. *Neuroimage* 86, 111–122. doi: 10.1016/j.neuroimage.2013.07.079

Daly, I., Scherer, R., Billinger, M., and Muller-Putz, G. (2015). FORCe: Fully online and automated artifact removal for brain-computer interfacing. *IEEE Trans. Neural Syst. Rehabil. Eng.* 23, 725–736. doi: 10.1109/TNSRE.2014.2346621

Dobkin, B. H. (2004). Strategies for stroke rehabilitation. *Lancet Neurol.* 3, 528–536. doi: 10.1016/S1474-4422(04)00851-8

Faisal, A. A., Selen, L. P., and Wolpert, D. M. (2008). Noise in the nervous system. *Nat. Rev. Neurosci.* 9, 292–303. doi: 10.1038/nrn2258

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recogn. Lett.* 27, 861–874. doi: 10.1016/j.patrec.2005.10.010

French, B., Thomas, L. H., Leathley, M. J., Sutton, C. J., McAdam, J., Forster, A., et al. (2007). Repetitive task training for improving functional ability after stroke. *Cochrane Database Syst. Rev.* 4:CD006073. doi: 10.1002/14651858.CD006073.pub2

Fritsch, B., Reis, J., Martinowich, K., Schambra, H. M., Ji, Y., Cohen, L. G., et al. (2010). Direct current stimulation promotes BDNF-dependent synaptic plasticity: potential implications for motor learning. *Neuron* 66, 198–204. doi: 10.1016/j.neuron.2010.03.035

Gonzalez Andino, S. L., Michel, C. M., Thut, G., Landis, T., and Grave de Peralta, R. (2005). Prediction of response speed by anticipatory high-frequency (gamma band) oscillations in the human brain. *Hum. Brain Mapp.* 24, 50–58. doi: 10.1002/hbm.20056

Grech, R., Cassar, T., Muscat, J., Camilleri, K., Fabri, S., Zervakis, M., et al. (2008). Review on solving the inverse problem in EEG source analysis. *J. Neuroengineering Rehabil.* 5, 792–800. doi: 10.1186/1743-0003-5-25

Grosse-Wentrup, M., Harmeling, S., Zander, T., Hill, J., and Scholkopf, B. (2013). "How to test the quality of reconstructed sources in independent component analysis (ICA) of EEG/MEG data," in *International Workshop on Pattern Recognition in Neuroimaging (PRNI)* (IEEE), 102–105. doi: 10.1109/PRNI.2013.35

Hadjiosif, A. M., and Smith, M. A. (2015). Flexible control of safety margins for action based on environmental variability. *J. Neurosci.* 35, 9106–9121. doi: 10.1523/JNEUROSCI.1883-14.2015

Hammer, E. M., Halder, S., Blankertz, B., Sannelli, C., Dickhaus, T., Kleih, S., et al. (2012). Psychological predictors of SMR-BCI performance. *Biol. Psychol.* 89, 80–86. doi: 10.1016/j.biopsycho.2011.09.006

Hammon, P. S., Makeig, S., Poizner, H., Todorov, E., and Sa, V. D. (2008). Predicting reaching targets from human EEG. *Signal Proces. Mag. IEEE* 25, 69–77. doi: 10.1109/MSP.2008.4408443

Hanslmayr, S., Aslan, A., Staudigl, T., Klimesch, W., Herrmann, C. S., and Bäuml, K.-H. (2007). Prestimulus oscillations predict visual perception performance between and within subjects. *NeuroImage* 37, 1465–1473. doi: 10.1016/j.neuroimage.2007.07.011

Haufe, S., Meinecke, F., Görgen, K., Dähne, S., Haynes, J.-D., Blankertz, B., et al. (2014). On the interpretation of weight vectors of linear models in multivariate neuroimaging. *NeuroImage* 87, 96–110. doi: 10.1016/j.neuroimage.2013.10.067

Hoogenboom, N., Schoffelen, J.-M., Oostenveld, R., and Fries, P. (2010). Visually induced gamma-band activity predicts speed of change detection in humans. *NeuroImage* 51, 1162–1167. doi: 10.1016/j.neuroimage.2010.03.041

Horschig, J. M., Jensen, O., van Schouwenburg, M. R., Cools, R., and Bonnefond, M. (2014a). Alpha activity reflects individual abilities to adapt to the environment. *NeuroImage* 89, 235–243. doi: 10.1016/j.neuroimage.2013.12.018

Horschig, J. M., Oosterheert, W., Oostenveld, R., and Jensen, O. (2014b). Modulation of posterior alpha activity by spatial attention allows for controlling a continuous brain–computer interface. *Brain Topogr.* 28, 852–864. doi: 10.1007/s10548-014-0401-7

Horschig, J. M., Zumer, J. M., and Bahramisharif, A. (2014c). Hypothesis-driven methods to augment human cognition by optimizing cortical oscillations. *Front. Syst. Neurosci.* 8:119. doi: 10.3389/fnsys.2014.00119

Hsu, S.-H., Mullen, T., Jung, T.-P., and Cauwenberghs, G. (2015). "Validating online recursive independent component analysis on EEG data," in *Neural Engineering (NER), 2015 7th International IEEE/EMBS Conference on*, (Montpellier: IEEE), 918–921. doi: 10.1109/NER.2015.7146775

Jensen, O., Bahramisharif, A., Oostenveld, R., Klanke, S., Hadjipapas, A., Okazaki, Y. O., et al. (2011). Using brain–computer interfaces and brain-state dependent stimulation as tools in cognitive neuroscience. *Front. Psychol.* 2:100. doi: 10.3389/fpsyg.2011.00100

Jerbi, K., Vidal, J., Mattout, J., Maby, E., Lecaignard, F., Ossandon, T., et al. (2011). Inferring hand movement kinematics from meg, EEG and intracranial EEG: from brain-machine interfaces to motor rehabilitation. *IRBM* 32, 8–18. doi: 10.1016/j.irbm.2010.12.004

Kindermans, P.-J., Tangermann, M., Müller, K.-R., and Schrauwen, B. (2014). Integrating dynamic stopping, transfer learning and language models in an adaptive zero-training erp speller. *J. Neural Eng.* 11:035005. doi: 10.1088/1741-2560/11/3/035005

Krauledat, M., Tangermann, M., Blankertz, B., and Müller, K.-R. (2008). Towards zero training for brain-computer interfacing. *PLoS ONE* 3:e2967. doi: 10.1371/journal.pone.0002967

Kwakkel, G., Kollen, B. J., and Krebs, H. I. (2008). Effects of robot-assisted therapy on upper limb recovery after stroke: a systematic review. *Neurorehabil. Neural Rep.* 22, 111–121. doi: 10.1177/1545968307305457

Lage, G. M., Ugrinowitsch, H., Apolinário-Souza, T., Vieira, M. M., Albuquerque, M. R., and Benda, R. N. (2015). Repetition and variation in motor practice: a review of neural correlates. *Neurosci. Biobehav. Rev.* 57, 132–141. doi: 10.1016/j.neubiorev.2015.08.012

Langhorne, P., Bernhardt, J., and Kwakkel, G. (2011). Stroke rehabilitation. *Lancet* 377, 1693–1702. doi: 10.1016/S0140-6736(11)60325-5

Lemm, S., Blankertz, B., Dickhaus, T., and Müller, K.-R. (2011). Introduction to machine learning for brain imaging. *NeuroImage* 56, 387–399. doi: 10.1016/j.neuroimage.2010.11.004

Lew, E. Y., Chavarriaga, R., Silvoni, S., and Millán, J. D. R. (2014). Single trial prediction of self-paced reaching directions from EEG signals. *Front. Neurosci.* 8:222. doi: 10.3389/fnins.2014.00222

Lodha, N., Naik, S. K., Coombes, S. A., and Cauraugh, J. H. (2010). Force control and degree of motor impairments in chronic stroke. *Clin. Neurophysiol.* 121, 1952–1961. doi: 10.1016/j.clinph.2010.04.005

Logar, V., Škrjanc, I., Belič, A., Karba, R., Brežan, S., Koritnik, B., et al. (2008). Gripping-force identification using EEG and phase-demodulation approach. *Neurosci. Res.* 60, 389–396. doi: 10.1016/j.neures.2007.12.009

Maeder, C. L., Sannelli, C., Haufe, S., and Blankertz, B. (2012). Pre-stimulus sensorimotor rhythms influence brain–computer interface classification performance. *Neural Syst. Rehabil. Eng. IEEE Trans.* 20, 653–662. doi: 10.1109/TNSRE.2012.2205707

Makeig, S., Kothe, C., Mullen, T., Bigdely-Shamlo, N., Zhang, Z., and Kreutz-Delgado, K. (2012). Evolving signal processing for brain–computer interfaces. *Proc. IEEE* 100, 1567–1584. doi: 10.1109/JPROC.2012.2185009

Meinel, A., Castaño-Candamil, J. S., Dähne, S., Reis, J., and Tangermann, M. (2015). "EEG band power predicts single-trial reaction time in a hand motor task," in *Proceedings of International IEEE Conference on Neural Engineering (NER)*, (Montpellier: IEEE), 182–185. doi: 10.1109/ner.2015.7146590

Meyer, T., Peters, J., Zander, T. O., Schölkopf, B., and Grosse-Wentrup, M. (2014). Predicting motor learning performance from electroencephalographic data. *J. NeuroEng. Rehabil.* 11, 24. doi: 10.1186/1743-0003-11-24

Millán, J. D. R., Rupp, R., Müller-Putz, G. R., Murray-Smith, R., Giugliemma, C., et al. (2010). Combining brain–computer interfaces and assistive technologies: state-of-the-art and challenges. *Front. Neurosci.* 4:161. doi: 10.3389/fnins.2010.00161

Müller, K.-R., Tangermann, M., Dornhege, G., Krauledat, M., Curio, G., and Blankertz, B. (2008). Machine learning for real-time single-trial EEG-analysis: from brain–computer interfacing to mental state monitoring. *J. Neurosci. Methods* 167, 82–90. doi: 10.1016/j.jneumeth.2007.09.022

Müller-Putz, G. R., Leeb, R., Tangermann, M., Höhne, J., Kübler, A., Cincotti, F., et al. (2015). Towards non-invasive hybrid brain-computer interfaces: framework, practice, clinical application and beyond. *Proc. IEEE* 103, 926–943. doi: 10.1109/JPROC.2015.2411333

Osu, R., Morishige, K.-i., Nakanishi, J., Miyamoto, H., and Kawato, M. (2015). Practice reduces task relevant variance modulation and forms nominal trajectory. *Sci. Rep.* 5:17659. doi: 10.1038/srep17659

Ovbiagele, B., and Nguyen-Huynh, M. N. (2011). Stroke epidemiology: advancing our understanding of disease mechanism and therapy. *Neurotherapeutics* 8, 319–329. doi: 10.1007/s13311-011-0053-1

Pichiorri, F., Morone, G., Petti, M., Toppi, J., Pisotta, I., Molinari, M., et al. (2015). Brain–computer interface boosts motor imagery practice during stroke recovery. *Ann. Neurol.* 77, 851–865. doi: 10.1002/ana.24390

Pistohl, T., Ball, T., Schulze-Bonhage, A., Aertsen, A., and Mehring, C. (2008). Prediction of arm movement trajectories from ecog-recordings in humans. *J. Neurosci. Methods* 167, 105–114. doi: 10.1016/j.jneumeth.2007.10.001

Ramos-Murguialday, A., Broetz, D., Rea, M., Läer, L., Yilmaz, Ö., Brasil, F. L., et al. (2013). Brain–machine interface in chronic stroke rehabilitation: a controlled study. *Ann. Neurol.* 74, 100–108. doi: 10.1002/ana.23879

Ramoser, H., Muller-Gerking, J., and Pfurtscheller, G. (2000). Optimal spatial filtering of single trial EEG during imagined hand movement. *Rehabil. Eng. IEEE Trans.* 8, 441–446. doi: 10.1109/86.895946

Reis, J., Schambra, H. M., Cohen, L. G., Buch, E. R., Fritsch, B., Zarahn, E., et al. (2009). Noninvasive cortical stimulation enhances motor skill acquisition over multiple days through an effect on consolidation. *Proc. Natl. Acad. Sci. U.S.A.* 106, 1590-1595. doi: 10.1073/pnas.0805413106

Sannelli, C., Dickhaus, T., Halder, S., Hammer, E. M., Müller, K.-R., and Blankertz, B. (2010). On optimal channel configurations for SMR-based brain-computer interfaces. *Brain Topogr.* 23, 186–193. doi: 10.1007/s10548-010-0135-0

Schubert, R., Haufe, S., Blankenburg, F., Villringer, A., and Curio, G. (2009). Now you'll feel it, now you won't: EEG rhythms predict the effectiveness of perceptual masking. *J. Cogn. Neurosci.* 21, 2407–2419. doi: 10.1162/jocn.2008.21174

Tangermann, M., Reis, J., and Meinel, A. (2015). "Commonalities of motor performance metrics are revealed by predictive oscillatory EEG components," in *Proceedings of the 3rd International Congress on Neurotechnology, Electronics and Informatics (NEUROTECHNIX 2015)* (Lisbon), 32–38. doi: 10.5220/0005663100320038

Timmermans, A. A., Seelen, H. A., Willmann, R. D., and Kingma, H. (2009). Technology-assisted training of arm-hand skills in stroke: concepts on reacquisition of motor control and therapist guidelines for rehabilitation technology design. *J. Neuroeng. Rehabil.* 6:1. doi: 10.1186/1743-0003-6-1

Tonin, L., Leeb, R., Sobolewski, A., and del R Millán, J. (2013). An online EEG BCI based on covert visuospatial attention in absence of exogenous stimulation. *J. Neural Eng.* 10:056007. doi: 10.1088/1741-2560/10/5/056007

van Beers, R. J., van der Meer, Y., and Veerman, R. M. (2013). What autocorrelation tells us about motor variability: insights from dart throwing. *PLoS ONE* 8:e64332. doi: 10.1371/journal.pone.0064332

van Dijk, H., Schoffelen, J.-M., Oostenveld, R., and Jensen, O. (2008). Prestimulus oscillatory activity in the alpha band predicts visual discrimination ability. *J. Neurosci.* 28, 1816–1823. doi: 10.1523/JNEUROSCI.1853-07.2008

van Ede, F., Köster, M., and Maris, E. (2012). Beyond establishing involvement: Quantifying the contribution of anticipatory α-and β-band suppression to perceptual improvement with attention. *J. Neurophysiol.* 108, 2352–2362. doi: 10.1152/jn.00347.2012

Waldert, S., Preissl, H., Demandt, E., Braun, C., Birbaumer, N., Aertsen, A., et al. (2008). Hand movement direction decoded from MEG and EEG. *J. Neurosci.* 28, 1000–1008. doi: 10.1523/JNEUROSCI.5171-07.2008

Winkler, I., Brandl, S., Horn, F., Waldburger, E., Allefeld, C., and Tangermann, M. (2014). Robust artifactual independent component classification for bci practitioners. *J. Neural Eng.* 11:035013. doi: 10.1088/1741-2560/11/3/035013

Wolf, S. L., Blanton, S., Baer, H., Breshears, J., and Butler, A. J. (2002). Repetitive task practice: a critical review of constraint-induced movement therapy in stroke. *Neurologist* 8, 325. doi: 10.1097/01.nrl.0000031014.85777.76

Wu, H. G., Miyamoto, Y. R., Castro, L. N. G., Ölveczky, B. P., and Smith, M. A. (2014). Temporal structure of motor variability is dynamically regulated and predicts motor learning ability. *Nat. Neurosci.* 17, 312–321. doi: 10.1038/nn.3616

Yang, L., Leung, H., Plank, M., Snider, J., and Poizner, H. (2014). "Alpha and beta band power changes predict reaction time and endpoint error during planning reaching movements," in *Biomedical Engineering and Informatics (BMEI), 2014 7th International Conference on*, (IEEE), 264–268.

Zaepffel, M., Trachel, R., Kilavik, B. E., and Brochier, T. (2013). Modulations of EEG beta power during planning and execution of grasping movements. *PLoS ONE* 8:e60060. doi: 10.1371/journal.pone.0060060

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Meinel, Castaño-Candamil, Reis and Tangermann. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 28 February 2017
doi: 10.3389/fnhum.2017.00078

Evaluation of a Dry EEG System for Application of Passive Brain-Computer Interfaces in Autonomous Driving

Thorsten O. Zander[1,2], Lena M. Andreessen[1,2], Angela Berg[1], Maurice Bleuel[1], Juliane Pawlitzki[1], Lars Zawallich[1], Laurens R. Krol[1,2] and Klaus Gramann[1,3]*

[1] *Biological Psychology and Neuroergonomics, Technical University of Berlin, Berlin, Germany,* [2] *Team PhyPA, Biological Psychology and Neuroergonomics, Technical University Berlin, Berlin, Germany,* [3] *Center for Advanced Neurological Engineering, University of California San Diego, San Diego, CA, USA*

We tested the applicability and signal quality of a 16 channel dry electroencephalography (EEG) system in a laboratory environment and in a car under controlled, realistic conditions. The aim of our investigation was an estimation how well a passive Brain-Computer Interface (pBCI) can work in an autonomous driving scenario. The evaluation considered speed and accuracy of self-applicability by an untrained person, quality of recorded EEG data, shifts of electrode positions on the head after driving-related movements, usability, and complexity of the system as such and wearing comfort over time. An experiment was conducted inside and outside of a stationary vehicle with running engine, air-conditioning, and muted radio. Signal quality was sufficient for standard EEG analysis in the time and frequency domain as well as for the use in pBCIs. While the influence of vehicle-induced interferences to data quality was insignificant, driving-related movements led to strong shifts in electrode positions. In general, the EEG system used allowed for a fast self-applicability of cap and electrodes. The assessed usability of the system was still acceptable while the wearing comfort decreased strongly over time due to friction and pressure to the head. From these results we conclude that the evaluated system should provide the essential requirements for an application in an autonomous driving context. Nevertheless, further refinement is suggested to reduce shifts of the system due to body movements and increase the headset's usability and wearing comfort.

Keywords: autonomous driving, passive BCI, EEG, usability, ERP

OPEN ACCESS

Edited by:
Mikhail Lebedev,
Duke University, USA

Reviewed by:
Nima Bigdely-Shamlo,
Qusp, USA
Tjeerd W. Boonstra,
University of New South Wales, Australia
Tomas Emmanuel Ward,
Maynooth University, Ireland

****Correspondence:***
Thorsten O. Zander
tzander@gmail.com

Received: *08 January 2016*
Accepted: *08 February 2017*
Published: *28 February 2017*

Citation:
Zander TO, Andreessen LM, Berg A, Bleuel M, Pawlitzki J, Zawallich L, Krol LR and Gramann K (2017) Evaluation of a Dry EEG System for Application of Passive Brain-Computer Interfaces in Autonomous Driving. Front. Hum. Neurosci. 11:78. doi: 10.3389/fnhum.2017.00078

INTRODUCTION

Driving has become a part of everyday life, which makes the drive to work or for recreational activities a simple routine task. However, the effects of the mental workload and effort required by driving often go unnoticed. A study by Borghini et al. (2014) found that mental workload, fatigue, and drowsiness are significantly increased while driving. Especially long periods of constant driving, as often performed by professional truck drivers, result in an accumulation of these effects over time, decreasing the driver's cognitive capabilities and driving performance, thus increasing the chances of traffic accidents.

The field of automotive human factors and ergonomics is concerned with minimizing safety risks depending on human performance in driving tasks. Today, many automations and small devices have found their way into cars in order to help reduce the mental workload required to operate the vehicle (Young and Stanton, 1997; Tadaka and Shimoyama, 2004; Ma and Kaber, 2005). A different approach aims to fully or at least partly automate the task of driving, so the human driver can be eliminated as a risk factor in most instances. The scientific field working toward this goal is called *Autonomous Driving* (Franke et al., 1998) and has grown more important over the past years.

One particular problem with autonomous driving is the question of responsibility: Who is accountable in case of an accident? Most countries still define the human driver of a car as the entity responsible for anything that happens while driving (Beiker, 2012). Therefore, experts believe it would be best to only automate some of the tasks that arise while driving, but to leave the most complex tasks to a human driver for the time being. According to Sukthankar et al. (1997), the task of driving consists of different levels, which are the strategic level (route planning), the tactical level (maneuver selection), and the operational level (maneuver operation). Automation of the lowest, operational level is thus legally the least complex, and also technically possible (Dickmanns and Zapp, 1987; Pomerleau, 1992). Driving along a highway could relatively easily be automated, but once the traffic situation changes, the human may be required to take over control. This approach thus requires an important exchange of information between the human driver and the automated system: The human must be timely and appropriately informed of the pending takeover. As stated by Llaneras et al. (2013), people tend to focus their attention on secondary tasks once the primary objective of driving has been taken over by automation. As a consequence, in a situation where the car drives autonomously, a signal given by the system to indicate the necessity for takeover might be missed, or might catch the human by surprise. This may result in loss of control over the vehicle.

As a solution to the above problem, the car could monitor the driver's mental state, and adapt the notification process to the current context. A completely attentive driver might quickly perceive and understand even simple signals, whereas for example a sleeping driver may need to be woken carefully by the car in advance of leaving the highway. *Passive brain-computer interfaces* (passive BCIs, Zander and Kothe, 2011) are promising approaches for such monitoring and automated adaptation (Zander et al., 2011). This technology enables real-time detection of mental conditions like fatigue, workload, and degree of relaxation (Blankertz et al., 2010; Gerjets et al., 2014), which offer a good estimate of whether or not the driver is ready to take over control of the car. But the passive BCI approach during autonomous driving is not limited to this. More general information—like mood or situational awareness—and also very specific information about the subjective interpretation of the current context—that might be reflected in the driver's brain as error responses—could be assessed by the passive BCI (Zander and Jatzev, 2012). This information could then be integrated in the autonomous decisions of the car. The car learns how the driver interprets the context and gains a degree of context-awareness by utilizing the driver's brain as a sensor.

Passive BCIs are commonly based on electroencephalography (EEG). Traditional EEG systems are relatively cumbersome to apply and use, requiring preparation of the skin, application of conductive gel, and cleaning of the cap afterwards. To make EEG applicable for non-scientific uses, e.g., to be used by drivers, its application and handling needs to be as easy as possible. This is why alternative electrode systems (e.g., described in Zander et al., 2011; Liao et al., 2012) are an important focus of autonomous driving related BCI research. Primarily, the use of gel is eliminated, and the caps containing the electrodes are made for quick application, resulting in less preparation time and, ideally, more comfortable for the wearer. Recent laboratory studies provided evidence of good signal quality, comparable to that of standard gel-based electrodes. It is still unclear however that the signal quality can be maintained in real-world contexts.

This study focused on evaluating the use and application of a dry electrode EEG system in the context of a running vehicle. It was assessed how easy it is for untrained person to apply the system on their own head, how well the electrodes can be positioned and remain in place, and whether the signal quality is sufficient for BCI usage when the system is self-applied. Two common features in the EEG, an N200-P300 ERP and the parietal alpha rhythm, were analyzed as examples of signals that potentially can be used in a passive BCI application. Furthermore, interference in the EEG signal resulting from usage inside a running car—a noisy environment—was investigated. Finally, wearing comfort over a prolonged period of time as well as general user acceptance were evaluated.

MATERIALS AND METHODS

Participants

Ten participants, five male, participated in the experiment. The mean age was 28 years ($SD = 3.4$). Two participants reported to have sensitive skin. All participants gave their written informed consent to participate in the study and were paid 20 euros as expense allowance. The overall duration of the experiment was on average 165 min ($SD = 39$ min.), including breaks.

Apparatus

Vehicle

The vehicle we used to evaluate the influence of vehicle-induced noise on the recorded EEG was a Volkswagen Touran (year of manufacture 2003). The car was stationary during the experiments, but had the engine running, the radio switched on (though muted), and the air conditioning enabled. A 7.6″ TFT-display was mounted to the right of the steering wheel near the center console.

Experimental Room

The experimental room used for baseline recordings was a non-frequented room at the TU Berlin with constant light, right next to the parked car. Diversions and disturbances were kept to a minimum.



FIGURE 1 | Overview of the used EEG system, the Brain Products actiCAP Xpress. Image courtesy of Brain Products GmbH.

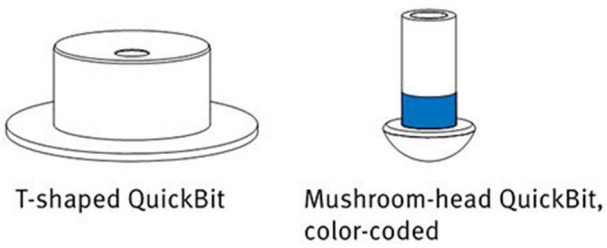


FIGURE 2 | The different QuickBit types provided with the actiCAP Xpress. Image courtesy of Brain Products GmbH.

Computer System

The EEG system was connected to a laptop (Sony Vaio Z, 2012) and EEG data was recorded using the *BrainVision Recorder, BrainVision RDA* (Brain Products GmbH, Munich, Germany), and *LabRecorder* (as part of the BCILAB framework, Delorme et al., 2010). The experimental paradigms were run using *SNAP*[1] (Iversen and Makeig, 2013). To analyze the data, we used the EEGLAB toolbox (Delorme and Makeig, 2004), an open source toolbox embedded in MATLAB. For classification we used the open source toolbox BCILAB (Kothe and Makeig, 2013), also embedded in MATLAB.

EEG System

The system examined in this study was the *Brain Products actiCAP Xpress* dry-electrode EEG system (see **Figure 1**) provided by Brain Products GmbH for the duration of the experiment. The system included 16 active data electrodes plus one reference and one ground electrode. Electrodes were applied to one of two differently-sized flexible caps, depending on the head circumference of the participant (52–58, or 58–64 cm). To ensure fixation on the participant's head, a chin belt was attached to the cap. Each cap provided 78 possible electrode positions most of the extended international 10% system, with additional options to set up regions of interest. We used electrode positions Fp1, Fp2, Fz, FC5, FC6, C3, C4, Cz, CPz, Pz, CP5, CP6, PO3, PO4, POz, and Oz.

To adjust the system to an individual participant, the electrodes can be extended to different shapes and sizes by attaching so-called *QuickBits* (see **Figure 2**). The kit used in the study came with six T-shaped flat tips (with a diameter of 7 mm) to be attached to the forehead and earlobes, as well as 32 mushroom-head tips for application on the scalp. These latter come in different lengths of 8, 10, 12, and 14 mm, which can be attached to the electrodes according to head shape and required pressure. This enabled a personalization of the system: Optimal sensor lengths for electrode positions can be noted, stored and re-applied in follow-up experiments.

Prior to applying the *actiCAP Xpress*, the electrodes were cleaned using a disinfectant spray. This was done even in case the electrodes and sensors had not been used before to remove dust and particles to improve connectivity.

The electrode array was connected to a *V-Amp* EEG signal amplifier (Brain Products GmbH, Munich, Germany), which in turn was connected to a laptop computer through a universal serial bus (USB) 2.0.

Experimental Procedure

Experimental Rationale

This study was designed to assess different requirements to an EEG system for application in real-world driving scenarios. We defined the following requirements: (1) self-applicability of the system, (2) impact of interfering noise signals inside a running vehicle on EEG signal quality, (3) stability of cap and electrode positions after context-related movements, and (4) usability and wearing comfort of the system.

The experiment was divided into four blocks covering these four issues, answering the following questions.

1. How easy and accurate is self-application of the system in comparison to application by another person? (Block I)
2. How strong is the effect of interfering signals in a running car on EEG recording? (Block II)
3. How do electrode positions change during typical body movements inside a car? (Block III)
4. How do participants rate the system's usability? (Block IV)

Figure 3 summarizes the experimental session. After arrival of the participant, the experiment was explained and a demographic survey was conducted. While the cap was personalized by the investigator by exchanging electrode tips where necessary, the participant was asked to read the instruction manual of the system, in preparation for Block I.

Block I: Self-application

Self-application of the cap, as opposed to having the cap fitted to you by a trained operator, may take a different amount of time and may affect the positioning of the electrodes and the signal quality. To estimate these effects, we compared cap application

[1]Simulation and Neuroscience Application Platform (SNAP). Available: https://github.com/sccn/SNAP.

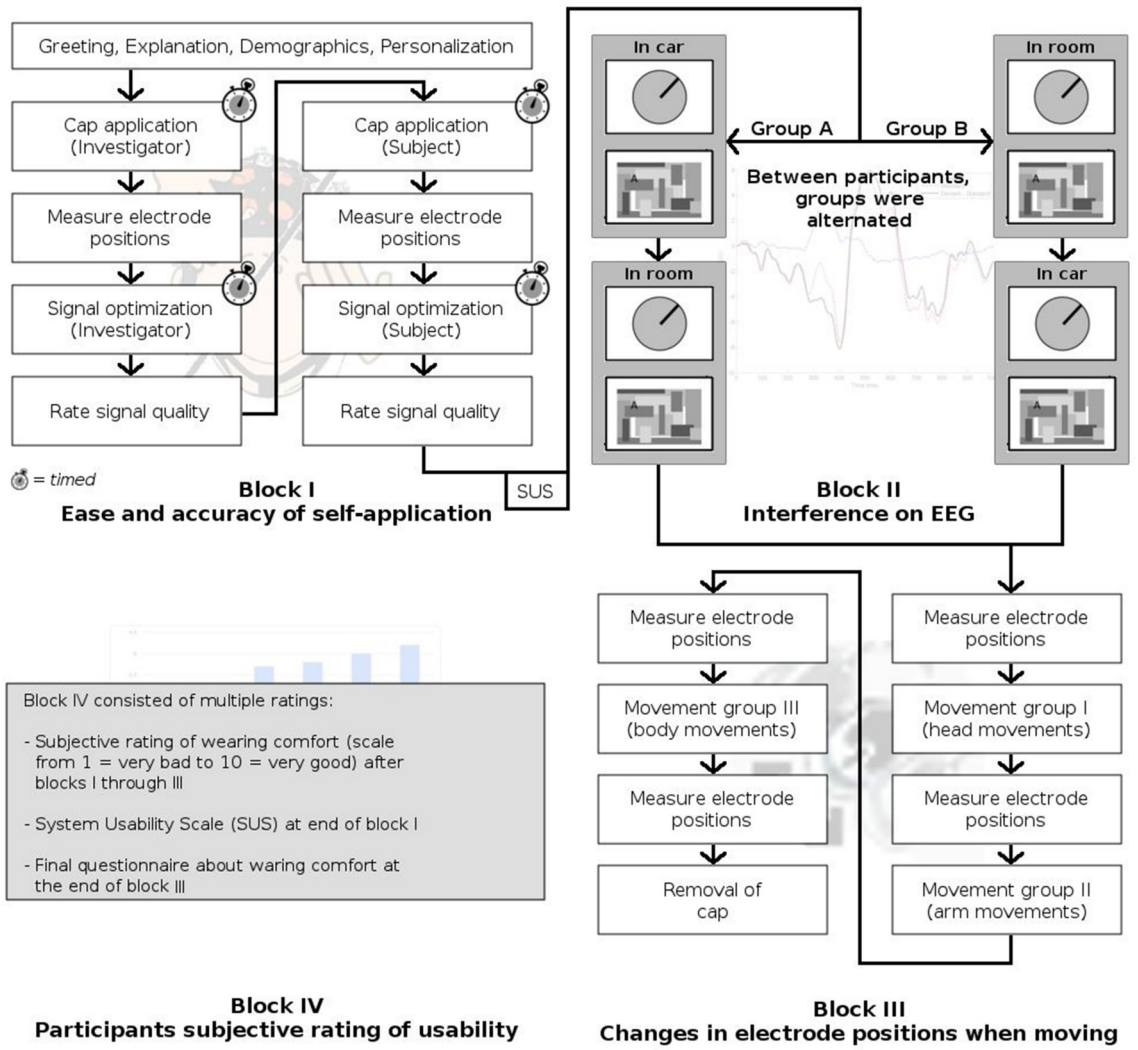


FIGURE 3 | Experiment timeline.

in two conditions: Application by the experimenter, and self-application by the participant. Customization of the cap was not included here, as it is assumed to be a one-time effort.

Participants were seated in the experimental room, in front of a laptop. A stopwatch was used to first measure the time required by the experimenter to apply the EEG cap to the participant's head.

Once the cap and ground/reference electrodes were in place, electrode positions were measured using the *Polaris Vicra* system (Northern Digital Inc., Waterloo, ON, Canada), allowing for measuring 3-dimensional electrode locations. We chose to record the 16 electrode positions, as well as the inion, the nasion and the left and right preauricular points. The latter three were used as coordinate references to allow the transformation of coordinates taken from different measuring sessions into one coordinate system to allow comparison (described below in the section "Analysis Procedures"). To achieve comparable, stable positions for the reference points in each measurement during the experiment, we marked them by a small dot on the respective positions on the participant's skin using a removable eudermic marker.

Following this, signal quality was optimized by relatively fine-grained adjustments to the electrodes. As the system did not provide an objective measure of signal quality or electrode contact (e.g., impedance), signal quality was assessed visually. The signal was monitored using the *BrainVision Recorder* software, with all 16 channels displayed at once, set to a resolution of 50 μV. A *display filter* was enabled, bandpass-filtering the visible signal from 0.1 to 40 Hz, not affecting the recording. The duration of this optimization was again timed using a stopwatch. The resulting signal quality was also recorded, as rated by the experimenter. The indication for signal quality was the visual form of the signal on the display, artifacts had to be recognized visually. The rating followed predefined guidelines and was done

on a 5-point scale with 5 meaning "perfect signal" and 1 meaning "no signal at all" (see **Figure 4**). This rating was done twice: Once for the signals with the display filter switched on, and once based on the unfiltered raw signal.

Following this, the cap was removed and participants, who read the instructions manual, were asked to put on the cap by themselves, after all of their questions about the procedure had been answered by the experimenter. Application time was again measured, as were the electrode positions and the resulting signal quality.

Block II: EEG Recording

For investigating signal quality in standard EEG analyses we chose the well-known N200 and P300 components of the visual event-related potential and the parietal alpha rhythm. Both time- and frequency domain parameters are well-examined phenomena in EEG research. Hence, clear expectations about morphology, topography and signal strength can be drawn, that build the baseline of comparison for our results.

In order to assess the EEG signal and the possible influence on it of the electromagnetically noisy environment that is the car, participants performed in two established experimental paradigms of BCI research (Zander et al., 2011), once in the experimental room, and once inside the car. The order of these two conditions was randomized between participants.

The first paradigm focused on the elicitation of visual event-related potentials (ERPs) using a standard oddball approach: An infrequent *deviant* stimulus sometimes appeared instead of the frequent *standard* stimuli (Duncan-Johnson and Donchin, 1977; see **Figure 5**). This is a common approach when researching ERPs referred to as the N200-P300 complex (Polich and Kok, 1995; Linden, 2005). ERP detection during autonomous driving can be useful, as they may allow a car to detect how drivers react cognitively to perceived stimuli/information.

On the screen, participants saw a circle divided by lines into 30° angles. First, a bar appeared, like a clock's arm pointing 12 o'clock. This bar then rotated clockwise in discrete steps, once every second. A standard stimulus had it rotate by 90°; a deviant consisted of an initial 60° rotation, followed by a 100 ms pause and a 15° counterclockwise rotation. After each deviant, the bar disappeared and reappeared at the 12 o'clock position.

10% of all stimuli were deviants. In total 400 trials were displayed (360 standard, 40 deviant).

The second paradigm focused on features in the spectral domain, specifically the parietal alpha rhythm. This feature is of special interest to autonomous driving, as parietal alpha can be used as an indicator of whether the participant is currently in a relaxed state or performing some mentally demanding task (Berka et al., 2007). It also is a standard example for features in the spectral domain.

The paradigm (see **Figure 6**) presented to the participant was designed to induce changes in parietal alpha activity by alternating between two states of mind: *Engaged* and *relaxed*. To engage the participant, a six-letter word was presented letter

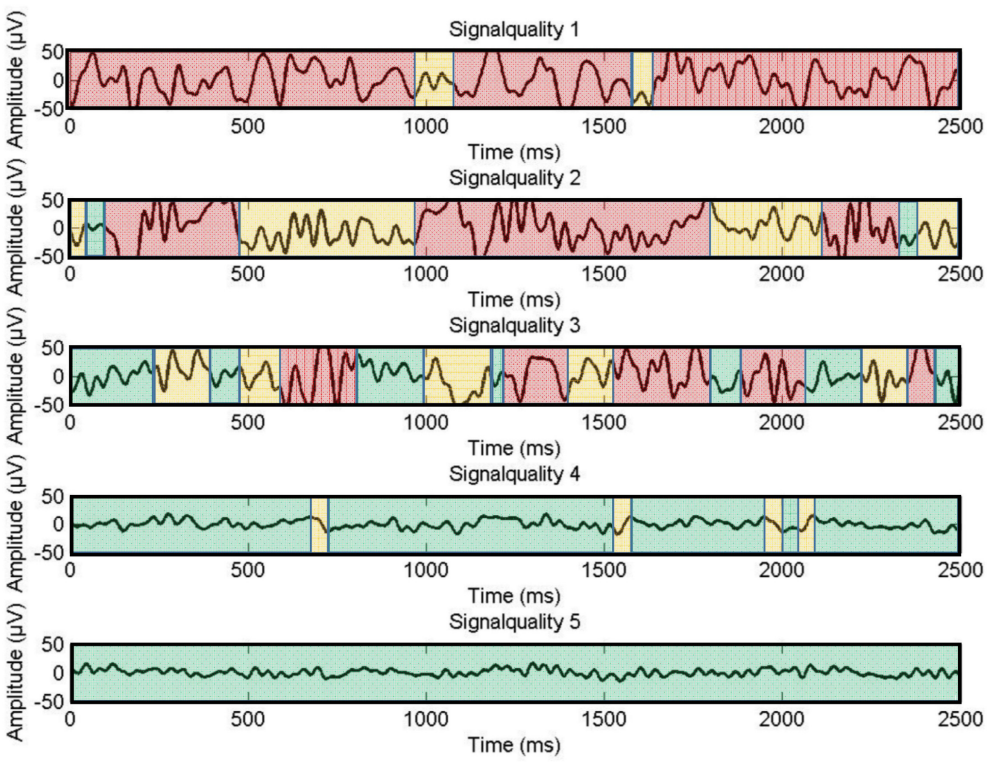


FIGURE 4 | Examples for signal quality ratings on a scale from one to five. Green colored parts indicate adequate signal quality, yellow parts moderate signal quality, and red parts unacceptable signal quality.

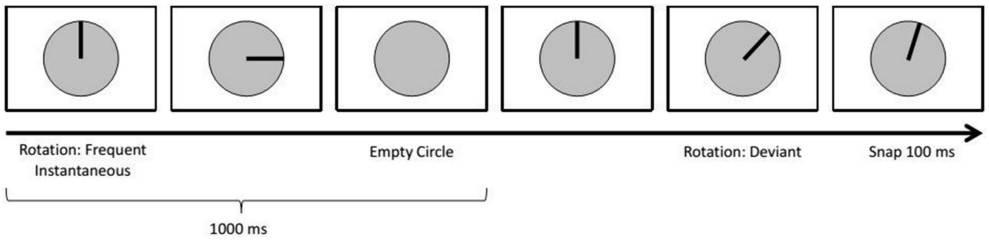


FIGURE 5 | Oddball Paradigm.

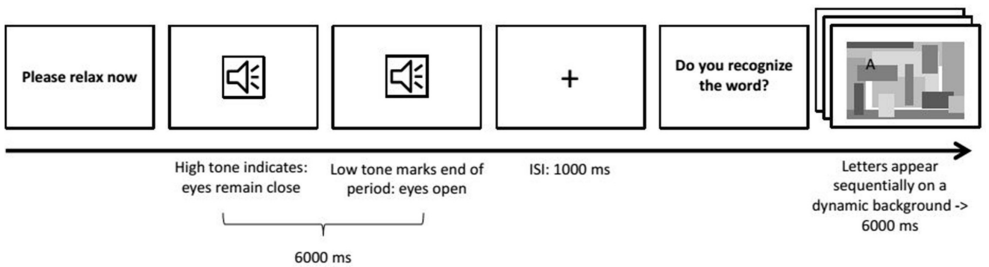


FIGURE 6 | Induced Alpha Paradigm.

by letter, with letters appearing on random locations on the screen amidst visual noise. Each letter was only visible for 1 s. Participants were instructed to read the word. After each engagement trial, the participant was instructed simply to relax for 6 s with their eyes open. This relaxation phase was introduced using an auditory signal and ended by a similar one with lower pitch.

There were 32 trials of each condition. The order of words in the engaged condition was randomized across participants.

These two paradigms were presented in fixed order to the participants in the two conditions (room vs. car).

Block III: Driving-Related Movements

The third block investigated the influence of movements on the position of the electrodes.

Electrode positions were recorded, using again the Polaris system mentioned earlier, at the start of this block. Participants then performed a series of three different types of driving-related movements inside the car, and the electrode positions were measured again after each group of movements. Because measurements were not done inside the car but in a nearby room, some walking was required. Electrode cables were bundled together and fixed to the participant's clothing in a relaxed way to minimize their strain on the cap while walking.

To make movements comparable between participants, we placed markers (sticky notes) at certain places in the car: One on the left rear window, one above the driver's seat, one in the legroom of the front passenger seat and one in the center of the rear bench seat. Before seating the participant in the driver's seat, the markers were shown to them. The EEG system was not connected to the amplifier during the movements. All instructions for different movements were given through pre-recorded audio files played back using a laptop and speakers inside the car.

Block IV: Usability

To assess the usability of the system, the participants were asked to fill out a questionnaire right after Block I. This questionnaire was the *System Usability Scale* (SUS; Brooke, 1986) was employed, also used in other BCI related studies prior to this one (Pasqualotto et al., 2011; Duvinage et al., 2012). SUS is a standardized questionnaire consisting of ten questions based on Likert scales with five options ranging from "strongly disagree" to "strongly agree." In total, SUS contains five positively and five negatively formulated questions about the system being assessed, for example "I think that I would like to use this system frequently" or "I found the system unnecessarily complex." From the answers given, a SUS score is calculated, ranging between 0 (worst possible system) and 100 (best possible system). This score has to be interpreted taking the individual context of system usage into account. In contrast to qualitative assessments, the SUS does not yield any insight into which usability problems exactly are present within the system. It provides however a quick and reliable way to determine whether or not major changes are necessary in order to make the system safe and comfortable to use.

Additionally, the participants were asked to rate the level of comfort wearing the system after each of the previously described experimental blocks (I–III) on a scale from 1 to 10, one meaning "extremely bad" and ten "very comfortable." We acquired these three subjective impressions to gather insight into how the system's perceived comfort changed over the course of the experiment.

To get an even deeper insight into the comfort of wearing the system, participants were asked to fill out another questionnaire after the third experimental block, after roughly 140 min of wearing the system almost constantly. We adapted a questionnaire for the evaluation of the wearing comfort for firemen helmets (Fabrizio and Cimolino, 2014), by only keeping questions deemed fitting to our context. All questions were rated on a five point Likert scale. In addition to these questions, we asked two yes-no questions: Whether or not the participant believed the cap had moved, and whether or not it induced the feeling of dents on their head. Finally, we asked the participants to mention any discomfort associated with wearing the system, like the feeling of pressure on the head, headaches, or nausea.

Analysis Procedures

Block I: Self-application

Comparison of time needed by the experimenter and the participant to apply the system and to adjust the electrodes was done by two-sample *t*-tests.

The signal quality ratings were subjected to a three-way mixed measures ANOVA with the two within-subject factors visual filters (no filters vs. 0.1–40 Hz bandpass) and electrode (Fp1 vs. Fp2 vs. vs. Fz vs. FC5 vs. FC6 vs. C3 vs. C4 vs. Cz vs. CPz vs. Pz vs. CP5 vs. CP6 vs. PO3 vs. PO4 vs. POz vs. Oz) and the between-subject factor applicant (investigator vs. participant).

Because a total of six different measurements of electrode positions were taken during the course of this experiment, these measurements were first transformed into one coordinate system to allow a unified comparison. To this end, all measurements were re-referenced to a mean head middle and radius, within participants, as follows.

1. All coordinates of recording j, $j = 1, ..., 6$ were referenced to the head midpoint hm_j, which is calculated with the recorded reference points (nasion n_j and left and right preauricular points, lp_j and rp_j) by

   a. Drawing a line through both preauricular points lp_j and rp_j:

      Calculate the slope by computing new coordinates

      $$(u_j)_i := (lp_j)_i - (rp_j)_i, \text{for } i = 1, 2, 3 \text{ denoting the scalars of the three-dimensional vector } u_j.$$

      Define the line by

      $$g_j := lp_j + r_j u_j \text{ with } r_j \text{ to be determined.}$$

   b. Construction of a plane H_j through n_j, which is perpendicular to the line g_j:

      Find the variables x, y, z to determine the plane equation for H_j

      $$H_j : (u_j)_1 x + (u_j)_2 y + (u_j)_3 z := e.$$

      To find e, insert the coordinates of the nasion reference point n_j into the equation

      $$H_j(n_j) : (u_j)_1 (n_j)_1 + (u_j)_2 (n_j)_2 + (u_j)_3 (n_j)_3 = e.$$

   c. For the purpose of finding the intersection of the line g_j with the plane H_j, insert the coordinates of g_j into the plane equation above and solve for r_j:

      $$H_j(g_j) : r_j = \frac{e - (u_j)_1 (lp_j)_1 - (u_j)_2 (lp_j)_2 - (u_j)_3 (lp_j)_3}{(u_j)_1^2 + (u_j)_2^2 + (u_j)_3^2}.$$

      Inserting r_j into the plane equation yields the head midpoint:

      $$hm_j = lp_j + r_j u_j.$$

2. After calculating the head midpoints hm_1 to hm_6, we compute the arithmetic average $\overline{hm}$ over all recordings as the final reference point in order to minimize the error of measurement in the system.
3. The deviation of the recorded head midpoint hm_j to $\overline{hm}$ is calculated for each recording:

   $$d_j := hm_j - \overline{hm}, \; j = 1, ..., 6.$$

4. Then, all recorded electrode positions $(ep_k)_j$, $k = 1, ..., 16$ are re-referenced to $\overline{hm}$ by addition with d_j and the euclidean distance $ed_{j_1 j_2}$ between different recordings j_1, j_2 is calculated:

   $$(d_{j_1 j_2})_i := ((ep_k)_{j_1} + d_{j_1})_i - ((ep_k)_{j_2} + d_{j_2})_i,$$
   $$ed_{j_1 j_2} := \sqrt{(d_{j_1 j_2})_1^2 + (d_{j_1 j_2})_2^2 + (d_{j_1 j_2})_3^2}$$

The value used for comparison of different recordings j_1, j_2 was this euclidean distance ed_{j_1, j_2}.

For Block I, recorded positions from the investigator-applied cap were compared to the positions from the self-applied cap. Mean differences of electrode positions were then compared to the expected value of no difference in positions using a one-sample *t*-test against zero.

Block II: EEG Recordings

Oddball paradigm: ERP analysis

EEG data was first preprocessed by applying a bandpass-filter from 1 to 30 Hz, retaining all frequencies relevant for later analyses. Then, epochs of 1100 ms were extracted, starting 100 ms before stimulus onset of the standard and deviant events. Baseline correction was performed with a 100 ms pre-stimulus interval.

To compare event-related activity between car and indoor recordings, amplitudes and latencies of the N200's and P300's were extracted.

First, the indoor condition was used as a baseline as it conforms to laboratory conditions. Inspection of the grand average revealed a global negative minimum at 300 ms over the centro-parietal lead (Pz) and a global positive maximum at 400 ms over the centro-central lead (Cz). Based on these peaks, a search window was defined around 300 ± 70 and 400 ± 70 ms to search for maxima in the individual averages. Once for each individual the global peaks were identified, the peaks on individual channels were identified using a ± 25 ms

window around the individual global peak. Mean amplitudes and latencies were extracted for all channels. This procedure resulted in a 4 x 16 vector for each participant, consisting of the mean amplitudes and the latencies of the two components at each channel.

For comparison of mean peak amplitudes two repeated measures ANOVAs were performed. Mean amplitudes from electrode Pz were used for the negativity and from Cz for the positivity. Each 2x2 ANOVA had the two within-participant factors recording condition (indoor vs. car) and stimulus (standard vs. deviant).

In order to examine disparities of mean peak latencies between conditions (indoor vs. car), mean difference peak latencies were calculated by subtracting the negative from the positive peak latency. The mean difference was taken per participant for the two conditions and subjected to a paired sample *t*-test.

To test for equivalence of EEG measures between recording conditions the two one-sided tests (TOST, Schuirmann, 1981, 1987; Westlake, 1981) procedure was applied to mean peak amplitudes and mean difference peak latencies with an epsilon of the standard deviation of the indoor condition, which was regarded as the control group (R-package "equivalence" May 14, 2016; V0.7.2). A *p*-value of 0.05 was taken as the significant threshold for all TOST.

Induced alpha paradigm: frequency analysis

To compare oscillatory features between car and indoor recordings, three different measures were taken: The power spectral density function covering 0.1–40 Hz, single measurements of the band power in the alpha band, and the time course of the alpha band power during the 6-s trials of the paradigm (engaged vs. relaxed).

Fluctuations in alpha power occur with a broader distribution over posterior areas of the scalp (Sauseng et al., 2005). Since we were interested in parietal alpha as potential indicator of mental load, analyses were restricted to five posterior electrodes, namely Pz, PO3, PO4, POz, and Oz. The data was bandpass filtered from 0.1 to 40 Hz and time epochs of 6 s were selected, covering each full trial.

Power spectral densities (PSD) were calculated for each entire epoch and averaged per participant, resulting in 2 x 2 x 5 PSD distributions for each participant (2 experimental conditions x 2 mental states x 5 channels). We used these participant-individual PSDs as well as the averaged PSDs over all participants (grand average), resulting in a total of 11 (2 x 5+1) PSD-distributions for each experimental condition.

Individual and grand average Pearson Correlation of the PSD in the frequency band of 0.1 Hz to 40 Hz were calculated for each electrode between indoor and car conditions and tested for significance using one sample *t*-tests against zero.

The alpha band (7–13 Hz) being of prime interest here, we also calculated a single bandpower value in this frequency range for each participant, electrode, and trial. We used epochs of 4 s length, starting 2 s after stimulus onset. Logarithmic variances of each trial per electrode of each participant were calculated and normalized with the maximum value of each electrode. These measures were then averaged over all trials, resulting in a normalized mean alpha band power for each participant under each experimental condition on the five investigated electrodes. Effects between recording conditions, stimuli and electrodes were investigated in a 2 x 2 x 5 ANOVA with the three within-participant factors recording condition (indoor vs. car), stimulus (standard vs. deviant) and electrode (Pz vs. PO3 vs. PO4 vs. POz vs. Oz). The factor electrode is a repeated measure here as EEG measures at one electrode depend on values measured by other electrodes. Again, the TOST procedure with an epsilon of the standard deviation of the indoor condition was applied to normalized mean alpha band power values to test for equivalence between recording conditions.

As a third measure, the time course of the band power in the alpha band was used. It was calculated by shifting a 500 ms window over each single trial and calculating the band power for each window position. To avoid leakage effects, the window was multiplied with a Gaussian bell curve of the same size. Afterwards the single-trial measurements were normalized with the mean of all band powers. The normalized measurements were averaged, resulting in 2 x 5 time courses for each participant (2 experimental conditions x 5 channels). As above, we also took the grand average into account, resulting in 11 time courses in total per experimental condition.

To examine the difference in the time course of the band power in the alpha range between conditions, Pearson Correlations were calculated for each participant, channel and condition.

BCI Analysis of both paradigms

BCILAB's built-in classification approaches were used to evaluate the offline single-trial accuracies as an estimate of potential online performance.

For the oddball paradigm, data was bandpass filtered from 0.1 to 15 Hz and downsampled to 100 Hz. Epochs of 800 ms were extracted starting at each stimulus marker. A windowed-means approach (Blankertz et al., 2011) was used to extract features, using 8 consecutive windows of 50 ms starting at 300 ms post-stimulus. As a classifier we used *linear discriminant analysis*, LDA (Webb, 2002). Mean ERP classification error rates of all eight participants were subjected to a paired samples *t*-test.

Logarithmic band power was used for feature extraction (Solis-Escalante et al., 2010; Zander et al., 2011) of the data of the second paradigm. This was applied to epochs of 6 s, as above. We performed a (10 x 10)-fold cross-validation, and classified using LDA. Mean classification error rates were subjected to a paired samples *t*-test.

Classification error rate results from both paradigms were subjected to a TOST procedure with an epsilon of the standard deviation of the indoor condition to test for equivalence between recording conditions.

Block III: Driving-Related Movements

Each of the three movement groups had one electrode position measurement before, and one after it. Mean differences of electrode positions prior to and after each movement group were compared to the expected value of no difference in positions using a one-sample *t*-test against zero.

Block IV: Usability

The *System Usability Scale* was interpreted following the guidelines set by Brooke (1986). To determine the resulting SUS score of the system, all given answers were weighted accordingly and added up. This resulted in a total score per participant, which then was multiplied by the factor 2.5.

After experimental blocks I to III, participants were asked to give a subjective estimate of how comfortable the system felt. The median of the comfort ratings of all participants was used as the overall comfort rating here. To test for differences between the three time points, a Wilcoxon Signed-rank test was applied. The wearing comfort questionnaire was evaluated descriptively.

RESULTS

Block I: Self-application

Application Time

A two-samples t-test indicated that the mean time needed for application of the cap did not differ significantly between experimenter ($M = 123.2$ s, $SD = 43.8$) and participants ($M = 104.9$ s, $SD = 49.0$), $t_{(9)} = 0.880$, $p = 0.391$, though showing a tendency that participants perform faster. Mean times needed for adjustment of electrodes also did not differ significantly between investigator ($M = 256.3$ s, $SD = 221.3$) and participants ($M = 310.2$ s, $SD = 285.1$), $t_{(9)} = 0.472$, $p = 0.642$, showing a tendency that experimenters are faster.

Electrode Signal

The three-way mixed measures ANOVA on signal quality ratings revealed no significant main effect of applicant, $F_{(1, 18)} = 0.341$, $p = 0.341$, $\eta^2 = 0.019$. The main effect of filter was significant, $F_{(1, 18)} = 66.861$, $p = 0.000$, $\eta^2 = 0.788$. Since the main effect of electrode violated the assumption of sphericity Greenhouse-Geisser corrected values were used. The main effect electrode was significant, $F_{(5.167, 93.012)} = 2.876$, $p = 0.017$ $\eta^2 = 0.138$. None of the interaction effects were significant, all $ps > 0.281$.

Electrode Positions

The t-test against zero performed on mean differences of electrode positions ($M = 13.76$ mm, $SD = 5.12$ mm) between investigator- and self-applied cap yielded significance, $t_{(9)} = 8.498$, $p = 0.00001$. The electrode positions varied most on the midline of the head, with 15.5 mm variation (averaged over all 10 participants) at Oz to 16.1 mm averaged variation at Fz. This could be due to the structure of the cap: It has two holes for the ears, so electrodes in this area are fixated more clearly than electrodes elsewhere. Electrodes on the forehead can be positioned up to 1 cm higher or lower without any obvious effects on the cap like inconvenience or ill-fittingness, so it was hard for both participants and investigators to position the cap correctly around the midline of the head (see **Figure 7**).

For Block I, recorded positions from the investigator-applied cap were compared to the positions from the self-applied cap. Mean differences of electrode positions were then compared to the expected value of no difference in positions using a one-sample t-test against zero.

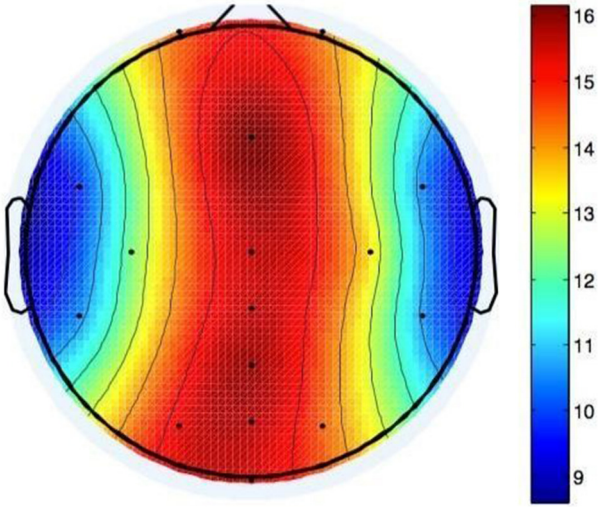


FIGURE 7 | Shifts in electrode positions after self application in mm compared to application by investigator.

Block II: EEG Recordings

Due to software problems on a laptop EEG data of two participants had to be excluded. Analyses of the EEG data were based on the remaining eight participants.

Oddball Paradigm: ERP Results

Grand average ERPs from the oddball paradigm are depicted in **Figure 8**. The repeated measures ANOVA performed on mean amplitudes of the negativity measure yielded significance for the main factor stimulus, $F_{(1, 7)} = 21.745$, $p = 0.002$, $\eta^2 = 0.756$. Amplitudes of the deviant stimuli ($M = -5.44$ μV, $SD = 6.21$ μV) were more negative than in standard stimuli ($M = -0.01$ μV, $SD = 2.66$ μV). The main factor environment was not significant, $F_{(1, 7)} = 0.101$, $p = 0.760$, $\eta^2 = 0.014$. There was also no significant interaction, $F_{(1, 7)} = 0.261$, $p = 0.625$, $\eta^2 = 0.036$. Results of a TOST procedure with an epsilon of the standard deviation of the indoor condition were not significant (*mean difference* = 0.145; *epsilon* = 3.95; *confidence-interval*: −6.79 to 7.08; $df = 7$; $p = 0.166$).

For the positivity measure there was no significant main effect of stimulus, $F_{(1, 7)} = 5.001$, $p = 0.060$, $\eta^2 = 0.417$. The main effect environment also was not significant, $F_{(1, 7)} = 2.767$, $p = 0.140$, $\eta^2 = 0.283$. The interaction between stimulus and environment was significant, $F_{(1, 7)} = 31.800$, $p = 0.001$, $\eta^2 = 0.820$. Amplitudes of the deviant trials were higher indoors ($M = 9.54$ μV, $SD = 9.05$ μV) than in the car ($M = 5.18$ μV, $SD = 10.57$ μV), while amplitudes in standard trials indoors ($M = 0.02$ μV, $SD = 1.25$ μV) were only slightly smaller than in the car ($M = 0.92$ μV, $SD = 2.27$ μV). Due to this significant interaction effect no TOST was performed.

Results from the t-test performed on mean peak latency differences of the indoor ($M = 85$ ms, $SD = 46.3$ ms) and the car condition ($M = 101.5$ ms, $SD = 75.1$ ms) were not significant ($p = 0.569$). The TOST procedure with an epsilon of the standard deviation of the indoor condition showed no significance for

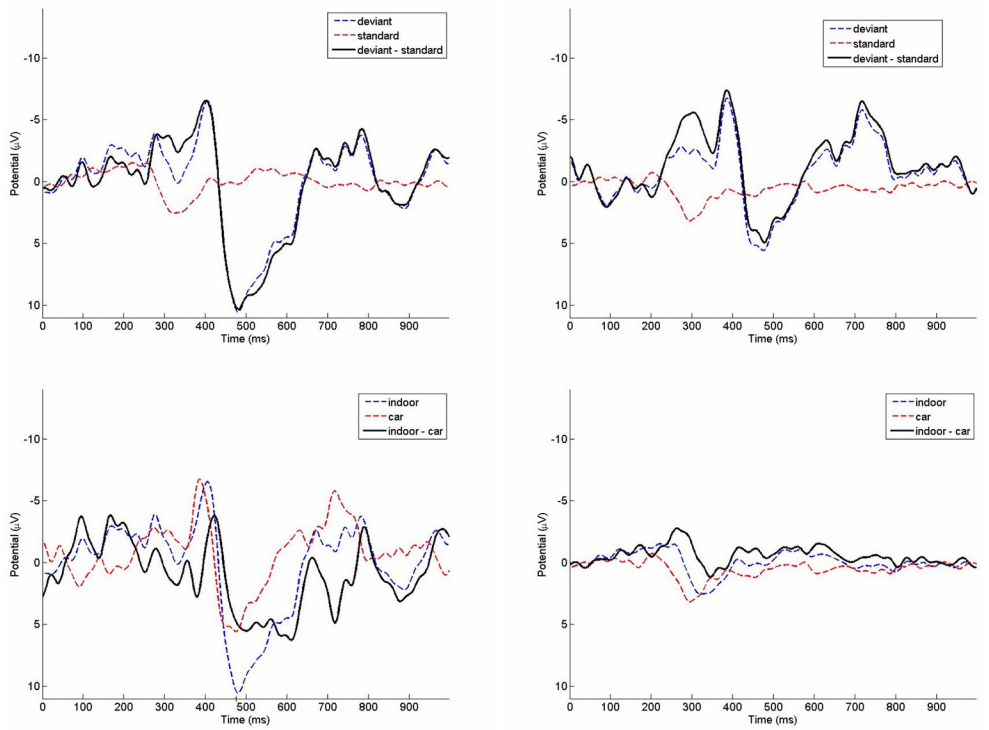


FIGURE 8 | Grand average ERPs of the indoor condition (top left) and the running car condition **(top right)** on channel Cz. Deviant **(bottom left)** and standard **(bottom right)** ERPs in comparison between indoor and car condition.

mean peak latency differences (*mean difference* = −16.5; *epsilon* = 46.3; *confidence-interval*: −68.8 to 35.8; *df* = 7; $p = 0.158$).

Induced Alpha Paradigm: Frequency Results

All individual correlation values for power spectral densities between conditions were higher than 0.79 on all five electrodes, with a mean correlation value of 0.97 ($SD = 0.046$). All *t*-tests of these correlations against zero were significant with $ps < 0.0001$. For the grand average, correlation values between indoor and car condition were both higher than 0.989, with a mean of 0.997 ($SD = 0.004$). *T*-tests against zero yielded significance ($ps < 0.0001$) for both conditions (engaged/relaxed).

The three-way repeated measures ANOVA with within-subject factors recording condition ($p = 0.061$), stimulus ($p = 0.177$), and electrode ($p = 0.24$) performed on mean alpha band powers was not significant on main or interaction effects, with non-significant interactions (all $ps > 0.272$). The TOST procedure with an epsilon of the standard deviation of the indoor condition assigned to mean alpha band powers showed significance on electrodes PO4 (*mean difference* = 0.049; *epsilon* = 0.129; *confidence-interval*: −0.031 to 0.128; *df* = 7; $p = 0.049$) and Oz (*mean difference* = 0.001; *epsilon* = 0.127; *confidence-interval*: −0.079 to 0.076; *df* = 7; $p = 0.009$). The TOST was not significant for electrodes PO3, POz, and Pz, all $ps > 0.340$.

Alpha band time course (see **Figure 9**) correlations between indoor and car condition yielded a mean correlation of $r = 0.27$ for the relaxed condition (Pz: $r = 0.43$, PO3: $r = 0.26$, PO4: $r = 0.29$, POz: $r = 0.30$, Oz: $r = 0.09$). Correlations in this condition were significant on all five electrodes for five participants ($ps < 0.00001$), on four electrodes for one participant ($ps < 0.005$), and for the other three participants on three electrodes ($ps < 0.021$). In the engaged condition the mean correlation of all participants was $r = 0.23$ (Pz: $r = 0.34$, PO3: $r = 0.19$, PO4: $r = 0.18$, POz: $r = 0.31$, Oz: $r = 0.14$). Tests yielded significance of correlations on all five channels for three participants ($ps < 0.043$). For three participants correlation was significant on four channels ($ps < 0.00001$) and for two participants on three electrodes ($ps < 0.00001$).

BCI Results of Both Paradigms

A paired samples *t*-test indicated that the error rates for ERP classification in the indoor condition ($M = 0.126$, $SD = 0.086$) did not differ significantly from the error rates in the car condition ($M = 0.145$, $SD = 0.116$), $t_{(7)} = -0.68149$, $p = 0.518$. Furthermore, the TOST procedure with an epsilon of the standard deviation over participants in the indoor condition confirmed significant equivalence classification results in the two recording conditions (*mean difference* = 0.018; *epsilon* = 0.086; *confidence-interval*: −0.032 to 0.069; *df* = 7; $p = 0.020$).

A paired samples *t*-test indicated that the error rates of band power classification for the indoor condition was lower ($M = 0.283$, $SD = 0.160$), but did not differ significantly from the error rates in the car condition ($M = 0.351$, $SD = 0.137$), $t_{(7)} = -1.608$, $p = 0.152$. The TOST procedure with an epsilon of the

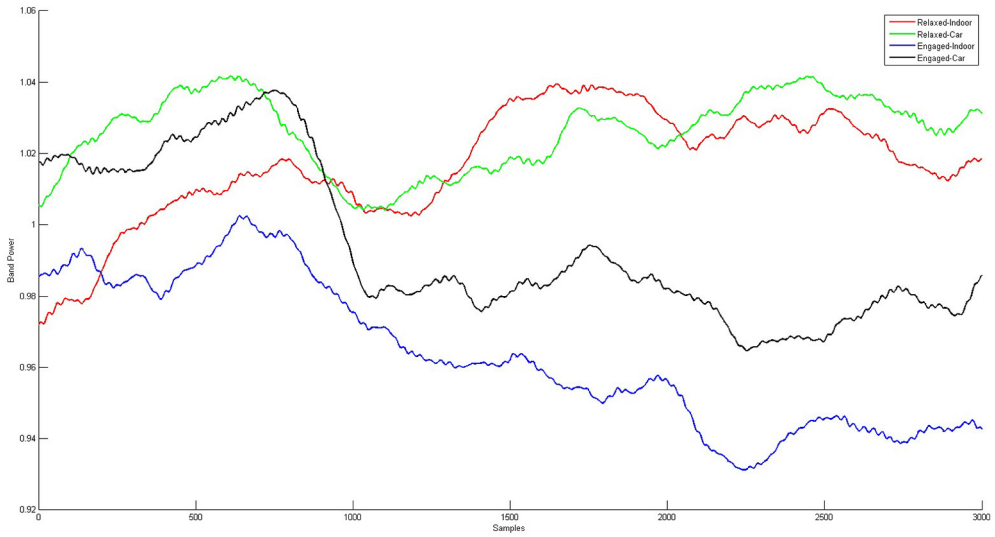


FIGURE 9 | Grand Averages of the alpha band time courses for relaxed and engaged conditions indoors and in the car. For the red and the green curve, displaying the relaxed conditions, a similar pattern starting 1 s after onset of stimulus presentation is observed. Similarities over time are also apparent for the engaged conditions, represented in the black and blue curve. Clear co-variation of indoor and in car alpha time courses for both relaxed and engaged conditions is proven by high correlation between the signals.

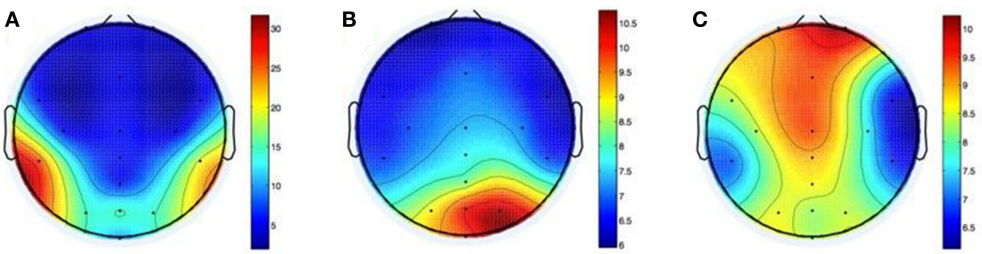


FIGURE 10 | Shifts in electrode positions after movements of the head (A), the arms **(B)**, and the whole body **(C)** in mm.

standard deviation over the participants in the indoor condition confirmed significant equivalence for classification results in the two recording conditions (*mean difference* = 0.066; *epsilon* = 0.162; *confidence-interval*: −0.012 to 0.144; $df = 7$; $p = 0.026$).

Block III: Driving-Related Movements

Figure 10 shows the shifts in electrode positions after each of the three groups of movements.

After head-related movements the difference between electrode positions ($M = 9.6$, $SD = 9.1$) differed significantly from zero, $t_{(9)} = 3.3237$, $p = 0.009$. The apparent lateralization of this effect (25.3 mm mean variation at CP5 vs. 19.6 mm at CP6) may be due to the direction of the shoulder check.

After performance of arm movements the mean difference between electrode positions ($M = 7.6$, $SD = 4.8$) differed significantly from zero, $t_{(9)} = 5.0241$, $p = 0.001$. Variations were located mainly to the right side of the head with a maximum of 10.5 mm mean variation at PO4. The cause for this may be the direction of the rotation and/or handedness of participants.

Mean electrode position differences after whole-body movements ($M = 8.4$, $SD = 6.4$) differed significantly from zero, $t_{(9)} = 4.1691$, $p = 0.002$. The greatest shift was on the forehead with 10.1 mm average variation on Fp2 and on the midline of the head (8.2 and 9.3 mm mean variation at POz and Fz). This could be caused by the cables, which were tied together, but interfered with the seatbelt nevertheless.

Block VI: Usability

The total SUS score of the system added up to 65. Following the official SUS score interpretation, this is slightly above the threshold for an acceptable system.

Due to minor delays during the experiments, the time points of the additional questionnaires varied slightly for each participant. On average, questions were answered after 60 (Block I), 122 (Block II), and 137.5 (Block III) min.

After the first 60 min, the system got a comfort rating of 7.5, which then decreased significantly over the next hour resulting in a rating of 3 after 122 min. In the following quarter of an hour needed for block III, the comfort rating stayed stable at 3.

A Wilcoxon signed-rank test showed that there was a significant difference between the first time point of the rating after 60 min ($Mdn = 7.5$) and the second rating after 122 min ($Mdn = 3$), ($W = 0$, $Z = -2.69$, $p = 0.008$). No valid Wilcoxon signed-rank test could be performed to compare the second and third ratings, because the number of effective samples was less than 6 after subtraction of ratings equaled zero for six participants ($W = 4$, $Z = -0.82$, $p = 0.625$). Rating scores of the first and the third rating again showed significant differences, ($W = 0$, $Z = -2.67$, $p = 0.008$).

The six examined items of wearing comfort of the system are summarized in **Figure 11**. A feeling of pressure on the head was rated as the most irritating with a mean score of 2.2. The overall impression of wearing comfort got a mean score of 2.7, and was therefore also perceived as bad. The overall weight of the system on the head was on average rated as the most pleasant aspect of it with a score of 4.2.

Furthermore, the wearing comfort questionnaire yielded the following insights. Seven participants complained about dents and chafe marks on their heads, four about headaches, and one each about neck pains, nausea, and dizziness. Moreover, one participant had the subjective impression that the system had moved over the course of the experiments. None of the participants reported skin irritations due to wearing the cap.

DISCUSSION

Block I: Self-application

We found that the participants were equally fast as the experimenter in applying the cap, and equally capable in optimizing signal quality. We thus conclude that this type of dry electrode EEG system can indeed be used by individual end-users. We should note, however, that there was no objective measure of when the application was finished; it was based on individual judgements of the experimenter.

We did not investigate the personalization of the cap by adjusting the length of each electrode pin, because this task needs to be done only once. Therefore, we did not investigate how easy it is to personalize the cap while wearing it. Personalization did, however, take up quite some time. We assume that the QuickBit approach would benefit from improvement: Continuously adjustable bits would probably simplify personalization and optimize the result.

While it is not surprising that the signal quality was rated better with active display filters, we had assumed that the signal quality would be better after adjustments by an expert operator than compared to that adjusted by the participant. This, however, was not the case: Participants reached a similar, sometimes even better signal quality. We assume the reason for this to be that participants had a better feeling for how hard, and where exactly the electrodes pressed against their heads, allowing them to fit them even better to the scalp than the experimenter could without the risk of harming the participant.

For the electrode positions, some variation in the measurements must be taken into account. The used system has known variations in measured data points, and for some electrodes (primarily at the back of the head), the measuring stylus may have moved slightly due to head shifts that were sometimes necessary for the measurement. This problem was addressed mathematically, as described above. It was also not possible to point the stylus exactly at the electrode's point of contact with the skin, but only at the electrode's body. It remains unclear, whether or to what extent the differences in electrode positions we measured, imply that the points of contact changed as well.

Block II: EEG Recordings

For the oddball paradigm ERP analysis revealed highly similar morphology of ERPs elicited by deviant stimuli in both recording conditions. We found highly significant effects for the negative

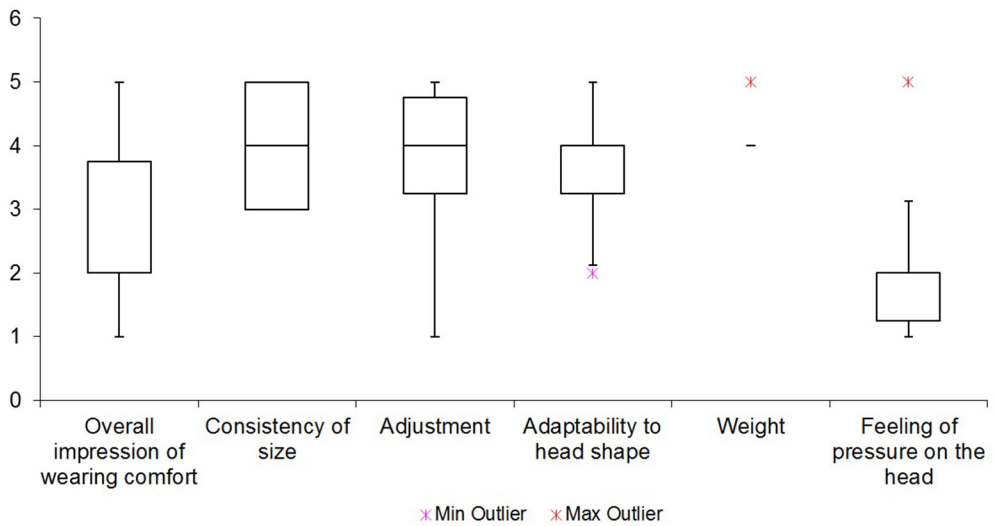


FIGURE 11 | Mean score of questions about wearing comfort.

peak in the ERP condition. The deviant trials were significantly different from the standard trials in both the indoor and the car condition, showing no difference between conditions. This is not the case for the positivity. The main effect is not significant. It should be mentioned though that we have a clear tendency into the right direction with a p value slightly missing the threshold criteria of 5%. Peaks of the P300 are reduced in the car environment as a result of other signals interfering with the recorded signal in the car. No significant differences were found between peak latencies between indoor and car recordings. We conclude that the main information carried in the signal is comparable for indoor and in car recordings, but its signal strength is attenuated slightly in the car condition.

For the alpha recordings, we have a slightly more complex case. We clearly see a correlation between conditions—alpha values show a similar development over time outside of and in the car. However, there is no significant difference between relaxed and engaged trials on average over all participants, which was expected from the experimental design. When we take a closer look at the individual values (see **Figure 12**), we see that some participants managed to get relaxed in the corresponding task, while others did not. This explains why we do not get significant main effects—several participants were not able to relax in the appropriate condition. This effect can be seen consistently on both conditions, inside and outside the car. However, we do perhaps see a tendency on the main effect of condition that, even though it's not significant, indicates a small change in alpha power between recordings inside and outside of the car.

For all comparisons that showed no significant difference between conditions an equivalence test was performed. Features of the ERP were not equivalent between conditions while spectral features were equivalent on some of the tested electrodes.

These results show that even though we do not have significant differences, the recoded data cannot be taken equivalent. For strict neurophysiological measurements it hence might be worth a consideration whether the tested headset should be used or not.

For ERP and spectral data classifications were not significantly different, and were furthermore clearly equivalent. We, hence, assume that the evaluated system measured the differences in cognitive states, well, in both conditions. Despite small morphological and power differences, classification results were comparable in both domains. Therefore, a BCI can be applied with equal reliability to data from both conditions.

The results we found on the EEG components examined here are as expected from the literature and replicate results from a previous comparison study (Zander et al., 2011). Therefore, we conclude that the dry electrode system investigated here provides comparable data to a conventional gel-based system when used in an autonomous driving context.

It still remains unclear whether the results can be fully transferred to a real-world autonomous driving context where the car would most likely be moving. A driving car would bring additional factors like increased vibration from the engine, jerks due to uneven roads, or inertial effects induced by direction changes. Moreover, the driving task itself could lead to additional artifacts, such as stress related sweating on the scalp and the user scratching their own skin. Also head movements against the headrest might lead to changes of electrode positions in a way that was not examined here. Another factor would be the radio not being muted in a real-world-driving scenario: Environmental noises between 70 and 120 decibels have been found to increase the amplitude of measured P300 events (Nam et al., 2008). Drivers will also be moving e.g. their heads and hands, which they minimized during data recording. This study however presents a first step in investigating the applicability of dry systems in a car environment, revealing initial insights in a scenario with controlled artifact activity. These results can form the basis for future studies in active driving study scenarios, where that control is further relaxed.

Block III: Driving-Related Movements

The results showed that the electrodes shifted in position when executing different driving-related movements.

The most significant shifts occurred during movements involving the head directly, primarily at the rear left of the head.

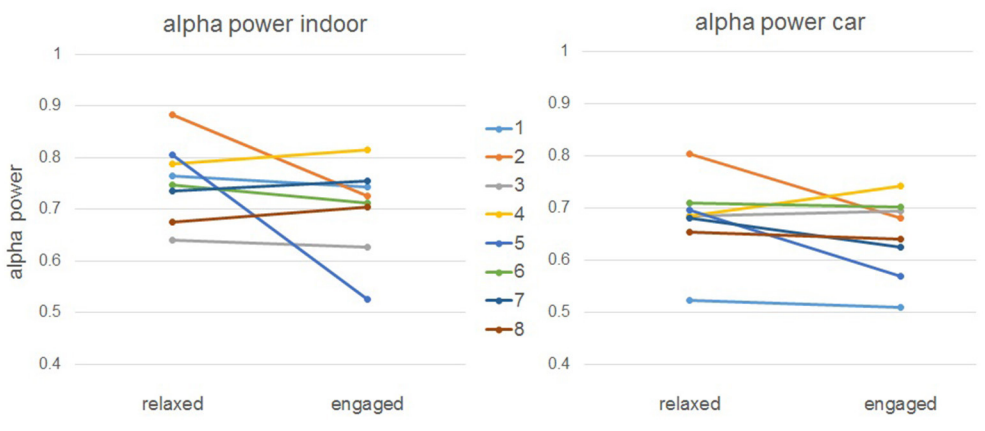


FIGURE 12 | Mean alpha power in relaxed and engaged trials for individual subjects.

We assume this was due to the shoulder check, which required a sudden, fast turn of the whole head to the left and back. We can, however, not be sure as to whether the shoulder check or the look at the ceiling had more effect on the electrodes positions since they were measured together as one movement group. Either way, the resulting differences may well-influence the quality of the data recorded by the system.

The performed arm movements had less impact on the electrode positions, though the shifts were still significant.

The third group of movements resulted in the least position changes for all electrodes although the participants had to move their whole upper body—including the head. The most pronounced shifts were observed at the right frontal area. The instruction to touch the marker in the legroom of the passenger seat might offer an explanation for this, as the head had to be moved rather far to the right and down. Also in the area around the left ear increased shifts in position were observed. Most likely, this was a result of fastening and unfastening the seatbelt which may have induced some strain in that area, maybe by pulling on the cables.

Finally, since the movements were always performed in the same order (head, arm, and body), order effects cannot be excluded.

For future use, the cap could be applied e.g., only after the seat belt has been fastened, which often requires some effort. Since the cables may also have caused some of the position shifts, a wireless system is preferable.

Block IV: Usability

The *System Usability Scale* is a general questionnaire to evaluate the usability of technical systems, and is not specifically designed for BCI systems. As SUS provided significant insights in other BCI-related studies, we decided to use it here as well (Duvinage et al., 2012; Käthner et al., 2013). Some questions however, especially about the interaction with the system, did not fit the current purpose and even confused some of the participants. The resulting SUS score might therefore not be entirely accurate, but, we believe, still provides a good indication about the overall usability of the system in an autonomous driving context.

The evaluation of the wearing comfort was better tuned to the current context and raised no questions from participants. The results showed that the first hour of using the system did not bother the participants much, which qualifies it for short-term usage at least. After the second hour of using the system, however, the subjective comfort ratings dropped significantly and participants began to complain about dents, slight headaches, neck pain, even nausea and dizziness, which clearly shows that the EEG system with the current cap design is not suitable for long-term use. We did not investigate recovery time: How long a break is needed, before the cap can be comfortably worn again? This remains an open question.

The most annoying features of the system, according to the participants, were its rather tight fit onto the head resulting in the feeling of pressure. The overall weight of the system was, in contrast, rated to be quite pleasant which might be caused by the flexible, thin material of the cap. Also, participants rated the adaptability of the cap as quite high. The cap was rated as being fixated well, thanks to the chin belt and the holes for the ears providing a lot of stability–only one participant had the feeling the cap had moved at all.

CONCLUSION

Concluding in brief, the EEG system allowed for technically sound recordings, even with car-induced interferences present. It thus appears to be suitable for passive BCIs in autonomous driving scenarios, allowing mental states to be detected in real time.

In only a few minutes, individuals were able to apply and adjust a pre-customized cap, with the help of a little mirror, like the rear view mirror of a car. A system to better support the evaluation of signal quality would be beneficial, however.

According to the system usability scale, the system is at the edge of acceptability in terms of usability. This may suffice for professional drivers, who likely stand to gain the most from autonomous driving and supportive systems, but room for improvement remains. In particular the reported discomfort after longer use is unacceptable. Here, major improvement is necessary to decrease pressure on the scalp so the system is no longer obstructive and uncomfortable, hindering the users from focusing on themselves and their tasks.

Seeing now that EEG technology has made clear progress toward ease of use and mobile scenarios, we can envision the application of passive BCIs in the context of autonomous driving. Passive BCIs can provide essential information about the driver's cognitive or affective state, which can be combined with other sensor data of the car. In that way, the car can adapt to, and make decisions informed by, individual aspects of the driver. As passive BCIs do not rely on directed or even conscious actions of the driver (Zander and Kothe, 2011), the car will still drive autonomously but gains an additional stream of information, pertaining to the subjective situational interpretation of the driver.

For example, we can clearly imagine applications improving safety and comfort. In cases where the driver is required to take over control, the communication of this requirement can be adapted to the current, actual state of the driver. Another scenario would be the detection of whether or not communicated alarm signals were perceived and processed by the driver. These are only a few, simple examples of a broad range of applications to be thought of.

Moreover the investigated system could be used in a broader field of scenarios and might be of special interest for the field of Mobile brain/body imaging (MoBI). The field's objective is to acquire neurophysiological recordings of human cognition in real world environments where subjects perform real-world tasks. A portable, wireless, high-quality data recording and fast to prepare dry contact system would prove useful for brain activity recordings on actively behaving participants (Gramann et al., 2011, 2014; De Sanctis et al., 2012).

The application of passive BCI during autonomous driving however provides an exemplary use case for technology that adapts to the (neuronal) state of its operator during automation

in general. Such Neuroadaptive Technology is a clear additional step toward closing the cybernetic loop (Pope et al., 1995).

ETHICS STATEMENT

The study involved standard EEG procedures covered in an ethic statement approved by the ethics committee of the Institute of Psychology and Ergonomics of the Berlin Institute of Technology. All participants gave written consent to their participation in the conducted study. They were provided with information on the purpose of the study, given the opportunity to ask questions and were informed that their participation was voluntary and they could end the experiment whenever they liked without a need to provide reasons. Participants also gave their consent for data recording, anonymous storage of that data, as well as its usage for publication.

AUTHOR CONTRIBUTIONS

All authors contributed substentially to the work presented here. Everybody was contributing to the drafting and revising of the documents and approved the final version. Everybody agreed to be accountable for the integrity and accuracy of the work. Specifically: TZ designed and supervised the experimental procedures, conducted and supervised the analyzes, interpreted the results for the context of autonomous driving. LK and KG were responsible for quality of writing and validation of results. Everybody below was involved in conducting the experiments and ensured data quality. LA was responsible for the statistical analyzes and integrity of the manuscript. JP and MB were responsible for the electrode localization and the related mathematical procedures. LZ: Was responsible for the programming and EEG and BCI analyzes. AB: Was responsible for evaluation of the questionnaires.

ACKNOWLEDGMENTS

We thank Brain Products GmbH (Munich, Germany) for providing us with the tested EEG System which made this research possible. We are also indebted to Prof. Dr.-Ing. Matthias Roetting and Mario Lasch, Chair for Human-Machinse Systems, TU Berlin for providing the research car and supporting us in all technical questions regarding the car.

REFERENCES

Beiker, S. A. (2012). *Symposium, Legal Aspects of Autonomous Driving, 52.* Santa Clara, CA, 1145.

Berka, C., Levendowski, D. J., Lumicao, M. N., Yau, A., Davis, G., Zivkovic, V. T., et al. (2007). EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. *Aviat. Space Environ. Med.* 78(Suppl. 1), B231–B244.

Blankertz, B., Lemm, S., Treder, M., Haufe, S., and Müller, K. R. (2011). Single-trial analysis and classification of ERP components—a tutorial. *Neuroimage* 56, 814–825. doi: 10.1016/j.neuroimage.2010.06.048

Blankertz, B., Tangermann, M., Vidaurre, C., Fazli, S., Sannelli, C., Haufe, S., et al. (2010). The berlin brain–computer interface: non-medical uses of BCI technology. *Front. Neurosci.* 4:198. doi: 10.3389/fnins.2010.00198

Borghini, G., Astolfi, L., Vecchiato, G., Mattia, D., and Babiloni, F. (2014). Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. *Neurosci. Biobehav. Rev.* 44, 58–75. doi: 10.1016/j.neubiorev.2012.10.00

Brooke, J. (1986). *SUS: A "Quick and Dirty" Usability Scale. Usability Evaluation in Industry*. London: Taylor and Francis

Delorme, A., Kothe, C., Vankov, A., Bigdely-Shamlo, N., Oostenveld, R., Zander, T. O., et al. (2010). "MATLAB-based tools for BCI research," in *Brain-Computer Interfaces* (London: Springer), 241–259.

Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21. doi: 10.1016/j.jneumeth.2003.10.009

De Sanctis, P., Butler, J. S., Green, J. M., Snyder, A. C., and Foxe, J. J. (2012). "Mobile brain/body imaging (MoBI): high-density electrical mapping of inhibitory processes during walking," in *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (San Diego, CA: IEEE), 1542–1545.

Dickmanns, E. D., and Zapp, A. (1987). "A curvature-based scheme for improving road vehicle guidance by computer vision," in *Cambridge Symposium_Intelligent Robotics Systems* (International Society for Optics and Photonics), 161–168.

Duncan-Johnson, C. C., and Donchin, E. (1977). On quantifying surprise: the variation of event-related potentials with subjective probability. *Psychophysiology* 14, 456–467. doi: 10.1111/j.1469-8986.1977.tb01312.x

Duvinage, M., Castermans, T., Petieau, M., Seetharaman, K., Hoellinger, T., Cheron, G., et al. (2012). "A subjective assessment of a P300 BCI system for lower-limb rehabilitation purposes," in *Engineering in Medicine and Biology Society (EMBC), Proceedings of the 2012 Annual International Conference of the IEEE*, (San Diego, CA: IEEE), 3845–3849.

Fabrizio, M., and Cimolino, U. (2014). *Persönliche Schutzausrüstung*. Düsseldorf: Hüthig Jehle Rehm.

Franke, U., Gavrila, D., Gorzig, S., Lindner, F.,Paetzold, F., and Wohler, C. (1998). Autonomous driving approaches downtown. *IEEE Intell. Syst.* 13, 40–48.

Gerjets, P., Walter, C., Rosenstiel, W., Bogdan, M., and Zander, T. O. (2014). Cognitive state monitoring and the design of adaptive instruction in digital environments: lessons learned from cognitive workload assessment using a passive brain-computer interface approach. *Front. Neurosci.* 8:385. doi: 10.3389/fnins.2014.00385

Gramann, K., Ferris, D. P., Gwin, J., and Makeig, S. (2014). Imaging natural cognition in action. *Int. J. Psychophysiol.* 91, 22–29. doi: 10.1016/j.ijpsycho.2013.09.003

Gramann, K., Gwin, J. T., Ferris, D. P., Oie, K., Jung, T. P., Lin, C. T., et al. (2011). Cognition in action: imaging brain/body dynamics in mobile humans. *Rev. Neurosci.* 22, 593–608. doi: 10.1515/RNS.2011.047

Iversen, J. R., and Makeig, S. (2013). "MEG/EEG Data Analysis Using EEGLAB," in *Magnetoencephalography: From Signals to Dynamic Cortical Networks*. eds Selma S. and Cheryl, J. A (Berlin; Heidelberg: Springer Verlag), 199–212.

Käthner, I., Ruf, C. A., Pasqualotto, E., Braun, C., Birbaumer, N., and Halder, S. (2013). A portable auditory P300 brain–computer interface with directional cues.*Clin. Neurophysiol.* 124, 327–338. doi: 10.1016/j.clinph.2012.08.006

Kothe, C. A., and Makeig, S. (2013). BCILAB: a platform for brain–computer interface development. *J. Neural Eng.* 10:056014. doi: 10.1088/1741-2560/10/5/056014

Liao, L. D., Lin, C. T., McDowell, K., Wickenden, A. E., Gramann, K., Jung, T. P., et al. (2012). "Biosensor technologies for augmented brain–computer interfaces in the next decades," in *Proceedings of the IEEE*, *100(Special Centennial Issue)*, 1553–1566.

Linden, D. E. (2005). The P300: where in the brain is it produced and what does it tell us?. *Neuroscientist* 11, 563–576. doi: 10.1177/1073858405280524

Llaneras, R. E., Salinger, J., and Green, C. A. (2013). "Human factors issues associated with limited ability autonomous driving systems: drivers' allocation of visual attention to the forward roadway," in *Proceedings of the 7th*

International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design (Bolton Landing: NY), 92–98.

Ma, R., and Kaber, D. B. (2005). Situation awareness and workload in driving while using adaptive cruise control and a cell phone. *Int. J. Ind. Ergon.* 35, 939–953. doi: 10.1016/j.ergon.2005.04.002

Nam, C. S., Johnson, S., and Li, Y. (2008). "Environmental Noise and P300-Based Brain-Computer Interface (BCI)," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Sage, CA; Los Angeles, CA: SAGE Publications), 803–807.

Pasqualotto, E., Federici, S., Simonetta, A., and Olivetti Belardinelli, M. (2011). "Usability of brain computer interfaces," in *11th European Conference for the Advancement of Assistive Technology: AAATE* (Maastricht).

Polich, J., and Kok, A. (1995). Cognitive and biological determinants of P300: an integrative review. *Biol. Psychol.* 41, 103–146. doi: 10.1016/0301-0511(95)05130-9

Pomerleau, D. (1992). *Neural Network Perception for Mobile Robot Guidance.* PhD thesis, Carnegie Mellon University.

Pope, A. T., Bogart, E. H., and Bartolome, D. S. (1995). Biocybernetic system evaluates indices of operator engagement in automated task. *Biol. Psychol.* 40, 187–195. doi: 10.1016/0301-0511(95)05116-3

Sauseng, P., Klimesch, W., Stadler, W., Schabus, M., Doppelmayr, M., Hanslmayr, S., et al. (2005). A shift of visual spatial attention is selectively associated with human EEG alpha activity. *Eur. J. Neurosci.* 22, 2917–2926. doi: 10.1111/j.1460-9568.2005.04482.x

Schuirmann, D. J. (1981). On hypothesis testing to determine if the mean of a normal distribution is contained in a known interval. *Biometrics* 37:617.

Schuirmann, D. J. (1987). A comparison of the two one-sided tests procedure and the power approach for assessing the equivalent of average bioavailability. *J. Pharmacokinet. Biopharm.* 15, 657–680. doi: 10.1007/BF01068419

Solis-Escalante, T., Müller-Putz, G., Brunner, C., Kaiser, V., and Pfurtscheller, G. (2010). Analysis of sensorimotor rhythms for the implementation of a brain switch for healthy subjects. *Biomed. Signal Process. Control* 5, 15–20 doi: 10.1016/j.bspc.2009.09.002

Sukthankar, R., Baluja, S., and Hancock, J. (1997). "Evolving an Intelligent Vehicle for Tactical Reasoning in Traffic," in *Proc. IEEE Int. Conf. on Robotics and Automation.* (New York, NY: IEEE), 519–524.

Tadaka, Y., and Shimoyama, O. (2004). "Evaluation of driving-assistance systems based on drivers' workload," in *Proceeding of the International Driving Symposium on Humand Factors in Driver Assessment, Training and Vehicle Design.*

Webb, A. R. (2002). *Statistical Pattern Recognition*. Chichester: John Wiley and Sons Ltd.

Westlake, W. J. (1981). Response to T.B.L. Kirkwood: bioequivalence testing – a need to rethink. *Biometrics* 37, 589–594.

Young, M. S., and Stanton, N. A. (1997). Automotive automation: Investigating the impact on drivers mental workload. *Int. J. Cogn. Ergon.* 1, 325–336.

Zander, T. O., and Jatzev, S. (2012). Context-aware brain–computer interfaces: exploring the information space of user, technical system and environment. *J. Neural Eng.* 9:016003. doi: 10.1088/1741-2560/9/1/016003

Zander, T. O., and Kothe, C. (2011). Towards passive brain–computer interfaces: applying brain computer interface technology to human–machine systems in general. *J. Neural Eng.* 8:025005. doi: 10.1088/1741-2560/8/2/025005

Zander, T. O., Lehne, M., Ihme, K., Jatzev, S., Correia, J., Kothe, C., et al. (2011). A dry EEG-system for scientific research and brain–computer interfaces. *Front. Neurosci.* 5:53. doi: 10.3389/fnins.2011.00053

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Zander, Andreessen, Berg, Bleuel, Pawlitzki, Zawallich, Krol and Gramann. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 21 June 2016
doi: 10.3389/fnhum.2016.00291

An Intelligent Man-Machine Interface—Multi-Robot Control Adapted for Task Engagement Based on Single-Trial Detectability of P300

Elsa A. Kirchner[1,2], Su K. Kim[2], Marc Tabie[2], Hendrik Wöhrle[2], Michael Maurus[2] and Frank Kirchner[1,2]*

[1] Research Group Robotics, Mathematic and Computer Science, University of Bremen, Bremen, Germany, [2] Robotics Innovation Center (RIC), German Research Center for Artificial Intelligence (DFKI GmbH), Bremen, Germany

Advanced man-machine interfaces (MMIs) are being developed for teleoperating robots at remote and hardly accessible places. Such MMIs make use of a virtual environment and can therefore make the operator immerse him-/herself into the environment of the robot. In this paper, we present our developed MMI for multi-robot control. Our MMI can adapt to changes in task load and task engagement online. Applying our approach of embedded Brain Reading we improve user support and efficiency of interaction. The level of task engagement was inferred from the single-trial detectability of P300-related brain activity that was naturally evoked during interaction. With our approach no secondary task is needed to measure task load. It is based on research results on the single-stimulus paradigm, distribution of brain resources and its effect on the P300 event-related component. It further considers effects of the modulation caused by a delayed reaction time on the P300 component evoked by complex responses to task-relevant messages. We prove our concept using single-trial based machine learning analysis, analysis of averaged event-related potentials and behavioral analysis. As main results we show (1) a significant improvement of runtime needed to perform the interaction tasks compared to a setting in which all subjects could easily perform the tasks. We show that (2) the single-trial detectability of the event-related potential P300 can be used to measure the changes in task load and task engagement during complex interaction while also being sensitive to the level of experience of the operator and (3) can be used to adapt the MMI individually to the different needs of users without increasing total workload. Our online adaptation of the proposed MMI is based on a continuous supervision of the operator's cognitive resources by means of embedded Brain Reading. Operators with different qualifications or capabilities receive only as many tasks as they can perform to avoid mental overload as well as mental underload.

Keywords: EEG, P300, machine learning, space robotics, teleoperation, task load, man-machine interaction, embedded brain reading

OPEN ACCESS

Edited by:
Klaus Gramann, Berlin Institute of Technology, Germany

Reviewed by:
Tamer Demiralp, Istanbul University, Turkey
Pieter-Jan Kindermans, TU-Berlin, Germany

****Correspondence:***
Elsa A. Kirchner
ekir@informatik.uni-bremen.de

Received: *05 December 2015*
Accepted: *31 May 2016*
Published: *21 June 2016*

Citation:
Kirchner EA, Kim SK, Tabie M, Wöhrle H, Maurus M and Kirchner F (2016) An Intelligent Man-Machine Interface—Multi-Robot Control Adapted for Task Engagement Based on Single-Trial Detectability of P300. Front. Hum. Neurosci. 10:291. doi: 10.3389/fnhum.2016.00291

1. INTRODUCTION

Human-robot interaction with semi-autonomous robots has to be improved to be safe and intuitive. This can be achieved by (1) building robots with advanced "on-board" solutions that support natural interaction behavior between human and robot (Kirchner et al., 2015) and (2) by developing intelligent man-machine interfaces (MMIs). Especially in cases of tele-operating robots at remote places the MMI has to be easy, intuitive and comfortable.

Usually only experienced people are chosen to remotely operate robotic systems (Cornellà et al., 2012), since their performance is robust. During remote control of several robots in a complex mission, task load and task engagement change tremendously over time, which can lead to mental over- or underload as well as fatigue. Therefore, an online-adaptable MMI can be applied to act on these changes. For this, reliable measures for online changes in the human's state must be detected (Allanson and Fairclough, 2004). Such realtime indicators have to consider theories about brain capacity and resources (Kahneman, 1973; Wickens, 1984, 1992, 2008), which propose that brain resources are limited and must be shared between tasks. Comprehensive work showed that certain patterns in the electroencephalogram (EEG), e.g., the amplitude of the event-related potential (ERP) P300 (Prinzel et al., 2003), or ratios of EEG power bands like alpha, beta or theta bands (Pope et al., 1995), can be used to measure the processing capability of the brain, mental workload and task demands. In earlier work from Pope et al. (1995) it is shown that an EEG-based index of user engagement and arousal could indeed be used to, i.e., adapt the level of system automation in response to changes in mental workload demands. It was found that especially the P300 is a reliable measure for changes in task load (Kok, 2001; Prinzel et al., 2003). Earlier work that examined the P300 in response to primary and secondary task demands showed that an increase in demands on the primary task resulted in fewer resources for the secondary task accompanied by a smaller P300 amplitude (Isreal et al., 1980). Many studies make use of the dual-task design (Isreal et al., 1980; Prinzel et al., 2003) to detect an increase in workload or task load in the primary task by analyzing the P300 amplitude evoked by the secondary task, e.g., listening to auditory stimuli presented in an oddball fashion (Prinzel et al., 2003) or P300 that is evoked by ignored probes (Kramer et al., 1995).

With the focus on online user state detection based on the analysis of brain activity, which is naturally evoked during human-machine interaction and deeply embedded into the systems control, embedded Brain Reading (eBR) was developed (Kirchner and Drechsler, 2013; Kirchner, 2014, 2015). The main focus of embedded Brain Reading is to passively infer on the human's intention to implicitly improve interfaces like an exoskeleton which is used for explicit interaction, such that the intended interaction or behavior can be supported best (Folgheraiter et al., 2012; Kirchner et al., 2013a,b, 2014). However, embedded Brain Reading can also be applied to passively infer on the users' neurophysiological state, such as their current workload or task load, to adapt an interface implicitly in such a way that the user is neither stressed nor bored (Kirchner et al., 2010, 2013b; Wöhrle and Kirchner, 2014a) which would both have negative impact on human-robot interaction. We already showed that eBR can utilize P300-related activity to infer, whether subjects recognize and will respond to important task messages, which were presented interleaved with task-irrelevant messages in an oddball fashion, while performing a complex interaction task like playing a labyrinth game (Kirchner et al., 2013b). In a later work we showed that eBR can indeed be applied to improve interaction in an application scenario in which subjects had to respond to warnings interleaved with task-irrelevant status messages while remotely controlling a robotic arm via an exoskeleton (Wöhrle and Kirchner, 2014a). In both cases, the information about the operator's capability of recognizing task-relevant warnings was used to adapt the developed MMI with respect to the timing of repetitions of task messages. To this end, the MMI was adapted before the operator would respond to the task message. In our previous work, subjects had to perform two tasks: controlling a machine and responding to task-relevant warnings. Thus, we did not make use of the primary and secondary task design just for the purpose of measuring task load on the user. The second task was indeed required to be performed by the user with the goal to estimate an operator's capability to perform two tasks at the same time. We also believe that even when using ignored probes to measure load on the user, i.e., workload (Kramer et al., 1995), any extra stimulation which is only added for the purpose of measuring load on the user will likely disturb the operator in a complex and demanding interaction task. Instead, we used the single-trial detectability of the naturally evoked P300 components in case that rare task-relevant stimuli were presented (i.e., warnings that anyway requested responses of the operator) and had to be answered as index of load, here, task load and task engagement. However, in many real world applications the occurrence of task-relevant target stimuli is likely not interleaved consistently with task-irrelevant stimuli as it was implemented in the previous studies by using the oddball design. Thus, it is of interest to investigate whether single-target stimuli successfully and reliably evoke P300 ERP components during human-machine interaction, as suggested by comprehensive work performed under controlled conditions of the single-stimulus paradigm (Mertens and Polich, 1997; Polich and Margala, 1997). Polich and Margala (1997) for example showed, that single-target stimuli evoke P300 components with similar characteristics as target stimuli presented in an oddball fashion as long as the probability and the inter target interval (ITI) were kept the same.

One research interest of the current work is therefore to investigate whether P300 ERP components are reliably evoked under application conditions in case of a single-stimulus presentation that was naturally embedded into a human-machine interaction task. We further investigate whether eBR can be used to adapt the frequency of task messages that are presented to the user by an MMI instead of modulating task repetitions as in a former work (Kirchner et al., 2013b; Wöhrle and Kirchner, 2014a). The adaptation of the MMI should again be performed online. However, the proposed MMI is designed for multi-robot control. Hence, an adaptation of the MMI with respect to the inferred task load and the users current task

engagement in preceding, still ongoing, tasks for other robots can be investigated. Again, task engagement or task load was inferred from P300-related ERP activity that is naturally evoked during interaction. Both a high task load and a high task engagement to a preceding task were expected to reduce the amplitude of P300-related activity evoked by a new task message. In the presented work, subjects performed only one type of task: controlling different robots with respect to different requested tasks. Hence, we break down dual-task execution into sequential and timely overlapping task execution to investigate the influence of task load and task engagement between subsequent tasks. We again show that it is not necessary to artificially add an extra task or probe, like in the dual task or ignored-probe design, to evoke P300-related activity for measuring task load and task engagement. Instead we directly infer the task load and task engagement of the operator from the P300-activity evoked by task messages.

Hence, our approach matches natural requirements on the user during robot control since it avoids to add potentially disturbing stimuli, like auditory stimuli, just for the goal to measure and adapt for task load.

We further present and describe the developed MMI, which makes use of a virtual control environment, i.e., a Cave Automatic Virtual Environment (CAVE) (**Figure 1**). This MMI can be adapted based on the changes in task engagement of the user measured by EEG, i.e., P300-related ERP activity. While the presentation of each task-relevant message was expected to evoke a P300 we further assumed that the amplitude of a single-trial P300 evoked by a new task message is reduced in case that the user is still involved in executing a previous task. This is due to the fact that mental resources are still bound to the previous task. The more frequently such task conflicts occurred the stronger we expected a reduction in averaged P300 peak amplitude. We further assumed that the expected changes in P300 amplitude were mainly caused by effects like task engagement or task load but not by target probability, since the inter-stimulus interval (ISI) between stimuli was very long. Polich (1990) showed by means of an auditory discrimination task that the target probability has no effect on P300 amplitude in case of longer ISIs, i.e., ISIs longer than 6–8 s (Polich, 2007). For longer ISIs, the probability effect (Tueting et al., 1970; Duncan-Johnson and Donchin, 1977) is missing since brain resources can be redirected fast enough to process a new target stimulus.

It is important to state that in the present work the level of task load and task engagement as well as the occurrence of task conflicts may strongly depend on different factors, e.g., the general capability of the user in controlling the robots, fatigue levels or secondary requirements on attention that are not related to the main task, i.e., distractions of any kind that may occur while the operator was controlling the robots. While the concept of workload is distinct from the concept of multiple resource theory (Wickens, 2008), both concepts do overlap in real world applications and it is not always clear what contributes most. Moreover, additional mechanisms like confusion, cooperation between task elements like ongoing task engagement to the preceding task and unwanted diversion of attention influence the allocation of brain resources (Wickens, 2008). Additionally, as known from educational research, changes in the motivational state influence perception of workload, task complexity and cognitive strategies (Kyndt et al., 2011). Real world applications are therefore not a good paradigm to decouple components and dimensions of influencing parameters, but they can be used as a test case on whether certain measures can be used to predict the general state and capacities of a subject. Since the goal of our study was to measure the current task engagement or task load of an operator and to use this measure to adapt an MMI continuously to avoid an overall state of overload, we took measures to avoid excessive workload.

In summary, the scope of this study was to artificially evoke task conflicts to (I) not only show that P300-related activity was naturally evoked when task messages were presented, but also that it was indeed modulated by generally high demands on the operator and by task engagement to previous tasks and (II) that the detectability of P300-related activity could be used to adapt an MMI with regards to task engagement and therefore enabling a kind of steady-state task involvement. This should result in higher subjective contentment and high overall task performance.

The paper is structured as follows. In Section 2 we describe the experimental setting, i.e., the developed MMI, the kind of human-machine interaction task which can be performed and the interaction tasks that the subjects had to solve, the experiments that were performed for this work, and data recording procedure. We further describe our research goals and hypotheses in more detail and describe the performed data processing and analysis. In Section 3 we describe our results with respect to behavioral, machine learning and ERP average analysis. Finally in Section 4 we will discuss the outcome of our work and its relevance for the improvement of MMIs for multi-robot control.

2. MATERIALS AND METHODS

2.1. Experimental Design

We developed an experimental setup in which a subject can control several simulated robots. For this, we designed a virtual environment using the in-house developed software "Machina Arte Robotum Simulans" (MARS) (Rommerman et al., 2009; DFKI - RIC, 2015), which can be run as a 3D environment in, e.g., a CAVE (see **Figure 1**), as a 2D environment on a standard personal computer and monitors or a multi-screen system (see **Figure 2**). In both environments the operator can use different input devices to control the robot, e.g., a 3D mouse, a wand, an exoskeleton or an eye tracking device. In the future, the developed virtual 3D environment will be used to control real robots. To allow this, we use a physical simulation with close to realistic physical simulations of the real robots developed at our institute. In this work a 2D multi-screen system was used as the environment and a wand was used as the interface to control the simulated robots in the simulated environment. The used wand is a hardware device and functions in a 3D environment similar to a mouse in a 2D environment. It is tracked in 3D space using an ultrasound-based tracking system combined with an IMU and has five buttons as well as a pressure-sensitive joystick as input options. We used the inertial-ultrasonic hybrid tracking device

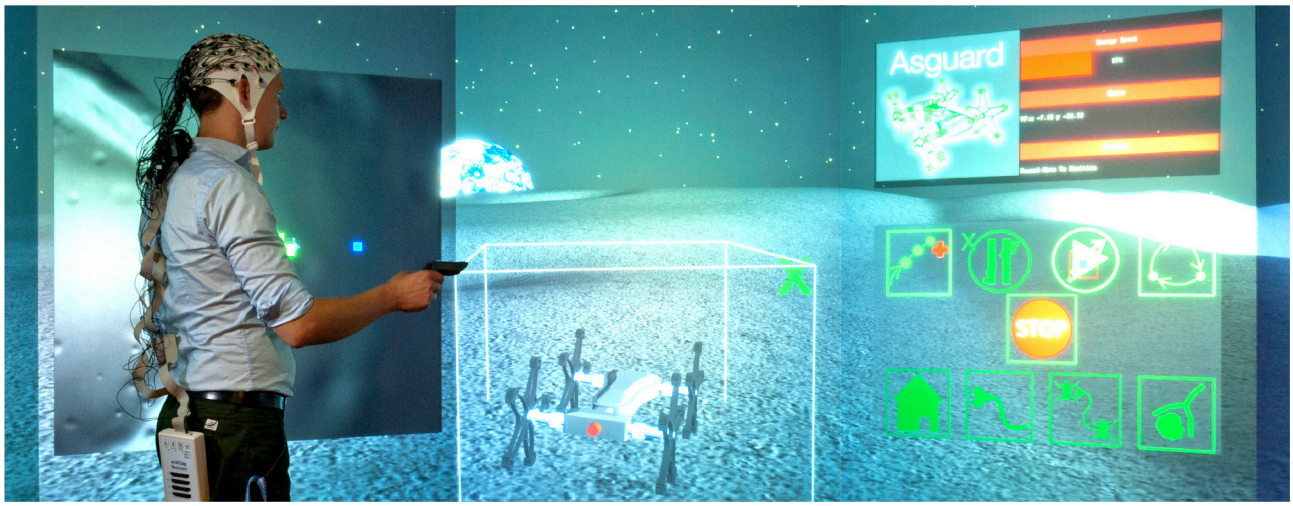


FIGURE 1 | Immersive virtual 3D multi-robot control using a CAVE supported by embedded Brain Reading (eBR).

InterSense IS-900 (Thales Visionix, Inc., Billerica, USA) in our experiments.

2.1.1. Human-Robot-Interaction

In general, the task of the operator in the multi-robot control environment (see **Figure 2**) was to supervise all robots and to assign new tasks to individual robots as indicated by messages presented to the user on the screen (see **Figures 3A,B** upper part for examples of different messages). Individual robots were labeled with different colors. Task messages were presented as icon based widgets supporting fast recognition by the operator. The operator used the interface to select a robot he or she wanted to control by either selecting the robot directly or by selecting the robot's icon in the upper part of the middle screen (see **Figure 3A**: 2). Moreover, information about the chosen system was presented to the operator on the right screen via an icon based information panel. Information such as the robot's name, its energy level, its current task as well as robot control commands were presented here (see **Figure 3A**: middle picture lower right corner). On the left monitor, tasks for the operator were listed as soon as the operator confirmed that he/she had seen the message by clicking on the appropriate robot icon on the monitor in the middle. By selecting the robot's icon with a double click, the virtual camera was additionally moved such that the chosen robot was in the focus of the operator. After selecting a robot, the operator can issue a task by clicking the corresponding robot control command icon. (see **Figure 3A**: 4). In case that an operator was not sure or did not recognize the robot to whom a task was assigned, he or she could select an unknown icon displaying a gray robot with a question mark (see **Figure 3A**). After clicking the unknown icon, all the missed tasks were displayed in the task list on the left screen. However, in the experiments presented here this gray robot button was disabled to force the subjects to focus on the task messages as much as possible. In case that a user did not recognize the task message correctly she or he had to wait for the automatic repetition of the task message.

2.1.2. Interaction Tasks

As mentioned in Section 2.1.1 the operator had to fulfill different tasks with the robots. Within the experiment there were three kinds of tasks with varying complexity:

- **Send message** The task with the lowest complexity is sending a message. This task can be solved by selecting the corresponding robot and clicking on the send-message icon within the robots control elements (see **Figure 3A** bottom number 4). An example of such a message for the green robot can be seen in the upper left part of **Figure 3B**.
- **Go to landmark** The task with a medium complexity is the navigation task. Within the experiment there are five different landmarks (for example see the cube labeled with 1 in **Figure 3A**). The goal of this task is to navigate the robot to one of these landmarks. Therefore the operator needs to select the robot and afterwards plan the path by creating waypoints. Waypoints will be put at the position of the cursor, when clicking a specific button. The robot will consecutively travel from waypoint to waypoint on straight lines. When the robot reaches the landmark the task is fulfilled. An example of such a message for the red robot with target position 3 can be seen in the upper middle part of **Figure 3B**.
- **Recharge robot** The most complex task is recharging the robot. Again the correct robot needs to be selected first. Afterwards the operator has to plan a path to the lander (see in the top right corner of **Figure 3A**). The path planning is realized as explained above in the "Go to landmark" section. After reaching the lander the robot needs to be selected again and the recharge icon from the robot's control elements needs to be activated by clicking on it. This task is more complex than the "Go to landmark" task due to a gap in between the two stages of the task and therefore the operator must track the

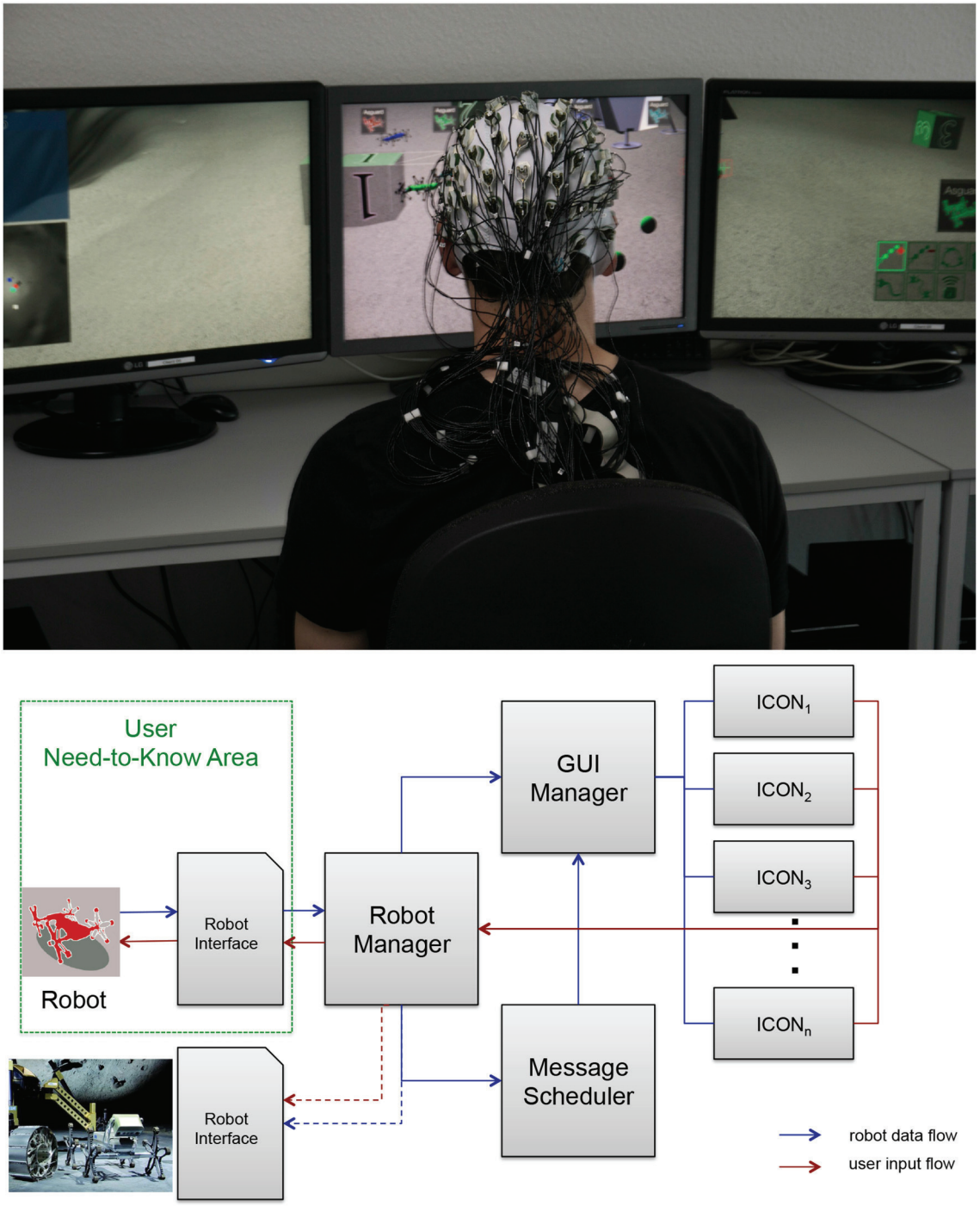


FIGURE 2 | Experimental setup. Upper part: Virtual multi-robot control in 2D using a multi-PC system supported by embedded Brain Reading (eBR). **Lower part:** Interaction is controlled by different software managers and schedulers. Widget-based icons are used to display information about the robots, messages for the user and to select robot commands. The user "Need-to-Know Area" is the part of the system visible to the user. The robot interface with connections to the real robots (depicted by dotted lines) is not yet implemented.

robot's state. The operator may also forget to click the recharge icon after the robot reached the lander. An example of such a message for the red robot can be seen in the upper right part of **Figure 3B**.

All tasks were pseudo-randomly chosen, such that no more than one task at a time was assigned per robot. When creating a new "Go to Landmark" task for a specific robot the robot's distance to the landmarks will be computed first. In order to solve the

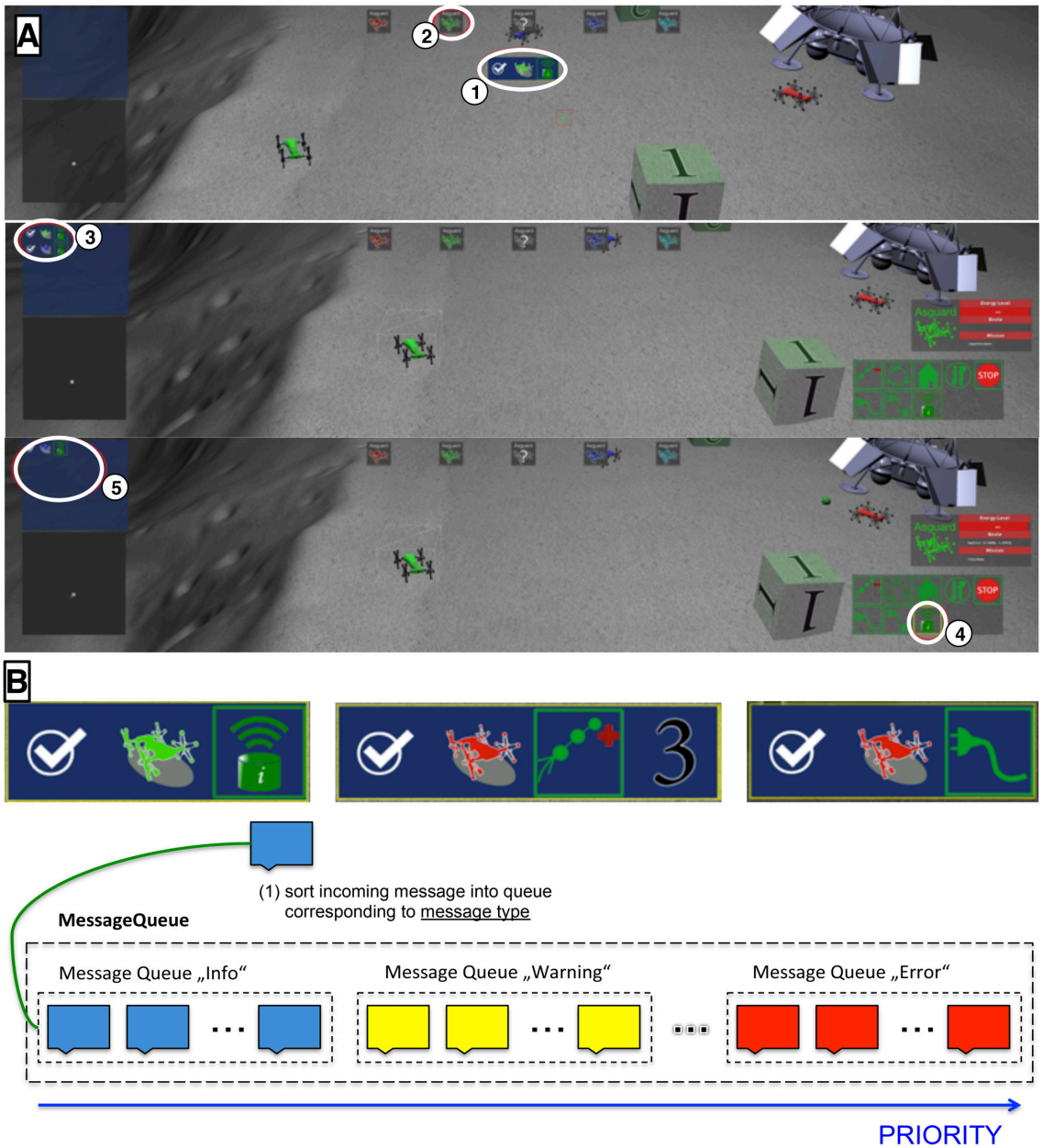


FIGURE 3 | Description of experimental setting and tasks performed by the operator. A (top) initial state, a task message was shown to the operator (1). The message contained information about the type of the task (e.g., send a message) and the corresponding robot (e.g., the green robot). The subjects had to confirm the task by clicking on a response button (2). **A (middle)** after the task was confirmed, it was shown in the task manager (3). **A (bottom)** when the green robot was selected, a menu with all possible control commands was shown. In this example, the mission could be accomplished by clicking on the send-message button of the control menu (4). When a task was accomplished, it was removed from the task manager (5). **(B)** The scenario contained three possible tasks, which were depicted by an intuitive symbol. All tasks were related to a specific robot, encoded by a colored symbol, see the following examples. **B (top left)** send a message with the green robot. **B (top middle)** send the red robot to waypoint 3. **B (top right)** recharge the red robot. Different robots (encoded by color) and different task messages were randomly combined. **B (bottom)** messages are sorted in order as they are presented. Some messages (repetitions of tasks) get a higher priority and will be presented earlier.

task the robot has to be in a specific radius around the chosen landmark. If the robot is already within the specific radius the new task would directly be solved when the robot is selected. In such a case the target landmark will be chosen among the other landmarks. Further, there was an automated mechanism which generated a "Recharge Robot" task in case that the energy level of a robot dropped bellow a certain value. This was necessary to ensure that a robot would remain fully functional. If a robot runs out of energy it would get stuck at its position and no more tasks could be solved by this robot.

When a message was presented requesting interaction the first response of the user like selecting the correct robot was counted as correct behavior. The message was not repeated. On the other hand, a predefined response time (in our experiments 13 s) and a predefined ISI was set for the operator. The predefined ISI was important for our experiments and research questions as will be explained in Section 2.4. Task messages were put into a message queue. To avoid unfair scheduling due to different urgency of information pending messages may change their priority over time (see **Figure 3B** lower part). So far it is implemented that a message is repeated as a warning in case that a complex task with longer duration is started, i.e., a robot is sent to a landmark, but does not arrive after a certain amount of time. Since the robot might have got stuck the warning is repeated with higher priority. To give the user an overview on initiated but still running tasks, they were visualized in a icon panel in the upper left corner of the left monitor in the order as they appeared with the newest depicted on the top (see **Figure 3A**: 5). As soon as a task was fulfilled the task message was removed.

2.2. Performed Experiments

Six subjects participated in the study. All subjects were male with normal or corrected to normal vision and aged between 20 and 38 years (mean: 28.74, SD: 6.92). All subjects were intensively trained in the scenario on a different day to get used to the tasks, i.e., to control the robots by using the developed MMI. On the same day of the study just before data recording subjects were asked to get comfortable with the scenario. The study consisted of 6 runs, performed in the same order. In each run, subjects had to complete 30 tasks. The response behavior was supervised and logged by the message scheduler (see **Figure 2** lower part).

In case no response was detected within 13 s after presentation of a task message, the same task message was again attached to the message queue. Since the queue is implemented as a FIFO (first in first out), the message is repeated after presentation of all other messages within the queue.

Task messages (**Figure 3** top illustration and **Figure 3B**) were presented for 1.1 s. The duration of presentation was determined by empirical tests with a different group of 4 subjects. The goal was to keep the duration of message presentation as short as possible to allow the evaluation of event-related activity in the EEG while ensuring that subjects were able to recognize and understand the presented messages.

2.2.1. Adaptation of the Inter-Stimulus Interval (ISI)

Between the 6 runs experimental conditions were varied with respect to the ISI (**Table 2.1**: EEG data). For runs 1 to 4 ISIs were fixed. We used two different ISIs: a long ISI (25 s) in runs 1 and 2 and a short ISI (15 s) in runs 3 and 4. In both cases an additional random jitter of ±5 s was added. Appropriate time intervals for long and short ISIs were empirically determined beforehand by tests with 4 subjects that were not involved in this study. The time interval for the short ISI was chosen such that the overall workload or overall task load caused by the message frequency was not too high. We were successful in empirically determining an appropriate time interval for short ISIs as supported by results of the evaluation of the NASA Task Load Index questionnaire (see Section 3.1.3). The time interval for the long ISI was empirically chosen to be clearly higher in the subjective perception of the 4 test subjects. A very low ISI could not be chosen, since we experienced that subjects easily gave up the run in cases of very short ISIs, i.e., with a duration of 5 s or even with a duration of 10 s. Further, no P300 was evoked under extremely stressful circumstances, as in runs with an ISI of 5 s. Moreover, to train the classifier qualitatively good training examples were required. And finally, we had to limit the number of runs and thus total experiment time to avoid overstraining the subjects.

For runs 5 and 6 the ISI was adapted online with respect to detectability of the P300 and related ERP activity. For the online detection of single-trial ERP activity a classifier was trained on examples from either runs 1 and 2 (for application in run 5) or on examples from runs 3 and 4 (for application in run 6) (see Section 2.8 for more details). Adaptation in runs 5 and 6 of the ISI was increased gradually (up to a maximum of 35 s in steps of 5 s) in case that an expected P300 was not detected two times in a row after a new task message or was decreased stepwise (down to a minimum of 5 s in steps of 5 s) in case that an expected P300 was detected two times in a row. For both adapted runs the ISI was preset to 25 s. We always startet with the fixed ISI condition with an ISI of 25 s in runs 1 and 2 to allow subjects to get comfortable with the control task. This was done since long training sessions just before the experimental session were not possible since they would have increased the total experiment time to an unacceptable long duration. For our experimental setting it was more important to record all runs in the same session to avoid between-session effects on the shape of the ERPs as well as the single-trial classification performance. Although subjects were intensively trained, they needed to readapt to the control of the robots, since the control task was very complex. Next, in runs 3 and 4 training data was recorded under the fixed ISI condition. We did not perform a run with adapted ISI right after the recording of training data with ISI 25 to keep both runs with adapted ISI close together and thus condition of the subjects similar. Further, interleaving runs with fixed and adapted ISIs were not performed, since this might have had an influence on the motivation of the subject during the recording of training data after a run with adapted ISI.

2.2.2. Ethics Statement

The study has been conducted in accordance with the Declaration of Helsinki and approved with written consent by the ethics committee of the University of Bremen. Subjects have given informed and written consent to participate.

2.3. Recorded Data

During each executed run EEG was recorded with 64 electrodes referenced against electrode FCz. An actiCap system (Brain Products GmbH, Munich, Germany) arranged as an extended 10–20 system was used for recording. Electrode impedance was kept below 5 kΩ. EEG signals were sampled at 5 kHz, amplified by two 32 channel BrainAmp DC amplifiers (Brain Products GmbH, Munich, Germany) and filtered with a low cutoff of 0.1 Hz and high cutoff of 1 kHz.

2.4. Research Goals & Hypotheses

The presented work addresses two different research goals with specific subgoals. (I) We want to show that a P300-related activity is naturally evoked when task messages are presented and recognized. (Ia) We investigate whether the evoked P300 is modulated by factors like demands on the operator or the operator's task engagement to previous tasks. (II) We want to show that single-trial detection of P300-related activity can be used to adapt the interaction with respect to the task engagement of the operator. (IIa) In particular, we investigate whether an individual balanced task involvement of the operator can be achieved by adaptation of the ISI resulting in a higher subjective contentment of the operator and in an individually optimized overall task performance.

By means of data recorded in runs 1–4 we investigated research goal (I). We artificially modulated the current task engagement (on the previous task) by presenting a new task. This was achieved by modulating the time interval between both consecutive tasks: long ISIs of 25 seconds in runs 1 and 2; short ISIs of 15 s in runs 3 and 4. Changes in P300 characteristics were investigated by averaged ERP analysis and machine learning methods. To support the usage of single-trial P300 detection we had to assure that the detection performance is adequately high and not too strongly influenced by ISI per se such that for very short ISIs possibly no P300 would be detectable in single-trial. For this, an *offline* machine learning analysis was performed first with training and test on runs with the same ISI. These results were used as a baseline for other experiments. This condition was called "baseline" condition. Using this analysis, we investigated whether P300-related activity is detectable in single-trial under application conditions and for different ISIs as well as how strongly different ISIs would influence classification performance.

Further, we investigated the effect of classifier transfer between runs with different ISIs. More precisely, a transfer of classifier between training runs (runs 1 and 2 or runs 3 and 4) and test runs (runs 5 and 6 with adapted ISI) was applied. This condition was called "transfer" condition. This offline analysis was relevant because under the *online* condition the classifier was transferred between different ISI conditions. Different ISIs were caused by the adaptation of the ISI under the *online* condition. Results allow to estimate the sensibility of the classifier for changes in ISI.

To achieve research goal (II) we adapted the developed MMI with respect to the current task engagement of the user to previous tasks when a new task was presented in runs 5 and 6 (**Table 2.3**: online stCL). Current task engagement was measured by the *online* single-trial classification of P300-related activity evoked by recognized target stimuli, i.e., task messages: (1) task engagement to a previous task was expected to be high in case that the P300-related activity was weakly evoked by a new task and thus not detected by a classifier, (2) task engagement to a previous task was expected to be low in case that P300-related activity was more strongly expressed and thus detected by a classifier. Note that in the online case each EEG trial after a presented first task message was classified, thus in case the operator completely missed a task message no P300 was expected to be evoked and could therefore not be detected. Hence, our approach did not only account for reduced P300 activity but also for missed P300 in case of missed target events.

To prove that the interaction of the user was improved by online adaptation of the ISI, we analyzed the total runtime, median reaction time and number of late responses and missed messages. We expected a reduction in total runtime by online adaptation of the ISI compared to the case of a fixed long ISI (ISI-25; runs 1 and 2). We did not expect a significant difference to be found for reaction times, since our approach would avoid user overload and responses were rather complex (see Section 2.1). However, we expected some late responses and missed messages in cases that the user was strongly involved in ongoing tasks when a new task was presented.

Our approach of online adaptation of the ISI allows to adapt an MMI with respect to the current task engagement or task load, improves user performance by equalizing the level of task engagement over all tasks and by selectively avoiding task overload. To further support this, we investigated the effect of an online adaptation of the ISI on averaged P300-related activity, i.e., we investigated whether expected changes related to task engagement in P300 amplitude could be found. For this evaluation, we compared averaged activity evoked in case of a fixed ISI of 25 s (runs 1 and 2) and a fixed ISI of 15 s (runs 3 and 4) with averaged P300-related activity evoked in runs 5 and 6.

Based on the research goals, we had three hypotheses: (1) The online adaptation of the ISI reduces total runtime if compared to the long fixed ISI condition (ISI of 25 s). (2) The modulation of the ISI influences amplitudes of averaged ERP. In particular, we expect differences between ISI types with respect to peak amplitudes of the averaged ERP. (3) The usage of historic data is feasible to detect P300 in the current data (e.g., a transfer of the classifier trained on historic data to the current data is possible).

2.5. Analysis of Subjects' Behavior

2.5.1. Analysis of Total Runtime

The total runtime was measured as the time between the first and the 30th task message within the experiment. This procedure was chosen since the total number of tasks differs slightly. This happens if the last task is from one of the categories "go to landmark" or "recharge robot" and if the adapted ISI is quite low. Solving one of these more complex tasks may take some time since the traveling distance can be rather long. Therefore, all robots may get one of these tasks. When one of the robots reaches its goal position the experiment is finished, but in this way more than 30 task messages could have been displayed to the user (see **Figure 5**).

For the statistical analysis, the value of total runtime was merged depending on the ISI type. This leads to three groups: ISI-25 (runs 1 and 2), ISI-15 (runs 3 and 4), and ISI-online adaptation (runs 5 and 6). The three ISI groups were compared by the Friedman test. For multiple comparison, the Wilcoxon signed-rank test was performed (the p-value was adjusted by the Bonferroni-Holm correction).

2.5.2. Analysis of Reaction Times

To calculate the reaction times, the EEG marker files were analyzed in order to deduce all important operator- and scenario-related events. Whenever a message was presented to the operator or the operator issued a control command this was marked in the EEG file. Based on the markers we calculated the reaction times, i.e., the amount of time the operator required to react to a task message by clicking on the correct response button for the robot. Only first task messages were considered in the analysis. Repetitions of task messages were not analyzed. The median of reaction time was calculated because of strong deviations and outliers. For a comparison with the ERP average analysis an additional analysis was performed considering only reaction times after target trials with ISIs that were used for the average analysis, i.e., target trials which belonged to one of the both groups: ISI-long or ISI-short (see **Table 1**). Note that for the ERP analysis not all trials could be used since in run 5 6.82% of the ISI-long trials and 13.33% of the short ISI trials and in run 6 18.57% of the ISI-long trials and 12.05% of the ISI-short trials contained artifacts and were discarded from analysis.

For the statistical analysis, the value of reaction time was merged depending on ISI type and this leads to three groups: ISI-25 (runs 1 and 2), ISI-15 (runs 3 and 4), and ISI-online adaptation (runs 5 and 6). The three ISI groups were compared by the Friedman test. For multiple comparison, the Wilcoxon signed-rank test was performed (the p-value was adjusted by the Bonferroni-Holm correction).

Additionally to median reaction times we calculated late responses after 15 s, and missed messages. EEG trials after messages with responses later than 15 s as well as missed message trials were not considered during training of the classifier (see Section 2.8).

2.5.3. Questionnaires

Before the experiments started, each subject was instructed to assess its skills related to the use of computers by filling out the "Computer usage questionnaire" (CUQ) (Schroeders and Wilhelm, 2011). For the statistical analysis, the Friedman test was performed to compare the patterns of computer usages between subjects. For multiple comparison, the Wilcoxon signed-rank test was performed (the p-value was adjusted by the Bonferroni-Holm correction). Furthermore, *after* each of the six runs of the experimental session, the subjects had to fill out the NASA Task Load Index (TLI) questionnaire (Hart and Staveland, 1988). For the statistical analysis, the value of task load index was merged depending on the ISI type and this leads to three groups: ISI-25 (runs 1 and 2), ISI-15 (runs 3 and 4), and ISI-online adaptation (runs 5 and 6). The three ISI groups were compared by the Friedman test. For multiple comparison, the Wilcoxon signed-rank test was performed (the p-value was adjusted by the Bonferroni-Holm correction).

2.6. Analysis of the MMI Behavior

The behavior of the MMI was analyzed by plotting the changes in the ISI for each subject in case of ISI adaptation (run 5 and 6, see **Figure 5**). **Figure 5** illustrates what kind of tasks were presented to the operator and which ISI was used, therefore the trace is the same as it was during the actual experiment. The purpose of this analysis was to give an impression of how "good" the adaptation worked and which ISI was most comfortable for the operator over the course of the run. For a comparison of the mean ISI between subjects, the mean ISI for each subject and run was calculated and the mean ISI of each run was compared between subjects by using the Friedman test. For a multiple comparison, the Wilcoxon signed-rank test was performed (the p-value was adjusted by the Bonferroni-Holm correction). Furthermore, we investigated whether the mean ISI is a useful indicator for the analysis of the MMI behaviors. To this end, the correlation between the mean ISI and the total runtime was calculated using the Spearman's rank correlation. We expected a positive correlation such that a longer ISI leads to a longer total runtime. In addition, we investigated *task type* as another factor with a potential effect on the total runtime. For example, the task types "go to landmark" and "charging robot" required a longer total runtime compared to the task type "send message." The frequency and order of task types were randomly chosen. Thus, differences in frequency of task types can in principle lead to differences in total runtime between subjects. However, we did not expect a strong correlation between task type and total runtime.

2.7. ERP-Average Analysis

Continuous EEGs were bandpass-filtered (0.1–30 Hz) and segmented into "target" trials from −100 to 1000 ms with respect to the stimulus onset (baseline correction: from −100 ms before the stimulus onset to 0 ms). As for the machine learning analysis only trials after the first task messages which have been responded to within a time period of 15 s were labeled as "target" trials when analyzing runs 1–4. For runs 5 and 6 again only trials with answered task messages were used as "target" trials and averaged as explained in **Table 1**. This procedure copies the procedure of the offline analysis. Trials after missed task messages were not averaged to exclude their contribution to the average ERP characteristic. We used a common average reference (CAR) and recalculated the data from channel FCz. For ERP average analysis only artifact-free segments were used (see **Table 1**). Artifact detection was performed semi-autonomously with a maximum amplitude of $-100\,\mu$V and $100\,\mu$V. We compared average artifact-free ERP activity evoked in runs with ISI-25 and ISI-15 as well as ISI-long and ISI-short. Trials for ISI-25 were conducted in runs 1 and 2 and trials for ISI-15 in runs 3 and 4. An adaptation of the ISI in runs 5 and 6 did not only result in various ISIs but also in individual ranges of ISIs for different users (see **Table 1**). Therefore, we individually divided the EEG segments of runs 5 and 6 into two ISI groups with respect to trials being evoked after short or long ISIs for each subject. For example, from the data of the subject depicted in **Figure 9** we merged examples after ISI-15 and ISI-20 to calculate average ERP activity after long ISIs and ISI-5 and ISI-10 to calculate average ERP activity after short ISIs (see **Table 1**). By means of this procedure, we could compare averaged P300-related activity

TABLE 1 | Number of artifact-free targets for each run and distribution over different ISIs.

Subject	Number of targets for each run					
	Run 1	**Run 2**	**Run 3**	**Run 4**	**Run 5**	**Run 6**
S1	26	25	28	27	32	21
S2	31	30	32	33	29	29
S3	23	19	9	27	23	17
S4	21	24	25	25	38	21
S5	23	19	25	23	25	22
S6	29	22	29	30	29	30
Average	25.50 ± 3.89	23.17 ± 4.17	24.67 ± 8.12	26.83 ± 4.02	29.33 ± 5.32	23.33 ± 5.09

Subject	Number of targets for all possible ISI-groups within runs 5 and 6					
	Run 5					
	ISI-05	**ISI-10**	**ISI-15**	**ISI-20**	**ISI-25**	**ISI-30**
S1	4	15	10	3	0	0
S2	14	7	5	2	1	0
S3	0	7	6	6	4	0
S4	0	2	10	20	6	0
S5	0	2	4	10	9	0
S6	8	12	3	4	2	0
Average	4.33 ± 5.72	7.50 ± 5.24	6.33 ± 3.01	7.50 ± 6.75	3.67 ± 3.39	0.00 ± 0.00
	Run 6					
S1	4	9	0	6	2	0
S2	1	12	11	3	2	0
S3	1	4	6	5	1	0
S4	0	1	2	6	7	5
S5	0	0	3	11	7	1
S6	19	6	3	0	2	0
Average	4.17 ± 7.41	5.33 ± 4.63	4.17 ± 3.87	5.17 ± 3.66	3.50 ± 2.74	1.00 ± 2.00
	Run 5 + Run 6					
Average	4.25 ± 6.31	6.42 ± 4.85	5.25 ± 3.49	6.33 ± 5.31	3.58 ± 2.94	0.50 ± 1.45

For average ERP analysis different ISIs were categorized in two ISI-groups: ISI-short (marked as red) and ISI-long (marked as blue).

for ISI-short and ISI-long of runs 5 and 6 with the activity evoked in runs 1 and 2 (fixed ISI of 25 ms: ISI-25) or runs 3 and 4 (fixed ISI of 15 ms: ISI-15) (**Table 2.2**). For peak detection, we selected a single window of the interval 0.3 –0.7 s after a "target" trial. The positive maximum peak was detected within the selected window.

For the statistical analysis of average ERP amplitude values with a sample size of 6 (i.e., 6 subjects), we performed the Wilcoxon signed-rank test to compare different ISI types (ISI-25 vs. ISI-15 and ISI-long vs. ISI-short).

2.8. Machine Learning Analysis

The data flow of the machine learning algorithm is depicted in **Figure 4A**. For the analysis the software framework pySPACE (Krell et al., 2013a) was used. First the continuous EEGs were processed by a DC removal filter, which is an online-capable method for centering the signal around zero. The normalized EEGs then were decimated from 5000 to 25 Hz.A cutoff frequency of 4 Hz was used for the anti-alias filter in the decimation process (Jansen et al., 2004; Ghaderi et al., 2014). Afterwards the EEGs were segmented into chunks of 1 s length. Chunks cut right after a *first* task message (*not* after repetitions of messages) were labeled as *"targets."* Within the training, these windows were only cut if the operator *responded* to the first task message within 15 s after presentation, in the online case *every* first task message was analyzed. We further cut *"standard"* windows of length 1 s while training. These windows were needed to train the used binary classifier. The standard windows

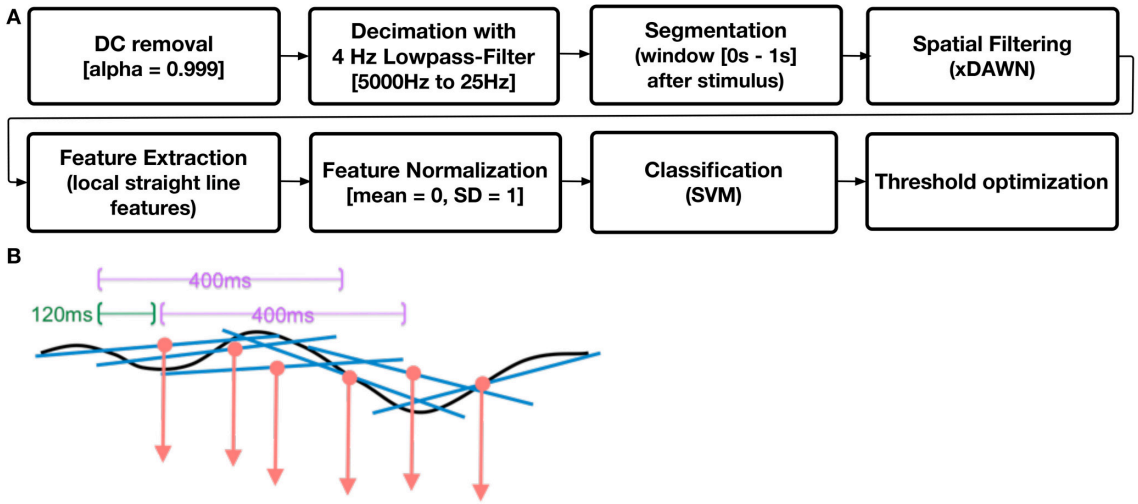


FIGURE 4 | Data Processing. (A) data flow for signal processing and single-trial classification. **(B)** example of an ERP (black line) being processed as local slopes of a straight line.

were cut every second with the constraint that no other action relevant for task recognition was performed in a range from $[-1, 1]$ s around the cut window. For the task recognition, actions such as the presentation of a task message or the response of the operator of one of these messages were used. The segments were further processed with the xDAWN spatial filter (Rivet et al., 2009). The xDAWN is a spatial filter especially designed for P300 detection. It (1) enhances the separability of the P300 ERP and noise and (2) reduces the dimensionality of the data. To achieve this, a set of filters maximizing the signal-to-signal-plus-noise ratio is computed on a training data set. The resulting filters can be used to create a set of pseudo-channels that contain the filtered signal. From the newly created pseudo channels the 8 most relevant channels were used for further processing.

As features we used local straight line features, i.e., polynomial features. To fit a polynominal function EEG data must be segmented (see **Figure 4B**). Earlier investigations showed that the longer the segments are chosen, the more coefficients are needed to keep the performance level high. For this paper every 120 ms, segments of length of 400 ms within the 1 s segments after stimulus onset were cut. Polynominal features of order one, i.e., straight lines were fitted to the 400 ms long segments of the ERP data with 120 ms steps to describe the ERP in terms of a series of slope values (see **Figure 4B**). Polynominal features of order one have been chosen since in former investigations of P300 ERP activity the highest value was obtained with this low coefficient. Previous analyses, too, as performed for example in Wöhrle and Kirchner (2014b) support our choice.

After this preprocessing a Support Vector Machine (SVM) (Chang and Lin, 2011) was used as classifier. During training the complexity of the SVM was optimized with a grid search and an internal five-fold cross validation. The possible complexities were 10^n with $n \in 0, -1, \ldots, -6$. Further a threshold optimization was applied (Metzen and Kirchner, 2011). Further a threshold optimization was applied (Metzen and Kirchner, 2011). After building the model of a SVM the decision boundary is defined as 0 and the two classes (here target and standard) are at the positive and negative side of the boundary. The threshold optimizations gives the opportunity to further improve the classification performance with respect to a given metric, here the balanced accuracy. The threshold is shifted into the negative or positive direction, in a way that for the training data the highest classification performance in terms of balanced accuracy is achieved.

We used the balanced accuracy (bACC), i.e., the mean of true positive rate (TPR) and true negative rate (TNR), as the performance metric due to the insensitivity of this metric to changes in class distribution (Krell et al., 2013b; Straube and Krell, 2014). Area under the curve (AUC) values were additionally calculated. Classification performance was compared between all conditions. For details see **Table 2.3**. Although the adaptation of the ISI was evaluated online (**Table 2.3**: online stCL), we additionally analyzed the data in the offline mode (**Table 2.4**: offline stCL). This procedure was chosen for reasons of fair comparison. While in the online mode data of two runs (runs 1 and 2 or runs 3 and 4) were used for training, this was not possible for evaluating the general P300 detectability in case of fixed ISIs since here only one run could be used for training while the other was used for testing. By means of the chosen offline approach we were able to analyze the no-transfer case (as baseline/control) and the transfer case equally.

For the statistical analysis on single-trial classification performance, two separate comparisons were performed by using the Wilcoxon signed-rank test. First, we compared two online cases: online P300 detection in run 5 vs. run 6 (see (e) vs. (f) in **Table 2.3**: online stCL). Here, two samples per subject were obtained for each online case. Altogether, we obtained a sample size of 12 (2 samples $\times$ 6 subjects) for each online case. Second,

TABLE 2 | Design for the recording of EEG data, evaluation design for ERP analysis and design for the analysis of single-trial classification performance (online/offline-mode).

Table 2.1. EEG data	Table 2.2. Evaluation design for ERP analysis	
(a) run 1: fixed ISI of 25 s	Average ERP in (a)	ISI-25: average of (a) and (b)
(b) run 2: fixed ISI of 25 s	Average ERP in (b)	
(c) run 3: fixed ISI of 15 s	Average ERP in (c)	ISI-15: average of (c) and (d)
(d) run 4: fixed ISI of 15 s	Average ERP in (d)	
(e) run 5: online adapted ISI	Average ERP in (e)	Various ISIs are grouped in short and long ISI
(f) run 6: online adapted ISI	Average ERP in (f)	for each subject: (e), (f), or average of (e) and (f)

Table 2.3. Online stCL		Table 2.4. Offline stCL							
Adapted ISI classifier transfer		Adapted ISI (e) transfer		ISI-25 (control) no transfer		Adapted ISI (f) transfer		ISI-15 (control) red no transfer	
Training	Test	Training	Test	Training	Test	Training	Test	Training	Test
ISI-25 (fixed ISI of 25 s)		(a)	(e)			(c)	(e)		
(a) + (b) merged	(e)	(b)	(e)			(d)	(e)		
			Mean (e)	(a)	(b)		Mean (e)	(c)	(d)
ISI-15 (fixed ISI of 15 s)		(a)	(f)			(c)	(f)		
(c) + (d) merged	(f)	(b)	(f)			(d)	(f)		
			Mean (f)	(b)	(a)		Mean (f)	(d)	(c)

ERP, event-related potentials; online stCL, online single-trial classification; offline stCL, offline single-trial classification; and ISI, inter-stimulus interval. Each run contained 30 *trials. For online single-trial classification,* 60 *trials (e.g., runs* 1 *and* 2*) were used to train a classifier and* 30 *trials (e.g., run* 5*) were used for evaluation. For offline single-trial classification,* 30 *trials were used for training and testing in both cases (no transfer/classifier transfer).*

two adapted ISI conditions were compared with two fixed ISI-conditions in offline mode depending on the type of training data (ISI-25 or ISI-15) used to train the classifier: (1) adapted ISI (e) vs. ISI-25 (control) (see in **Table 2.4**: offline stCL) and (2) adapted ISI (f) vs. ISI-15 (control) (see in **Table 2.4**: offline stCL). In the offline analysis, the number of training examples for the fixed ISI conditions (run 1 or run 2 / run 3 or run 4, see **Table 2.4**) was half the number of training examples used for the adapted ISI conditions in case of online evaluation (run 5 or run 6, see **Table 2.3**). For a fair comparison between the adapted and fixed ISI-condition, only one run (run 1 or run 2) was used to train the classifier to test it on run 5, and the mean of classification performance obtained by using run 1 or run 2 for training was calculated in the case of the *adapted ISI(e)* (see **Table 2.4** (e) in offline stCL). Similarly, in the case of the *adapted ISI(f)*, only one run (run 3 or run 4) was used to train the classifier to test it on run 6 and the mean of classification performance obtained by using run 3 or run 4 for training was calculated (see **Table 2.4** (f) in offline stCL). Each adapted and fixed condition has two samples per subject. Altogether, we obtained a sample size of 12 (2 samples $\times$ 6 subjects) for each condition.

3. RESULTS

3.1. Behavior of Subjects

3.1.1. Total Runtime

Figure 5 shows how the ISI changed over one run based on the inferred task load and task engagement of the user measured by P300 detectability. Subjects reported that the online adaptation made them feel to have just the right task frequency. This indicates that online adaptation of the MMI has a positive effect on the interaction. The finding was supported by the results of the behavioral analysis of the total runtime (see **Figure 6**). The online adaptation of the ISI reduced total runtime significantly if compared to the ISI-25 condition [$p < 0.001$]. Moreover, there was no significant difference in total runtime between the case of online adaptation of ISI and the case of ISI-15 condition [$p = n.s.$].

3.1.2. Reaction Time

Figure 7A shows the median reaction time for individual subjects over all runs. It can be seen that median reaction times are very similar over all conditions and runs for each subject. When merging the two runs of each condition (ISI-25, ISI-15, and ISI-adapt) we found no significant difference between ISI types. However, when analyzing median reaction time individually for ISI-long and ISI-short groups of runs 5 and 6 as performed for average ERP analysis it can be seen that the reaction time on task messages presented after short ISIs showed a higher variance compared to task messages presented after long ISIs (see **Figure 7B**).

A descriptive analysis of the sum of late responses and missed messages per subject for each run is visualized in **Figure 8**. It can be seen that for some subjects the number of late responses and missed messages was higher than for others (subjects 3 and 4).

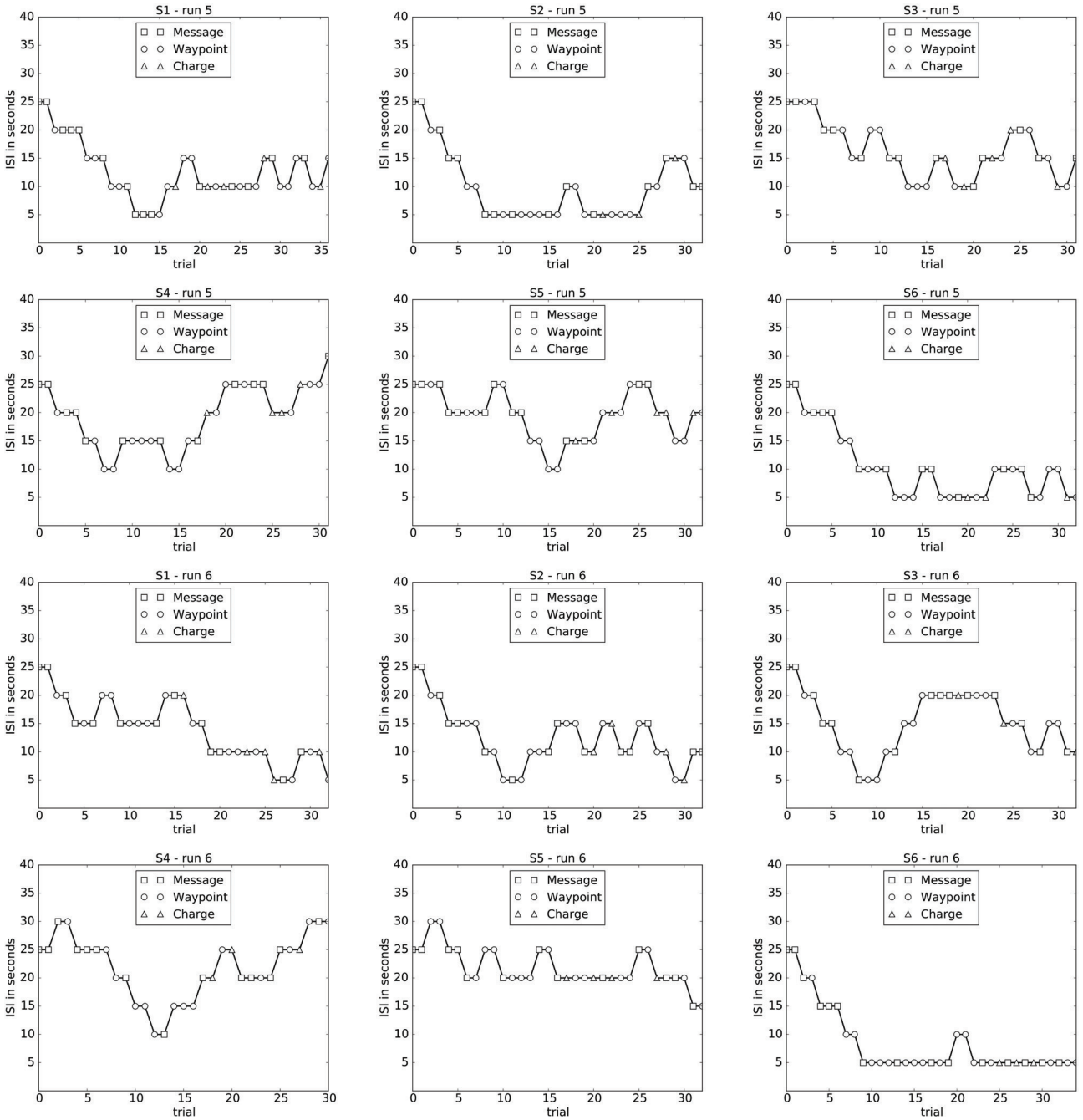


FIGURE 5 | Changes in ISI over each run in case of the adapted ISI condition (runs 5 and 6) for each subject are depicted.

Table 3 provides information about the number of late responses, missed messages and the sum of both as depicted in **Figure 8**.

3.1.3. Questionnaires

The analysis of the "computer usage questionnaire" shows a significant difference between subjects, especially subject 4 differed significantly from the other subjects [$p < 0.03$]. The analysis of the "NASA Task Load Index (TLI) questionnaire" shows no significant differences between runs [$p = n.s$].

3.2. Behavior of MMI

Figure 5 depicts the changes of the ISI for both adapted runs (runs 5 and 6) for each subject. It can be seen that the adaptation of the ISI is very individual for each subject and even for each run. While for some subjects and runs, as for subject 2 in run 5, the ISI goes down to the minimum of 5 s and stays there for almost 20 trials, for other subjects the ISI is not reduced that much (see for example subject 5 for both runs).

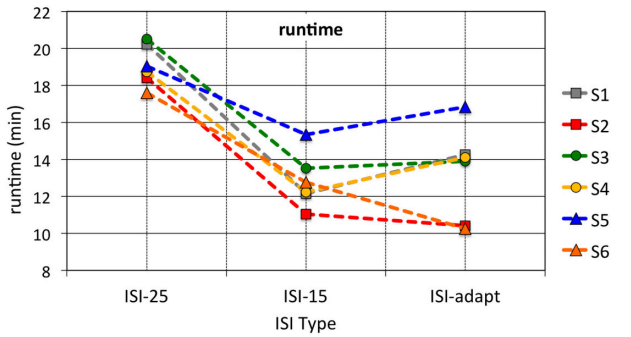


FIGURE 6 | Mean runtime for different ISI conditions. The means of both runs for each ISI type are depicted. The median across all subjects for each ISI type was 17.59 for ISI-25, 11.49 for ISI-15 and 12.45 for ISI-adapt.

In most cases the ISI gradually decreases just to later increase. However, there are exceptions from these findings. For example subject 1 shows a reduction of ISI at the end of run 6 and subject 6 stays with a low ISI during both runs. For all subjects the ISI starting with 25 s was reduced to a lower mean ISI with average values of 14.67 and 15.62 s (runs 5 and 6) (see **Table 5**). Moreover, we could also find differences in the mean ISI between subjects. For example, while the mean ISI for subject 4 and subject 5 is around 19 and 22 s (runs 5 and 6), the mean ISI for subject 6 is at 10.45 and 8.43 s (runs 5 and 6) and for subject 2 at 9.85 and 12.42 s (runs 5 and 6). The mean ISI for Subject 4 and subject 5 was significantly higher compared to the other subjects [$p < 0.017$]. Furthermore, the mean ISI correlated strongly with the total runtime [$r = 0.874$, $p < 0001$], but not the task type (e.g., send message, go landmark, etc.).

3.3. Average P300-Related Activity

As shown in **Figures 9**,**10**, we observed differences in averaged ERP shape depending on the ISI condition (short/long ISI). Note that the ISI in case of long ISIs and short ISIs differ for both average analysis conditions (fixed-ISI condition and adapted-ISI condition, see **Table 5**). While for ISI-long average analysis condition the ISI is set to 25 s, ISI-long for the adapted-ISI condition is around 19 s. Similar differences can be found for the ISI-short average analysis condition (fixed short ISI: 15 s versus adapted ISI around 10 s). The peak amplitude of the averaged P300-related activity was not significantly reduced in case of ISI-15 (runs 3 and 4) compared to ISI-25 condition (runs 1 and 2) [$p = n.s.$]. However, we observed a significant reduction in averaged P300 amplitude in run 5 and run 6 for short ISI groups compared to long ISI groups [$p < 0.04$]. Furthermore, there was a significant difference between ISI-15 and ISI-short [$p < 0.04$], but not between ISI-25 and ISI-long [$p = n.s.$].

3.4. Online P300 Detectability

Finally, we achieved high classification performances in both the online and offline analysis. In the online evaluation, we found no significant difference between both online runs [adapted ISI (e) vs. adapted ISI (f): bACC of 0.77 vs. bACC of 0.78, $p = n.s.$, see adapted ISI (e) vs. adapted ISI (f) in **Table 4.1**]. In the offline evaluation, classification performance obtained by using the classifier trained on ISI-25 statistically differed from classification performance obtained in case of no transfer [ISI-25 vs. adapted ISI: bACC of 0.84 vs. bACC of 0.75: $p < 0.003$, see adapted ISI (e) vs. ISI-25 in **Table 4.2**]. However, we found no significant difference in classification performance when using the classifier trained on ISI-15 compared to the case of no transfer (ISI-15) [ISI-15 vs. adapted ISI: bACC of 0.80 vs. bACC of 0.79: $p = n.s.$, see adapted ISI (f) vs. ISI-15 in **Table 4.2**]. There was no significant difference between the online and offline evaluation for the case of ISI-adaptation [adapted ISI (e) in **Table 4.1** vs. adapted ISI (e) in **Table 4.2**: $p = n.s.$; adapted ISI (f) in **Table 4.1** vs. adapted ISI (f) in **Table 4.2**: $p = n.s.$]. In summary, we found a transfer effect on classification performance in case that the classifier was trained on data from the ISI-25 runs. However, such an effect was missing when the classifier was trained on data from the ISI-15 runs. It must be emphasized that the classification performance was very similar in case of both classifier transfer analyses, i.e., adapted ISI (e) and adapted ISI (f) (see **Table 4.1**).

4. DISCUSSION

4.1. Improvement of Interaction

Supporting our hypothesis (1) behavioral data showed that total runtime in runs with adapted ISI was significantly shorter compared to an unadapted condition with an ISI of 25 s. Although there was no significant difference between the adapted ISI and the fixed shorter ISI of 15 s the mean total runtime was still very low considering the fact that runs with ISI adaptation did start at an ISI of 25 s. Significant differences in the total runtime between runs with adapted ISI and the fixed shorter ISI of 15 s were not expected, since the time needed until a task was performed by a robot does (although not strongly) depend on the type of task. For example, sending data was very fast and instant while reaching a certain landmark could take a long time depending on the current position of the robot and the landmark. Thus, some deviation in runtime depending on the kind of tasks that had to be performed by the robot, was expected. On the other hand, we did not choose subjects with a certain qualification but chose subjects independent of their experience in robot control or video gaming. Thus, we expected differences in the subjects' performances resulting in different "suitable" ISIs and hence also in different total runtimes. Important was that a significantly shorter runtime could be achieved compared to the fixed ISI-25 condition under which all the subjects could perform the tasks without being stressed.

Besides, the goal was not to reduce the total runtime to a minimum but to adapt the ISI with respect to the demands of the user of the MMI. Indeed, for some subjects the mean ISI was reduced to mean values around 10 s while for other subjects, i.e., subjects 4 and 5, the ISI was clearly above 15 s (around 19 s, see Section 3.2). On the other hand, even for subjects for whom the ISI was not reduced that much, mean ISI was clearly below 25 s, supporting our presupposition from the 4 test subjects that were not included in this study that a fixed ISI of 25 s ensures that all subject can easily perform the tasks but will probably make the subjects feel bored. An interesting

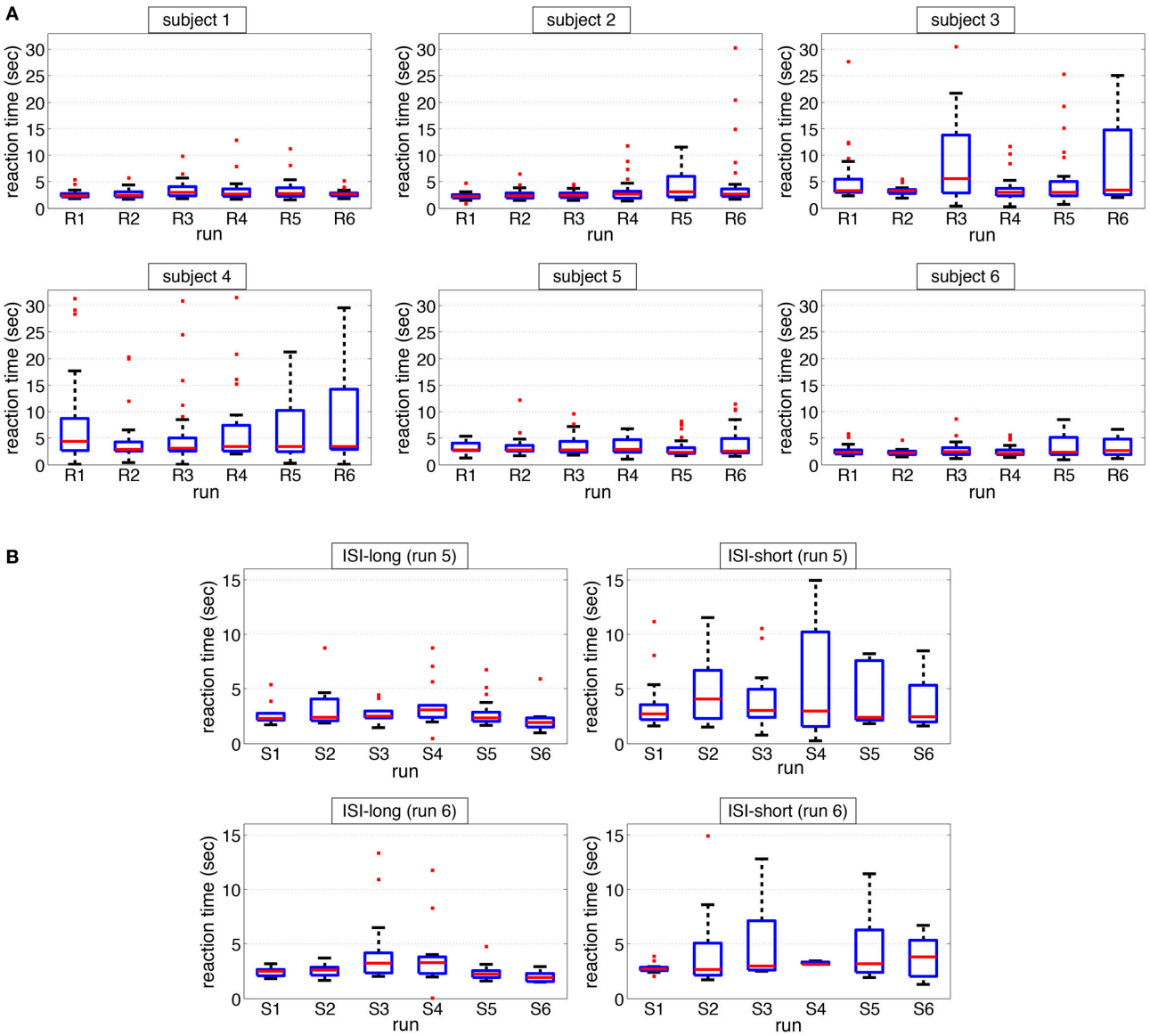


FIGURE 7 | Median response time. (A) median reaction times for each run and each subject are depicted. **(B)** median reaction times for each run and each subject sorted with respect to trials with short and long ISI as defined for average ERP analysis are depicted.

finding is that subject 4 for which the ISI was reduced only to a still high value (around 19 s) significantly differed from the other subjects with respect to computer usage as evaluated by the "Computer usage questionnaire" (CUQ). This finding supports our assumption that the MMI could be adapted based on the detectability of the P300 to support the user with respect to her or his general capabilities. Note that subject 4 showed the lowest classification performance in both runs compared to the other subjects (although no significant differences between subjects could be found, see **Table 4**). Moreover, subject 4 had a high amount of late responses and missed messages (see **Figure 8**). Another interesting finding is that the median reaction time does not significantly differ between subjects. This finding suggests that in our application behavioral data is probably not a good indicator for task load. Moreover, it shows that using our approach subjects were exposed to an appropriate workload. In summary, the results suggest that by using the developed MMI utilizing embedded Brain Reading, the MMI cannot only be adapted to the general capabilities of the user (e.g., experienced or rather inexperienced in computer usage) but also to the changes in task load over time.

4.2. Changes in the Characteristic of Average P300 Depending on the ISI

Applying average ERP analysis, we were able to show that during a complex multi-robot control task a P300-related activity is

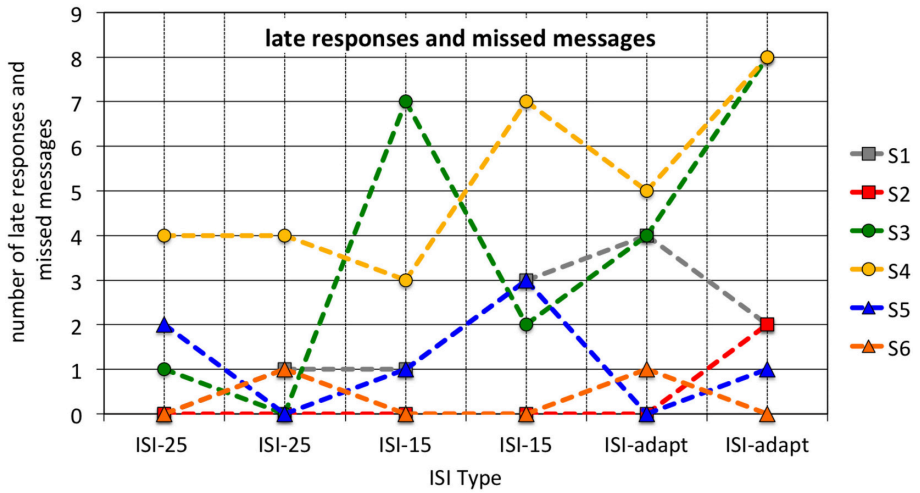


FIGURE 8 | Sum of late responses and missed messages for each run and each subject is depicted.

TABLE 3 | Number of tasks with late or no response in runs 5 and 6.

	Run 5			Run 6		
Subject	**Late**	**Missed**	**Total**	**Late**	**Missed**	**Total**
S1	4	0	4	2	0	2
S2	0	0	0	2	2	4
S3	4	0	4	8	0	4
S4	5	0	5	8	0	8
S5	0	1	1	1	0	1
S6	1	0	1	0	2	2

evoked by task messages which are presented to the operator. This finding is the most important basis for our approach to adapt an MMI based on P300 detectability. As expected we found *no* significant differences in the averaged-peak P300 amplitude for both fixed ISI conditions. This supports earlier findings that the ISI has no influence on the P300 amplitude in case of long ISIs (longer than 6–8 s as found by Polich, 2007). More importantly this finding supports our assumption that on both fixed ISI conditions the general workload on the subjects was rather modest and comparable. Hence, any found differences in the P300 peak amplitude should be caused by changes in the current task load and task engagement. This finding is supported by the fact that in case of an ISI adaptation the average P300 peak amplitude was significantly reduced for trials after short ISIs compared to trials after long ISIs.

Our results from the average ERP analysis support hypothesis (2): we could show differences in the P300 peak amplitude for average conditions with a high task load (averaged ERP activity after ISI-long in adapted ISI condition) compared to average conditions with low task load (averaged ERP activity after ISI-long in adapted ISI condition).

The finding that the peak amplitude of the average P300 activity after trials with ISI-short (adapted ISI condition) is significantly smaller compared to the peak amplitude of the average P300 activity of both fixed ISI conditions (ISI-25 and ISI-15) suggests that for all subjects the MMI was indeed adapted to achieve the best performance without enhancing the workload too much such that no P300 would be evoked. Tests on 4 subjects (not included in this study) showed that in cases in which the workload was too high no P300 was evoked on average or could not be detected in single-trial while subjects reported that they were very stressed and could not perform the tasks. Hence, the MMI is adapted such that subjects perform best while avoiding an excessive general workload. Some subjects were able to keep their performance high with a short ISI all through the experiment while others did not. For the latter, the MMI was again adapted to longer ISIs reducing the task load back to normal. The task load and thus the general workload being modest under the adapted condition after long ISIs is supported by the finding that the average P300 peak amplitude evoked after long ISI trials under the adapted ISI condition is comparable to the average P300 peak amplitude under the fixed ISI conditions (ISI-25 and ISI-15). This was even the case although the mean long and short ISI differed strongly between subjects (see **Tables 1**, **5**). Based on these findings we suggest that the P300 ERP is indeed a good indicator for the current and individually different task load of a subject while controlling the robots.

4.3. Detectability of P300 in Single-Trial

The results of the offline machine learning analysis support that the P300-related activity which was evoked by task messages can be detected in single-trial even in case that the classifier is transferred between different ISI conditions. Thus, the results support hypothesis (3).

When comparing online classification with offline classification a performance drop can be observed. This can be explained as follows: In the online case each first message was classified independently of having been responded to. Therefore, trials after missed task messages which likely did

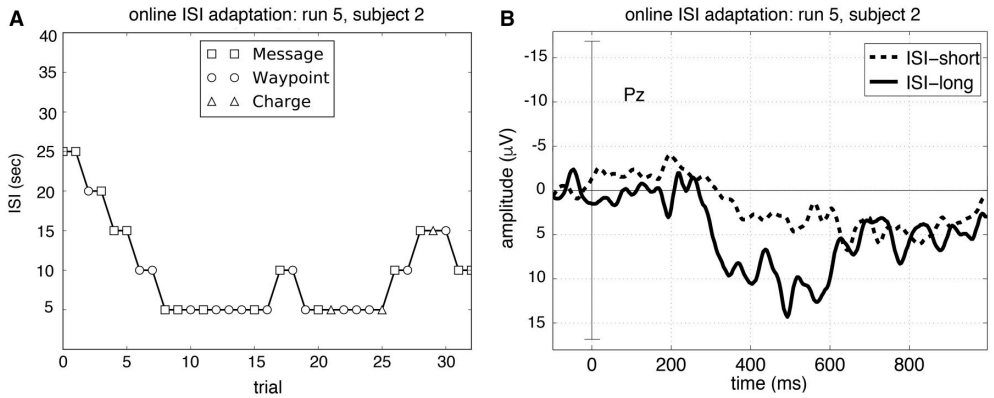


FIGURE 9 | Adaptation of the ISI over one run and the evoked averaged ERP activity at Pz for one subject: (A) depicts online adaptation of ISI in case of using the classifier trained on data with ISI of 25 s (i.e., training data: ISI-25, test data: run 5, see (e) in Table 2.3: online stCL) and (B) the corresponding averaged ERP curve evoked during the same run (Table 2.2: ERP analysis). Only artifact-free trials were used: 7 trials for ISI of 15 s and ISI of 20 s; 21 trials for ISI of 5 s and ISI of 10 s. Different types of tasks (tasks of type: message, way point and charging, see **Figure 3** for details) had to be solved by the subjects.

not contain a P300 were classified, leading to "false negative" results. It was therefore expected that classification performance was lower for the online case, since the approach is sensitive to missed targets. The small difference between online and offline results support that the MMI was well designed such that only few target events (messages) were completely missed (see also **Table 3**).

Besides this, in both transfer cases similar classification performance can be achieved. Hence, for an application it is not that relevant for the classification on which data a classifier is trained. While we found no significant differences between subjects for online classification performance it is noticeable that subject 4 had the worst classification performance in both runs compared to the other subjects (Discussion see Section 4.1).

4.4. P300 Detectability as Index for Task Load or Task Engagement

By reducing the ISI to way shorter ISIs compared to the ISI-15 condition (see **Table 5**) we strongly enhanced the task load and likelihood of conflicts since subjects might still be engaged in a former task when a new task message was presented. This is supported by two findings: (1) the higher variance in reaction time found for the ISI-short group (based on grouping for average analysis) and (2) the smaller average P300 evoked after short ISI trials in the adapted ISI condition (see **Figure 10**). Likely, subjects were still involved in a previous task and often could therefore respond to a new task only with a delay.

We found a similar effect in a previous study (Kim and Kirchner, 2012). In this previous study, subjects played a labyrinth game and had to respond to target stimuli which were presented in an oddball design. However, subjects were not allowed to respond to target events right away. We asked the subjects to steer the ball in a save corner first before answering a target event. When analyzing the average P300 potential we grouped the data with respect to reaction time such that the first group consisted of EEG trials with only short reaction times up to 1.4 s, for the second group trials were added which had reaction times up to 1.6 s, for the third group up to 1.8 s, the fourth up to 2.0 s, and the fifth up to 7.0 s. Although keeping the trials with short reaction times up to 1.4 s for the second group and up to 1.6 s for the third group, we still found descriptive differences in average peak amplitude of the P300 component between all groups with highest amplitude for the group of 1.4 s and lowest for the group of 7.0 s. When classifying between standard and target trials we found significant differences between the group of 1.4 s compared to all other groups with the exception of group 1.4 s compared to group 1.6 s and significant differences between the group of 7.0 s compared to all other groups with highest classification performance of 0.85 for the group of 1.4 s and lowest classification performance of 0.76 for the group of 7.0 s. These results suggest that ongoing task engagement, i.e., playing the labyrinth game, reduced the P300 evoked by a new target stimulus tremendously and would also reduce classification performance.

4.5. Summary and Outlook

In summary, our results show that complex interaction between humans and robotic systems can be improved by the application of an MMI adapted by eBR. The time between tasks can be adjusted such that a reduction of run time compared to a safe mode is possible. The strength of adaptation does further correlate with the experience of the user. Thus, the MMI can be adapted to the needs of the user within a range of workload that can otherwise not be resolved. Our approach shows that EEG activity like the P300-related activity that is naturally evoked during interaction can be used to adapt an MMI with respect to online changes in task load or task engagement of an operator. Thus, the dual-task design (with a primary and usually artificially introduced secondary task) that is often applied to infer on current processing capacity of the brain must not be applied to adapt for task engagement. The ERP activity can be used rather naturally, similar to approaches that make use of ratios of EEG power bands (Pope et al., 1995) while being specific to certain

TABLE 4 | Online and offline classification performance.

Table 4.1. Online single trial classification performances (cf. Table 2-3. Online stCL)

	bACC		AUC	
	Adapted ISI (e) classifier transfer	Adapted ISI (f) classifier transfer	Adapted ISI (e) classifier transfer	Adapted ISI (f) classifier transfer
S1	0.7646	0.7586	0.8481	0.8343
S2	0.8403	0.8177	0.8692	0.8596
S3	0.7892	0.7422	0.8516	0.8681
S4	0.6486	0.7083	0.7620	0.7365
S5	0.7981	0.7631	0.8790	0.8621
S6	0.7931	0.9021	0.9292	0.9375
Mean	0.7723	0.7820	0.8565	0.8497

Table 4.2. Offline single trial classification performances (cf. Table 2.4. Offline stCL)

	bACC				AUC			
	Adapted ISI (e) transfer	ISI-25 no transfer	Adapted ISI (f) transfer	ISI-15 no transfer	Adapted ISI (e) transfer	ISI-25 no transfer	Adapted ISI (f) transfer	ISI-15 no transfer
S1	0.7057	0.8086	0.7595	0.7725	0.8063	0.8815	0.7979	0.7966
S2	0.7690	0.9536	0.8366	0.9361	0.8183	0.9873	0.8870	0.9604
S3	0.6987	0.7568	0.7912	0.7406	0.8643	0.8923	0.476	0.8159
S4	0.6772	0.7310	0.7195	0.7187	0.7670	0.7451	0.7170	0.8459
S5	0.7722	0.8135	0.7685	0.7650	0.8054	0.8942	0.8193	0.8745
S6	0.8625	0.9600	0.8843	0.8864	0.9037	0.9692	0.9549	0.9045
Mean	0.7476	0.8373	0.7933	0.8032	0.8275	0.8951	0.8373	0.8663

TABLE 5 | Mean ISIs in case of online ISI-adaptation (runs 5 and 6).

	Mean ISI in sec.					
	Table 5.1		Table 5.2			
Subject	Run 5	Run 6	Run 5		Run 6	
			ISI-long	ISI-short	ISI-long	ISI-short
S1	12.7 ± 5.01	13.94 ± 5.47	16.15	8.95	21.25	8.46
S2	9.85 ± 5.97	12.42 ± 5.09	16.47	6.67	16.07	9.62
S3	16.25 ± 4.84	15.15 ± 5.43	22.00	12.31	17.27	9.00
S4	19.22 ± 5.46	21.94 ± 5.63	21.15	14.17	22.69	13.33
S5	19.55 ± 4.33	21.82 ± 3.44	22.37	13.33	25.63	8.57
S6	10.45 ± 6.08	8.43 ± 5.95	17.86	8 .00	15.00	6.20
Average	14.67 ± 4.29	15.62 ± 5.35	19.33 ± 2.84	10.57 ± 3.10	19.65 ± 4.19	9.20 ± 2.33

Table 5.1*: mean over all trials.* ***Table 5.2****: mean over a selected group of trials with ISI-short and ISI-long as defined for average ERP analysis (see* ***Table 1****).*

stages of information processing (Prinzel et al., 2003). Hence, for the user, our approach of measuring brain states and task engagement remains invisible and avoids any possible additional load on the user, since the task itself is used to measure task load, without any additional task.

In the future, we will have a closer look at the long term effect of adaptation of the ISI compared to a high task load condition, i.e., ISI of 10 s or even lower. For this, it is required to avoid the recording of extra training data since this requires a considerable amount of time. The total time for one experiment (6 runs) was already between three to 4 h including preparation. Thus, for a long term study, preparation and especially training of the classifier must be kept to a minimum. This can be achieved by using zero-training approaches (Krauledat et al.,

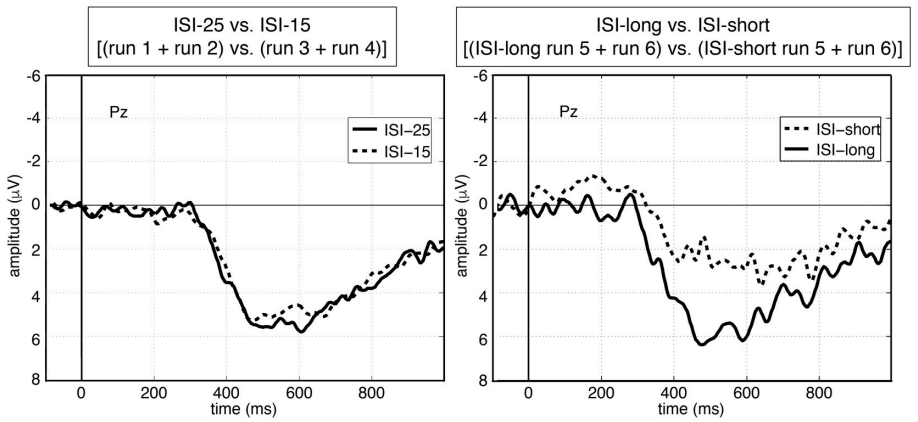


FIGURE 10 | Averaged ERP activity over all subjects at electrode Pz under offline condition (left side) and under online condition (right side). Grand averages over all subjects are depicted. Each run contained 30 trials. Only artifact-free trials were used: 292 trials for ISI-25 and 313 trials for ISI-15, 139 trials for ISI-long and 164 trials for ISI-short.

2008; Kindermans et al., 2012) or by using old training data from either previous recordings of the same subject or other subjects (Lotte and Guan, 2010; Devlaminck et al., 2011; Samek et al., 2013). To reduce transfer effects (between sessions and between subjects) adaptive algorithms for the spatial filter (Rivet et al., 2011; Ghaderi and Straube, 2013), the classifier (Li et al., 2008; Lu et al., 2009; Tabie et al., 2014) or both (Wöhrle et al., 2015) can be applied. Moreover, we want to investigate whether adaptive measures can be used to even improve the classification performance and the support for the user as we could already show for the prediction of movement onsets (Tabie et al., 2014). Finally, we will investigate transferability of the final approach to a mobile analysis system which makes use of hardware accelerators as already tested for the current application. Even for an adaptive approach hardware accelerators have shown to be feasible for the detection of both the P300 event-related potential (Wöhrle et al., 2013b,a, 2014a) and the movement-related ERP activity (Wöhrle et al., 2014b).

AUTHOR CONTRIBUTIONS

EK developed concepts for the MMI and for data evaluation, interpreted the results and wrote most of the manuscript. She further contributed to data recording, evaluation and statistical design. SK performed the analysis of ERP averages and of questionnaires and the statistic evaluation of classification performance, ERP results and behavior data. She further wrote parts of the manuscript and supported data acquisition. MT conducted the experiments, designed the online processing flow, did the machine learning analysis and evaluated the ISI changes. He further wrote parts of the manuscript. HW conducted the experiments, designed the online processing flow and performed behavioral analysis with respect to reaction time and late reaction time, selected the questionnaires and wrote parts of the manuscript. MM conducted the experiments, adjusted the MMI to match the experiments needs and wrote parts of the manuscript. FK contributed to the concept of the MMI, critically discussed the research goals, and revised and improved the manuscript. All authors gave their final approval of the version to be published and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

FUNDING

This work was funded by research grants from the German Federal Ministry for Economic Affairs and Energy (grant FKZ 50 RA 1011, FKZ 50 RA 1012 and grant FKZ 50 RA 1301).

ACKNOWLEDGMENTS

We want to thank Elke Neubauer for her help in data acquisition, Alexander Dieterle and Johannes Teiwes for their help during the development of the MMI, and Nils Eckardt for helping us in digitizing behavioral data as preparation of analysis.

REFERENCES

Allanson, J., and Fairclough, S. (2004). A research agenda for physiological computing. *Interact. Comput.* 16, 857–878. doi: 10.1016/j.intcom.2004.08.001

Chang, C.-C., and Lin, C.-J. (2011). LIBSVM: a library for support vector machines. *ACM Trans. Intell. Syst. Technol.* 2, 27:1–27:27. doi: 10.1145/1961189.1961199

Cornellă, J., Zerbato, D., Giona, L., Fiorini, P., and Sequeira, V. (2012). "Dynamics simulation for the training of teleoperated retrieval of spent nuclear fuel," in *ICRA* (St. Paul, MN), 5012–5017.

Devlaminck, D., Wyns, B., Grosse-Wentrup, M., Otte, G., and Santens, P. (2011). Multisubject learning for common spatial patterns in motor-imagery BCI. *Comput. Intell. Neurosci.* 2011, 1–9. doi: 10.1155/2011/217987

DFKI - RIC (2015). *Mars - a Cross-Platform Simulation and Visualization Tool*. Available online at: http://rock-simulation.github.io/mars/, last visited on 04.11.2015.

Duncan-Johnson, C. C., and Donchin, E. (1977). On quantifying surprise: the variation of event-related potentials with subjective probability. *Psychophysiology* 14, 456–467.

Folgheraiter, M., Jordan, M., Straube, S., Seeland, A., Kim, S. K., and Kirchner, E. A. (2012). Measuring the improvement of the interaction comfort of a wearable exoskeleton. *Intern. J. Soc. Robot.* 4, 285–302. doi: 10.1007/s12369-012-0147-x

Ghaderi, F., Kim, S. K., and Kirchner, E. A. (2014). Effects of eye artifact removal methods on single trial P300 detection, a comparative study. *J. Neurosci. Methods* 221, 41–47. doi: 10.1016/j.jneumeth.2013.08.025

Ghaderi, F., and Straube, S. (2013). "An adaptive and efficient spatial filter for event-related potentials," in *Signal Processing Conference (EUSIPCO), 2013 Proceedings of the 21st European* (Marrakech), 1–5.

Hart, S. G., and Staveland, L. E. (1988). Development of nasa-tlx (task load index): Results of empirical and theoretical research. *Adv. Psychol.* 52, 139–183.

Isreal, J., Chesney, G., Wickens, C., and Donchin, E. (1980). P300 and tracking difficulty: evidence for multiple resources in dual-task performance. *Psychophysiology* 17, 259–273.

Jansen, B., Allam, A., Kota, P., Lachance, K., Osho, A., and Sundaresan, K. (2004). An exploratory study of factors affecting single trial P300 detection. *IEEE Trans. Biomed. Eng.* 51, 975–978. doi: 10.1109/TBME.2004.826684

Kahneman, D. (1973). *Attention and Effort*. Englewood Cliffs, NJ: Prentice Hall.

Kim, S. K., and Kirchner, E. A. (2012). "Preliminary results on p300 detection using machine learning when modulating task reaction time," in *Proceedings of the 18th Annual Meeting of the Organization for Human Brain Mapping, OHBM-2012* (Beijing).

Kindermans, P.-J., Verschore, H., Verstraeten, D., and Schrauwen, B. (2012). "A P300 BCI for the masses: prior information enables instant unsupervised spelling," in *Advances in Neural Information Processing Systems 25 (NIPS-2012)*, eds P. Bartlett, F. Pereira, C. Burges, L. Bottou, and K. Weinberger (Lake Tahoe), 719–727.

Kirchner, E. A. (2014). *Embedded Brain Reading*. PhD thesis, University of Bremen, Bremen.

Kirchner, E. A. (2015). "Intrinsische intentionserkennung in technischen systemen," in *GI-Edition: Lecture Notes in Informatics, Ausgezeichnete Informatikdissertationen 2014*, ed S. Hölldobler (Bonn: Bonner Köllen Verlag), 121–130.

Kirchner, E. A., Albiez, J., Seeland, A., Jordan, M., and Kirchner, F. (2013a). "Towards assistive robotics for home rehabilitation," in *Proceedings of the 6th International Conference on Biomedical Electronics and Devices (BIODEVICES-13)*, eds M. F. Chimeno, J. Solé-Casals, A. Fred, and H. Gamboa (Barcelona: SciTePress), 168–177.

Kirchner, E. A., de Gea Fernandéz, J., Kampmann, P., Schröer, M., Metzen, J. H., and Kirchner, F. (2015). "Intuitive interaction with robots - technical approaches and challenges," in *Formal Modeling and Verification of Cyber Physical Systems*, ed R. D. U. Kühne (Heidelberg: Springer Verlag GmbH), 224–248.

Kirchner, E. A., and Drechsler, R. (2013). A formal model for embedded brain reading. *Indust. Robot* 40, 530–540. doi: 10.1108/IR-01-2013-318

Kirchner, E. A., Kim, S. K., Straube, S., Seeland, A., Wöhrle, H., Krell, M. M., et al. (2013b). On the applicability of brain reading for predictive human-machine interfaces in robotics. *PLoS ONE* 8:e81732. doi: 10.1371/journal.pone.0081732

Kirchner, E. A., Tabie, M., and Seeland, A. (2014). Multimodal movement prediction - towards an individual assistance of patients. *PLoS ONE* 9:e85060. doi: 10.1371/journal.pone.0085060

Kirchner, E. A., Wöhrle, H., Bergatt, C., Kim, S.-K., Metzen, J. H., Feess, D., et al. (2010). "Towards operator monitoring via brain reading – an EEG-based approach for space applications," in *Proceedings of the 10th International Symposium on Artificial Intelligence, Robotics and Automation in Space* (Sapporo), 448–455.

Kok, A. (2001). On the utility of P3 amplitude as a measure of processing capacity. *Psychophysiology* 38, 557–577. doi: 10.1017/S0048577201990559

Kramer, A. F., Trejo, L. J., and Humphrey, D. (1995). Assessment of mental workload with task-irrelevant auditory probes. *Biol. Psychol.* 40, 83 – 100.

Krauledat, M., Tangermann, M., Blankertz, B., and Müller, K.-R. (2008). Towards zero training for brain-computer interfacing. *PLoS ONE* 3:e2967. doi: 10.1371/journal.pone.0002967

Krell, M. M., Straube, S., Seeland, A., Wöhrle, H., Teiwes, J., Metzen, J. H., et al. (2013a). pySPACE - a signal processing and classification environment in Python. *Front. Neuroinformat.* 7:40. doi: 10.3389/fninf.2013.00040

Krell, M. M., Tabie, M., Wöhrle, H., and Kirchner, E. A. (2013b). "Memory and processing efficient formula for moving variance calculation in EEG and EMG signal processing," in *Proc. International Congress on Neurotechnology, Electronics and Informatics (NEUROTECHNIX 2013)* (Vilamoura: SciTePress), 41–45.

Kyndt, E., Dochy, F., Struyven, K., and Cascallar, E. (2011). The direct and indirect effect of motivation for learning on students approaches to learning through the perceptions of workload and task complexity. *Higher Educ. Res. Dev.* 30, 135–150. doi: 10.1080/07294360.2010.501329

Li, Y., Guan, C., Li, H., and Chin, Z. (2008). A self-training semi-supervised SVM algorithm and its application in an EEG-based brain computer interface speller system. *Pattern Recogn. Lett.* 29, 1285–1294. doi: 10.1016/j.patrec.2008.01.030

Lotte, F., and Guan, C. (2010). "Learning from other subjects helps reducing brain-computer interface calibration time," in *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on* (Dallas, TX), 614–617.

Lu, S., Guan, C., and Zhang, H. (2009). Unsupervised brain computer interface based on intersubject information and online adaptation. *Neural Syst. Rehabilit. Eng. IEEE Trans.* 17, 135–145. doi: 10.1109/TNSRE.2009.2015197

Mertens, R., and Polich, J. (1997). P300 from a single-stimulus paradigm: passive versus active tasks and stimulus modality. *Electroencephalogr. Clin. Neurophysiol.* 104, 488–497.

Metzen, J. H., and Kirchner, E. A. (2011). "Rapid adaptation of brain reading interfaces based on threshold adjustment," in *Proceedings of the 2011 Conference of the German Classification Society (GfKl-2011)*, 138.

Polich, J. (1990). P300, probability, and interstimulus interval. *Pyschophysiology* 27, 396–403.

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clin. Neurophysiol.* 118, 2128–2148. doi: 10.1016/j.clinph.2007.04.019

Polich, J., and Margala, C. (1997). P300 and probability: comparison of oddball and single-stimulus paradigms. *Intern. J. Psychophysiol.* 25, 169–176.

Pope, A. T., Bogart, E. H., and Bartolome, D. S. (1995). Biocybernetic system evaluates indices of operator engagement in automated task. *Biol. Psychol.* 40, 187–195.

Prinzel, L. J., Freeman, F. G., Scerbo, M. W., Mikulka, P. J., and Pope, A. T. (2003). Effects of a psychophysiological system for adaptive automation on performance, workload, and the event-related potential p300 component. *Human Fact.* 45, 601–614. doi: 10.1518/hfes.45.4.601.27092

Rivet, B., Cecotti, H., Perrin, M., Maby, E., and Mattout, J. (2011). Adaptive training session for a P300 speller brain–computer interface. *J. Physiol. Paris* 105, 123–129. doi: 10.1016/j.jphysparis.2011.07.013

Rivet, B., Souloumiac, A., Attina, V., and Gibert, G. (2009). xDAWN algorithm to enhance evoked potentials: application to brain-computer interface. *IEEE Trans. Biomed. Eng.* 56, 2035–2043. doi: 10.1109/TBME.2009.2012869

Rommerman, M., Kuhn, D., and Kirchner, F. (2009). "Robot design for space missions using evolutionary computation," in *Evolutionary Computation, 2009. CEC '09. IEEE Congress on* (Trondheim), 2098–2105.

Samek, W., Meinecke, F., and Müller, K.-R. (2013). Transferring subspaces between subjects in brain-computer interfacing. *Biomed. Eng. IEEE Trans.* 60, 2289–2298. doi: 10.1109/TBME.2013.2253608

Schroeders, U., and Wilhelm, O. (2011). Computer usage questionnaire: Structure, correlates, and gender differences. *Comput. Hum. Behav.* 27, 899–904. doi: 10.1016/j.chb.2010.11.015

Straube, S., and Krell, M. M. (2014). How to evaluate an agent's behaviour to infrequent events? – reliable performance estimation insensitive to class distribution. *Front. Comput. Neurosci.* 8:43. doi: 10.3389/fncom.2014.00043

Tabie, M., Wöhrle, H., and Kirchner, E. A. (2014). "Runtime calibration of online eeg based movement prediction using emg signals," in *Proceedings of the 7th International Conference on Bio-inspired Systems and Signal Processing (BIOSIGNALS-14)* (Angers: SciTePress), 284–288.

Tueting, P., Sutton, S., and Zubin, J. (1970). Quantitative evoked potential correlates of the probability of events. *Psychophysiology* 7, 385–394.

Wickens, C. D. (1984). "Processing resources in attention," in *Varieties of Attention*, eds R. Parasuraman and D. Davies (London; Oxford; Boston, MA; New York, NY; San Diego, CA: Academic Press), 63–101.

Wickens, C. D. (1992). *Engineering Psychology and Human Performance (2nd. Edn.)* New York, NY: HarperCollins.

Wickens, C. D. (2008). Multiple resources and mental workload. *Hum. Fact.* 50, 449–455. doi: 10.1518/001872008X288394

Wöhrle, H., and Kirchner, E. A. (2014a). "Online classifier adaptation for the detection of p300 target recognition processes in a complex teleoperation scenario," in *Physiological Computing Systems*, volume 8908 of *Lecture Notes in Computer Science*, H. P. da Silva, A. Holzinger, S. Fairclough, and D. Majoe (Berlin; Heidelberg: Springer), 105–118.

Wöhrle, H., and Kirchner, E. A. (2014b). "Online detection of P300 related target recognition processes during a demanding teleoperation task. in *Proceedings of the International Conference on Physiological Computing Systems, (PhyCS 2014)*, (Lissabon: SCITEPRESS Digital Library), 13–19.

Wöhrle, H., Krell, M. M., Straube, S., Kim, S. K., Kirchner, E. A., and Kirchner, F. (2015). An adaptive spatial filter for user-independent single trial detection of event-related potentials. *IEEE Trans. Biomed. Eng.* 62, 1696–1705. doi: 10.1109/TBME.2015.2402252

Wöhrle, H., Teiwes, J., Kirchner, E. A., and Kirchner, F. (2013a). "A framework for high performance embedded signal processing and classification of psychophysiological data," in *APCBEE Procedia. International Conference on Biomedical Engineering and Technology (ICBET-2013), 4th, May 19-20* (Kopenhagen: Elsevier), 6066.

Wöhrle, H., Teiwes, J., Krell, M. M., Kirchner, E. A., and Kirchner, F. (2013b). "A dataflow-based mobile brain reading system on chip with supervised online calibration," in *Proc. International Congress on Neurotechnology, Electronics and Informatics (NEUROTECHNIX 2013)* (Vilamoura: ScitePress), 46–53.

Wöhrle, H., Teiwes, J., Krell, M. M., Seeland, A., Kirchner, E. A., and Kirchner, F. (2014a). "Reconfigurable dataflow hardware accelerators for machine learning and robotics," in *ECML/PKDD-2014 PhD Session Proceedings* (Nancy), 129–138.

Wöhrle, H., Teiwes, J., Tabie, M., Seeland, A., Kirchner, E. A., and Kirchner, F. (2014b). "Prediction of movements by online analysis of electroencephalogram with dataflow accelerators," in *Proc. International Congress on Neurotechnology, Electronics and Informatics (NEUROTECHNIX 2014)* (Rome: ScitePress), 31–37.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer PK and handling Editor declared their shared affiliation, and the handling Editor states that the process nevertheless met the standards of a fair and objective review.

Copyright © 2016 Kirchner, Kim, Tabie, Wöhrle, Maurus and Kirchner. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 23 November 2015
doi: 10.3389/fnhum.2015.00636

Perception and Cognition of Cues Used in Synchronous Brain–Computer Interfaces Modify Electroencephalographic Patterns of Control Tasks

Luz María Alonso-Valerdi[1,2], Francisco Sepulveda[1] and Ricardo A. Ramírez-Mendoza[2]*

[1] Brain-Computer Interfaces (BCI) Group, School of Computing Science and Electronic Engineering, University of Essex, Colchester, UK, [2] Escuela de Ingeniería y Ciencias, Tecnológico de Monterrey – Campus Ciudad de México, Mexico City, Mexico

A motor imagery (MI)-based brain–computer interface (BCI) is a system that enables humans to interact with their environment by translating their brain signals into control commands for a target device. In particular, synchronous BCI systems make use of cues to trigger the motor activity of interest. So far, it has been shown that electroencephalographic (EEG) patterns before and after cue onset can reveal the user cognitive state and enhance the discrimination of MI-related control tasks. However, there has been no detailed investigation of the nature of those EEG patterns. We, therefore, propose to study the cue effects on MI-related control tasks by selecting EEG patterns that best discriminate such control tasks, and analyzing where those patterns are coming from. The study was carried out using two methods: standard and all-embracing. The standard method was based on sources (recording sites, frequency bands, and time windows), where the modulation of EEG signals due to motor activity is typically detected. The all-embracing method included a wider variety of sources, where not only motor activity is reflected. The findings of this study showed that the classification accuracy (CA) of MI-related control tasks did not depend on the type of cue in use. However, EEG patterns that best differentiated those control tasks emerged from sources well defined by the perception and cognition of the cue in use. An implication of this study is the possibility of obtaining different control commands that could be detected with the same accuracy. Since different cues trigger control tasks that yield similar CAs, and those control tasks produce EEG patterns differentiated by the cue nature, this leads to accelerate the brain–computer communication by having a wider variety of detectable control commands. This is an important issue for Neuroergonomics research because neural activity could not only be used to monitor the human mental state as is typically done, but this activity might be also employed to control the system of interest.

Keywords: brain–computer interface, motor imagery, classification accuracy, electroencephalographic patterns, human factors

OPEN ACCESS

Edited by:
Klaus Gramann, Berlin Institute of Technology, Germany

Reviewed by:
Reinhold Scherer, Graz University of Technology, Austria
Josef Faller, Graz University of Technology, Austria

****Correspondence:***
Luz María Alonso-Valerdi
luzalonsoval@gmail.com

Received: *17 August 2015*
Accepted: *06 November 2015*
Published: *23 November 2015*

Citation:
Alonso-Valerdi LM, Sepulveda F and Ramírez-Mendoza RA (2015) Perception and Cognition of Cues Used in Synchronous Brain–Computer Interfaces Modify Electroencephalographic Patterns of Control Tasks. Front. Hum. Neurosci. 9:636. doi: 10.3389/fnhum.2015.00636

INTRODUCTION

A brain–computer interface (BCI) is a system that enables humans to interact with their environment by translating their brain signals into control commands for a device of interest (Graimann et al., 2010). The mechanism of a BCI system fundamentally consists of two steps: (1) detecting and decoding the user intentions for controlling the system and (2) maintaining a continuous user-system communication. The *user intentions* of controlling a BCI system are changes in the user brain signals that are regulated through control tasks. A control task can be based on exogenous or endogenous paradigms (Jackson and Mappus, 2010). Particularly, the endogenous paradigm is based on the quantification of brain oscillations that are modulated via cognitive tasks such as motor imagery (MI). The *user-system communication* is established by a control interface. A control interface can be synchronous or asynchronous (Hassanien and Azar, 2015). In a synchronous interface, the user-system communication is allowed only in fixed time windows; whereas in an asynchronous interface, the user initiates the communication with the system at will. Both synchronous (Obermaier et al., 2003; Leeb et al., 2006; Maeder et al., 2012; Bamdadian et al., 2014) and asynchronous (Scherer et al., 2007; Galán et al., 2008; Lotte et al., 2008; Tseng et al., 2015) systems have been developed over the past few years. For real-world applications, the prototyping of asynchronous systems is preferred because these allow users to interact naturally with their environment. The relevance of synchronous systems cannot, however, be ignored, even in real applications. The cueing process facilitates the early and accurate detection of the user control tasks, despite the user ability for modulating his/her brain signals. This, in turn, raises confidence, persistency, and autonomy in the users toward the mastery of BCI skills. Furthermore, the identification of MI onset allows to analyze prior and post periods, which have been associated with the improvement of BCI performance and the recognition of the user cognitive state (Maeder et al., 2012; Bamdadian et al., 2014; Gutierrez et al., 2015).

As brain signals are modulated by neural networks that modify their degree of synchronization according to the sensory–cognitive input, it is not surprising that control tasks (particularly those based on MI) contain much more information than only that related to the user intention of controlling the system (Kropotov, 2010). MI-related control tasks are a source of information that has been exploited not only to generate control commands for a target device, but also to enhance BCI performance, to predict classification accuracy (CA), or to determine the user mental state. For example, Pfurtscheller and Neuper (2001), and then Obermaier et al. (2003), reported that left- and right-hand MIs were correctly discriminated as early as 250 ms after the onset of a specific visual cue. They attributed the early discrimination to the cue properties, concluding that the control tasks were the result of conscious (MI) and unconscious (visual stimulation) processes over the sensory–motor area of the brain. Furthermore, in a later and more detailed study, Pfurtscheller et al. (2008) found that distinct short-lasting brain patterns appeared within a time window of about 500–750 ms after cue onset. Those brain patterns produced different features for different imaginary movements (hands and feet), facilitating and accelerating the discrimination of MI-related control tasks in naïve subjects. Another example is the study carried out by Grosse-Wentrup and Scholkopf (2012) in which high gamma range (55–85 Hz) between two fronto-parietal networks were used to predict BCI performance on a trial-to-trial basis. Additional and important examples are two studies, respectively, undertaken by Maeder et al. (2012) and Bamdadian et al. (2014). Those researchers demonstrated that the user performance in classical synchronous BCIs can be predicted by quantifying the modulation of the brain signals on pre-cue stages because they reflected somehow the user cognitive state. Finally, and more recently, Scheel et al. (2015) found that visual and auditory cues provoked significant differences of the peak amplitude of movement-related cortical potentials in synchronous BCIs. They also found that potentials from the auditory-cue paradigm had a wider spatial distribution than those from the visual cue.

Overall, all aforementioned studies support the view that brain patterns extracted from MI-related control tasks can provide much more information than that used to control a target device. In particular, the cue effects on MI-related control tasks have been studied. Researchers in the field have shown that both perception (e.g., sensory–cognitive processing of the cue) and cognition (e.g., imaginary motor activity) are reflected on the brain signals wherefrom BCI control tasks are extracted. Studying the influence of human factors on BCI control tasks may help to design a more versatile human–machine interaction for this type of systems because of the active (e.g., extraction of control commands for manipulating a target device) and passive (e.g., monitoring of the level of attention of an individual) use of the brain signals. This work could have further applications in Neuroergonomics, where neural activity is registered in order to monitor human mental state. Making use of neural activity in an active and passive way may be much more fruitful.

So far, it has been shown that brain patterns before and after cue onset can reveal the user cognitive state and enhance the discrimination of MI-related control tasks. However, there has been no detailed investigation of the nature of those brain patterns. We, therefore, propose to study the cue effects on MI-related control tasks by selecting the brain patterns that best discriminate such control tasks, and analyzing where those patterns are coming from in order to answer two questions:

(1) If different cues provoke significant changes on MI-related control tasks, can different cues improve BCI performance as Scheel et al. (2015) suggested in their research? and
(2) If MI-related control tasks are defined by motor activity *per se* and the cue in use, do brain patterns proceed from sources (recording sites, frequency bands, and time intervals) not only associated with motor activity, but also related to the sensory–cognitive processing of the cue?

The present study was conducted as follows. First, brain activity was registered by means of electroencephalography (EEG). Second, the frequently used stimulation modalities (SMs) for cueing in training sessions were applied. These were auditory (Nijboer et al., 2008) and visual (Boostani et al., 2007) stimuli.

In addition, a bimodal cue (combination of auditory and visual stimuli) was included in the study because previous investigations in sensory encoding (Basar et al., 1999; Isoğlu-Alkaç et al., 2007) have shown that simultaneous presentation of auditory, visual, and somatosensory stimuli significantly enhances sensory responses. Third, as preparation and imagination of movements evoke similar neural desynchronization events over the sensory–motor areas (Neuper et al., 2006) and both of them are widely used as control task, the two motor activities were included in the study. Finally, given that brain oscillations occur in a wide range of EEG recording sites, frequency bands, and time intervals (Kropotov, 2010); brain patterns were analyzed using two methods: standard and all-embracing. The standard method was restricted to the well-established motor activity sources (Pfurtscheller et al., 2007), while the all-embracing method involved all the available EEG information.

MATERIALS AND METHODS

Experimental Procedure

Participant Recruitment and General Instructions

Nine participants (four females and five males) took part in this study, which was previously authorized by the Ethics Committee of the University of Essex. All of them were aged between 28 and 41 years. None of them reported auditory impairments, seven of them had normal vision, and two of them had corrected-to-normal vision. Eight of the nine reported to be right-handed and only one was left-handed.

The participants were informed about the experimental procedure and signed a consent form. Only two of the nine had previously engaged in cognitive tasks related to imagination of movements. At the beginning of the experiment, every participant was carefully instructed as follows:

- Get ready to imagine the movement of the hand that indicates a track playing "left" or "right" (audio), an arrow pointing to left or right (visual), or both of them (bimodal).
- Imagine yourself opening and closing your formerly pointed hand as soon as you listen to an increasing tone, see a green bulb, or perceive both of them.
- Stop the MI process and relax as soon as you listen to a decreasing tone, see a red bulb, or perceive both of them.

Organization of the Experiment

In order to collect sufficient EEG data, the participants attended two sessions. The sessions lasted 48 min each and followed an identical procedure. Every session consisted of six runs and one run had 50 trials. One trial took from 8500 to 9500 ms (**Figure 1**), resulting in runs of ~8 min. Within each trial, there were three phases: MP (0–2500 ms), MI (2500–6000 ms), and relaxing (6000–8500 ± 1000 ms). In the latter phase, a random variation of 1000 ms was included to reduce expectation effects.

As there were three SMs (audio, visual, and bimodal) and both hands (left and right) were involved, there were six categories of trials: audio-left, audio-right, visual-left, visual-right, bimodal-left, and bimodal-right. Each of these categories was randomly presented 50 times and distributed over the six runs. We thereby obtained 12 conditions (six categories of trials × two control tasks) and one condition had 100 trials (2 sessions × 50 trials).

Timing Protocol

The duration of the cues was standardized to 500 ms in accordance with sensory recognition and reaction time studies (Teichner, 1954; Shelton and Kumar, 2010). The movement preparation (MP) was adjusted to 2000 ms, which is the necessary period to achieve readiness in the neural networks over the sensory–motor area (Jeannerod, 2006; Neuper et al., 2006). The MI was limited to 3000 ms, as is commonly done in synchronous BCIs. The relaxation span varied from 2000 to 3000 ms, guaranteeing a proper recovery of the longest desynchronization process, i.e., the alpha one (Pfurtscheller et al., 1996). See **Figure 1**.

EEG Data Collection

The EEG signals were recorded by means of Biosemi equipment (Amsterdam, The Netherlands), the integration of ActiveTwo system and ActiView software (Honsbeek et al., 1998). The ActiveTwo system was configured to acquire the signals within a bandwidth between DC and 400 Hz, and at a sampling frequency of 2048 Hz. The ActiView software was programed to decimate the signals at 512 Hz. Such configuration limited the effective digital bandwidth to 104 Hz by default.

The EEG signals were sensed via 61 active electrodes, plus driven-right-leg and common-mode-sense electrodes. The 61 active electrodes were mounted on a head-cap labeled as stated in the 10/10 system. The other two electrodes were only used for referencing electrically the ActiveTwo system, but they were not recorded. In addition, three external electrodes were included for recording the eye movements (EOG). Two of them (EOG_L and EOG_R) were placed 1 cm below and above the lateral canthus of the left and right eyes, respectively. The third one was placed on the right mastoid (M_R) for referencing EEG and EOG signals (**Figure 2**). At the end of the experiments, we gathered 18 datasets (9 participants × 2 sessions).

Data Analysis

The datasets of one participant were excluded from the study. Those showed electrode-pop artifacts over the occipital area of the scalp. There were then 16 datasets for the study purposes.

Processing of Continuous EEG Data

To attenuate the interference in the EEG channels, these were processed by using the open-access toolbox for electrophysiological signal processing, EEGLAB (Delorme and Makeig, 2004). First, every channel from each of the 16 datasets was processed as follows: (1) referencing against M_R, (2) high-pass filtering at 0.1 Hz using a Butterworth filter of order 4, (3) low-pass filtering at 41 Hz using a Butterworth filter of order 7, and (4) down-sampling from 512 to 256 Hz. Second, every dataset was scanned to eliminate discontinuities and detect high-impedance electrodes. Up to three electrodes under this condition were detected per dataset. Third, independent component analysis was applied to each dataset for rejecting artifacts such as EOG and electrocardiography. Only EEG channels without high-impedance difficulties were involved

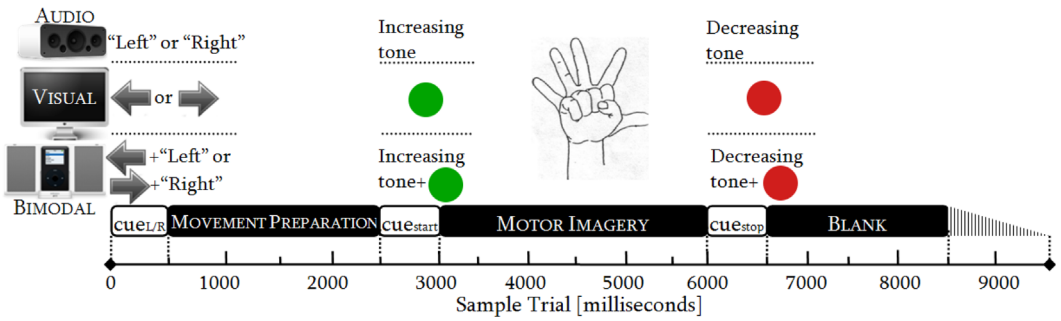


FIGURE 1 | Timing protocol: audio, visual, and bimodal stimuli for cueing MP and MI.

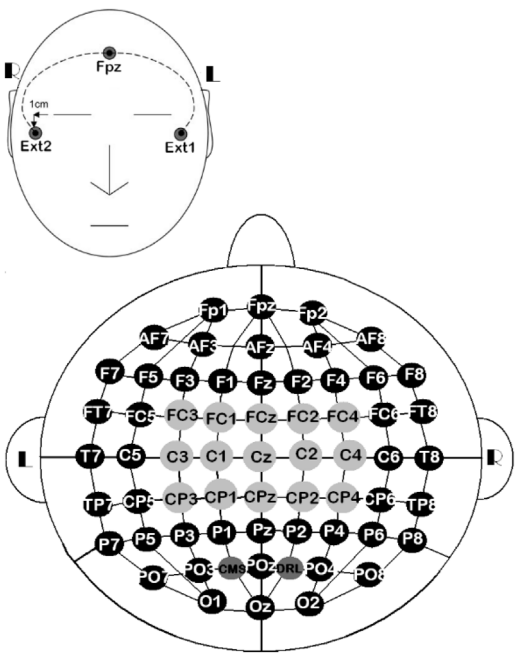


FIGURE 2 | 10/10 EEG layout of 61 channels, along with three external electrodes. All the EEG channels (■, ▒) were used for the all-embracing method, 15 of them (▒) were used for the standard method, EOG_L/EOG_R were used for recording eye movements, and M_R was used for referencing EEG and EOG signals.

in such analysis. EOG_L and EOG_R channels were used to identify all the independent components related to EOG activity. Finally, the EEG channels with high-impedance difficulties were replaced by interpolating their nearest neighboring channels as reported by Gargiulo et al. (2010). See **Figure 3**.

Processing of the Control Tasks (MP and MI)

The processing of the control tasks was carried out through the *miBCI* software[1], package published by Alonso-Valerdi and Sepulveda (2015). From every EEG channel of the 16 datasets, the control tasks were extracted in line with the cue onset (**Figure 1**). The MP and MI were thus 2500 and 3500 ms long, respectively. Having obtained the EEG signals of interest, they were spatially filtered via large Laplacian in order to obtain more localized electrical activity (Dornhege et al., 2007).

Feature Extraction

It is well-established that MP and MI provoke neural desynchronization with peak power around 10 and 20 Hz (Neuper et al., 2009). As band power (BP) estimation has been validated as a stable and consistent method for quantifying EEG power changes due to motor activity (Neuper et al., 2005), this was selected as feature extractor. BP estimation was applied in line with the methods described below.

Standard Method

Previous investigations have empirically established the following criteria to effectively discriminate hand imaginary movements. First, 18 central recording sites have been validated as the maximum number of EEG channels for satisfactory classification (Ramoser et al., 2000). Second, narrow frequency bands around the maxima 10 and 20 Hz have been widely used in synchronous BCI systems (Pfurtscheller et al., 2007; Neuper et al., 2009). Third, it has become common practice to discard 1 s post-cue, wherein evoked potentials are typically detected (Boostani et al., 2007). With these criteria in mind, we laid down the *standard method*. This method was based on 15 central recording sites (**Figure 2**), four frequency bands, and EEG segments taking place 1 s post-cue. The frequency bands were established as follows: lower alpha (α_L) from 8 to 10 Hz, upper alpha (α_U) from 10 to 12 Hz, lower beta (β_L) from 16 to 20 Hz, and upper beta (β_U) from 20 to 24 Hz.

All-Embracing Method

EEG signals are regulated by brain oscillators that adjust their state of synchrony according to sensory (e.g., cue decoding) and cognitive (e.g., MP and MI) events. These oscillators are neural networks that enter into synchrony in a wide range of resonant frequencies (from 0 up to about 80 Hz) and over specific periods of time (Krause, 2003). In view of this fact, we extended the scope of the standard method by establishing the

[1]Available at https://github.com/LuzAlondra/BrainComputerInterfaces/tree/master/MI-based_BCIsystem

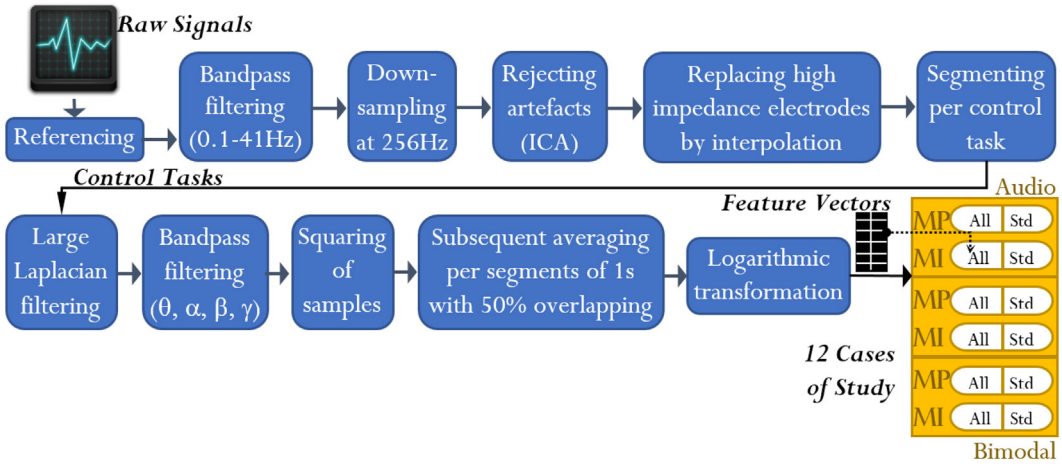


FIGURE 3 | EEG signal analysis: from raw signals to feature extraction. At the end of the analysis, feature vectors for 12 cases of study resulted. Note that the term "All" refers to all-embracing method, and the term "Std" refers to standard method.

all-embracing method. This method was based on 61 recording sites (**Figure 2**), seven frequency bands, and whole trace of MP and MI. In addition to the previously mentioned bands, the following ones were also considered: lower theta (θ_L) from 4 to 6 Hz, upper theta (θ_U) from 6 to 8 Hz, and gamma (γ) from 39 to 41 Hz. These bands were included in the analysis on the basis of the following evidence. Theta band rhythms resonate at the frequency band 4–8 Hz and emanate from the frontal midline due to audio–visual information encoding, attention demands, memory retrieval, and cognitive load. Moreover, these rhythms enhance after practice on the cognitive tasks at hand. They are more prevalent when the subject is focused and relaxed, and prolonged activity is related to selective attention (Basar et al., 1999; Krause, 2003; Kropotov, 2010). The upper theta band (6–8 Hz) generally reflects levels of alertness (Pineda, 2005). On the other side, gamma band rhythms oscillate near 40 Hz during sensory encoding, perceptual–cognitive functions, and motor behaviors. These rhythms are phase-locked to the stimulus and short-lasting, and appear 100 ms post-stimulus in sensory–motor tasks (Pfurtscheller and Lopes da Silva, 1999; Ward, 2003; Altermaller et al., 2005).

Bearing in mind the criteria of standard and all-embracing methods, we can now briefly describe the feature extraction based on BP. The MP/MI signals were first filtered through Butterworth band-pass filters of order 7, with cut-off frequencies defined by the afore-stated bands. Afterwards, the signals were squared per sample and segmented by using time windows of 500 ms length with 50% overlapping rate. Finally, the resulting time segments (herein denoted by δ_n) were averaged and logarithmically transformed (refer to **Figure 3**), obtaining nine features per MP signal and 13 features per MI signal.

By the standard method, there were 15 channels and 4 frequency bands under consideration. In addition, three time segments [δ_1 (0–500 ms), δ_2 (250–750 ms), and δ_3 (500–1000 ms)] were discarded. Hence, vectors of 300 features for MP and vectors of 540 features for MI were obtained. By the all-embracing method, vectors of 3843 features for MP and vectors of 5551 features for MI were similarly obtained.

Feature Selection and Classification

After the feature extraction, there were 24 types of feature vectors that proceeded from three SMs, two control tasks, two hands, and two methods. These feature vectors were grouped by merging left and right MIs. Having obtained 12 different cases of study (**Figure 3**), Davies–Bouldin indexes (DBIs) were determined in each case to increasingly sort the corresponding features (Sepulveda et al., 2004; Kovács et al., 2005). DBI is a method for measuring the linear separability among m classes (Equation 1). This metric is based on comparing the similarity (R) among classes. Such similarity is determined by the class dispersion (s) and the distance (d) between centroids (Eq. 2). The class dispersion is the average distance between every element (τ) in the class and the centroid of the class (v). See Eq. 3. Thereby, the features within each vector were ranked from the most to the least suitable feature in terms of linear separability between two classes: left and right (Kovács et al., 2005). Note that smaller DBIs correspond to major linear separability.

$$\mathrm{DBI} = \frac{1}{m}\sum_{i=1}^{m} R_i \tag{1}$$

where $R_i = \max_{\substack{j=1,\ldots,m \\ i \neq j}} (R_{ij})$ and $i = 1,\ldots,m$

$$R_{ij} = \frac{s_i + s_j}{d_{ij}} = \frac{s_i + s_j}{d(v_i, v_j)} \tag{2}$$

$$s_i = \frac{1}{T_i}\sum d(\tau, v_i) \tag{3}$$

where T_i is the number of features in class i.

After ranking the features, a classification process took place in order to select the appropriate number of features that best

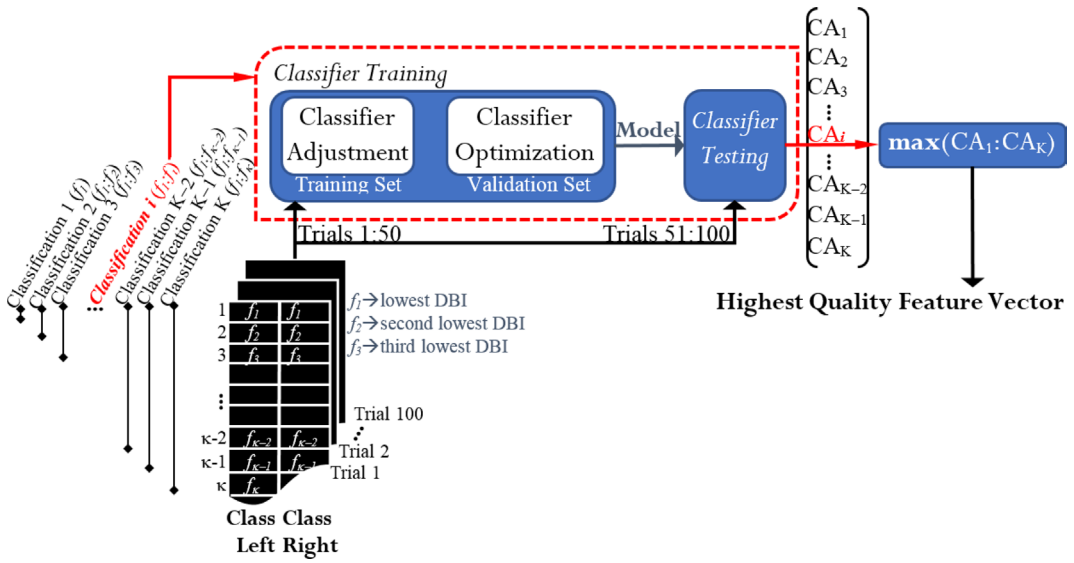


FIGURE 4 | Feature selection and classification: obtaining the highest quality feature vectors (HQFVs). For each case of study, a HQFV was obtained. A HQFV was the feature vector that reached the maximum CA and wherein each feature (*f*) was ranked according to DBI.

discriminated between left and right. If there were two classes and κ denoted the total number of features in each vector, K classifications were run for each case of study (**Figure 4**). From the K resulting CAs, the feature vector yielding the maximum performance was selected from each case of study. Thereby, we obtained 12 feature vectors for every participant. They were called the highest quality feature vectors (HQFVs). Note that the term "maximum performance" refers to 1.5 times the inter-quartile range plus the upper quartile of the general distribution of all the CAs obtained at the end of the process. As a result, any peak value that was beyond the 99% of the distribution was discarded.

Every classification process[2] was based on Fisher discriminant analysis (FDA) and consisted of two phases: training and testing. The 100 available trials per cluster were distributed half and half; that is, 50 trials (session 1) for training and 50 trials (session 2) for testing. The classifier was trained via 10-fold cross validation. That is, 50 training trials were split into a training set and a validation set. Through the validation set, the model was optimized by adjusting the regularization term that generally avoids overfitting problems due to the large number of features in use (Bishop, 2006). Once the classifier had been trained, this was tested by the rest of the trials and the percentage of the total number of correct predictions was estimated. The resulting CA in the testing phase was the parameter for acquiring the HQFVs. See **Figure 4**.

[2]Before undertaking the classification process, note that the features were normalized using normalization and standardization methods developed by *mlpy* (high-performance Python package for predictive modelling avaliable at http://mlpy.sourceforge.net/) developers, to avoid BP estimates in greater numeric ranges dominate those in smaller numeric ranges.

In this study, all statistical analyses were performed using the non-parametric method Kruskal–Wallis one-way ANOVA, and significance levels were set at 5%.

Statistical Evaluation of the HQFVs

The features of the HQFVs proceeded from specific recording sites (e.g., C3, Cz, or C4), frequency bands (e.g., α_L, α_U, β_L, or β_U) and time windows (δ_n). The origin of a feature in any of these three domains (location, frequency, and time) was referred as to *feature source*. On this basis, the HQFVs were statistically evaluated in accordance with those three domains and under two parameters: index of dispersion (ID) and *mode*.

The ID was calculated by using Eq. 4 and was an approach to quantify how spread a HQFV was over the feature sources in each domain. In Eq. 4, k is the number of feature sources in the domain of interest, f_i is the number of occurrences of each feature source, and N is the total number of features in the HQFV under analysis (Norman and Streiner, 2008). Note that ID is 0 when all the features fall into one feature source. By contrast, it is 1 when the features are equally divided among the k feature sources.

$$\mathrm{ID} = \frac{k\left(N^2 - \sum f_i^2\right)}{N^2\left(k-1\right)} \tag{4}$$

The *mode* was the central tendency of a HQFV, i.e., the most frequently occurring feature source in the domain at hand. Having gathered the *modes* of all the HQFVs, these were graphically represented via a 2D-histogram (*modal* distribution) for each domain. In every 2D-histogram, the number of occurrences of each *mode* (f_{mode}) was normalized by diving f_{mode} by N.

RESULTS

Classification Accuracy of the HQFVs

The CAs reached by the HQFVs are arranged in **Figure 5**. This figure indicates that there is no significant difference of CAs among SMs ($p = 0.935$). The figure also indicates that there is a significant increase of CAs ($p = 1.11 \times 10^{-16}$) between standard and all-embracing methods for both control tasks (MP and MI) and the three SMs (audio, visual, and bimodal). Finally, the figure shows that CAs ($p = 0.707$) between MP and MI are comparable for the three SMs and the two methods in use.

Index of Dispersion of the HQFVs in Location, Frequency, and Time

The IDs of the HQFVs, which were obtained from the standard method, are presented in **Figures 6A–C**. In location and time, the HQFVs are generally spread over all the feature sources showing IDs above 0.52 and 0.7, respectively. By contrast, IDs range between 0 and 1 in frequency. The IDs of the HQFVs, which resulted from the all-embracing method, are provided in **Figures 6D–F**. These IDs are above 0.85, 0.5, and 0.87 in location, frequency, and time, respectively. The statistical comparison of the IDs between both methods in location, frequency, and time resulted in the following p-values: 1.461×10^{-9}, 0.049, and 4.767×10^{-7}. Note that all the remarks mentioned in this section apply to MP and MI.

Modal Distribution of the HQFVs in Location, Frequency, and Time

Standard Method

Figure 7 presents the *modal* distribution of the HQFVs over the following feature sources: (a) 15 recording sites, (b) 4 frequency bands, and (c) 5/9 time windows for MP/MI. With regard to the location domain, **Figure 7A** shows that *modes* from audio cues mainly tend toward FC3 and C3, while those from visual cues mostly tend to FC2, FC4 (only applicable for MI), and C4. *Modes* from bimodal cues are essentially distributed among FC3, C3, FC4, and C4. In all the cases, MI displays greater tendencies than MP.

We can see from **Figure 7B** that the overriding band for the three SMs is α_U. In this case, the highest and the lowest tendencies are reached by *modes* from visual and bimodal cues, respectively. The MI control task shows a second dominant band. Such dominant band for *modes* from audio and visual cues is β_L, while that for *modes* from bimodal cues is β_U. The MP control task only shows a second dominant band for *modes* from bimodal cues, which is β_L. As in the location domain, MI reveals stronger tendencies in comparison with MP.

Lastly, **Figure 7C** provides the *modal* distribution in time. Keeping in mind that MP only involved five time windows (from δ_5 to δ_9), we can see that *modes* from the three SMs are evenly distributed along most of them. Although MI involved the nine time windows, the *modes* from the three SMs are mostly distributed across δ_5 and δ_9 as well. In both cases, the major *modal* tendencies for audio, visual, and bimodal cues are, respectively, the following: δ_7/δ_8 (1500–2000 ms), $\delta_6/\delta_8/\delta_9$ (1250–2500 ms), and δ_5/δ_9 (1000–1500 ms and 2000–2500 ms). There is additionally a relevant *modal* distribution over δ_{13} for the three SMs in MI, regardless of the decreasing trend of the foregoing time windows.

All-Embracing Method

Figure 8 provides the *modal* distribution of the HQFVs over the following feature sources: (a) 61 recording sites, (b) 7 frequency bands, and (c) 9/13 time windows for MP/MI. With respect to the location domain, **Figure 8A** indicates that *modes* from the three SMs are distributed over about 40% of the feature sources in both control tasks. Specifically, *modes* from audio cues are distributed among 24 of the 61 recording sites. From those, 62% are on central areas, 25% are on parieto-occipital areas, and 13% are on frontal areas. *Modes* from visual cues are also distributed among 24 of the 61 recording sites. However, those are differently spread. Over

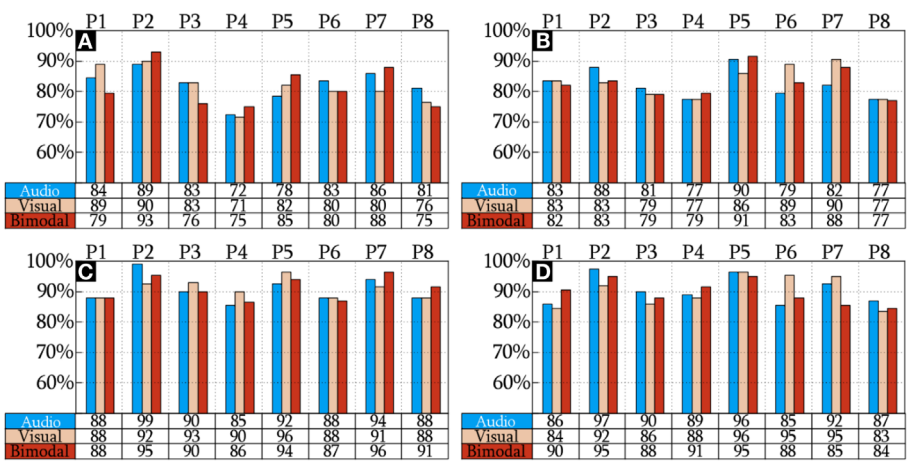


FIGURE 5 | CA of the HQFVs. MP and MI analyzed using the standard method are illustrated in **(A,B)**, respectively. MP and MI analyzed using the standard method are shown in **(C,D)**, respectively.

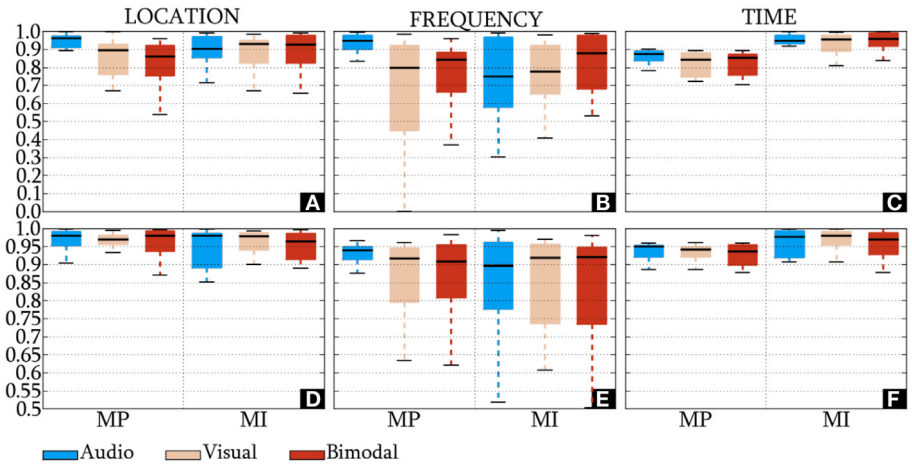


FIGURE 6 | IDs of the HQFVs. For the standard method, the indexes of the HQFVs over feature sources in location, frequency, and time are depicted in **(A–C)**, respectively. For the all-embracing method, the indexes of the HQFVs over feature sources in location, frequency, and time are presented in **(D–F)**, respectively.

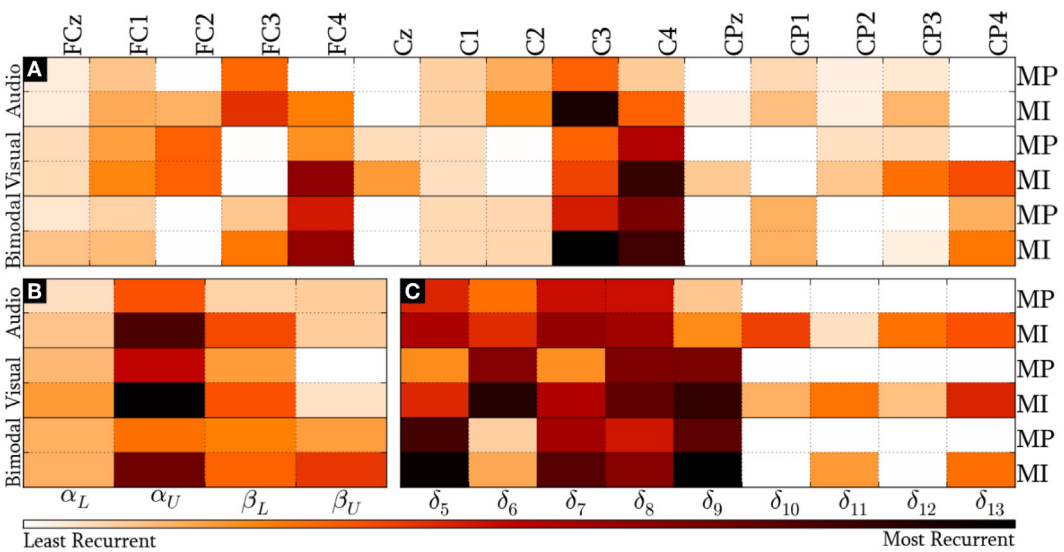


FIGURE 7 | *Modal* distribution of the HQFVs that resulted from the standard method. The *modal* distribution is illustrated in three domains: **(A)** 15 recoding sites, **(B)** 4 frequency bands, and **(C)** 5/9 time windows for MP/MI.

half of them are distributed between frontal and parieto-occipital areas (33% and 21%, respectively), while less than half of them are related to central areas (46%). *Modes* from bimodal cues are distributed among 27 of the 61 recording sites. From those, 56% are on central areas, 37% are on parieto-occipital areas, and 7% are on frontal areas.

Figure 8B illustrates the prevalence of α_U band in the *modes* from the three SMs in both MP and MI. The figure also reveals the secondary but not insignificant role of β_L band. The *modal* distribution between θ_L and α_L bands is moderate for the three SMs, whereas that between θ_U and β_U bands is negligible for the three SMs. Furthermore, the *modal* distribution over γ band is considerable for bimodal cues. In all these cases, MI shows much higher tendencies than MP.

Last but not least, **Figure 8C** depicts the *modal* distribution in time, considering that MP only involved 9 of the 13 time windows. It can be seen from this figure that the *modes* from audio cues are spread across δ_1 and δ_9 (0–2500 ms), while those from visual and bimodal cues are spread across δ_1 and δ_7 (0–2000 ms). Particularly for MI control tasks, the *modes* from audio, visual, and bimodal cues strongly tend toward δ_3 (500–1000 ms), δ_4 (750–1250 ms), and δ_2/δ_3 (250–1000 ms), respectively. In addition, there is an unexpected *modal* tendency to δ_{13} for the three SMs in MI.

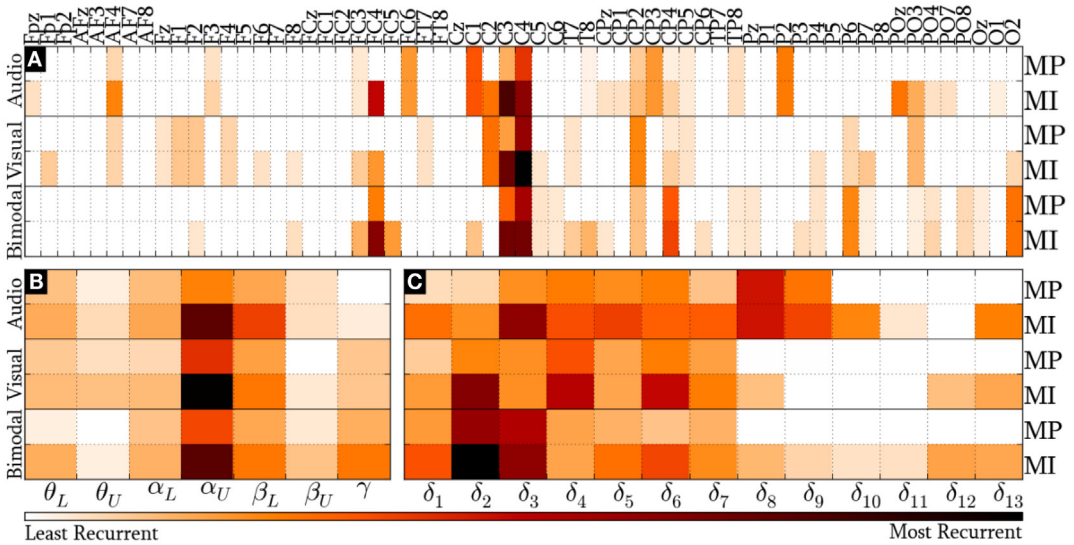


FIGURE 8 | *Modal* distribution of the HQFVs that resulted from the all-embracing method. The *modal* distribution is illustrated in three domains: **(A)** 61 recording sites, **(B)** 7 frequency bands, and **(C)** 9/13 time windows for MP/MI.

DISCUSSION

This paper set out with the aim of analyzing the cue effects on the discriminability of MP and MI control tasks. The analysis was carried out using two methods: standard and all-embracing. The standard method was based on feature sources, where the modulation of brain signals due to MP/MI is typically detected. For the all-embracing method, the scope of the standard method was extended by including a wider variety of feature sources, where not only motor activity is reflected. The analysis was limited to the HQFVs, i.e., the feature vectors that yielded the highest CAs during a DBI-FDA process. The following is a discussion of the most relevant results of the analysis.

Classification Accuracy of the HQFVs

On the question of improving CA by using different cues, we found that there was no significant difference in the discrimination of MI-related control tasks triggered by three SMs: audio, visual, and bimodal. However, there was a significant increase in the CA of control tasks analyzed under the all-embracing method over those analyzed using the standard method. These findings demonstrated that an unbiased approach in location, frequency, and time leads to a better performance, but different cues do not make a difference. Although Scheel et al. (2015) suggested that different stimuli might improve the CA of the control tasks at hand, this study has been unable to demonstrate that. Nevertheless, it is a fact that more distinguishable EEG patterns are extracted from no MI-related sources. Possibly, other type of stimuli could improve the differentiation of MI-related control tasks. In accordance with the findings of Pfurtscheller and Neuper (2001) and Obermaier et al. (2003), control tasks are result of conscious and unconscious processes. As different stimuli may evoke different unconscious processes, more differentiable EEG patterns could be found. However, this needs further investigation.

Another important result was the analogous performances of MP and MI for the three SMs and the two methods. It is well established that both control tasks generate similar event-related oscillations, but the "no-go" signal accompanying MI is frequently overlooked (Krepki et al., 2007). An imaginary movement activates motor areas of the brain almost to the same extent as a real one, except for the visible contractions. This means that the neural commands for muscular contractions are blocked at some level of the motor system by an active inhibitory mechanism. This questions whether MI in motor-disabled people takes place like that in healthy ones, or it is rather a real movement process (Jeannerod, 2006). In addition, the use of MI as control task involves the development of an electromyographic detector so as to eliminate undesirable muscular activity un- or consciously triggered by healthy BCI users. Based on these two factors and given that both control tasks achieved analogous performances, MP may be a better option for BCI systems.

Index of Dispersion of the HQFVs in Location, Frequency, and Time

For both methods, the distribution of the HQFVs over the available feature sources was much more even in location and time than in frequency. This indicates that the most gainful features for discriminating between left and right MIs mainly proceeded from the entire set of recording sites and the total duration of the control task, but only from a specific frequency band (α_U). The finding is in agreement with that of other studies (Pfurtscheller and Neuper, 2001; Neuper et al., 2009) which showed that the correct discrimination between left and right started 250–500 ms after cue onset and where the most discriminating frequency band

was the α_U. With reference to the location domain, although this finding differs from some published studies (Ramoser et al., 2000; Leeb et al., 2006), it is consistent with those of Meckes et al. (2004) and Sepulveda et al. (2004). They suggested giving attention to non-motor locations, even when the mental task of interest was movement related.

The current study also found that the HQFVs tended to be more widely spread over the feature sources in the all-embracing method than in the standard method. It is worth mentioning that the inclusion of more feature sources increased the diversity of HQFVs. This result corroborate the ideas of Pfurtscheller and Neuper (2001) and Obermaier et al. (2003), who suggested that control tasks are result of the mental effort demanded by the control task (conscious process) and the sensory–cognitive processing of the cue (unconscious process).

Modal Distribution of the HQFVs in Location, Frequency, and Time

The *modes* of the HQFVs fundamentally tended toward the expected sources (Pineda, 2005). These were the C3/C4 recoding sites and the α_U/β_L frequency bands. The *modes* also revealed clear tendencies toward feature sources that reflected the nature of the cue in use. Before going on to discuss this further, it is necessary to mention that the *modal* tendencies were much greater in MI than in MP. The reason for this is not clear, but it may be due to the mental effort involved in each control task. MP is such an intention, whereas MI is a dynamic process that goes through many of the central phases of actual movements.

Standard Method

The *modes* from audio cues tended to the left hemisphere, where some language-related functions take place, whereas those from visual cues tended to the right hemisphere, where visual perception is processed (Kropotov, 2010). Being the bimodal cue, a composition of audio and visual cues, the corresponding *modal* tendency was to both hemispheres. This result suggests that the most discriminating features were defined not only by the MP/MI mechanisms but also by the sensory–cognitive processing of the cue in use. With respect to the frequency domain, the involvement of high frequency bands took importance successively in *modes* from audio, visual, and bimodal cues. This result may be related to previous work of Giannitrapani [whose work is cited in Kropotov (2010)], who found that high beta activity (21–33 Hz) increased when the stimulus structure complexity also increased. It is possible, therefore, that the cue complexity had played a significant role in the discrimination process of features as well. Regarding the time domain, the highest tendencies took more time (after the cue onset) to appear in *modes* from audio than from bimodal cues. Hence, it is also possible to hypothesize that more informative features were found earlier when a more direct cue was employed.

All-Embracing Method

The *modes* from audio cues mostly tended to central recording sites, where auditory evoked potentials are typically recorded (Proverbio and Zani, 2003), and to δ_3 (500–1000 ms) time window, where brain rhythms normally respond to the recognition and/or retrieval of acoustic stimuli (Krause, 2006). On the other hand, the visual stimulation is registered around 200 ms post-stimulus as a response to modulations of alpha band rhythms over parieto-occipital areas and beta band rhythms over fronto-parieto-occipital areas (Kropotov, 2010; Andreassi, 2013). This may be a reason why *modes* from visual cues tended to fronto-parieto-occipital recording sites, α_U band, and δ_2 (250–750 ms) time window. Finally, the *modes* from bimodal cues displayed a well-balanced distribution between central and fronto-parieto-occipital recording sites and between δ_2 (250–750 ms) and δ_3 (500–1000 ms). This finding confirms that bimodal stimuli require feature sources that are also required by audio and visual stimuli separately (Isoğlu-Alkaç et al., 2007). In the frequency domain, the remarkable tendency of these *modes* toward γ band accords with previous findings of Ward (2003), who found that the sensory decoding around 250 ms post-stimulus is reflected in modulation of γ band rhythms.

In the three SMs, one unanticipated finding was the minor role occupied by β_U band that is well-known as one of the major contributors in the discrimination process of MI activity. A possible explanation for the small contribution of this band is that neural desynchronization around 20 Hz has been considered as a harmonic response of desynchronization around 10 Hz, whereas the one around 16 Hz is an authentic response to motor activity (Pfurtscheller et al., 1996). Moreover, Pfurtscheller et al. (1999) found that the most discriminating frequency components throughout MI-related tasks were found within the α_U band in three of four subjects, while those were found within the β_U band only in one subject.

Lastly, it is worth noting the underlying tendency of *modes* from visual and bimodal cues toward the δ_9 (2000–2500 ms) time window in the standard method. There was also another clear tendency of *modes* from the three SMs toward the δ_{13} (3000–3500 ms) time window in both methods. For visual and bimodal stimulation, we believe that gaze fixation at the screen center provoked by cues "left"/"right" could have driven the participants to anticipate the upcoming cue "start." Similarly, the cue "start" appearance caused the cue "stop" expectation. For audio stimulation, once the participants had received the cue "start," and owing to the likeness between increasing and decreasing tones (cues "start" and "stop," respectively), the anticipation of the audio cue "stop" was likely to have arisen. This speculation is supported by the findings of Scherer et al. (2008), who found that the involuntary expectations for the approaching cues provoked false control commands during virtual navigation. Another interesting tendency of *modes* of the three SMs is toward δ_2 (250–750 ms) and δ_3 (500–1000 ms) time windows in the all-embracing method. These results are in agreement with the findings of Pfurtscheller et al. (2008), who showed that distinct short-lasting brain patterns appeared within a time window of about 500–750 ms after cue onset.

All these interpretations must be, however, taken with caution. More research on this topic need to be undertaken because these findings can only be conclusive in early training sessions. The effects observed in this study could diminish or vanish, either

in further training sessions or in online applications. Another source of uncertainty is associated with the ambiguity of multivariate classifiers (such as FDA) to determine the brain regions, frequency, and time intervals where cognitive processes are reflected. Haufe et al. (2014) demonstrated that backward models (e.g., multivariate classifiers) combine information from different channels to separate the brain patterns of different classes. These models may, however, give significant weight to channels unrelated to brain processes of interest. By contrast, forward models (e.g., blind source separation) explain how the measure data are generated from the neural sources, providing a neurophysiological meaning of the outcomes. Furthermore, Haufe et al. showed that brain patterns were much smoother and covered more diverse cognitive-related areas, when those patterns were obtained via forward methods. These findings are of particular interest due the nature of our study. It seems that the present results are limited by the methods applied to select the features vectors. Possibly, by transforming the backward model in use (DBI-FDA process) into a forward model such as proposed Haufe et al. (selection of brain patterns according to the neurophysiological contribution of each EEG channel), a clearer feature distribution over unrelated MI sources could have been achieved.

Implications on Neuroergonomics Research

This is a key issue for Neuroergonomics research because neural activity could not only be used to monitor the human mental state, but this might be also employed to control a system of interest. In fact, Myrden and Chau (2015) have suggested to develop a BCI system on the basis of an overt adaptation to keep user mental within the optimal region, and a covert adaptation that automatically adjusts BCI parameters according to such mental state.

An important application may be on driver modeling and vehicle simulation environments (Xu et al., 2015). These two areas of research have been of interest to develop driver assistance systems for safer driving and intelligent transportation. For example, EEG signals of a driver can be used to model the driver neuromuscular dynamics (Bi et al., 2015) and, thus, improving the performance of a driver simulator. Such EEG signals can also be employed to detect the fatigue (Wang et al., 2014) and level of attention (Wang et al., 2015) of the driver to activate the driver simulator and, hence, preventing driving accidents. Furthermore, the performance of the driver simulator can be improved by analyzing the human reaction to traffic cues such as car horn, direction indicators, and traffic lights. All of these cues produce specific EEG patterns on the driver brain signals as has been shown in this study.

CONCLUSION

The findings of this study have provided a new understanding of how MI-related control tasks used to control a BCI system may become modified by their preceding cues. Although previous investigations have somehow studied the cue effects on MI-related control tasks; in this study, we have shown that the CA of those control tasks does not depend on the type of cue in use. Moreover, we found that the EEG patterns that best differentiate MI-related control tasks emerge from recording sites, frequency bands, and time windows well defined by the perception and cognition of the cue in use. An implication of this study is the possibility of obtaining different control commands that could be detected with the same accuracy. Since different cues trigger control tasks that yield similar CAs, and those control tasks produce EEG patterns differentiated by the cue nature, this leads to accelerate the brain–computer communication by having a wider variety of detectable control commands. This is an important issue for Neuroergonomics research because neural activity could not only be used to monitor the human mental state, but this might be also employed to control the system of interest.

ACKNOWLEDGMENTS

LA-V wishes to express her sincere gratitude to the National Council of Science and Technology of Mexico (CONACyT) for providing funding support to accomplish this research project (grant number: 306714). Authors also thank Alexander Bedborough (language advisor of the International Academy at the University of Essex) for his editorial help with the manuscript.

REFERENCES

Alonso-Valerdi, L. M., and Sepulveda, F. (2015). *Implementation of a Motor Imagery Based BCI System Using Python Programming Language*. Angers: SCITEPRESS Digital Library.

Altermaller, E. O., Munte, T. F., and Gerloff, C. (2005). "Neurocognitive functions and the EEG," in *Electroencephalography: Basic Principles, Clinical Applications and Related Fields*, 5th Edn, eds Niedermeyer E. and Lopes da Silva F. (Philadelphia, PA: Lippincott Williams and Wilkins), 661–683.

Andreassi, J. L. (2013). *Psychophysiology: Human Behaviour and Physiological Response*, 4th Edn. New York, NY: Psychology Press.

Bamdadian, A., Guan, C., Ang, K. K., and Xu, J. (2014). The predictive role of pre-cue EEG rhythms on MI-based BCI classification performance. *J. Neurosci. Methods* 235, 138–144. doi:10.1016/j.jneumeth.2014.06.011

Basar, E., Basar-Erouglu, C., Karakas, S., and Schumann, M. (1999). Are cognitive processes manifested in event-related gamma, alpha, theta and delta oscillations in the EEG? *Neurosci. Lett.* 259, 165–168. doi:10.1016/S0304-3940(98)00934-3

Bi, L., Wang, M., Wang, C., and Liu, Y. (2015). Development of a driver lateral control model by integrating neuromuscular dynamics into the queuing network-based driver model. *IEEE Trans. Intelligent Transportation Syst.* 16(5), 2479–2486.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. New York, NY: Springer.

Boostani, R., Graimann, B., Moradi, M. H., and Pfurtscheller, G. (2007). A comparison approach toward finding the best feature and classifier in cue-based BCI. *Med. Biol. Eng. Comput.* 45, 403–412. doi:10.1007/s11517-007-0169-y

Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21. doi:10.1016/j.jneumeth.2003.10.009

Dornhege, G., Krauledat, M., Müller, K.-R., and Blankertz, B. (2007). *Toward Brain-Computer Interfacing*. Cambridge, MA: MIT Press.

Galán, F., Nuttin, M., Lew, E., Ferrez, P. W., Vanacker, G., Philips, J., et al. (2008). A brain-actuated wheelchair: asynchronous and non-invasive brain-computer

interfaces for continuous control of robots. *Neurophysiol. Clin.* 119, 2159–2169. doi:10.1016/j.clinph.2008.06.001

Gargiulo, G., Calvo, R. A., Bifulco, P., Cesarelli, M., Jin, C., Mohamed, A., et al. (2010). A new EEG recording system for passive dry electrodes. *Neurophysiol. Clin.* 121, 686–693. doi:10.1016/j.clinph.2009.12.025

Graimann, B., Allison, B., and Pfurtscheller, G. (2010). "Brain-computer interfaces: a gentle introduction," in *Brain-Computer Interfaces*, 1st Edn, ed. Graimann B. (Berlin: Springer), 1–27.

Grosse-Wentrup, M., and Scholkopf, B. (2012). High gamma-power predicts performance in sensorimotor-rhythm brain-computer interfaces. *J. Neural. Eng.* 9, 96–046001. doi:10.1088/1741-2560/9/4/046001

Gutierrez, D., Ramirez-Moreno, M. A., and Lazcano-Herrera, A. G. (2015). "Assessing the acquisition of a new skill with electroencephalography," in Proceedings of the 7th International IEEE EMBS Neural Engineering Conference, Montpellier.

Hassanien, A. E., and Azar, A. T. (2015). *Brain-Computer Interfaces: Current Trends and Applications.* 1st Edn. Berlin: Springer-Verlag.

Haufe, S., Meinecke, F., Görgen, K., Dähne, S., Haynes, J. D., Blankertz, B., et al. (2014). On the interpretation of weight vectors of linear models in multivariate neuroimaging. *Neuroimage* 87, 96–110. doi:10.1016/j.neuroimage.2013.10.067

Honsbeek, R., Kuiper, T., and van Rijn, C. M. (1998). *BIOSEMI*. Available at: http://www.biosemi.com/products.htm

Isoğlu-Alkaç, U., Kedzior, K., Keskindemirci, G., Ermutlu, N., and Karamursel, S. (2007). Event-related potentials to visual, auditory, and bimodal (combined auditory-visual) stimuli. *Int. J. Neurosci.* 117, 259–273. doi:10.1080/00207450500534118

Jackson, M. M., and Mappus, R. (2010). "Neural control interfaces," in *Brain-Computer Interfaces: Applying Our Minds to Human-Computer Interaction*, 1st Edn, eds Tan D. S. and Nijholt A. (London: Springer-Verlag), 21–35.

Jeannerod, M. (2006). *Motor Cognition: What Actions Tell the Self*. 5th Edn. Oxford: Oxford University Press.

Kovács, F., Legány, C., and Babos, A. (2005). "Cluster validity measurement techniques," in *6th International Symposium of Hungarian Researchers on Computational Intelligence,* Budapest.

Krause, C. M. (2003). "Brain electric oscillations and cognitive processes," in *Experimental Methods in Neuropsychology*, ed. Hugdahl K. (Amsterdam: Springer), 111–130.

Krause, C. M. (2006). Cognition- and memory-related ERD/ERS responses in the auditory stimulus modality. *Prog. Brain Res.* 159, 197–207. doi:10.1016/S0079-6123(06)59013-2

Krepki, R., Curio, G., Blankertz, B., and Müller, K.-R. (2007). Berlin brain-computer interface – the HCI communication channel for discovery. *Int. J. Hum. Comput. Stud.* 65, 460–477. doi:10.1016/j.ijhcs.2006.11.010

Kropotov, J. (2010). *Quantitative EEG, Event-Related Potentials and Neurotherapy*, 1st Edn. San Diego, CA: Academic Press – Elsevier.

Leeb, R., Keinrath, C., Friedman, D., Guger, C., Scherer, R., Neuper, C., et al. (2006). Walking by thinking: the brainwaves are crucial, not the muscles. *Presence* 15, 500–514. doi:10.1162/pres.15.5.500

Lotte, F., Renald, Y., and Lécuyer, A. (2008). "Self-paced brain-computer interaction with virtual worlds: a quantitative and qualitative study out of the lab," in *4th International Brain Computer Interface Workshop and Training Course*, Graz.

Maeder, C. L., Sannelli, C., Haufe, S., and Blankertz, B. (2012). Pre-stimulus sensorimotor rhythms influence brain-computer interface classification performance. *IEEE Trans. Neural Syst. Rehabil. Eng.* 20, 653–662. doi:10.1109/TNSRE.2012.2205707

Meckes, M. P., Sepulveda, F., and Conway, B. A., (2004). "1st order class separability using EEG-based features for classification of wrist movements with direction selectivity," in *Engineering in Medicine and Biology Society*, 2004. IEMBS '04. 26th Annual International Conference of the IEEE, San Francisco, CA, 4377–4379.

Myrden, A., and Chau, T. (2015). Effects of user mental state on EEG-BCI performance. *Front. Hum. Neurosci.* 9:308. doi:10.3389/fnhum.2015.00308

Neuper, C., Grabner, R. H., Fink, A., and Neubauer, A. C. (2005). Long-term stability and consistency of EEG Event-Related (de-) synchronization across different cognitive tasks. *Neurophysiol. Clin.* 116, 1681–1694. doi:10.1016/j.clinph.2005.03.013

Neuper, C., Scherer, R., Wriessnegger, S., and Pfurtscheller, G. (2009). Motor imagery and action observation: modulation of sensorimotor brain rhythms during mental control of a brain-computer interface. *Neurophysiol. Clin.* 120, 239–247. doi:10.1016/j.clinph.2008.11.015

Neuper, C., Wörtz, M., and Pfurtscheller, G. (2006). ERD/ERS patterns reflecting sensorimotor activation and deactivation. *Prog. Brain Res.* 159, 211–222. doi:10.1016/S0079-6123(06)59014-4

Nijboer, F., Furdea, A., Gunst, I., Mellinger, J., McFarland, D. J., Birbaumer, N., et al. (2008). An auditory brain-computer interface (BCI). *J. Neurosci. Methods* 167, 43–50. doi:10.1016/j.jneumeth.2007.02.009

Norman, G. R., and Streiner, D. L. (2008). *Biostatistics: The Bare Essentials*. 3rd Edn. Shelton: People's Medical Publishing House.

Obermaier, B., Müller, G., and Pfurtscheller, G. (2003). Virtual keyboard controlled by spontaneous EEG activity. *IEEE Trans. Neural. Syst. Rehabil. Eng.* 11, 422–426. doi:10.1109/TNSRE.2003.816866

Pfurtscheller, G., Müller-Putz, G. R., Alois Schlögl, Graimann, B., Scherer, R., Leeb, R. (2007). "Graz-brain-computer interface: state of research," in *Toward Brain-Computer Interfacing*, eds G. Dornhege, J. D. R. Millán, T. Hinterberger, D. J. McFarland, and K. R. Müller (Cambridge: The MIT Press), 65–84.

Pfurtscheller, G., and Lopes da Silva, F. H. (1999). "Functional meaning of event-related desynchronization (ERD) and synchronization (ERS)," in *Handbook of Electroencephalography and Clinical Neurophysiology*, eds Pfurtscheller G. and Lopes da Silva F. H. (Amsterdam: Elsevier), 51–65.

Pfurtscheller, G., and Neuper, C. (2001). Motor imagery and direct brain-computer communication. *Proc. IEEE* 89, 1123–1134. doi:10.1109/5.939829

Pfurtscheller, G., Neuper, C., Ramoser, H., and Müller-Gerking, J. (1999). Visually guided motor imagery activates sensorimotor areas in humans. *Neurosci. Lett.* 269, 153–156. doi:10.1016/S0304-3940(99)00452-8

Pfurtscheller, G., Scherer, R., Müller-Purtz, G. R., and Lopes de Silva, F. H. (2008). Short-lived brain state after cued motor imagery in naive subjects. *Eur. J. Neurosci.* 28, 1419–1426. doi:10.1111/j.1460-9568.2008.06441.x

Pfurtscheller, G., Stancák, A., and Neuper, C. (1996). Post-movement beta synchronization. A correlate of an idle motor area? *Electroencephalogr. Clin. Neurophysiol.* 98, 281–293. doi:10.1016/0013-4694(95)00258-8

Pineda, J. A. (2005). The functional significance of mu rhythms: translating "seeing" and "hearing" into "doing". *Brain Res. Rev.* 50, 57–68. doi:10.1016/j.brainresrev.2005.04.005

Proverbio, A. M., and Zani, A. (2003). "Electromagnetic manifestations of mind and brain," in *The Cognitive Electrophysiology of Mind and Brain*, eds Zani A. and Proverbio A. M. (San Diego, CA: Academic Press), 13–40.

Ramoser, H., Müller-Gerking, J., and Pfurtscheller, G. (2000). Optimal spatial filtering of single trial EEG during imagined hand movement. *IEEE Trans. Rehabil. Eng.* 8, 441–446. doi:10.1109/86.895946

Scheel, H., et al. (2015). "Influence of external cues on synchronized brain-computer interface based on movement related cortical potentials," in *2015 7th International IEEE/EMBS Conference on Neural Engineering (NER)*, Montpellier.

Scherer, R., Lee, F., Schlogl, A., Leeb, R., Bischof, H., and Pfurtscheller, G. (2007). "EEG-based interaction with virtual worlds: a self-paced three class brain-computer interface," in *BRAINPLAY 07 Brain-Computer Interfaces and Games Workshop at ACE (Advances in Computer Entertainment)* (Graz: Advances in Computer Entertainment).

Scherer, R., Lee, F., Schlogl, A., Leeb, R., Bischof, H., and Pfurtscheller, G. (2008). Toward self-paced brain-computer communication: navigation through virtual worlds. *IEEE Trans. Biomed. Eng.* 55, 675–682. doi:10.1109/TBME.2007.903709

Sepulveda, F., Meckes, M. P., and Conway, B. A. (2004). "Cluster separation index suggests usefulness of non-motor EEG channels in detecting wrist movement direction intention," in *IEEE Conference on Cybernetics and Intelligent Systems*, Singapore, 943–947.

Shelton, J., and Kumar, G. P. (2010). Comparison between auditory and visual simple reaction times. *Neurosci. Med.* 1, 30–32. doi:10.4236/nm.2010.11004

Teichner, W. H. (1954). Recent studies of simple reaction time. *Psychol. Bull.* 51, 128–149. doi:10.1037/h0060900

Tseng, K. C., Lin, B. S., Wang, A. M. K., and Lin, B. S. (2015). Design of a mobile brain computer interface based smart multimedia controller. *Sensors* 15, 5518–5530. doi:10.3390/s150305518

Wang, H., Zhang, C., Shi, T., Wang, F., and Ma, S. (2014). Real-time EEG-based detection of fatigue driving danger for accident prediction. *Int. J. Neural Syst.* 25, 1–14. doi:10.1080/00207721.2014.953798

Wang, Y. K., Jung, T. P., and Lin, C. T. (2015). EEG-based attention tracking during distracted driving. *IEEE Trans. Neural Syst. Rehabil. Eng.* 23(6), 1085–1094.

Ward, L. M. (2003). Synchronous neural oscillations and cognitive processes. *Trends Cogn. Sci.* 7, 553–559. doi:10.1016/j.tics.2003.10.012

Xu, L., Hu, J., Jiang, H., and Meng, W. (2015). Establishing style-oriented driver models by imitating human driving behaviors. *IEEE Trans. Intell. Transp. Syst.* 99, 1–9. doi:10.1109/TITS.2015.2459722

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Alonso-Valerdi, Sepulveda and Ramírez-Mendoza. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 27 April 2016
doi: 10.3389/fnhum.2016.00187

The Brain Is Faster than the Hand in Split-Second Intentions to Respond to an Impending Hazard: A Simulation of Neuroadaptive Automation to Speed Recovery to Perturbation in Flight Attitude

Daniel E. Callan[1,2]*, *Cengiz Terzibas*[2], *Daniel B. Cassel*[3], *Masa-aki Sato*[4] and *Raja Parasuraman*[5]

[1] Center for Information and Neural Networks, National Institute of Information and Communications Technology, Osaka University, Osaka, Japan, [2] Multisensory Cognition and Computation Laboratory, Universal Communication Research Institute, National Institute of Information and Communications Technology, Kyoto, Japan, [3] Locomobi Inc., Toronto, ON, Canada, [4] Neural Information Analysis Laboratories, Advanced Telecommunications Research Institute, Kyoto, Japan, [5] Center of Excellence in Neuroergonomics, Technology, and Cognition, George Mason University, Fairfax, VA, USA

OPEN ACCESS

Edited by:
Thorsten O. Zander, Technical University of Berlin, Germany

Reviewed by:
Sebastien Helie, Purdue University, USA
Josef Faller, Graz University of Technology, Austria

***Correspondence:**
Daniel E. Callan
dcallan@nict.go.jp

In Memoriam:
The present manuscript wishes to be the authors' personal honor to the memory of Raja Parasuraman, as a tribute to his scientific competence, and the contribution that he gave to our research projects in these years of proficient collaboration.

Received: *06 November 2015*
Accepted: *12 April 2016*
Published: *27 April 2016*

Citation:
Callan DE, Terzibas C, Cassel DB, Sato M and Parasuraman R (2016) The Brain Is Faster than the Hand in Split-Second Intentions to Respond to an Impending Hazard: A Simulation of Neuroadaptive Automation to Speed Recovery to Perturbation in Flight Attitude. Front. Hum. Neurosci. 10:187. doi: 10.3389/fnhum.2016.00187

The goal of this research is to test the potential for neuroadaptive automation to improve response speed to a hazardous event by using a brain-computer interface (BCI) to decode perceptual-motor intention. Seven participants underwent four experimental sessions while measuring brain activity with magnetoencephalograpy. The first three sessions were of a simple constrained task in which the participant was to pull back on the control stick to recover from a perturbation in attitude in one condition and to passively observe the perturbation in the other condition. The fourth session consisted of having to recover from a perturbation in attitude while piloting the plane through the Grand Canyon constantly maneuvering to track over the river below. Independent component analysis was used on the first two sessions to extract artifacts and find an event related component associated with the onset of the perturbation. These two sessions were used to train a decoder to classify trials in which the participant recovered from the perturbation (motor intention) vs. just passively viewing the perturbation. The BCI-decoder was tested on the third session of the same simple task and found to be able to significantly distinguish motor intention trials from passive viewing trials (mean = 69.8%). The same BCI-decoder was then used to test the fourth session on the complex task. The BCI-decoder significantly classified perturbation from no perturbation trials (73.3%) with a significant time savings of 72.3 ms (Original response time of 425.0–352.7 ms for BCI-decoder). The BCI-decoder model of the best subject was shown to generalize for both performance and time savings to the other subjects. The results of our off-line open loop simulation demonstrate that BCI based neuroadaptive automation has the potential to decode motor intention faster than manual control in response to a hazardous perturbation in flight attitude while ignoring ongoing motor and visual induced activity related to piloting the airplane.

Keywords: neuroadaptive automation, brain computer interface, brain machine interface, neuroergonomics, decoding, independent component analysis, MEG, aviation

INTRODUCTION

Safe and effective performance in many occupational settings is critically dependent on people making timely and correct split-second decisions to avoid an impending hazard. Consider a speeding driver having to swerve to avoid hitting a child running unexpectedly onto the roadway; a nurse having to administer defibrillation to a patient having sudden cardiac arrest; or a pilot having to execute a rapid maneuver to recover from a stall or other abrupt perturbation during high-speed flight. Although some drivers, nurses, and pilots may be sufficiently skilled to make quick decisions and avoid mishaps in these situations, there are many conditions—fatigue, stress, mind wandering, task overload, to name a few—that can degrade human performance so that a correct and timely response is not possible and an accident may result.

One approach to this problem is to enhance human performance in such time-critical situations by decoding a person's neural activity associated with the intention to act. Once intention has been detected, one could provide appropriate feedback to the human operator or trigger computer aiding. Brain activity precedes motor action, so if neural signals associated with the intention to act could be successfully decoded in real time, one could use the decoded output to aid the human user. Using computer technology to augment human performance based on an assessment of human operator cognitive states is termed adaptive automation (Parasuraman et al., 1992; Scerbo, 2008; Parasuraman and Galster, 2013). *Neuroadaptive* automation is when neural signals are used to assess operator state, an approach that has been successful in mitigating human performance decrements in a variety of cognitive tasks (Byrne and Parasuraman, 1996; Prinzel et al., 2000; Scerbo et al., 2003; Wilson and Russell, 2007; Ting et al., 2010; Durantin et al., 2015; Gateau et al., 2015). Such an approach is consistent with the field of passive Brain Computer Interfaces (BCI), also referred to as Brain Machine Interfaces (BMI), in which user neural states are monitored in order to enhance human interaction with external devices (Blankertz et al., 2010; Lotte et al., 2013).

There is extensive research on the use of BCIs to support partially or fully disabled persons to control devise such as computers and prosthetic limbs (Reiner, 2008), and a smaller but growing literature on their use for healthy individuals so as to enhance human-system interaction (Zander and Kothe, 2011; Lotte et al., 2013). Comparatively little work has been conducted comparing the effects of neuroadaptive automation or passive BCIs to human performance in time-critical (split-second) decision-making situations [For related research see the studies by Haufe et al. (2011, 2014) and Kim et al. (2015) concerned with detection of braking intention by EEG]. In particular, when a critical event has to be detected and responded to quickly, can one decode the associated neural states of the human operator to achieve a faster response than the operator's manual action? We can rephrase the question as follows: given that the brain is faster than the hand (or foot or other effector), can one solve the problem that human manual actions are sometimes too sluggish to avoid a mishap when very little time is available by using the decoded brain activity to respond to a critical hazardous event?

We addressed this issue in the present study by examining whether neural signals could be decoded to enhance human performance in a time-critical decision-making task. We chose a decision-making situation that is encountered in aviation tasks: responding quickly to an in-flight perturbation, such as turbulence, micro-bursts, severe windshear, structural damage (e.g., from trim tab failure, bird strike, etc.). While such perturbations can occur in many types of flight, they can be a major contributor to mishaps in military aviation, given the greater exposure to risky situations requiring split-second decision-making, such as low-level flight over terrain, or high G-force maneuvers (Knapp and Johnson, 1996; Moroze and Snow, 1999; Nakagawa et al., 2007). When flying at high speed and very close to terrain, a savings of even a few milliseconds in responding to a perturbation can represent the difference between life and death (Haber and Haber, 2003).

Decoding neural states corresponding to cognitive states has been the object of considerable attention in the neuroimaging literature. A major approach to the problem has been to apply pattern-classification algorithms to multi-voxel functional MRI data in order to decode information representation in a participant's brain (Kamitani and Tong, 2005, 2006; Norman et al., 2006; Nishimoto et al., 2011; Poldrack, 2011; Shibata et al., 2012; Callan et al., 2014; Christophel and Haynes, 2014; Hutzler, 2014). However, the relatively low temporal resolution of fMRI and other neuroimaging methods based on cerebral hemodynamics renders them unsuitable for decoding neural states associated with split-second decision-making. Instead, electroencephalography (EEG) or magnetoencephalography (MEG) provide methods with sufficient temporal resolution to decode neural states associated with rapid decision-making. In the present study we used MEG as our primary source of neural activity, but also conducted an fMRI study to allow for better localization of MEG activity to brain areas.

A number of studies have applied pattern classification methods to neural signals in order to decode specific cognitive states. Typically these approaches train the classifier on part (e.g., half) of the neuroimaging obtained during performance of a cognitive task and then evaluate the effectiveness of the classifier on the remaining (untrained) half of the data (Garrett et al., 2003; Wilson and Russell, 2007; Baldwin and Penaranda, 2012). This is certainly an accepted criterion for evaluating how well a particular decoding algorithm works in a particular domain of human performance. But a stricter test is necessary if such neural BCI-decoders are to be useful in a general way. The more stringent test would involve application of a trained classifier to untrained data taken from a different task in the same general cognitive domain. Such a strategy, if successful, can provide for a more generalizable test of the efficacy of neural state decoding for a given cognitive function. We used this approach in the present study by training a MEG classifier during performance of a simple flight task involving a perturbation and applying it to a more

complex flight task involving similar types of perturbations during ongoing piloting. It is important to note that the ongoing piloting task uses the same control stick (controlled by articulation of the hand, wrist, and arm) as that needed to recover from the perturbation. It is therefore necessary for the BCI-decoder to be able to distinguish between brain activity responding to changes in the visual field and motor intention that are a result of piloting from brain activity responding to changes in the visual field and motor intention that arise from a perturbation (even though the BCI-decoder has not been specifically trained to do so). We additionally evaluated the generalizability of a trained model across participants' by using the weights of the model of the best participant to test performance over the trials of the complex flying task of the remaining participants.

Several MEG and EEG studies have identified neural correlates of visual and motor responses that suggest our goal of predicting motor intention to a visually presented hazard prior to actual movement is possible. Single trial response times to visual coherent motion onsets were predicted by MEG activity from 150 to 250 ms before the manual response of the observer (Amano et al., 2006). While the focus of the Amano et al. (2006) study is on the onset of visual perception, not on motor intention, it does provide a potential link between response time and the identification of the perceptual event. In a study investigating neural correlates of speeded motor responses to a visual stimulus it was found that larger low-theta complexes in EEG preceded more rapid button presses (Delorme et al., 2007). It has also been found that self-paced motor intention of reaching direction can be successfully decoded prior to movement onset (62.5 ms with 76% classification performance) using slow cortical potentials (0.1–1 Hz) recorded by EEG (Lew et al., 2014). In addition, research conducted on the detection of braking intention in simulated (Haufe et al., 2011; Kim et al., 2015) and real (Haufe et al., 2014) driving using EEG was able to make predictions about 130 ms earlier than the corresponding behavioral responses. The real-world task set out in our experiment to be able to predict motor intention to a visual hazard in the presence of complex ongoing motor control and a dynamically changing visual field goes beyond what was investigated in these previous studies. Nevertheless, we do believe that these studies taken together suggest that there may potentially be some features present in the MEG (and EEG) signal that can be decoded prior to movement onset in response to a visually presented hazard even under the robust conditions set out in our experiment.

There have been previous studies (Blankertz et al., 2002; Parra et al., 2003) using online BCI to detect error-related potentials to reduce error-rate and improve overall performance. While these methods are promising they utilize data that occurs after the response is made and are thus not applicable to our objective of detecting motor intention prior to movement. It is the goal of our study to utilize an off-line BCI-decoder to evaluate the feasibility of using real-time neuroadaptive automation to enhance piloting performance by reducing response time to recover from an impending hazard (see **Figure 1**).

MATERIALS AND METHODS

Participants

A total of seven right–handed adults participated in this study. Five (three females and two males) were glider pilots from local university clubs. The two participants (males) that were not pilots had considerable experience with driving or flying related video games. The age of the participants ranged from 19 to 40 years with a mean of 23.9 years and SE = 2.7 years. All participants gave written informed consent for experimental procedures approved by the ATR Human Subject Review Committee in accordance with the principles expressed in the Declaration of Helsinki.

Experimental Tasks

Two different tasks were used, a simple piloting task of level flight over the ocean and a more complex piloting task through the Grand Canyon. We used the first task to develop a method for decoding of neural states associated with response to a perturbation and the second task to investigate the generalizability of the method to a related but more complex situation. In both tasks the participant was given a first-person unobstructed view from the airplane (the view was as if from a camera in the front of the aircraft, see **Figures 2A–D**). The aircraft model simulated was an F22—Raptor using the X-plane flight simulator (Version 9.75, Laminar Research). The data for various flight parameters (elevator, aileron, rudder deflections, pitch, roll, yaw, heading, speed, dive rate, structural g-forces, latitude, longitude, altitude, etc.) and the control stick (NATA Technologies MRI and MEG compatible) deflections were collected at a mean sampling rate of 400 Hz using a UDP Matlab interface. The experimental conditions could be controlled via Matlab by using the UDP interface to give commands to the flight simulator.

Simple Piloting Task

This task had four conditions, two involving the presence or absence of a perturbation, and two in which the participant had the choice to either pilot the plane or passively watch the screen without moving the control stick [see Supplementary Videos 1–4; (1) fly_perturbation; (2) fly_noperturbation; (3) watch_perturbation; (4) watch_noperturbation; The participants viewed the 1st person perspective given on the left side of the video]. The primary task required the participant to pull back on the control stick (causing an upward elevator deflection resulting in the plane to climb) as rapidly as possible in response to a perturbation in attitude (orientation of the plane with respect to the horizon) causing the plane to dive at a steep rate (see **Figures 2A,B**). The participant was instructed to hold the control stick but not to move it until after the perturbation occurred. The perturbation consisted of instantaneous maximum downward deflection of the elevator for 200 ms causing the plane to enter a steep dive. The trial started with the plane flying at an altitude of 107 m above sea level at a speed of 1040 kph. The perturbation occurred on 67% of the trials at a random time between 2 and 4 s (randomly determined) after the beginning of the trial. If the plane descended to 30 m above sea level the

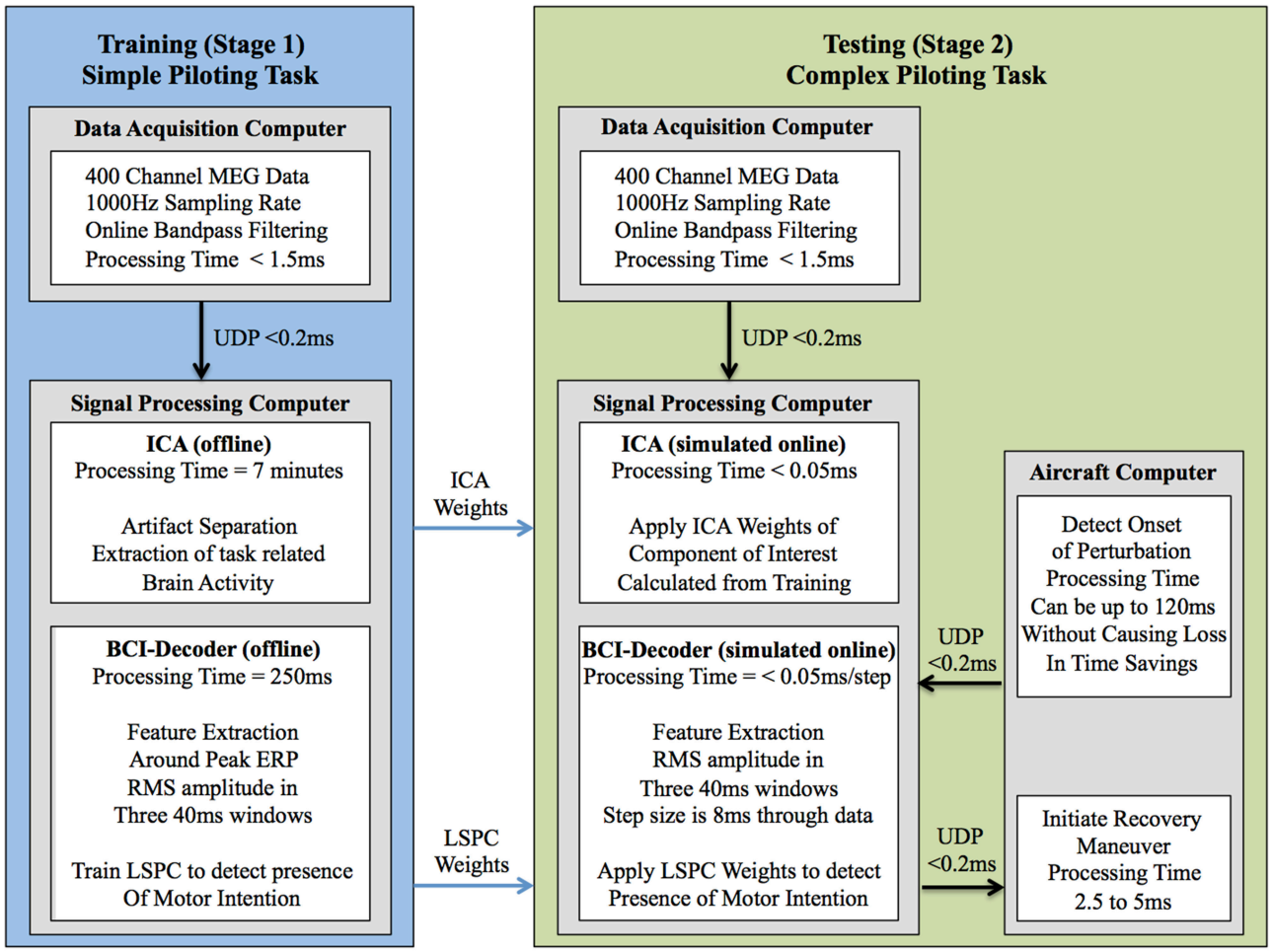


FIGURE 1 | Outline of processing procedures for the implementation of the hypothesized neuroadaptive automation to speed recovery to perturbation in flight attitude. The goal of this system is to speed up response time for the aircraft to recover from a perturbation by decoding the motor intention of the pilot. In this way the pilot is always in control of the aircraft rather than relying on automation in which the pilot is out of the loop. It should be noted that all processing was done offline and that the online parts of the system were simulated. The processing times for each of the procedures if ran in real-time online are given. As we were carrying out an offline simulation to determine the feasibility of signal processing and the BCI-decoder performance during training and testing stages for implementation in a real-time neuroadaptive automation system the aircraft computer was not actually implemented in this study. The system is theoretically able to work in real time with only a 5–7.5 ms loss in time savings because the weights of the ICA and BCI-decoder are determined before hand and applied to the online system. The aircraft computer is a necessary part of the neuroadaptive automation system that receives commands from the BCI-decoder to implement the recovery maneuver (in this case upward elevator deflection). The aircraft computer can also send information to the BCI-Decoder that can signal the onset of potential perturbations to the airplane. This information can be used to reduce the occurrence of false-alarms made by the system (executing upward elevator deflection when there is no actual perturbation or motor intention to recover). The aircraft computer can use up to 120 ms (time of the processing window for the BCI-decoder) to determine the presence of a non-pilot initiated perturbation in attitude without causing a loss in the time savings afforded by the neuroadaptive automation system. ICA, Independent Component Anayalsis; BCI, Brain Computer Interface; LSPC, Least Squares Probabilistic Classification; UDP, Universal Datagram Protocol.

simulator was paused before the plane crashed into the ocean. At the end of each trial the simulator was paused for 1.5–2.5 s (randomly determined). The timing was the same for trials in which there was no perturbation. Before the beginning of each trial participants chose by button press whether they were going to pilot the plane or passively watch without moving the joystick. Participants were instructed to try to make about twice as many piloting trials as passive trials. The rational for having the participant select whether they were to fly or watch rather than to direct them which condition it was by instruction was to better insure that they were actually doing the task correctly. If given visual directed instructions, participants would often try to recover from the perturbation even when they were instructed to just watch. Allowing participants to choose which condition to fly or watch helped to alleviate this problem. In this study, for the simple piloting task, only the trials containing a perturbation (fly_perturbation and watch_perturbation) were used to train the BCI-decoder. Please see the section under Decoding Pilot

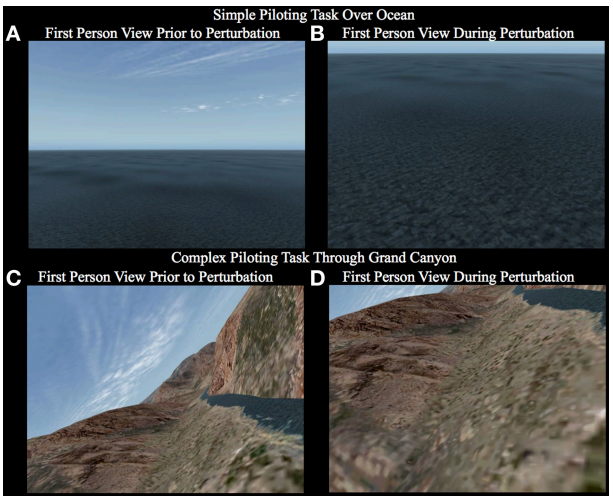


FIGURE 2 | First person view the participant observes while carrying out the simple piloting task over the ocean (A,B) and the complex piloting task (C,D). The first panel for each task **(A,C)** shows a representative image of what the view may appear like prior to the perturbation. The second panel for each task **(B,D)** shows a representative image of what the view may appear like during the perturbation. Notice that in the simple piloting task over the ocean **(A,B)** the bank angle is always level, whereas, in the complex piloting task the bank angle is continuously changing based on the control stick inputs to maintain the goal of tracking the river (See Supplementary Videos 1–6).

Intention below for the rational. After the button was pressed there was a delay of 0.8–1.3 s (randomly determined) before the trial begins. The passive trials were the same as the piloting trials with the exception that the plane would pause when it reached an altitude of 30 m above the ocean, which occurred at a mean time of 1.3 s after the onset of the perturbation.

There were 90 trials per session. On average there were 40 piloting perturbation trials, 20 piloting no-perturbation trials, 20 passive viewing perturbation trials, and 10 passive viewing no-perturbation trials. The actual number of trials for each condition was dependent on the participant's choice to pilot or passively view. The percentage of perturbation trials (67%) was experimentally determined and presented randomly within each of those conditions. Each session was ∼13 min. Bad trials (plane did not fly straight and level until time of perturbation) were removed from the analysis. Additionally, trials with response times slower than 700 ms from the onset of the perturbation were removed from the analysis.

Complex Piloting Task

This task involved flying through the Grand Canyon and consisted of two conditions: perturbation (67% of trials) or no perturbation (33%) [See Supplementary Videos 5, 6; (5) Grand_Canyon_perturbation; (6) Grand_Canyon_noperturbation]. Unlike the simple flying task, the participant was always required to pilot the plane. There were no passive viewing conditions. In the complex task the participant was constantly required to move the elevator and ailerons of the plane with the control stick to track the river through the Grand Canyon. The perturbation was caused by an instantaneous maximum downward deflection of the elevator for 200 ms. Depending on the attitude (particularly the angle of bank–roll) of the plane, the perturbation would cause a rapid departure from the trajectory of flight toward the ground and/or one of the cliffs (see **Figures 2C,D**). The plane started each trial at approximately 30 m above ground level at a speed of 1135 KPH. As in the simple task, the perturbation occurred between 2 and 4 s (randomly determined) after the beginning of the trial. There was also a pause for 1.5–2.5 s (randomly determined) at the end of each trial. Unlike the simple task, in which the participant specified by button press whether they were going to pilot the plane or passively watch, in the complex task every trial was a piloting trial. The instructions on the screen denoted that the participant could push the button when they were ready to begin the trial. After the button was pressed there was a delay of 0.8–1.3 s (randomly determined) before the trial began. Unlike the simple task, in the complex task the plane was allowed to crash into the ground or cliff. Upon a crash the system would pause the screen. There were 90 trials total in the complex piloting task. There were 60 perturbation trials and 30 no-perturbation trials. The order was randomly determined. Each session was approximately 14 min. Trials in which the plane crashed before the onset of the perturbation were removed from the analysis.

Functional MRI

Our goal to develop a classifier of operator intention to undertake a rapid action to avoid a perturbation was to use a neuroimaging method with high temporal resolution, such as EEG or MEG. We used MEG in the current study, but in order to bolster our ability to localize MEG activity to intracortical sources, we also conducted an fMRI study of the same piloting tasks in order to establish seeds for conducting source localization analyses of MEG data. In the fMRI experiment participants underwent two sessions of the simple piloting task. Visual presentation of the flight simulation was projected by mirrors to a screen behind the head coil that could be viewed by the participant by a mirror mounted on the head coil. An fMRI compatible control stick (NATA technologies) was used by the right hand of the participant to control the elevator (back = pitch up; forward = pitch down) and aileron (roll left and right) deflections. Trigger timing of the fMRI scanning was directly read into one of the flight parameters of the flight simulator by means of a National Instruments Hi Speed USB NI USB-9162 BNC analog to digital converter.

A Siemens Verio 3T scanner was used to obtain functional $T2^*$ weighted images with a gradient echo-planar imaging sequence (echo time 30 ms; repetition time 2500 ms; flip angle 80°). A total of 40 interleaved axial slices were acquired with a 4 × 4 × 4 mm voxel resolution covering the cortex and cerebellum. A single run consisted of approximately 340 scans. (The number varied depending on the randomized time and how long the participant took to make a button response to start the trial). The first three scans were discarded. Structural T2 images, later used for normalization, were also collected using the same axial slices as the functional images with a 1 × 1 × 4 mm resolution.

Images were preprocessed using SPM8 (Wellcome Department of Cognitive Neurology, UCL). Echo planar images EPI were unwarped and realigned. The T2 image was co-registered to the mean EPI image. The T2 images were acquired during the same fMRI session as the EPI images with the same slice thickness. Since the head was in approximately the same position it is thought that this will facilitate coregistration. The EPI images were then spatially normalized to MNI space ($3 \times 3 \times 3$ mm voxels) using a template T2 image and the co-registered T2 image as the source. Normalization was done using the T2 image rather than EPI because we believe it gives better results due to better spatial resolution. The images were smoothed using an $8 \times 8 \times 8$ mm FWHM Gaussian kernel. Regional brain activity was assessed using a general linear model employing an event-related analysis in which the onset times were convolved with a hemodynamic response function. High pass filtering (cutoff period 128 s) was carried out to reduce the effects of extraneous variables (scanner drift, low frequency noise, etc.). Auto-regression was used to correct for serial correlations. The six movement parameters were used as regressors of non-interest in the analysis to account for biases in head movement correlations present during the experimental conditions. Anatomical T1 weighted images were acquired with a $1 \times 1 \times 1$ voxel resolution for use in constructing source models for localizing brain activity recorded by MEG.

MEG

In the MEG experiment participants underwent three sessions of the simple piloting task and one session of the complex piloting task. The first two sessions of the simple piloting task were used for training the decoding algorithm. The third session of the simple piloting task was used to evaluate the effectiveness of the trained algorithm in decoding neural states when participants perform the same task. As discussed previously, however, an effective classifier should be able to decode not only neural states on the same task that it has been trained on, but on more complex versions of the task that the classifier has not been trained on—that is, whether the classifier can achieve transfer. Accordingly, we also assessed the effectiveness of the classifier in decoding neural activity preceding detection and response to a perturbation in the complex piloting task. Visual presentation of the flight simulation was projected to a mirror to a screen above the participant's head. An fMRI compatible control stick (NATA technologies) was used by the right hand of the participant to control the elevator (back = pitch up; forward = pitch down) and aileron (roll left and right) deflections. Trigger timing for the start of each trial and the start of the perturbation was registered by a photodiode placed on the screen. A small white square was constantly presented on the lower center part of the screen (out of the view of the participant) at the start of each trial and at the onset of the perturbation the small square turned black for 20 ms. The light intensity change was detected by the photo diode and written directly to one of the extra channels on the MEG.

The data was recorded using a Yokogawa 400 channel MEG supine position system. Head movement was restrained by using a strap across the forehead. A sampling rate of 1000 Hz was used with input gain of $\times 5$ and an output gain of $\times 100$. The trials were segmented 1000 ms before and after the onset of the perturbation. For trials with no perturbation the timing of the virtual perturbation was given by the photodiode and used as the onset point for segmentation. The data were down sampled from 1000 to 250 Hz and filtered using a causal Butterworth online bandpass filter from 2 to 100 Hz. Only bad trials in which there was a machine failure in the flight simulator causing the plane to verge from a straight and level course (for the simple piloting task) or bad trials in which the plane crashed before the onset of the perturbation (for the complex piloting task) were removed from the data. Besides bandpass filtering there was no manual or automated artifact cleaning of the data prior to independent component analysis. Infomax independent component analysis (EEGLAB, Delorme and Makeig, 2004) with principal component analysis PCA reduction to 64 components was conducted on the first two sessions of the simple piloting task (processing time was approximately 7 min). ICA has been shown to be well suited for separation of artifact and task related components (Delorme et al., 2007). The weights derived from the ICA were used to calculate component activation waveforms for the trials in sessions one and two. They were also used to calculate component activation waveforms for the trials in sessions three of the simple piloting task and the session of the complex piloting task. There were two reasons that the weights from the first two sessions were used to calculate the activation waveforms of the later sessions: first, we did not want to bias the classification results of the BCI-decoder used for training by including the test data of the later sessions; and second, we wanted to simulate conditions required to run the BCI-decoder as if we were running it online in real-time. The independent components showing evoked responses to the averaged perturbation piloting trials were considered for training of the BCI-decoders. Each participant had one evoked potential related component with an ICA spatial filter showing a prominent sinc and source (See **Figures 3A,B**). All preprocessing steps described above were automated except for the selection of the independent components showing evoked responses, which was done by visual inspection of the averaged activation waveform and the ICA spatial filter. This step can also be automated if desired.

MEG Source Localization Analysis

Source localization analysis involves the following steps: (1) Determining the position of the head (brain) within the MEG device, (2) Segmentation of the cortex of the brain, (3) Estimation of the leadfield model on the vertex points of the segmented cortex, and (4) Current source estimation on the cortex.

1. Five coils attached to the participant's head (one behind each ear, and three across the forehead) were used to determine the position of the head within the MEG. The positions of the five coils on the participant's head were measured by the Polhemus FastSCAN Cobra system. This system allows for a 3D laser scan of the face as well as the coordinate location for the five markers to be obtained. Matlab software (part of the VBMEG toolbox) was used to register the coordinate space of the 3D face image to the participant's anatomical T1 MRI structural image. Once the position of the five coils in reference to the

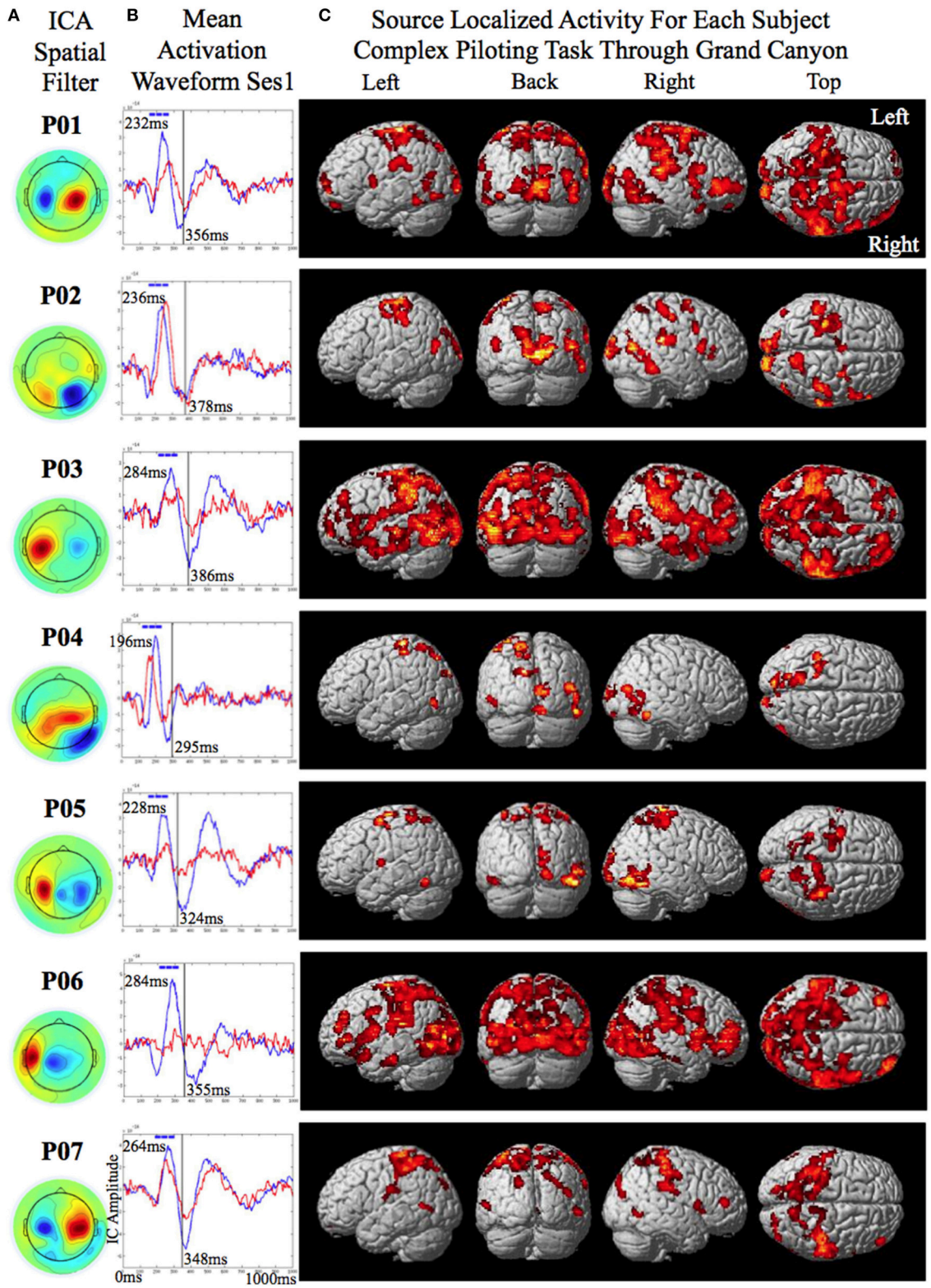


FIGURE 3 | Source localized activity for each participant (P01–P07 denotes participant identification number). (A) On the left the independent component analysis spatial filters for the MEG channels are shown for each participant. **(B)** The mean activation waveform for session one with the peak latency given in the upper corner for each participant. The blue boxes over the peak denote the three 40 ms windows the decoder was trained on. The mean response time is denoted by the gray line in the plot. The corresponding value is shown to the bottom right of this line for each participant. **(C)** The estimated current using variational Baysian multimodal encephalography VBMEG is shown rendered on the surface of the brain for each participant.

MEG sensors are known the position of these sensors can be registered in the coordinate space of the participant's T1 MRI structural image.
2. The cortex was segmented from the brain using FreeSurfer software (Fischl et al., 1999). This software was also used to make an inflated model of the cortex used for display.
3. The segmented vertex points of the cortex were used to estimate the leadfield matrix using the Sarvas equation (Sarvas, 1987).
4. Current source on the surface of the cortex was estimated using variational Bayesian multimodal encephalography (VBMEG) that uses fMRI information as a prior constraint in the analysis. See the articles by Sato et al. (2004) and Yoshioka et al. (2008) for a detailed description as well as the tested accuracy of source localization for VBMEG.

The VBMEG analysis used the fMRI *t*-values of the contrast of the perturbation piloting condition vs. the no perturbation piloting condition on the simple piloting task. The results of the SPM analysis for the contrast for each participant (using a threshold of $p < 0.05$ uncorrected, with a spatial extent of 50 voxels, and masking out activity in the cerebellum and subcortical areas) were projected onto the brain of their own T1 image using their individual specific normalization parameters. For one of the participants for which fMRI data was not collected the random effects analysis across all of the participants for the same contrast (using a threshold of $p < 0.05$ uncorrected, with a spatial extent of 50 voxels, and masking out activity in the cerebellum and subcortical areas) was used as a prior and projected back to the individuals T1 image using their specific normalization parameters. The fMRI *T*-values were then mapped to the vertex points of the segmented brain serving as prior information for the VBMEG analysis. A lenient uncorrected threshold of $p < 0.05$ was used to ensure that sufficient vertex points of the brain were given prior information for the VBMEG analysis. Using a conservative threshold corrected for multiple comparisons for the fMRI analyses may considerably restrict the extent of prior information for the VBMEG analysis.

The activation waveforms of the trials from all conditions and sessions for both the simple and complex tasks were projected to the MEG sensor space (400 channel) using the weights of the independent components as determined from the ICA on the trials from the first two sessions of the simple piloting task. The mean activity of the trials for each condition was used in the VBMEG analysis. The noise model, serving as a baseline, was calculated using the activity from the no perturbation passive viewing condition. The VBMEG analysis estimated current activity over the entire cortex using a variance magnification factor = 500 and a confidence parameter = 500 [these parameters are such that they give somewhat less weight to the fMRI prior activity in determining the location of the source activity Sato et al. (2004) and Yoshioka et al. (2008)]. The time period for current estimation was 250 samples and the time step for the next period was 100 samples. The output of the analysis was the mean estimated current across trials for each cortical vertex point for each condition.

To determine the location of current on the brain thought to underlie the response to the perturbation and to be able to compare the results across participants the following procedure was used for each participant using data from the complex piloting task: For each of the vertex points (there were ∼6000 for each participant), the root-mean-squared RMS current was determined for perturbation recovery and for a baseline period prior to perturbation: The RMS current for perturbation recovery was calculated from 12 ms before and 8 ms after the new mean response time (utilizing performance of the adaptive automation BCI-decoder—see Results section). The RMS current for the baseline period was calculated from 400 ms before onset of the perturbation to just before the onset of the perturbation. The current for each vertex point was normalized by subtracting 20 times the mean RMS current of the baseline period (across all vertex points) from the RMS current of perturbation recovery for each vertex point and then dividing by the maximum RMS current across all vertex points. The normalized current of the vertex points that were greater than zero were projected to the standard template brain ($2 \times 2 \times 2$ mm) (given in SPM8) using the MNI coordinates determined during segmentation by Free Surfer. The resulting images were smoothed using a FWHM $8 \times 8 \times 8$ mm Gaussian kernel. Because smoothing may cause activity to be spread to regions that were not originally active a threshold was used such that only voxels showing mean RMS values greater than the lowest value of the original smoothed voxels (corresponding to the original projected vertex points) were considered to be significant (using a spatial extent threshold of 100 voxels). The intersection of active voxels across all seven participants was used to define common activity. The SPM Anatomy Toolbox v1.8 (Eickoff et al., 2005) was used to determine the labels of active brain regions.

Decoding Pilot Intention

We developed a method to decode the participant's intention to perform an action in response to a perturbation by training a classifier on neural data taken from the first two sessions of the simple piloting task. The classifier was then evaluated by testing its ability to decode participant intention on the third session of the same task. As a more stringent test of classifier performance—an examination of its transfer generalizability—we then examined its ability to decode intention in the complex piloting task. It should be noted that this classifier represents an open loop simulation of a BCI in order to test the feasibility of such a method for real-time implementation of neuroadaptive automation using a closed loop BCI-decoder. See **Figure 1** for a depiction of the hypothesized neuroadaptive automation system implemented in this study.

The training of the classifiers was conducted using trials from the first two sessions of the simple piloting task. The two classes to be decoded were presence of perturbation while piloting the plane vs. presence of perturbation while passively viewing. The reason for selecting these contrasts to train on was because we wanted to ensure that the BCI-decoder was not just picking up the visual evoked response induced by the perturbation but was extracting activity related to the attentional components of

the response to the perturbation in relation to the intention to recover from the change in attitude. Rather simple features were used for decoding in the hope that they would generalize across sessions, tasks, and participants. The first step in calculating the features used for decoding was to determine the time point of the absolute peak in the mean evoked response (that was less than 300 ms) to the onset of the perturbation in the selected independent component of the perturbation piloting trials of the first session of the simple task used for training (the peak time for the participants was as follows: S1 = 232 ms; S2 = 236 ms; S3 = 284 ms; S4 = 196 ms; S5 = 228 ms; S6 = 284 ms; S7 = 264 ms; mean = 246 ms) (See **Figure 3B** for the mean activation waveforms for the Fly and Watch conditions for each participant from session 1). For each trial RMS amplitude was calculated within two consecutive 40 ms windows prior to the time of the peak of the averaged evoked potential and one 40 ms window after (These windows are depicted as blue bars at the top of the mean activation waveforms in **Figure 3B** for each participant). The RMS amplitude values in these three windows served as the features for decoding for the perturbation piloting trials. To help in generalization and to bias the classifier to make fewer false alarms the perturbation passive trials used three separate time points to extract the features (120, 60, and 0 ms before the onset of the perturbation). This had the effect of increasing the number of training trials for the perturbation passive condition by three. Since there were originally half as many perturbation passive trials than perturbation piloting trials this increased the training ratio to be about 1.5 times as many trials for the perturbation passive condition to that of the perturbation piloting condition. The greater number of training trials and the greater variability for the perturbation passive condition is used to increase the ability to reject trials that are not from the perturbation piloting condition (reduce false-alarms) and increase the noise variability with regards to timing such that the classifier may more readily generalize to the complex task in which the attitude of the plane (and thus the visual image on the screen) is constantly changing as a result of the continuous piloting task of tracking the river in the Grand Canyon. The reason we did not use the no perturbation piloting task as one of the conditions to train the BCI-decoder on is that it would likely just extract the visual evoked response to the perturbation piloting task and not extract the attention related component of the motor intention for attitude recovery that we are interested in determining. Given the continuous changes in attitude of the plane while maneuvering on the complex task a BCI-decoder that is based on visual evoked perturbations from the simple task may result in a large number of false alarms.

The BCI-decoder was trained on approximately 80 trials of the perturbation piloting task and 120 trials of the perturbation passive task using the Matlab Least Squares Probabilistic Classification (LSPC) toolbox (Sugiyama et al., 2010). LSPC uses a linear combination of kernel function to model the class-posterior probability. Regularized least-squares fitting of the true class-posterior probability is used to learn its parameters (Sugiyama et al., 2010). The use of least-squares fitting to determine a linear model allows for a global solution to be made analytically providing a considerable speedup in computational time. The default parameters were used in training of the LSPC models (see Matlab code: Sugiyama et al., 2010). The time required to train the classifier is approximately 0.25 s. Prior to training the features for the trials were normalized by subtracting the mean and dividing by the standard deviation. The mean and standard deviation from the training trials were used to normalize the trials used for testing. The first test set consisted of trials from the session of the simple piloting task that was not used during training. There were approximately 40 perturbation piloting trials and 20 perturbation passive trials to be classified using the train LSPC model. Balanced accuracies (Brodersen et al., 2010) are reported to account for biases in unequal number of trials in the two conditions to be classified. The test data consisted of features computed at the time point specified by the peak of the evoked response determined from the training data. No information about the distribution of the test data was used. The BCI-decoder treats each test trial as independent. One hundred BCI-decoders were trained and then tested using trials from the simple piloting task. The primary parameter that is random for training of the model for each BCI-decoder is the order of the trials in the training cross validation steps. The BCI-decoder with the best performance as determined by balanced accuracy was used to test the trials from the complex piloting task.

The goal for the BCI-decoder in the complex piloting task was to be able to detect the intention to recover from a perturbation in attitude faster than by movement of the control stick by the hand. The selected LSPC model trained on the simple piloting task was used for testing of the complex piloting task. Additionally the same parameters (mean and standard deviation) used during training were also used on the test set for normalization of the features. For the perturbation piloting trials and the no perturbation piloting trials the LSPC model began testing at time point zero when the perturbation started. The window for the BCI-decoder was 120 ms encompassing the three 40 ms time windows in which the RMS amplitude was calculated. Therefore, the earliest time the perturbation could be detected was at 120 ms. The 120 ms time window tested by the BCI-decoder was incremented in 8 ms steps through 1000 ms of the data for each trial. The earliest point at which the BCI-decoder detected the presence of a perturbation piloting trial was the point at which the adaptive automation would be implemented to recover attitude. The time between detection by the BCI-decoder and the onset of the control stick by the hand to recover from the perturbation in attitude was used to evaluate the time benefit (time savings) of the implementation of the adaptive automation. The trial was only considered a hit if the BCI-decoder predicted time was faster than the actual movement time of the control stick. To determine the statistical significance of the BCI time savings, BCI-decoders were trained using 1000 random permutations of the labels and each was tested on the complex piloting task. All 1000 permuted models used for evaluation had less than 25% false positives. This criterion was used in order to be comparable to the false positive performance of that of the BCI models trained with correct labels. The perturbation time benefit were calculated for each of the 1000 permuted models and used as a distribution to compare against the model trained with the actual labels.

In order to evaluate the generalizability of a single model across participants the weights of the model of the participant with the best performance were used to test the trials of the complex flying task of the remaining six participants. Performance measures including BCI time savings were determined using the same method as applied when using each participants corresponding model to test the trials of the complex piloting task (see above).

Procedure

Participants underwent the fMRI and MEG sessions on separate days. The fMRI experiment was used to calculate a prior for the MEG source localization analysis using Variational Bayes Multimodel Encephalography (VBMEG). Six of the seven participants participated in the fMRI experiment. One participant only did the MEG experiment. All of the participants that participated in the fMRI experiment did it prior to the MEG experiment. MRI anatomical T1 scans were acquired for all seven participants and used to make models for source localization analysis using VBMEG. All analyses were conducted using Matlab software unless otherwise stated.

RESULTS

Behavioral Performance

The response times (RTs) for each of the participants to initiate pull back on the control stick in reaction to a perturbation causing the plane to enter a steep dive for both the simple piloting task and the complex piloting task are presented in **Table 1**. The RT in the complex task (median = 436.5) was not found to be significantly higher than in the simple task (median = 368.5), $p = 0.0781$ (df = 6). However, there is a tendency in this direction. The number of trials the plane crashed into the ground/cliff, as well as the number of bad trials (resulting from machine failures and/or movement before the onset of the perturbation for the simple task and crashes before the onset of the perturbation for the complex task) are also presented in **Table 1**. As **Table 1** indicates, these numbers were relatively small, but were greater in the complex task. It should be noted that bad trials were excluded from analysis and not used for calculation of response times or training/testing of the BCI-decoders. In some cases on the complex task there were crashes after the onset of the perturbation. These trials were not excluded from analysis.

Source Localization

The smoothed RMS current centered around the time of the perturbation on the complex piloting task of the activation waveform of the projected task related independent component rendered on the surface of the brain (See Methods for details of source localization analysis) is displayed for each participant along with the corresponding independent component spatial map in **Figure 3C**. There was some degree of variability in the extent to which different brain regions were active across participants (**Figure 3C**, **Table 2**). As can be seen in **Figure 4** and **Table 3** brain regions that were commonly active for all participants include the pre-central gyrus (including premotor cortex), post- central gyrus, the superior parietal lobule, the primary visual cortex, and human occipital cortex visual motion processing area V5 (hOC5). It should be noted that while source localization is interesting in determining the brain regions associated with the independent component upon which decoding is made it is not a necessary step in implementation of the proposed neuroadaptive automation brain-machine interface.

BCI-Decoder Performance

The results of the performance of the BCI-decoder are presented in **Tables 4–9**. The performance of the best (as determined by the highest balanced accuracy score) out of 100 BCI-decoders tested on the novel sessions of the simple piloting task is presented in **Table 4** for each participant. The average over all 100 BCI-decoders is given in the table for comparison. The BCI-decoder for six of the seven participants showed significant differences ($p < 0.05$) in being able to distinguish between perturbation piloting trials and perturbation passive viewing trials. The mean balanced accuracy performance was approximately 70%. Certainly the selection of the best BCI-decoder out of 100 trained biases these results, however, it was our goal to find the model that may best extract attentional information related to the intention of recovering from the perturbation in attitude. In this respect

TABLE 1 | Response time to pull back on control stick after start of perturbation for the two training sessions and two test sessions.

ID	RT train Ses1 (ms)	BT train Ses1	RT train Ses2 (ms)	BT train Ses2	RT test simple (ms)	BT test simple	RT test complex (ms)	BT test complex	CT test complex
1	356.4	1	387.7	2	370.4	1	357.9	0	0
2	377.7	0	427.0	2	452.6	3	436.5	8	9
3	386.5	2	371.4	0	384.4	0	454	2	6
4	295.1	0	303.7	0	305.5	0	359.8	2	6
5	323.9	0	336.9	0	342.3	0	462.2	1	1
6	355.0	0	386.3	9	368.5	3	480.4	5	0
7	348.4	0	347.1	0	337.8	0	424.1	4	9
Group mean	349.0	0.43	365.7	1.86	365.9	1.0	425.0	3.14	10.33

ID, Participant identification number; RT, Response Time; Ses, Session; BT, Bad Trial; CT, Crash Trial.

TABLE 2 | MNI coordinates of clusters of brain activity for each participant.

Brain region	P1 MNI	P2 MNI	P3 MNI	P4 MNI	P5 MNI	P6 MNI	P7 MNI
Orbital Gyrus	−22,58, −4 −12,54, −16		−34,56, −6 48,36, −10			26,45, −13	−6,56, −14 22,46, −12
SFG, MFG			−14,44,32 −32,40,36			−38,40,30 42,51,7	
IFG BA44	−52,4,10 56,16,22	54,14,10 56,8,34	−49,5,6			−50,10,6 58, −1,7	54,18, −2
IFG BA45	45,44,2	58,30,0	−50,25, −1			−48,36,4 56,26,2	48,44,12
SFG SMA BA6	1, −16,66	−4,0,58	−4, −3,68 11, −13,67				
PMC BA6	−35, −25,68 46, −8,48	−34, −23,68 28, −8,64	−30, −27,68	−36, −28,68	−38, −8,64 22, −31,76	−26, −21,70 58,4,31	−33, −25,67
Pre−CG BA4	−48, −16,56 42, −10,46	−35, −25,58	−35, −25,54 12, −27,74	−32, −28,69	−42, −18,54 8, −34,76 38, −30,61	−9, −42,75 −28, −27,59 9, −38,75	−30, −38,70
Post−CG BA1,2,3	−18, −36,76 −48, −30,58	−24, −36,72 −52, −32,54 46, −32,60	−40, −41,61 44, −22,60	−30, −34,70 −44, −34,58	−30, −44,64 −62, −8,10 38, −32,60	−35, −35,67 −47, −26,55 53, −24,55 62, −24,26	−40, −44,58 56, −20,46 44, −30,60
SPL	−10, −76,52 −22, −82,48 12, −56,68 26, −50,66	10, −70,56 24, −56,56	−38, −48,57 −8, −65,58 9, −64,61 41, −45,58	−14, −54,66 −10, −90,34 16, −68,56	−24, −60,62 20, −66,60	−16, −74,48 18, −66,48	−16, −70,62 −16, −86,38 13, −66,62
IPC	−54, −70,16 −56, −24,30 60, −22,34	−46, −74,18 46, −70,14 68, −30,20	−52, −44,38 48, −43,35 58, −30,40			−58, −39,35 57, −34,40	−58, −18,28 60, −30,30 46, −82,20
hOC5 (V5) MT, IOG	52, −67, −1	−44, −74,18 56, −62,4	−40, −70, −2 54, −67,13	−52, −72, −2 50, −70,0	52, −62, −16	−44, −72,5 53, −69,1	−52, −70,0 56, −64, −2
hOC4	−38, −72, −12		32, −71, −1		46, −80, −15	−32, −77,0 40, −72, −12	
Cuneus Calcarine Gyrus BA17,18	−12, −104,4 −24, −102, −6 20, −98,22 14, −98,0 16, −100, −14	−4, −72,2 −10, −94,26 −10, −100,8 10, −70,2 16, −96,20	−8, −68,4 −8, −94, −8 18, −92, −14	−2, −80, −2 8, −82,10 12, −102,4	16, −94,16 20, −98, −10	−10, −98,2 12, −90,0 6, −86,16	−4, −72,0 10, −70,0
MTG	−56, −64, −2		−48, −24, −10 −56, −59,8 60, −26, −16 54, −63, −2		55, −66, −2		
ITG	−54, −50, −18 48, −50, −12	60, −48, −12	−46, −60, −10 −52, −32, −20 −58, −24, −20	54, −54, −20	50, −56, −16 −44, −62, −16	−44, −63,0 −56, −20, −24 60, −46,0	
Temporal Pole			−50,6, −15			−48,12, −20 58, −2, −6	

P, participant identification number; IFG, Inferior Frontal Gyrus; SFG, Superior Frontal Gyrus; SMA, Supplementary Motor Area; PMC, Premotor Cortex; Pre-CG, Pre Central Gyrus; Post-CG, Post Central Gyrus; SPL, Superior Parietal Lobule; IPC, Inferior Parietal Cortex; hOC5 (V5), Human Occipital Cortex Visual motion processing area V5; MT, Middle Temporal Cortex overlaps area hOC5; IOG, Inferior occipital gyrus; MTG, Middle Temporal Gyrus; ITG, Inferior temporal gyrus. MNI coordinates of Clusters of root-mean-squared RMS current 12 ms before and after the mean time in which the decoder detected motor intention to the presence of a perturbation. The threshold of significant RMS current activity at a specific vertex point was set at 20x the mean baseline RMS current from −400 to 0 ms across all vertex points. A spatial extent threshold of 100 voxels was used on the smoothed projection into MNI space.

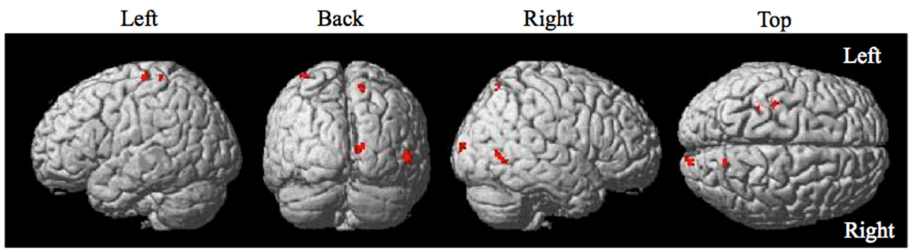


FIGURE 4 | Source localized activity common to all participants. Activity is present in the left pre- central gyrus, the left post central gyrus, the right superior parietal lobule, the right primary visual cortex V1, and the right visual motion processing area V5.

TABLE 3 | MNI coordinates of clusters of brain activity common across all participants.

Brain region	All seven participants MNI
Pre-CG, PMC BA4,6	−36,−26,68
Post-CG BA1,2,3	−30,−38,68
SPL 7P 7A Precuneus	12 −68,60
Cuneus (V1) BA17	16,−96,10
hOC5(V5)	52,−62,0

Pre-CG, Pre Central Gyrus; Post-CG, Post Central Gyrus; SPL, Superior Parietal Lobule; hOC5 (V5), Human Occipital Cortex Visual motion processing area V5.
MNI coordinates of Clusters of root-mean-squared RMS current 12 ms before and after the mean time in which the decoder detected motor intention to the presence of a perturbation that are common across all seven participants. The threshold of significant RMS current activity at a specific vertex point was set at 20x the mean baseline RMS current from −400 to 0 ms across all vertex points.

we feel justified in selecting the best model trained and tested on the simple piloting task to use for testing in an unbiased manner on the complex piloting task. Although training the BCI-decoder to distinguish between the perturbation piloting and no perturbation piloting trials on the simple piloting task may have provided better performance when testing on the novel session from the same task it is likely that the model would have learned the response to the visual aspects of the perturbation rather than the neural activity related to the intention to recover attitude.

As discussed above the model with the highest balanced accuracy on the test session of the simple piloting task was used to test the session of the complex piloting task. The goal was to simulate the use of a brain computer interface in real time that would initiate the use of adaptive automation to initiate recovery from a perturbation in attitude faster than could be done by moving the control stick by the hand. In accomplishing this goal the BCI-decoder was used on a 120 ms window starting at the time of the perturbation and stepping through the data in 8 ms steps. The BCI-decoder was also tested on trials in which there was no perturbation within the same time region in which the perturbation may have occurred. This point was determined randomly during the experiment and triggered on the MEG trace using a photodiode (see Methods). Bad trials were eliminated from the analysis (see **Table 1**). The first instance of the classification by the BCI-decoder as a perturbation piloting trial is the time point at which the adaptive automation is initiated. Only trials in which the BCI-decoder is faster than the movement of the controls stick are counted as hits (true positives). The results of the classification performance for the complex piloting task are presented in **Table 5**. Because unequal number of trials existed for perturbation piloting and no perturbation piloting trials balanced accuracies were used (Brodersen et al., 2010). All seven participants showed significant classification performance above chance even with the additional criteria that the detection of a perturbation piloting trial had to be *before* movement of the control stick. In these cases where there was classification of a perturbation trial after control stick movement, the trial was counted as a miss (false negative). The ratio of correct rejections (true negatives) to false alarms (false positives) was greater than the ratio of hits (true positives) to misses (false negatives). The mean balanced accuracy across participants was 73%. **Table 6** shows the performance on the complex piloting task of the six subjects tested using the weights from the model of the best participant. The results indicate that the balanced accuracies of all six participants showed significant classification performance above chance (**Table 6**). While there were significant differences in hit rate and false alarm rate between the generalized and own model tests there were no significant differences, using the Wilcoxon signed rank test, in the primary performance measures including balanced accuracy, d′, and a′ (a′ a prime or area under the curve was calculated by the method given in Macmillan and Creelman (1991) (See **Table 3B**).

The improvement in response time afforded by the use of the neuroadaptive automation is given in **Table 7**. In trials in which there was a miss, neuroadaptive automation was not employed and the original response time was used. The mean response time difference was calculated from the original onset time minus the onset of the neuroadaptive automation for all perturbation piloting trials. The mean improvement in response time across participants was a reduction from 425 to 353 ms under neuroadaptive automation, or an average of 72 ms improvement. **Figures 5A,B** depicts the decoded response times plotted on the single trial activation waveforms of the adaptive automation (black circles) for participant 1 (best performer) and 3 (median performer), respectively. The single trials are arranged in increasing order of behavioral response time from bottom to top (white line). The significance of the time savings was

TABLE 4 | Novel test session classification performance: simple piloting task over ocean: detect perturbation piloting vs. perturbation passively watch.

ID	bacc_mean	bacc_ppi1	bacc_ppi2	bacc_p	tp	fn	fp	tn	hr	far	d′	a′
1	61.5 (58.4)	48.6	73.5	0.039	21	16	6	13	0.57	0.32	0.65	0.70
2	58 (51.2)	45.9	69.3	0.093	15	21	5	16	0.42	0.24	0.50	0.67
3	73.1 (72.4)	61.1	83.0	0.023	26	15	2	14	0.63	0.13	1.49	0.85
4	73.8 (67.1)	62.2	83.7	0.009	27	13	3	16	0.68	0.16	1.46	0.85
5	76.3 (75.0)	67.3	84.4	0.0002	23	17	0	19	0.58	0.05	1.86	0.89
6	81.1 (78.8)	70.9	89.4	0.0001	27	10	1	18	0.73	0.05	2.23	0.91
7	64.8 (62.6)	52.9	75.6	0.009	21	19	4	16	0.53	0.20	0.90	0.76
Group mean	69.8 (66.5)	58.4	79.8	0.016	23	16	3	16	0.59	0.16	1.30	0.80

ID, Participant identification number; bacc_mean, Balanced accuracy mean in percent; bacc_ppi, Posterior probability intervals; bacc_p, p value; TP, true positive (hit); FN, false negative (miss); FP, false positive (false alarm); TN, true negative (correct rejection); HR, hit rate; FAR, false alarm rate; d′, d prime; a′, a prime.
In the case when the FAR = 0 calculation of FAR is made by adding 1 to the original FP and TN values.
The performance scores are for the best out of 100 BCI-decoders trained on the first two sessions and tested on the novel simple piloting task session. The average balanced accuracy for all 100 BCI-decoders is given in parentheses for comparison.
The bacc_p value for the group mean is the Wilcoxon signed rank test over the bacc_mean values for the 7 subjects that the values are greater than 50.

TABLE 5 | Novel test session classification performance: complex piloting task through Grand Canyon: detect perturbation piloting vs. no perturbation piloting.

ID	bacc_mean	bacc_ppi1	bacc_ppi2	bacc_p	tp	fn	fp	tn	hr	far	d′	a′
1	85.6	78.3	91.5	0.0001	47	13	1	29	0.78	0.03	2.62	0.93
2	66.6	58.1	74.6	0.02	22	30	2	28	0.42	0.07	1.31	0.81
3	74.7	65.5	82.7	0.0001	38	20	4	26	0.66	0.13	1.51	0.85
4	76.7	67.1	85.1	0.0001	46	12	7	23	0.79	0.23	1.55	0.86
5	66.1	56.3	75.1	0.001	32	27	6	24	0.54	0.20	0.95	0.76
6	79.4	70.1	87.4	0.0001	45	10	6	24	0.82	0.20	1.75	0.88
7	63.7	54.5	72.2	0.003	24	32	4	26	0.43	0.13	0.93	0.76
Group mean	73.3	64.3	81.2	0.016	36	21	4	26	0.63	0.14	1.52	0.84

ID, Participant identification number; bacc_mean, Balanced accuracy mean in percent; bacc_ppi, Posterior probability intervals; bacc_p, p value; TP, true positive (hit); FN, false negative (miss); FP, false positive (false alarm); TN, true negative (correct rejection); HR, hit rate; FAR, false alarm rate; d′, d prime; a′, a prime.
The bacc_p value for the group mean is the Wilcoxon signed rank test over the bacc_mean values for the seven participants that the values are greater than 50.

TABLE 6 | Generalization of performance using best subjects weights: novel test session classification performance: complex piloting task through Grand Canyon: detect perturbation piloting vs. no perturbation piloting.

ID	bacc_mean	bacc_ppi1	bacc_ppi2	bacc_p	tp	fn	fp	tn	hr	far	d′	a′
1	–	–	–	–	–	–	–	–	–	–	–	–
2	57.7	51.9	63.8	0.007	9	43	0	30	0.17	0.03	0.91	0.79
3	73.8	64.9	81.6	0.0001	35	23	3	27	0.60	0.10	1.54	0.85
4	68.6	60.3	76.3	0.0001	27	31	2	28	0.47	0.07	1.41	0.82
5	63.4	54.8	71.4	0.002	23	36	3	27	0.39	0.01	1.00	0.77
6	80.8	71.9	88.3	0.0001	43	12	4	26	0.78	0.13	1.89	0.89
7	64.3	56.0	72.0	0.0007	21	35	2	28	0.38	0.07	1.18	0.79
Group mean	68.1 (71.2)	60.0 (61.9)	75.6 (79.5)	0.032 (0.032)	26* (35)	30* (21)	2* (5)	28* (25)	0.47* (0.61)	0.07* (0.16)	1.32 (1.33)	0.82 (0.82)

ID, Participant identification number; bacc_mean, Balanced accuracy mean in percent; bacc_ppi, Posterior probability intervals; bacc_p, p value; TP, true positive (hit); FN, false negative (miss); FP, false positive (false alarm); TN, true negative (correct rejection); HR, hit rate; FAR, false alarm rate; d′, d prime; a′, a prime.
The bacc_p value for the group mean is the Wilcoxon signed rank test over the bacc_mean values for the six subjects that the values are greater than 50.
*The number in parentheses are the group mean values of the original decoder excluding sub01. *Denotes $p < 0.05$ on paired Wilcoxon signed rank test for the comparison between the original decoder and the one trained with sub01 model over the six participants.*

TABLE 7 | Improvement in response time by adaptive automation complex piloting task through the Grand Canyon.

ID	N	TP	FP	Org Mean (ms)	Org Se (ms)	BCI mean (ms)	BCI Se (ms)	BCI rt diff mean (ms)	BCI rt diff se (ms)	Perm rt diff Mean (ms)	Perm P
1	60	47	1	357.9	5.2	283.1	9.8	74.8	7.8	5.2	0.002
2	52	22	2	436.5	13.8	369.7	15.6	66.7	14.3	6.4	0.023
3	58	38	4	454.0	10.2	405.2	13.6	48.7	7.9	4.5	0.014
4	58	46	7	359.8	7.3	277.0	10.6	82.8	10.4	4.9	0.001
5	59	32	6	462.2	6.6	404.2	12.7	58.0	11.8	3.1	0.003
6	55	45	6	480.4	7.3	341.4	14.1	138.9	15.4	13.5	0.002
7	56	24	4	424.1	9.3	388.0	13.3	36.1	8.6	7.4	0.041
Group mean	59	36	4	425.0	8.5	352.7	12.8	72.3	10.9	6.4	0.016

The p-value in the last column denotes the significance of the time savings improvement of the BCI adaptive automation over the original joystick based response times based on permutation testing of 1000 models trained with random labels.

ID, Participant identification number; N, Number of Perturbation Piloting Trials; TP, True Positives (hits); FP, False Positives (false alarms); rt diff, response time difference; Org, Original; se, standard error; Perm, Permuted; BCI Brain Computer Interface.

The Perm P-value for the group mean is the paired Wilcoxon signed rank test comparing the BCI rt diff values to the Perm rt diff values for the seven participants.

TABLE 8 | Generalization of performance using best subjects weights: improvement in response time by adaptive automation complex piloting task through the Grand Canyon.

ID	N	TP	FP	Org Mean (ms)	Org Se (ms)	BCI mean (ms)	BCI Se (ms)	BCI rt diff mean (ms)	BCI rt diff se (ms)	Perm rt diff Mean (ms)	Perm P
1	–	–	–	–	–	–	–	–	–	–	–
2	52	9	0	436.5	13.8	408.7	15.6	27.8	10.7	7.9	0.026
3	58	35	3	454.0	10.2	407.3	13.5	46.7	7.7	5.5	0.009
4	58	27	2	359.8	7.3	327.9	9.3	31.9	7.9	4.1	0.001
5	59	23	3	462.2	6.6	427.7	11.2	34.5	9.5	2.9	0.001
6	55	43	4	480.4	7.3	347.9	15.8	132.5	15.7	17.8	0.002
7	56	21	2	424.1	9.3	388.7	11.8	35.4	9.2	2.8	0.006
Group mean	56	26* (35)	2* (5)	436.2	9.1	384.7* (364.3)	12.9 (13.3)	51.5* (71.9)	10.1 (11.4)	6.8 (6.3)	0.032 (0.032)

The p-value in the last column denotes the significance of the time savings improvement of the BCI adaptive automation over the original joystick based response times based on permutation testing of 1000 models trained with random labels.

ID, Participant identification number; N, Number of Perturbation Piloting Trials; TP, True Positives (hits); FP, False Positives (false alarms); rt, response time; Org, Original; se, standard error; Perm, Permuted; BCI Brain Computer Interface.

The Perm P-value for the group mean is the paired Wilcoxon signed rank test comparing the BCI rt diff values to the Perm rt diff values for the six subjects.

*The number in parentheses are the group mean values of the original decoder excluding sub01. *Denotes $p < 0.05$ on paired Wilcoxon signed rank test for the comparison between the original decoder and the one trained with sub01 model over the six participants.*

evaluated by comparing the neuroadaptive automation response time difference (to that of the control stick response) relative to the distribution of response time differences of over 1000 models trained with randomly permuted labels (See Methods Section). The p value was computed by the number of times the models with permuted labels had larger response time differences than the BCI trained with the correct labeling over the 1000 permuted models (see **Table 7**). **Table 8** shows the time savings of the six participants tested using the weights from the model of the best participant on the complex piloting task. The same permutation technique as discussed above was used to evaluate statistical significance. While all participants showed a significant difference in time savings even using a model trained by a different participant, the time savings were significantly ($p < 0.05$; paired Wilcoxon signed rank test) greater when using their own model (median = 62.4 ms; mean = 71.9 ms) vs. the generalized model (median = 35.0 ms; mean = 51.5 ms).

DISCUSSION

The present study examined whether it is possible to decode neural signals associated with the intention to act in response to an impending hazard. Using MEG, the results showed that neural activity could be decoded so as to decrease the time needed to respond to the hazard, compared to manual action. As such, the results demonstrate that neuroadaptive automation can be implemented to speed up intentional action when there is very little available to respond.

There has been extensive prior research showing the effectiveness of both neuroadaptive automation (Byrne and

TABLE 9 | Flight characteristics of F22 on Grand Canyon task.

ID	Avg altitude at time of perturbation (m)	Avg airspeed at time of perturbation (Km/h)	Avg airspeed at time of perturbation (m/s)	Descent rate DR mean (m/s)	Descent rate DR greatest (m/s)	Avg response Improvement by BCI over all trials (ms)	BCI Savings in Descent Altitude based on Mean DR (m)	BCI Savings in Descent Altitude based on Greatest DR (m)
1	81.4	1122	311.6	69.1	87.4	75	5.2	6.5
2	79.8	1082	300.7	65.4	94.7	67	4.4	6.3
3	80.2	1117	310.3	56.2	83.7	49	2.8	4.1
4	79.8	1115	309.8	79.7	103.2	83	6.6	8.6
5	82.4	1117	310.2	71.8	124.6	58	4.2	7.2
6	82	1122	311.7	75.4	111	139	10.5	15.4
7	79.9	1113	309.1	69.6	93.3	36	2.5	3.4
Group mean	80.8	1113	309.1	69.6	99.7	72	5.2	7.4

ID, Participant identification number; DR, Decent Rate; Avg, average.

The climb/descent rate is variable depending on the attitude of the plane at time of perturbation. The values given are the (1) mean of the maximum slope of descent calculated over a 200 ms period across trials and (2) the greatest maximum slope of descent calculated over a 200 ms period across trials.

It is important to note that time and distance to ground saved by earlier elevator engagement is not only the savings in less descent toward ground but also allows for gain in altitude relative to time because of earlier climb.

Parasuraman, 1996; Wilson and Russell, 2007; Ting et al., 2010) and passive BCI (Blankertz et al., 2010; Zander and Kothe, 2011) in enhancing human performance. However, the present study represents the first successful attempt to show that decoded neural activity can be used to potentially speed up split-second decision making in response to an impending hazard on a novel complex task that neither the participant or the classification model has been trained on. While the brain is indeed faster than the hand in responding to a hazard, its activity must be accurately decoded so as to accrue a savings in time. In the piloting task used in the present study, the mean savings in response time was 72 ms (ranging from 36.1 to 138.9 ms). Although this may seem relatively small, in situations where humans are moving at high speed toward a hazard, as in driving or piloting, the savings may be sufficient to avert disaster.

To put a time savings of 72 ms in context, consider the flight characteristics of a F22 aircraft on the complex piloting task. **Table 9** gives the response times to the in-flight perturbation for each participant [**Figures 5A,B** depicts the decoded response times plotted on the single trial activation waveforms of the adaptive automation (black circles) for participant 1 (best performer) and 3 (median performer), respectively]. Even with an average improvement of 72 ms in response time this can result in an average savings of up to 7.4 m of lost altitude as a result of earlier initiation of recovery in attitude to the perturbation. This could make a difference between a successful and failed attempt to avoid a collision. It should be noted that the large variability in savings between participants is likely a result of the quality of the MEG data in terms of separating task related activity from artifacts rather than expertise on the task. There was no apparent relationship between the savings afforded by the simulated neuroadaptive automation and manual response time on the task. It is known that there is considerable variability in the quality of MEG and EEG data across individuals that impacts successful BCI performance (Lotte et al., 2013).

It must be acknowledged, however, that the improvement in response time using neuroadaptive automation comes at the expense of making false alarms on a small number of trials. As in any automated alarm system, the tradeoff between correct early warning (hits) and false alarms has to be considered when setting the decision criterion of the alarm (Swets, 1973; Parasuraman and Riley, 1997). It may be possible in some cases to adjust the criteria of the BCI-decoder to make less false alarms at the expense of making less hits as well and reducing the overall response time improvement afforded by the neuroadaptive automation. For example in the study by Blankertz et al. (2002) the classifier was trained such that it was optimal under the constraint that the false positive (false alarm) rate attains a preset value.

The presence of a false alarm by the BCI-decoder could be somewhat problematic. Without some type of system that would identify externally induced perturbations from changes in attitude induced by the pilot in flight the neuroadaptive automation would initiate a recovery maneuver. Which in this case is to reverse the pitch down elevator deflection caused by the perturbation. Without the presence of a real perturbation, if the plane was in level flight and the BCI decoder made a false-alarm the neuroadaptive automation would cause the plane to make an abrupt climb. With respect to the pilot, this would constitute a pitch up perturbation. The goal of the hypothesized neuroadaptive automation is not to take control away from the pilot but rather to speed up the response of the pilot's motor intentions to unexpected flight conditions such as perturbation of attitude. While the use of detecting error-related potentials to decrease error rate has been successful in some implementations (Blankertz et al., 2002; Parra et al., 2003) it, unfortunately, is not likely to be of benefit in detecting motor intention to improve response time. This is because the relevant features for detecting the error-related potential on a single trial basis is after the response is made. One way to possibly keep the pilot in the loop and reduce the effects of

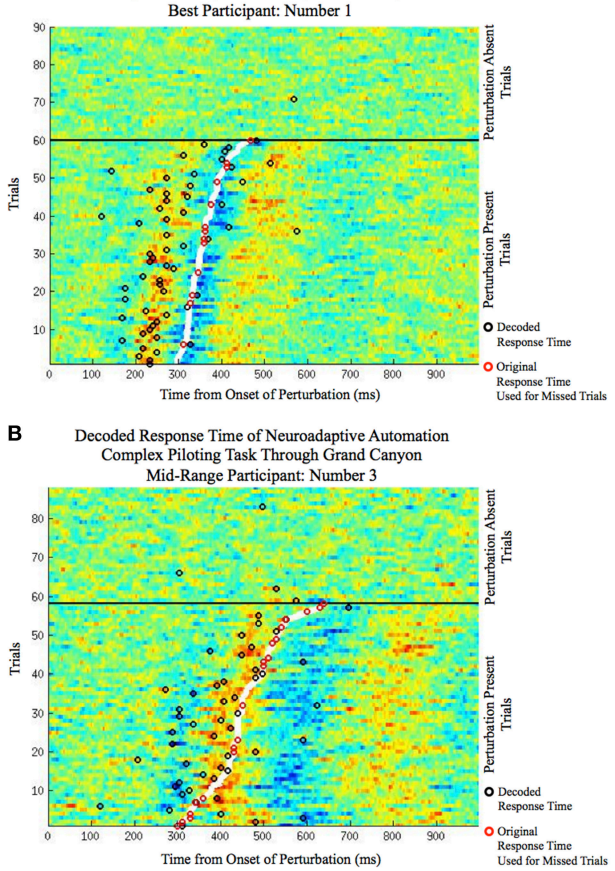


FIGURE 5 | The decoded response time (circles) plotted on the single trial activation waveforms (ranging from: red: large positive amplitude to blue: large negative amplitude) of the selected independent component of the simulated neuroadaptive automation on the complex piloting task for (A) the best participant (P01) and (B) the middle range participant in terms of classification performance (P03). Both perturbation absent trials (top of each plot) and perturbation present trials (bottom of each plot) are shown. The perturbation present trials are arranged in order of fastest manual response time (bottom) to the slowest (top). The manual response times are denoted by the thick white line for the perturbation present trials. The decoded response time, by the simulated neuroadaptive automation (BCI classifier), of each trial is denoted by a black circle. For perturbation present trials the black circles denote hits when their time is faster than the manual response time (white line). For perturbation absent trials the black circles denote false alarms. A red circle is shown over the original response time in the case when the simulated neuroadaptive automation failed to classify the trial as a hit (misses) or in which it was slower than the original response time (slow responses).

false-alarms is to engage the neuroadaptive automation for only a couple hundred milliseconds and immediately disengage it in response to opposite deflection of the flight controls by the pilot. This would reduce the detrimental effects of false-alarms and at the same time would speed up response to recover from true perturbations in the case of hits. Given that the BCI-decoder is extracting motor intention related activity it would be interesting to determine whether the pilot actually notices the engagement of the neuroadaptive automation in the case of hits or rather just feels that they are really fast in reacting. The extent to which pilot-automation induced oscillations arise and offset the beneficial affects of time savings of the neuroadaptive automation need to be investigated using closed-loop implementation of the system during flight simulation (It should be noted that our study reported here only uses an open-loop BCI decoder tested offline to test the feasibility of implementation in neuroadaptive automation).

Although the BCI-decoder was trained using a specified window (120 ms) centered at the time of the peak evoked response prior to movement onset to detect a perturbation causing a pitch down attitude while in straight and level flight over the ocean (simple visual field) it was able to generalize to a novel complex flight condition in which the pilot maneuvered the plane through the Grand Canyon. In this complex condition the orientation of the perturbation with respect to the horizon is dependent on the roll angle (bank angle) of the plane at the time of the perturbation. The magnitude of the perturbation reflected in negative deflection in the pitch axis is dependent on the planes attitude (pitch, roll, yaw axes), speed, airflow over the flight surfaces, and the time in which it takes for the pilot to initiate recovery (the longer it takes the bigger the perturbation effect). It is impressive the BCI-decoder is able to generalize to the novel complex flight condition given that the nature of the perturbation and the corresponding visual aspects of the scene and ongoing motor control are quite different from the training situation. As we envision the closed-loop operational neuroadaptive automation system it would not need to know the magnitude of the perturbation (although this information may be available by flight instruments) as its job is to only initiate recovery based on the decoded motor intention of the pilot. It is up to the pilot to appropriately control the plane within the first couple hundred milliseconds after the neuroadaptive automation has been initiated. As it stands now the system is only set up to recover from a pitch down attitude. Ideally, we would like a system that could recover from a perturbation in attitude to any of the axes (pitch, roll, yaw) or combinations thereof. By comparing data from flight instruments that precisely measure attitude of all axes of the plane and pilot directed control movements the neuroadaptive automation could initiate the proper combination of control surface deflections to recover from various types of non-pilot induced perturbations. It would be interesting to test whether our system would generalize to other types of perturbation in attitude even though it was only trained on a pitch down perturbation. While this system using constraints determined by flight instruments may work in the case of perturbations it may not be effective in situations involving collision avoidance (e.g., with another aircraft or bird, etc.). In these situations it would be necessary to additionally build a BCI-decoder that can determine the desired direction of motor intention as it relates to the flight controls governing the attitude of the plane. This task may be difficult to accomplish within the framework of achieving the desired time savings to initiate recovery as fast as possible.

For the complex flying task no information was given concerning the timing of the peak of the event related evoked

response to the onset of the perturbation used during training. Rather, the 120 ms window of the BCI-decoder progressed through the data in 8 ms steps until it identified an occurrence of a perturbation. The initial time window for the perturbation present trials started at the onset of the perturbation (the onset of the perturbation absent trials was randomly determined). However, there is no implicit information in this starting time that would reference the time of the evoked response upon which the BCI-decoder was trained. The presence of false alarms for the perturbation absent trials may be problematic for application of neuroadaptive automation working in a continuous manner given that the occurrence of true perturbations is quite sparse. As was discussed above, one way to reduce the number of false alarms made by the BCI-decoder is to only attempt to decode motor intention at points in which a perturbation is detected by flight instruments and then the appropriate recovery maneuver is applied by the neuroadaptive automation. One could implement a system that automatically recovers from a perturbation without regards to the pilot's intention ("Automation"). However, this is not the intention of the neuroadaptive automation proposed here for which the goal is to always keep the pilot intentions in control of the aircraft.

Previous research conducted on detection of driver braking intention, using EEG (Haufe et al., 2011, 2014; Kim et al., 2015), is relevant to the discussion of our results. In their studies as well as in ours simple amplitude based features related to the onset of the visual event were used for decoding the onset of movement intention. The visual event signaling the onset to move in the Haufe et al. (2011, 2014) and Kim et al. (2015) studies was the flashing of the brake light on the car just in front of the one the participant was driving. In our study the visual event signaling the onset to move was the changes in the optic flow field and the change in the position of the horizon (sky relative to ground) (See **Figure 2**). The finding that the best participant's decoding model generalizes to the remaining six subjects on detecting perturbation on the complex flying task with significant, although reduced, time savings (See **Tables 6, 8**), does suggest that the features selected by the model are not individual specific but are to some degree common across participants. As it stands now at least one session of the simple flying task is necessary to extract the task related independent component that help in artifact extraction. However, the finding that the BCI-decoder generalizes across participants (See **Tables 6, 8**) is promising in future attempts to make a generic system that does not require training.

There are three important aspects that distinguish our study from that of previous studies investigating braking intention.

The first is that our test condition was on a novel task that was fairly different from the one the BCI-decoder was trained on rather than just using a subset of trials on the same task for testing as is commonly done in decoding studies (Garrett et al., 2003; Wilson and Russell, 2007; Haufe et al., 2011, 2014; Baldwin and Penaranda, 2012; Callan et al., 2015; Kim et al., 2015). Our study demonstrates that a BCI-decoder trained on a simple task can generalize to a more complex one characteristic of real world conditions with significant performance in identifying perturbation events (mean bacc = 73%, $p < 0.05$; mean $d' = 1.52$; mean $a' = 0.84$; **Table 5**) with a significant time savings of 72 ms (**Table 7**).

The second is that the testing session (complex piloting task) requires that the participant use the same control stick to recover from the perturbation as used to maneuver the plane tracking above the river. Under these conditions it is necessary for the BCI-decoder to be able to distinguish brain activity related to the perturbation and the intention to move from ongoing changes in the visual field and motor intention required to pilot the plane. This is substantially different from decoding of movement intention of the foot from the accelerator to the brake in response to a flashing light. In order to extract neural activity related to movement intention in response to a perturbation, rather than that just related to the visual event, the BCI-decoder was trained to distinguish between trials in which the participant was to pull back on the control stick in response to a perturbation vs. just passively viewing the perturbation. All but one of the subjects showed significant classification performance in identifying movement intention trials from passive viewing trials on the test session (mean bacc = 69.8%, $p < 0.05$; mean $d' = 1.30$; mean $a' = 0.80$; See **Table 4**). The ability of the BCI-decoder to be able to identify cases of motor intention in response to identical visual events likely contributes to its ability to distinguish between variations in brain activity in response to changes in the optic flow pattern and movement intention in response to a perturbation rather than changes in the optic flow pattern induced by piloting while maneuvering through the Grand Canyon.

The third is the difference in response time for emergency braking, that is approximately 650 ms (Haufe et al., 2011, 2014; Kim et al., 2015), compared to pulling back on the stick to recover from a perturbation, which took approximately 437 ms for the complex flying task and 369 ms for the simple flying task. One reason why the time savings in the braking studies [up to 222 ms using combined EEG and EMG (Haufe et al., 2014)] is larger than in our study (72 ms) may be attributed to the longer response time for emergency braking (over 200 ms longer). The mean peak of the event related potentials used as the target range to train the BCI-decoders in our study was 246 ms (See **Figure 3B**). Because of the relatively fast response times the slower event related potentials could not be used for decoding because they occur after the behavioral response has already been given. The mean response time for the complex flying task for the adaptive automation is 352.7 ms compared to the original of 425.0 ms. The mean time in which decoding performance reached an area under the curve (A') value of 0.8 was also around 350 ms in the emergency braking studies (Haufe et al., 2011, 2014; Kim et al., 2015). It should be mentioned that the improvement in response time afforded by the adaptive automation in our study for some of the participants allowed them to have almost superhuman performance on this piloting task.

Given that the perturbation we employed abruptly alters the optic flow field we predict that visual motion processing areas as well as brain regions involved with motor intention (premotor cortex, motor cortex, somatosensory cortex, parietal cortex) are involved in decoding the decision for rapid movement in response to an impending hazard. While there is considerable

variability in the extent and location of brain activity of the selected independent component used for the BCI-decoder for each participant there are regions that are commonly activated across the participants (See **Figures 3**, **4** and **Tables 2**, **3**). Consistent with our predictions all subjects showed some degree of activity in visual motion processing areas (hOC5, MT, IOG), as well as the premotor cortex, pre-central gyrus (motor cortex), post-central gyrus (somatosensory cortex), and parietal cortex (superior parietal lobule) (See **Figures 3**, **4** and **Tables 2**, **3**). The visual cortex (BA17,18) also showed some degree of activity from all subjects (See **Figures 3**, **4** and **Tables 2**, **3**). Our findings are consistent with fMRI research investigating action intentions from preparatory brain activity (Gallivan et al., 2011). In the Gallivan et al. (2011) study, decoded activity from voxels in multiple parietal, premotor, and motor regions was found to successfully predict intended future grasp and reach movements. A study using electrocortiocography (ECoG) revealed that in addition to motor and premotor activity somatosensory activity also precedes voluntary movement (Sun et al., 2015). The finding of predominantly caudal rather than rostral dorsal premotor cortex activation found for most participants in our study (See **Figures 3**, **4** and **Tables 2**, **3**) is interesting as it relates to studies showing that that action intention is processed more caudally and attention is processed more rostrally in the premotor cortex (Boussaoud, 2001).

A potential limitation of our study is the low number of participants. However, the primary aim of our study is to show the feasibility of the proposed approach for the development of neuroadaptive automation and to determine limitations that need to be addressed in future research. In our study the results from each individual participant are given. Even when individually tailored models were trained specifically on data from that participant there is some degree of variability in performance at predicting presence/absence of a perturbation (ranging from 63.7 to 85.6%, See **Table 5**), and the corresponding time savings (ranging from 36.1 to 138.9 ms, See **Table 7**), as well as the pattern of brain activation (See **Figures 3**, **4** and **Tables 2**, **3**). In the future, it may be interesting to investigate why some participants have better predictive models than others. These results strongly suggest that best performance will be achieved by individually tailored systems rather than using a generalized system that works across individuals. The drawback of individually tailored systems is the time necessary to train the system including ICA and the BCI-decoder. While this study does demonstrate that it is potentially possible to enhance response time by using an off-line BCI-decoder in these select participants it will be necessary to test a larger sample to see how well they generalize to the population in general and to determine factors predicting model efficacy.

Given that the mean time savings is 72 ms in the simulated off-line open loop neuroadaptive automation system demonstrated here, it is important to discuss whether the processing time would be of sufficient speed to be used in a real-time closed loop neuroadaptive automation system (see **Figure 1**). The Yokogawa 400 channel MEG system at ATR is set up with a real-time processing system. The hardware and software for acquiring MEG channel data in real time and analog to digital conversion includes the following: National Instruments A/D Converter boards (6 Boards: 80 channels per board) can convert 400 MEG channels plus additional channels (EEG, EOG, triggers, etc.) sampled at 1000 Hz. To get high temporal precision that is stable the National Instruments real-time operating system "Pharlab" is used on a dedicated computer. Pharlab carries out filtering operations on 400 channel MEG and sends the analog to digital converted MEG channel data via UDP to a different computer for further processing in ∼1.5 ms. The application of the ICA weight matrix of the selected independent component to the 400 channel MEG data as well as the weight matrix of the BCI-decoder over the computed activation waveform is <0.1 ms. The ICA and BCI-decoder can operate in such a short time because the weights have been trained ahead of time based on data from previous sessions. Therefore, the data acquisition, preprocessing, and BCI-decoding can all be accomplished in <2 ms in the real-time system. The X-Plane flight simulator is running at around 400 Hz. It takes ∼2.5–5 ms for the flight simulator computer to receive the command from the BCI-decoder computer over UDP and initiate the directed command. Based on the specifications of the system at ATR the loss in time savings afforded by the simulated neuroadpative automation resulting from processing time would be ∼4.5–7 ms. This would still leave a mean time savings afforded by the neuroadaptive automation ranging from 65 to 67.5 ms, which could be of substantial benefit in hazardous time critical situations. Most of the delay (resulting in reduction of time savings) is in the processing speed of the flight simulator, which theoretically could be improved if using dedicated hardware and software in a real aircraft.

While this system is specific to the MEG setup at ATR it is possible to make such a dedicated real-time system that will work with EEG that can be used in real-world settings. In order for the system to be feasible in real aircraft it will be necessary to use a more portable technology such as EEG. The signal processing techniques used in this study together with automatic subspace reconstruction have been shown to be able to separate artifacts from brain related activity in flight even in an open cockpit biplane (Callan et al., 2015). It is uncertain whether moving from 400 channels to 64 or 20 channels with an EEG setup will have a large effect on system performance. Source localization would likely be considerably worse in the case of EEG especially with 20 channels compared with that of MEG. The number of channels will also play an important part in the ICA in the number of brain and artifact components that can be separated. In future research we will test an EEG based closed-loop version of this neuroadaptive automation system on a motion platform based flight simulator to determine its feasibility and additional processing that may be necessary if it were to be realized in actual manned or unmanned aircraft.

CONCLUSION

Our study explores the potential that neuroadaptive automation may have in facilitating human performance. Our goal is to

develop a system that enhances performance to super human levels during normal hands on operation of an airplane (vehicle) by reducing the response time by directly extracting from the brain the movement intention in response to a hazardous event. This approach differs considerably from those that utilize BCI to maneuver a vehicle by hands-off control by such methods as decoded mental imagery or attention related steady state visual evoked potentials (Blankertz et al., 2010; LaFleur et al., 2013). These applications of BCI, although impressive, are severely limited in performance compared to normal hands on control with the addition of greater workload as well as divided attention away from the task at hand (It should be noted however, that these types of BCI are of extreme benefit when the normal channels of motor control are impaired). Advantages of the neuroadaptive automation BCI implementation proposed here, afforded by the use of only brain activity naturally occurring during the perceptual motor task, include improved performance with no additional workload or attentional demands for the pilot (operator), as well as no training by the pilot to fit the BCI. However, it should be noted that human training protocols for utilizing BCI are likely to improve performance (Lotte et al., 2013). Our proposed BCI-decoder works continuously over time without any a-priori knowledge of when a perturbation may occur. In addition it was shown to be able to generalize to more complex tasks and differentiate between motor intention to an unexpected perturbation from that used during normal maneuvering. Future research needs to test the proposed neuroadaptive automation online using EEG in motion based flight simulators as well as in real airplanes to evaluate its real-world performance. It is interesting to conjecture whether the participants will notice when the neuroadpative automation is active or will they just think they are responding really fast. This research adds to the growing field of neuroergonomics and specifically to aviation cerebral experimental sciences. Our results, using an off-line BCI-decoder, suggest that indeed neuroadaptive automation can be implemented that is faster than the hand. The data can be shared with interested scientists upon request.

AUTHOR CONTRIBUTIONS

DEC, CT, and DBC designed and conducted experiment. DEC, CT, DBC, MS, and RP analyzed results of the experiment. DEC and RP wrote the manuscript.

FUNDING STATEMENT

Funding for this research was supported in part by a contract with the National Institute of Information and Communications Technology, Japan, entitled, "Multimodal integration for brain imaging measurements" and by the Center for Information and Neural Networks, National Institute of Information and Communications Technology. MS was supported by the National Institute of Information and Communications Technology. Additional support to RP was given by the Air Force Office of Sponsored Research grant FA9550-10-1-0385.

ACKNOWLEDGMENTS

We would like to thank Eiji Nawa for his assistance with this research. We would also like to thank MEG and fMRI technicians Yasuhiro Shimada, Ichiro Fujimoto, Hiroaki Mano, and Hironori Nishimoto at the Brain Activity Imaging center at ATR as well as Yuka Furukawa for assisting in running the experiments. We would especially like to thank Yasushi Morikawa and Erika Matsumoto for helping in recruitment of glider pilots who participated in this study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fnhum.2016.00187

REFERENCES

Amano, K., Goda, N., Nishida, S., Ejima, Y., Takeda, T., and Ohtani, Y. (2006). Estimation of the timing of human visual perception from magnetoencephalography. *J. Neurosci.* 26, 3981–3991. doi: 10.1523/JNEUROSCI.4343–05.2006

Baldwin, C. L., and Penaranda, B. N. (2012). Adaptive training using an artificial neural network and EEG metrics for within- and cross-task workload classification. *Neuroimage* 59, 48–56. doi: 10.1016/j.neuroimage.2011.07.047

Blankertz, B., Schäfer, C., Dornhege, G., and Curio, G. (2002). "Single trial detection of EEG error potentials: a tool for increasing BCI transmission rates," in *Artificial Neural Networks—ICANN 2002* (Berlin; Heidelberg: Springer), 1137–1143. doi: 10.1007/3-540-46084-5_184

Blankertz, B., Tangermann, M., Vidaurre, C., Fazli, S., Sannelli, C., Haufe, S., et al. (2010). The Berlin brain- computer interface: non-medical uses of BCI technology. *Front. Neurosci.* 4:198. doi: 10.3389/fnins.2010.00198

Boussaoud, D. (2001). Attention versus intention in the primate premotor cortex. *Neuroimage* 14, S40–S45. doi: 10.1006/nimg.2001.0816

Brodersen, K. H., Ong, C., and Buhmann, J. (2010). "The balanced accuracy and its posterior distribution," in *Twentieth International Conference on Pattern Recognition IEEE* (Istanbul), 3121–3124. doi: 10.1109/icpr.2010.764

Byrne, E. A., and Parasuraman, R. (1996). Psychophysiology and adaptive automation. *Biol. Psychol.* 42, 249–268. doi: 10.1016/0301-0511(95)05161-9

Callan, D., Mills, L., Nott, C., England, R., and England, S. (2014). A tool for classifying individuals with chronic back pain: using multivariate pattern analysis with functional magnetic resonance imaging data. *PLoS ONE* 9:e98007. doi: 10.1371/journal.pone.0098007

Callan, D. E., Durantin, G., and Terzibas, C. (2015). Classification of single-trial auditory events using dry-wireless EEG during real and motion simulated flight. *Front. Syst. Neurosci.* 9:11. doi: 10.3389/fnsys.2015.00011

Christophel, T. B., and Haynes, J-D. (2014). Decoding complex flow-field patterns in visual working memory. *Neuroimage* 91, 43–51. doi: 10.1016/j.neuroimage.2014.01.025

Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics. *J. Neurosci. Methods* 134, 9–21. doi: 10.1016/j.jneumeth.2003.10.009

Delorme, A., Sejnowski, T., and Makeig, S. (2007). Improved rejection of artifacts from EEG data using high-order statistics and independent component analysis. *Neuroimage* 34, 1443–1449. doi: 10.1016/j.neuroimage.2006.11.004

Durantin, G., Dehais, F., and Delorme, A. (2015). Characterization of mind wandering using fNIRS. *Front Syst. Neurosci.* 9:45. doi: 10.3389/fnsys.2015.00045

Eickoff, S. B., Stephan, K. E., Mohlberg, H., Grefkes, C., Fink, G. R., Amunts, K., et al. (2005). A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *Neuroimage* 25, 1325–1335. doi: 10.1016/j.neuroimage.2004.12.034

Fischl, B., Sereno, M., and Dale, A. (1999). Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. *Neuroimage* 9, 195–207. doi: 10.1006/nimg.1998.0396

Gallivan, J. P., McLean, D. A., Valyear, K. F., Pettypiece, C. E., and Culham, J. C. (2011). Decoding action intentions from preparatory brain activity in human parieto-frontal networks. *J. Neurosci.* 31, 9599–9610. doi: 10.1523/JNEUROSCI.0080-11.2011

Garrett, D., Peterson, D. A., Anderson, C. W., and Thaut, M. H. (2003). Comparison of linear, nonlinear, and feature selection methods for EEG signal classification. *IEEE Trans. Neural Syst. Rehabil. Eng.* 11, 141–144. doi: 10.1109/TNSRE.2003.814441

Gateau, T., Durantin, G., Lancelot, F., Scannella, S., and Dehais, F. (2015). Real-time state estimation in a flight simulator using fNIRS. *PLoS ONE* 10:e0121279. doi: 10.1371/journal.pone.0121279

Haber, R. N., and Haber, L. (2003). "Perception and attention during low-altitude high-speed flight," in *Principles and Pracice of Aviation Psychology,* eds P. S. Tsang and M. A. Vidulich (Boca Raton, FL: CRC Press), 21–68.

Haufe, S., Kim, J.-W., Kim, I.-H., Sonnleitner, A., Schrauf, M., Curio, G., et al. (2014). Electrophysiology-based detection f emergency braking intention in real-world driving. *J. Neural Eng.* 11:056011. doi: 10.1088/1741-2560/11/5/056011

Haufe, S., Treder, M. S., Gugler, M. F., Sagebaum, M., Crio, G., and Blankertz, B. (2011). EEG potentials predict upcoming emergency braking during simulated driving. *J. Neural Eng.* 8:056001. doi: 10.1088/1741-2560/8/5/056001

Hutzler, F. (2014). Reverse inference is not a fallacy per se: cognitive processes can be inferred from functional imaging data. *Neuroimage* 84, 1061–1069. doi: 10.1016/j.neuroimage.2012.12.075

Kamitani, Y., and Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nat. Neurosci.* 8, 679–685. doi: 10.1038/nn1444

Kamitani, Y., and Tong, F. (2006). Decoding seen and attended motion directions from activity in the human visual cortex. *Curr. Biol.* 16, 1096–1102. doi: 10.1016/j.cub.2006.04.003

Kim, I.-H., Kim, J.-W., Haufe, S., and Lee, S.-W. (2015). Detection of braking intention in diverse situations during simulated driving based on EEG feature combination. *J. Neural Eng.* 12:016001. doi: 10.1088/1741-2560/12/1/016001

Knapp, C. J., and Johnson, R. (1996). F-16 Class A mishaps in the U.S. Air Force, 1975-93. *Aviat. Space Environ. Med.* 67, 777–783.

LaFleur, K., Cassady, K., Doud, A., Shades, K., Rogin, E., and He, B. (2013). Quadcopter control in three-dimensional space using a noninvasive motor imagery-based brain-computer interface. *J. Neural Eng.* 10, 1–15. doi: 10.1088/1741-2560/10/4/046003

Lew, E. Y., Chavarriaga, R., Silvoni, S., and Millán, J. (2014). Single trial prediction of self-paced reaching directions from EEG signals. *Front. Neurosci.* 8:222. doi: 10.3389/fnins.2014.00222

Lotte, F., Larrue, F., and Mühl, C. (2013). Flaws in current human training protocols for spontaneuous Brain-Computer Interfaces: lessons learned from instructional design. *Front. Hum. Neurosci.* 7:568. doi: 10.3389/fnhum.2013.00568

Macmillan, N., and Creelman, D. (1991). *Detection Theory: A User's Guide*. New York, NY: Cambridge University Press.

Moroze, M. L., and Snow, M. P. (1999). *Causes and Remedies of Controlled Flight into Terrain in Military and Civil Aviation*. Technical Report. Air Force Research Laboratory Wright-Patterson AFB, OH 45433.

Nakagawa, V. B., Montgomery, R. W., and Wood, K. J. (2007). Aircraft accidents and incidents associated with visual effects from bright light exposures during low-light flight operations. *Optometry* 78, 415–420. doi: 10.1016/j.optm.2006.11.016

Nishimoto, S., Vu, A., Naselaris, T., Benjamini, Y., Yu, B., and Gallant, J. (2011). Reconstructing visual experiences from brain activity evoked by natural movies. *Curr. Biol.* 21, 1641–1646. doi: 10.1016/j.cub.2011.08.031

Norman, K. A., Sean, M., Polyn, S. M., Detre, G. J., and Haxby, J. V. (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends Cogn. Sci.* 10, 424–430. doi: 10.1016/j.tics.2006.07.005

Parasuraman, R., Bahri, T., Deaton, J. E., Morrison, J. G., and Barnes, M. (1992). *Theory and Design of Adaptive Automation in Aviation Systems*. Technical Report, Code 6021, Naval Air Development Center (Warminster, PA).

Parasuraman, R., and Galster, S. (2013). Sensing, assessing, and augmenting threat detection: behavioral, neuroimaging, and brain stimulation evidence for the critical role of attention. *Front. Hum. Neurosci.* 7:273. doi: 10.3389/fnhum.2013.00273

Parasuraman, R., and Riley, V. (1997). Humans and automation: use, misuse, disuse, abuse. *Hum. Factors* 39, 230–253.

Parra, L. C., Spence, C. D., Gerson, A. D., and Sajda, P. (2003). Response error correction-a demonstration of improved human-machine performance using real-time EEG monitoring. *Neural Syst. Rehabil. Eng. IEEE Trans.* 11, 173–177. doi: 10.1109/TNSRE.2003.814446

Poldrack, R. (2011). Inferring mental states from neuroimaging data: from reverse inference to large-scale decoding. *Neuron* 72, 692–697. doi: 10.1016/j.neuron.2011.11.001

Prinzel, L. J., Freeman, F. G., Scerbo, M. W., Mikulka, P. J., and Pope, A. T. (2000). A closed-loop system for examining psychophysio- logical measures for adaptive task allocation. *Int. J. Aviat. Psychol.* 10, 393–410. doi: 10.1207/S15327108IJAP1004_6

Reiner, R. (2008). "Neurorehabilitation robotics and neuroprosthetics," in *Neuroergonomics; The Brain at Work,* eds. R. Parasuraman and M. Rizzo (New York, NY: Oxford University Press), 346–359.

Sarvas, J. (1987). Basic mathematical and electromagnetic concepts of the biomagnetic inverse problem. *Phys. Med. Biol.* 32, 11–22. doi: 10.1088/0031-9155/32/1/004

Sato, M., Yoshioka, T., Kajiwara, S., Toyama, K., Goda, N., Doya, K., et al. (2004). Hierarchical Bayesian estimation for MEG inverse proble m. *Neuroimage* 23, 806–826. doi: 10.1016/j.neuroimage.2004.06.037

Scerbo, M. W. (2008). "Adaptive automation," in *Neuroergonomics; The Brain at Work,* eds R. Parasuraman and M. Rizzo (New York, NY: Oxford University Press), 239–252.

Scerbo, M. W., Freeman, F. G., and Mikulka, P. J. (2003). A brain-based system for adaptive automation. *Theor. Issues Ergon. Sci.* 4, 200–219. doi: 10.1080/1463922021000020891

Shibata, K., Chang, L., Kim, D., Nanez, J., Kamitani, Y., Watanabe, T., et al. (2012). Decoding reveals plasticity in V3A as a result of motion perceptual learning. *PLoS ONE* 7:e44003. doi: 10.1371/journal.pone.0044003

Sugiyama, M., Hachiya, M., Simm, J., and Nam, H. (2010). Least-squares probabilistic classifier: a computationally efficient alternative to kernel logistic regression. *IEICE Trans. Inform. Syst.* 93-D, 2690–2701. doi: 10.1587/transinf.E93.D.2690

Sun, H., Blakely, T. M., Darvas, F., Wander, J. D., Johnson, L. A., Su, D. K., et al. (2015). *Sequential activation of* premotor, primary somatosensory and primary motor areas in humans during cued finger movements. *Clin. Neurophysiol.* 126, 2150–2161. doi: 10.1016/j.clinph.2015.01.005

Swets, J. A. (1973). The relative operating characteristic in psychology. *Science* 182, 990–1000.

Ting, C., Mahfouf, M., Nassef, A., Linkens, D. A., Panoutsos, G., Nickel, P., et al. (2010). Real-time adaptive automation system based on identification of operator functional state in simulated pro- cess control operations. *IEEE Trans. Syst. Man Cybern. Part A Syst. Hum.* 40, 251–262. doi: 10.1109/TSMCA.2009.2035301

Wilson, G. F., and Russell, C. A. (2007). Performance enhancement in an uninhabited air vehicle task using psychophysiologically determined

adaptive aiding. *Hum. Factors* 49, 1005–1018. doi: 10.1518/001872007X249875

Yoshioka, T., Toyama, K., Kawato, M., Yamashita, O., Nishina, S., Yamagishi, N., et al. (2008). Evaluation of hierarchical Bayesian method through retinotopic brain activities reconstruction from fMRI and MEG signals. *Neuroimage* 42, 1397–1413. doi: 10.1016/j.neuroimage.2008.06.013

Zander, T. O., and Kothe, C. (2011). Towards passive brain–computer interfaces: applying brain–computer interface technology to human–machine systems in general. *J. Neural Eng.* 8:025005. doi: 10.1088/1741-2560/8/2/025005

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Callan, Terzibas, Cassel, Sato and Parasuraman. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 13 October 2016
doi: 10.3389/fnhum.2016.00519

Efficient Workload Classification based on Ignored Auditory Probes: A Proof of Concept

Raphaëlle N. Roy[1,2], Stéphane Bonnet[1,3], Sylvie Charbonnier[1,2] and Aurélie Campagne[1,4]*

[1] Université Grenoble Alpes, Grenoble, France, [2] Gipsa-Lab, Centre National de la Recherche Scientifique, Grenoble, France, [3] CEA LETI, Grenoble, France, [4] Laboratoire de Psychologie et NeuroCognition, Centre National de la Recherche Scientifique, Grenoble, France

Mental workload is a mental state that is currently one of the main research focuses in neuroergonomics. It can notably be estimated using measurements in electroencephalography (EEG), a method that allows for direct mental state assessment. Auditory probes can be used to elicit event-related potentials (ERPs) that are modulated by workload. Although, some papers do report ERP modulations due to workload using attended or ignored probes, to our knowledge there is no literature regarding effective workload classification based on ignored auditory probes. In this paper, in order to efficiently estimate workload, we advocate for the use of such ignored auditory probes in a single-stimulus paradigm and a signal processing chain that includes a spatial filtering step. The effectiveness of this approach is demonstrated on data acquired from participants that performed the Multi-Attribute Task Battery – II. They carried out this task during two 10-min blocks. Each block corresponded to a workload condition that was pseudorandomly assigned. The easy condition consisted of two monitoring tasks performed in parallel, and the difficult one consisted of those two tasks with an additional plane driving task. Infrequent auditory probes were presented during the tasks and the participants were asked to ignore them. The EEG data were denoised and the probes' ERPs were extracted and spatially filtered using a canonical correlation analysis. Next, binary classification was performed using a Fisher LDA and a fivefold cross-validation procedure. Our method allowed for a very high estimation performance with a classification accuracy above 80% for every participant, and minimal intrusiveness thanks to the use of a single-stimulus paradigm. Therefore, this study paves the way to the efficient use of ERPs for mental state monitoring in close to real-life settings and contributes toward the development of adaptive user interfaces.

Keywords: workload, classification, auditory evoked potentials, spatial filtering

OPEN ACCESS

Edited by:
Thorsten O. Zander, Technical University of Berlin, Germany

Reviewed by:
Virginia R. De Sa, Vanderbilt University, USA
Hasan Ayaz, Drexel University, USA

****Correspondence:***
Raphaëlle N. Roy
roy.raphaelle@gmail.com

Received: *04 October 2015*
Accepted: *30 September 2016*
Published: *13 October 2016*

Citation:
Roy RN, Bonnet S, Charbonnier S and Campagne A (2016) Efficient Workload Classification based on Ignored Auditory Probes: A Proof of Concept. Front. Hum. Neurosci. 10:519. doi: 10.3389/fnhum.2016.00519

INTRODUCTION

Mental workload is frequently defined as task difficulty and the associated mental effort (Gevins and Smith, 2007). It is therefore of critical interest to better assess this state to the human factor community who aims at developing smart technologies that enhance operator's safety and performance. The impact of workload on behavior has been extensively documented. Participants' reaction time is known to increase linearly with the increase in the number of items to memorize

(Sternberg, 1969), as well as with the number of tasks to perform in parallel (Cain, 2007). However, behavioral responses are not always enough for mental state monitoring (MSM) systems, mainly due to their latency of occurrence, and to the fact that some mental states are not necessarily or systematically reflected by a specific response. Physiological data give more insight into the operator's state, especially electroencephalography (EEG), a method that allows for direct mental state assessment. The use of physiological markers derived from the cerebral activity for human factor purposes has given rise to a new field: neuroergonomics (Parasuraman et al., 2012).

Amongst the various markers derived from the EEG activity, event-related potentials (ERPs) are frequently used for MSM. ERPs correspond to the EEG activity that is temporally locked to the appearance of a given stimulation, or probe. Although, ERPs only allow for a discontinuous evaluation of the operator's mental state -unlike frequency measures, according to Roy et al. (2016) frequency measures are very sensitive to mental fatigue and vigilance states whereas ERPs are more robust to these states. Therefore, ERPs may be more suitable for ecological settings. Moreover, the literature describes numerous workload-related ERP modulations, such as early and late components' amplitude decreases. Hence, the P300 component's amplitude is reduced by an increase in workload (Kok, 2001; Schultheis and Jameson, 2004; Gomarus et al., 2006; Holm et al., 2009; Friedrich et al., 2011), and so is the N1, N2, and P2 components' amplitude (Kramer et al., 1995; Ullsperger et al., 2001; Gomarus et al., 2006; Allison and Polich, 2008; Miller et al., 2011; Boonstra et al., 2013). In the specific context of simulated flight, the P300 component's amplitude elicited by auditory probes has also been shown to decrease when the primary task's complexity increases (Natani and Gomer, 1981; Kramer et al., 1987; Sirevaag et al., 1993).

In order to determine an operator's mental state to modify the behavior of a system, one needs to compute an index or a class of workload to be fed as an input. This can be done using machine learning algorithms developed for brain–computer interfaces (BCIs). When those algorithms are used for applications that are not directed toward the voluntary control of an effector, those systems are often referred to as passive BCIs (Zander and Kothe, 2011; van Erp et al., 2012). Although, the number of publications regarding mental workload assessment has drastically increased this decade, only a few articles actually propose a classification based on ERPs. Brouwer et al. (2012) used seven electrodes and achieved 64% of correct binary classifications. Recently, it was proved that ERP spatial filtering could significantly enhance workload classification (Mühl et al., 2014; Roy et al., 2015). The authors achieved 72 and 98% of correct classifications using respectively a Fisher spatial filtering (FSF; Hoffmann et al., 2006) and a canonical correlation analysis filtering (CCA; Hotelling, 1936). However, these authors used task-dependent probes, i.e., items that were paramount for the task at hand, which is therefore quite unrealistic for real-life settings. Very recently, Roy et al. (2016) showed that ERPs elicited by visual task-independent probes could be used for mental workload estimation. They inserted a basic detection task in a Sternberg memory task and used the ERPs elicited by the targets to classify the workload level of the memory task. They reached 91% of correct binary classifications by filtering the ERPs using a CCA. This is very promising, however the probes, although task-independent, still required an overt answer from the participant. This kind of dual task setting can therefore lead to decreased attentional engagement to the primary task, which seems rather unwelcome for operators' monitoring in hazardous work situations (e.g., driving, plant monitoring, custom control). Hence, the best approach to use ERPs in ecological settings would be a stimulation paradigm with task-independent and ignored probes. And as the ultimate goal should be to develop systems based on minimally intrusive probes, these stimulations should be as scarce as possible. As reported by Mertens and Polich (1997), the ERPs elicited in a single-stimulus paradigm by visual or auditory probes are a viable alternative to the traditional oddball procedure, although late components' amplitude is reduced when the stimuli are ignored compared to when they are counted or await a motor response. The authors even report that auditory probes elicit ERPs that are more robust to response type. That is to say that ignored auditory stimuli generate early and late components which amplitude is quite similar to that of stimuli awaiting an active answer. This makes them very good candidates for the features to use in a mental workload estimation procedure.

This study intends to provide an evaluation of the efficiency of a workload estimation based on the ERPs elicited by infrequent, task-independent and ignored auditory probes. Workload was modulated by modulating the number of tasks to perform in parallel with the Multi-Attribute Task Battery – II (MATB; Comstock and Arnegard, 1992). A single-stimulus paradigm was used to elicit ERPs which were then spatially filtered with a CCA and classified. The performance of this processing chain was also compared to that of a simpler chain without spatial filtering. The contributions of this paper are threefold: (1) to assess the validity of the single-stimulus paradigm for effective mental workload estimation; (2) to assess the relevance of a processing chain that includes a spatial filtering step in order to classify accurately the auditory evoked potentials (AEPs) of those ignored, infrequent probes; (3) to assess the relevance of both the stimulation paradigm and the processing chain for an ecologically valid task, the MATB.

MATERIALS AND METHODS

This research was promoted by Grenoble's clinical research direction (France) and was approved by the French ethics committee (ID number: 2014-A00040-47) and the French health safety agency (B140052-31).

Experimental Setup

Eight healthy right-handed volunteers (three females; 29.9 years old ± 5.9) performed two 10-min experimental blocks of the Multi-Attribute Task Battery-II, the last version of task developed by NASA to study divided attention and multitasking (Comstock and Arnegard, 1992; **Figure 1**). In this experimental setup, each block corresponded to a different workload level (low/high), which was pseudo-randomly assigned. In the low

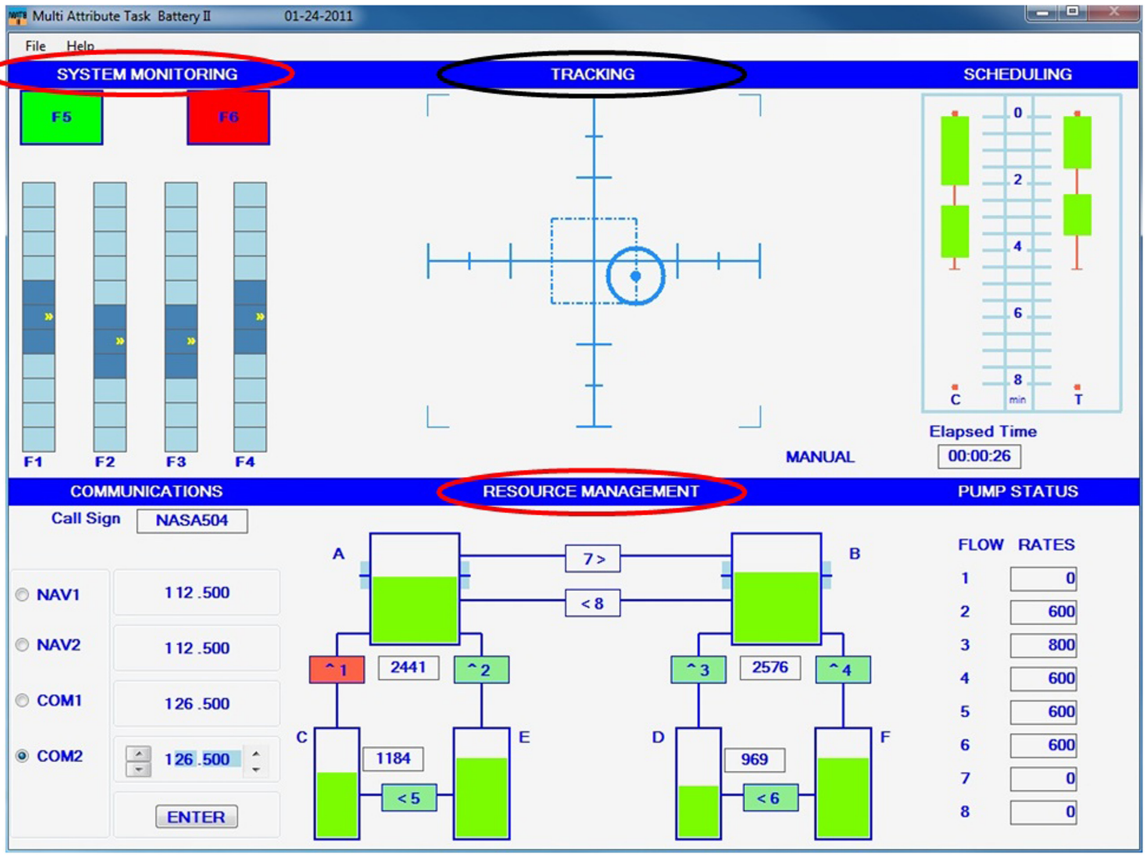


FIGURE 1 | Multi-Attribute Task Battery – II. The participants performed two out of three of the circled sub-tasks during the low workload condition (system monitoring and resource management tasks), and an additional third task during the high workload condition (tracking task).

workload condition, the participants performed two monitoring tasks using the keyboard, i.e., the system monitoring and the resource management tasks. The system monitoring task was presented in the upper left window of the display. As explained in the article of Comstock and Arnegard (1992), the demands of monitoring gages and warning lights were simulated here. The participants had to respond to the absence of the green light, the presence of the red light, and to monitor the four moving pointer dials deviation from midpoint. Regarding the resource management task, it simulated the demands of fuel management. The participants had to maintain tanks A and B at 2500 units each. This was done by turning on or off any of the eight pumps, which can sometimes fail.

In the high workload condition, they had an additional tracking task to manage in parallel. The tracking task was located in the upper middle window and simulated the demands of manual control. The participants had to keep the target at the center of the window using the joystick. Therefore, in both the low and high workload conditions perceptual, attentional, and decision making processes are recruited, along with motor preparation and performance. The difference between the low and the high workload conditions only stems in the additional workload imposed by the additional task.

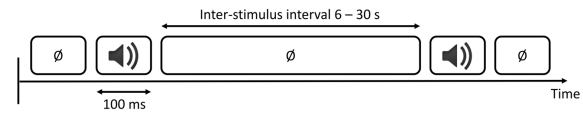


FIGURE 2 | Single-stimulus paradigm using ignored and infrequent auditory probes.

In addition to the visual stimulations induced by the MATB-II, the participants received auditory stimuli. They were instructed to ignore these auditory stimuli and to focus on the task at hand. These stimuli were sent by the Eprime software (E-prime Psychology Software Tools, Inc., Pittsburgh, PA, USA) into their Sennheiser audioset. In a similar fashion to the single-stimulus paradigm of Allison and Polich (2008), they consisted of 100 ms 1000 Hz pure tones (10 ms rise/fall, 65 dB SPL), with a random 6–30 s inter-tone interval (**Figure 2**). A minimum of 30 stimulations per block were presented.

Data Acquisition

Data acquisition was performed at the IRMaGe Neurophysiology facility (Grenoble, France). The participants' answers to the Rating Scale Mental Effort questionnaire (RSME; Zijlstra, 1993)

and their resource management task root mean square (RMS) error scores were recorded, as well as their EEG activity using an Acticap® (Brain Products, Inc.) equipped with 32 Ag-AgCl unipolar active electrodes that were positioned according to the 10–20 system. The reference and ground electrodes used for acquisition were those of the Acticap, i.e., FCz for the reference, and AFz for the ground. The electro-oculographic activity was also recorded using two electrodes positioned at the eyes outer canthi, and two respectively above and below the left eye. Impedance was kept below 10 kΩ for all electrodes. The signal was amplified using a BrainAmp^{TM} system (Brain Products, Inc.) and sampled at 500 Hz with a 0.1 Hz high-pass filter and a 0.1 μV resolution. Participants were instructed to limit eye and body movements during the task.

Signal Processing

The processing chain is detailed in **Figure 3**. In a general manner, the raw data was preprocessed, then spatially filtered, and lastly classified. Details are given in the following sub-sections regarding each step of this chain. It should be noted that the same processing chain was replicated without the spatial filtering step in order to evaluate if spatial filtering enhances the discriminability of the two workload levels.

Preprocessing

The digital EEG signal was band-pass filtered between 1 and 40 Hz, and re-referenced to a common average reference. The signal was then epoched starting 100 ms before and ending 600 ms after the auditory stimulation. Next, artifacts related to ocular movements (saccades and blinks) were corrected using the signal recorded from the electrooculographic electrodes (EOG) and the Second Order Blind Identification algorithm (SOBI; Belouchrani et al., 1997). This algorithm was chosen to perform the source decomposition because thanks to its assumption of non-correlation –and not mutual independence- it has been shown to be more suitable for electrophysiological data by Congedo et al. (2008). In order to get closer to a system that could be implemented on-line in a real-life setting, the two sources that were the most correlated to the EOG activity were canceled. All trials were kept for analysis. The AEPs were then extracted by subtracting a 100 ms baseline (i.e., mean signal amplitude) to the 600 ms segment that starts at the onset of the stimulation. Lastly, the data was decimated to 100 Hz using a five-point moving average.

Spatial Filtering

Then, the preprocessed data $\mathbf{X}$ (N_s – number of samples × N_e – number of channels) were spatially filtered, resulting in the signal $\mathbf{Z} = \mathbf{WX}$ (N_s – number of samples × N_f – number of spatial filters). Each column of the matrix $\mathbf{W}$ contains a spatial filter with its spatial pattern in the corresponding column of $\mathbf{A} = (\mathbf{W}^{-1})^{\mathrm{T}}$. In this paper, we use CCA as a spatial filtering method. As Spüler et al. (2014) detailed it , in a two-class scenario the CCA filters are computed in order to maximize the correlation between the EEG signal $\mathbf{X}$ and the matrix $\mathbf{Y} = \mathbf{D}_1\mathbf{P}_1 + \mathbf{D}_2\mathbf{P}_2$ that contains the time replication of the average ERP responses $\mathbf{P}_i$ for each class. The matrix $\mathbf{D}_i$ is a Toeplitz binary matrix that indicates the stimulation onset for the ith class (Rivet et al., 2009).

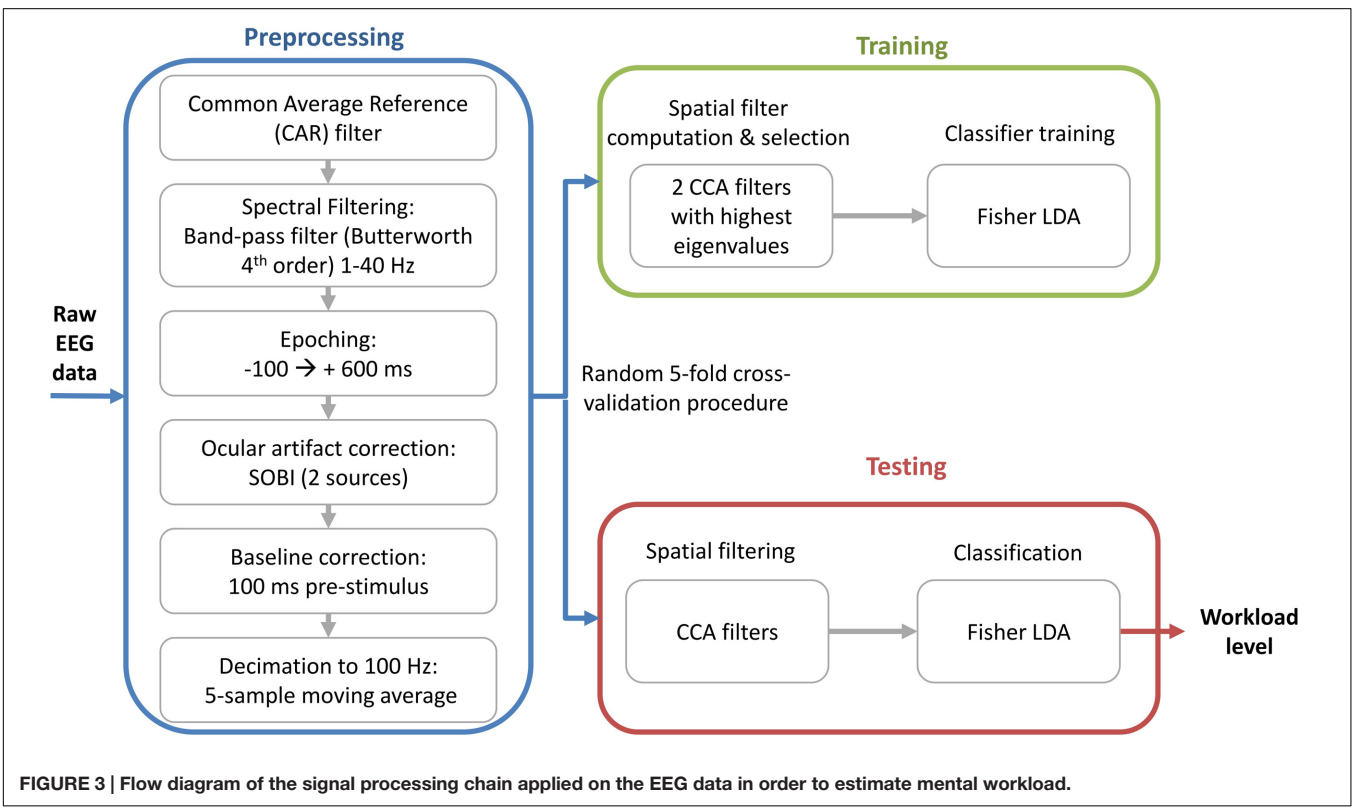


FIGURE 3 | Flow diagram of the signal processing chain applied on the EEG data in order to estimate mental workload.

Several methods have been proposed to solve CCA by computing orthonormal bases for the data matrices either by QR or singular value decomposition – SVD (Björck and Golub, 1973).

The CCA spatial filters were computed using the training data only. Then, the spatial filters with the two highest associated canonical correlations were selected. When these filters are applied on the testing data, the feature vector for the j^{th} trial is given by the column concatenation $\mathbf{f}_j = \text{vec}(X_j[\mathbf{w}_1\mathbf{w}_2])$ with dimension 120x1 (i.e., 60 samples × 2 virtual electrodes). In order to have the same number of features for both processing chains (with and without spatial filtering), for the chain without spatial filtering the feature vector was composed of the concatenated signals of the C3 and Pz electrodes (chosen visually using the average spatial patterns presented in Section "Spatial patterns").

Classification

A single-trial classification was performed on the feature vector **f** using a Fisher linear discriminant analysis (FLDA), with a shrinkage estimation of the covariance matrices (Schäfer and Strimmer, 2005). As explained by Blankertz et al. (2011), this estimation method allows the use of LDA with high dimensional features and gives good results that can generalize well (Blankertz et al., 2011). We used a random fivefold cross-validation procedure. The spatial filters were learned on the training set, and applied on the testing set. In the same way, the shrinkage estimation was learned on the training set. The performance of the processing chains was assessed based on their intra-subject binary classification accuracy.

Statistical Analyses

Statistical analyses were carried out on all results, i.e., subjective results from the RSME questionnaire, N1, P1, N2, P2, and P3 peak amplitude and latency from the AEP components, and classification results obtained using the processing chains with and without spatial filtering. All results were compared between themselves using repeated measures ANOVAs and Tukey *post hoc* tests. The significance level was set at 0.05.

RESULTS

Behavioral and Subjective Data

In a similar manner to Fournier et al. (1999), behavioral responses were standardized within each participant by dividing their response times to the resource management tasks by their proportion of correct responses. There was a significant effect of workload on this performance score ($t = 2.99$, $p < 0.05$), the participants' performance was significantly degraded in the high workload condition compared to the low workload condition (m1_perf = 0.33; sd1_perf = 0.12; m2_perf = 0.43; sd2_perf = 0.12). Moreover, the participants reported having furnished a significantly bigger effort in the high workload condition than in the low workload one [$F(1,7) = 38.04$, $p < 0.01$; m1_RSME = 45.5; sd1_RSME = 18.2; m2_RSME = 71.6; sd2_RSME = 24.3].

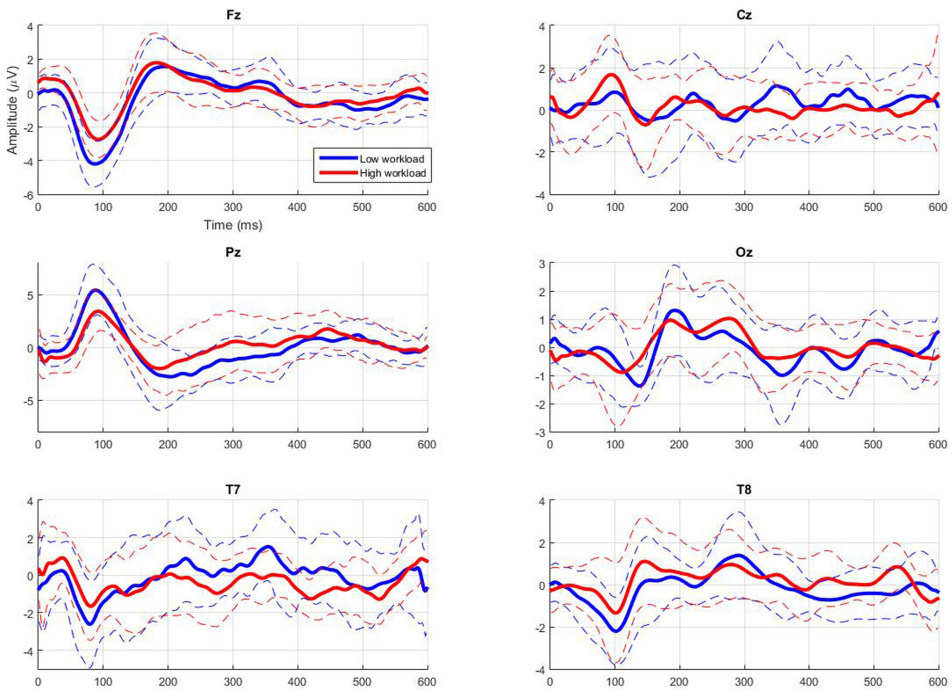


FIGURE 4 | Grand average (in bold) and standard deviation (dotted line) of the auditory evoked potentials (AEPs) elicited by the ignored infrequent auditory probes depending on workload condition at major midline electrode sites (Fz, Cz, Pz, and Oz), as well as at auditory processing relevant sites (T7 and T8).

Auditory Evoked Potentials

Figure 4 gives the grand-average AEPs across participants at major median electrode sites (Fz, Cz, Pz, and Oz), as well as at electrode sites located close to the auditory cortex (T7 and T8). **Figure 5** also gives the individual AEPs for the eight participants at the Pz electrode site (chosen to illustrate the results that follow regarding early components). The typical components reported to be modulated by workload can be noticed, i.e., N1, P1, P2, N2, and P3 (Kramer et al., 1995; Kok, 2001; Ullsperger et al., 2001; Schultheis and Jameson, 2004; Gomarus et al., 2006; Allison and Polich, 2008; Holm et al., 2009; Miller et al., 2011; Boonstra et al., 2013). However, the statistical analyses revealed only few significant results at the group level, which is understandable given the mostly overlapping variance of both signals (see standard deviations in **Figure 4**). Indeed, with increasing workload there were only trends at the Pz electrode for a decrease in amplitude of the P1 component ($p = 0.11$; **Figure 5**) and for a decrease in latency of the N1 component ($p = 0.07$). Moreover, when workload increased there was a significant decrease in latency of the P2 component at all electrode sites [$F(1,7) = 6.74$, $p < 0.05$].

Spatial Patterns

The topographical representation of each of the two CCA spatial patterns obtained for each participant using the processing chain proposed in this paper are presented respectively in **Figures 6** and **7** (average across training folds). The first spatial pattern reveals that in order to better discriminate workload levels, our first selected spatial filter enhances the activity from centro-parieto-occipital regions -consistent with attentional processing, while the second one enhances the activity from temporal regions -consistent with auditory processing- as well as prefrontal areas which could be related to ocular activity.

Filtered EEG Signal

The filtered EEG signals obtained from the testing sets for each participant using the first and second CCA filters are presented respectively in **Figures 8** and **9** (grand average across cross-validation folds). Both filters seem to mainly enhance the ERP activity of the low workload condition while decreasing it for the high workload condition from around 50 to 400 ms. This is particularly true for participants 3 and 4 for the early components. The signal's polarity fluctuates in a different manner depending on the participant and the filter, however, a general pattern emerges. Particularly, for both filters we can see an enhancement in the low workload condition of the amplitude of the early components that peak between 80 and 250 ms, be it in the negative or in the positive range. Therefore it seems that the filters act in a way so that they enhance the relevance of early auditory evoked components but not so much of later components.

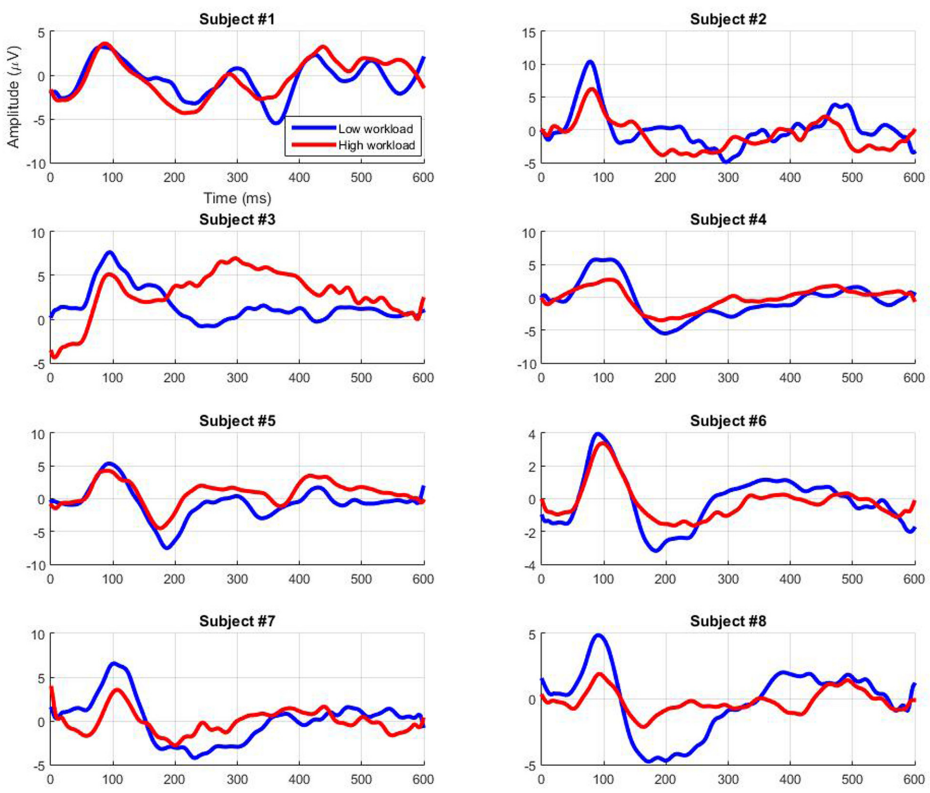


FIGURE 5 | Auditory evoked potentials elicited by the ignored infrequent auditory probes depending on workload condition at the Pz electrode site for all participants (grand average across trials).

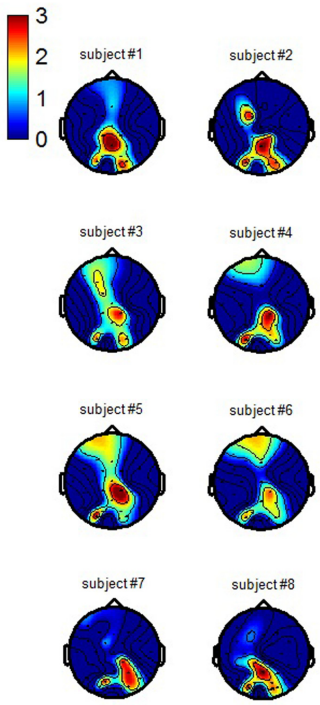


FIGURE 6 | Individual CCA spatial patterns of the first filter used to enhance the discrimination of the workload condition (grand average across training folds).

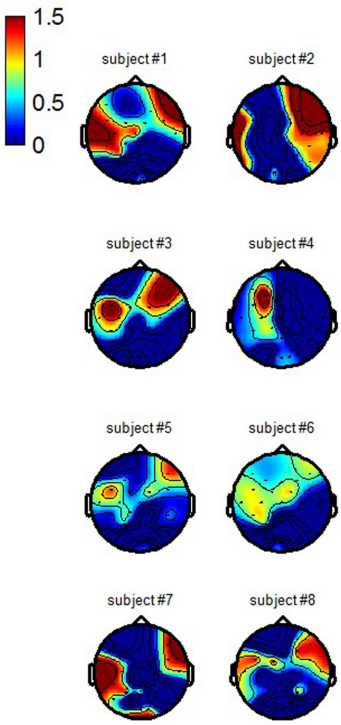


FIGURE 7 | Individual CCA spatial patterns of the second filter used to enhance the discrimination of the workload condition (grand average across training folds).

Classification Accuracy

The workload level classification results obtained using the single-stimulus paradigm and both the processing chain that includes a CCA spatial filtering step and the simpler processing chain without spatial filtering are given by **Figure 10** for each participant. There was a significant effect of the type of processing chain [$F(1,7) = 39.90, p < 0.001$]. Indeed, the chain that included the CCA spatial filtering step gave higher classification results than the one that didn't. The mean percentage of correct binary classification across the eight participants was 90.51% ($\pm$ 10.7 SD) and 71.49% ($\pm$ 15.9 SD) respectively for the processing chains with and without spatial filtering. Using the chain that included the CCA filtering, the performance was optimal for participant 4 with a classification accuracy of 100% and a null standard deviation, and the lowest performance was obtained for participant 7 with a classification accuracy of 80% and a very large standard deviation of 21.73.

DISCUSSION

Studies have demonstrated that workload modulates the ERPs elicited by attended or ignored auditory probes in a classical oddball paradigm involving deviant and standard tones (Kramer et al., 1995). Allison and Polich (2008) had also demonstrated this phenomenon using only infrequent standard tones (i.e., single-stimulus paradigm). However, to our knowledge, there was no literature regarding effective workload *classification* based on ignored auditory probes. Indeed, no signal processing chain had been applied to estimate workload in an automatic way from the ERPs of ignored auditory stimuli. Hence, this study was intended to bring new light on the potential use of ignored infrequent task-independent probes to efficiently and *automatically* assess mental workload in a minimally intrusive way. In order to do so, a single-stimulus paradigm similar to that of Allison and Polich (2008) was used, along with a processing chain that included a CCA spatial filtering step. The participants rated their effort as significantly higher for the high workload condition than for the low one and also exhibited a decrease in performance in the high workload condition compared to the low workload condition akin to that observed by Fournier et al. (1999). Their ERPs revealed only trends for a decrease in P1 amplitude and N1 latency, as well as a significant decrease in P2 latency. These results are in line with the literature regarding resource allocation processes. In a general manner, the amplitude of the ERP components that occur within the first 250 ms following stimulus onset has been demonstrated to be influenced by attentional capacity allocated to the eliciting stimulus and task operations (for a review see Kok, 1997). For instance, the P1 component amplitude is larger in active relative to passive viewing conditions (Fu et al., 2010). As for the N1 and the

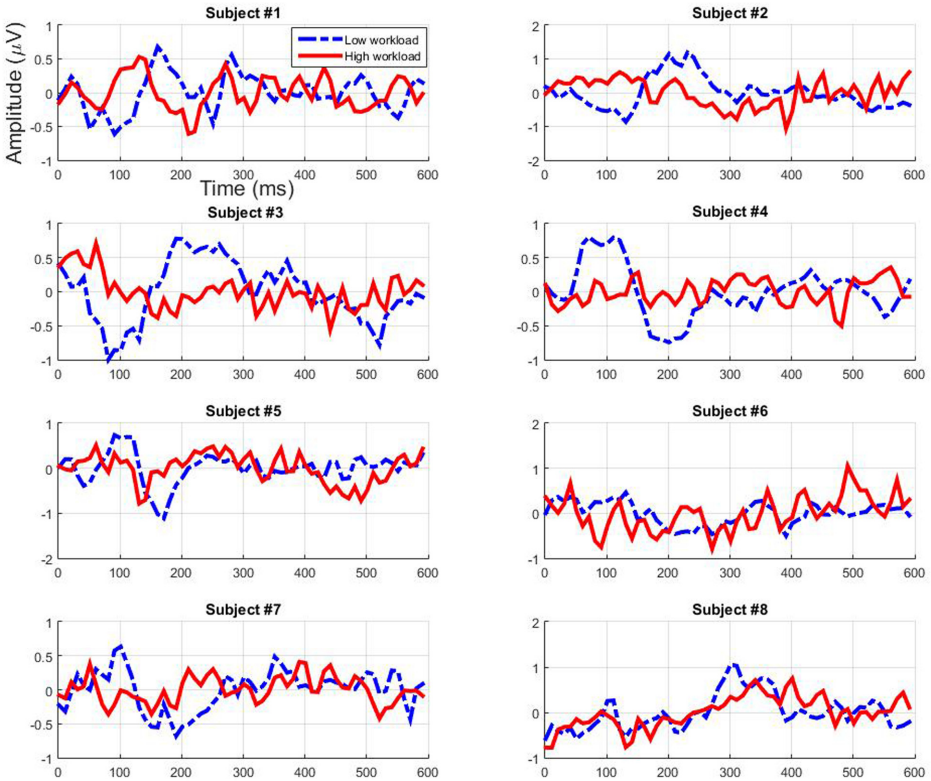


FIGURE 8 | Individual filtered test data – using the first CCA filter – depending on workload condition (grand average across cross-validation folds).

P2 latency, it has also been shown to decrease with a decrease in allocated attentional resources (Okita, 1979; Callaway and Halliday, 1982). However, more differences in amplitude were expected based on articles that describe workload modulations for ERPs elicited by task-dependent stimuli, specifically on late ERP components' amplitude (Kok, 2001; Ullsperger et al., 2001; Schultheis and Jameson, 2004; Gomarus et al., 2006; Holm et al., 2009; Miller et al., 2011; Boonstra et al., 2013). Yet, Kramer et al. (1995) had found that only the early components of the ERPs elicited by ignored task-irrelevant probes were relevant to perform a non-intrusive workload assessment, and that the late P300 component was not a good marker for such a goal. Our results confirm theirs as to which components are significantly modulated by workload using ignored auditory probes.

Despite the few significant results obtained at the group level regarding AEP components' amplitude, very accurate mental workload estimations were obtained using a signal processing chain that included a CCA spatial filtering step with at least 80% of correct binary classification accuracy for all participants, and an average of 90.51%. This result is in line with the literature that shows that classifiers can reveal statistical differences when standard statistical tests between ERPs do not (Noh and de Sa, 2014). Also, here the use of the CCA spatial filtering step significantly enhanced the estimation performance, as already demonstrated by Roy et al. (2015), and also reduced the variance in the results. Besides, this is only slightly lower than what Roy et al. (2016) obtained using task-independent visual probes -91%. This is very promising given that here, opposite to their protocol, the task-independent probes required no overt response and were ignored by the participants. Moreover, the probes used in this experiment are auditory while they were visual in their protocol. Lastly, those results are also higher than that obtained by previous studies that classified raw or spatially filtered ERPs elicited by task-dependent probes (Brouwer et al., 2012; Mühl et al., 2014). Therefore, the use of the single-stimulus paradigm coupled to a processing chain that includes a spatial filtering step allows a precise estimation of mental workload for a task that is very close to an actual work task. A limitation to this study is the number of trials, although in ecological settings it will be difficult to use more probes and to remain minimally intrusive. Nevertheless, according to Combrisson and Jerbi (2015), if we have more than 20 trials our performance should be over 70% in order to account for a significant detection with a $p < 0.05$ significance rate. Here, using the spatial filtering step we obtained at least 80% of correct detections, and an average of 90.51% with a minimum of 30 trials per condition. Therefore, we can say that our results were significantly above chance and that our method is quite efficient.

The spatial patterns of the selected CCA filters revealed that an enhancement of temporal and centro-parietal activity allowed reaching such high classification results. This is in accordance with the auditory nature of our probes. It is interesting to note that the activity that was enhanced by the spatial filters in the

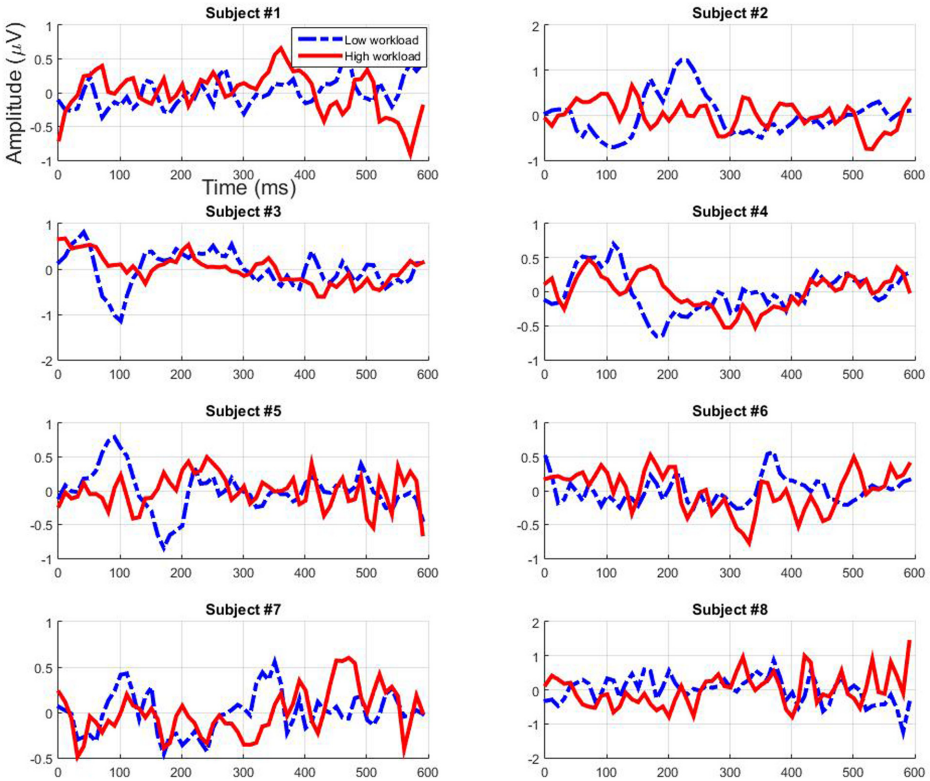


FIGURE 9 | Individual filtered test data – using the second CCA filter – depending on workload condition (grand average across cross-validation folds).

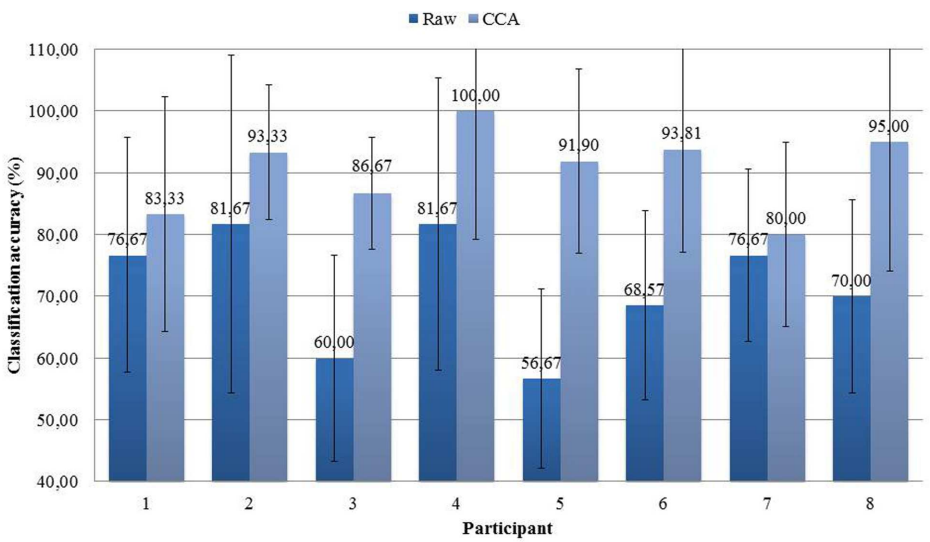


FIGURE 10 | Workload level classification accuracy reached for each participant using the raw data of two electrodes or a CCA spatial filtering step (mean and standard deviation of the percentage of correct binary classification across the random fivefold cross-validation procedure).

previously mentioned study of Roy et al. (2016) who used task-independent visual probes originated from the occipital sites, in accordance with the visual nature of their stimuli. In our study, given that the probes were auditory, we observed a specific enhancement of the activity from the temporal electrode sites. The signal from the centro-parietal sites was also enhanced in

their study as in ours. These sites are known to be involved with attentional processing, and more generally with resource engagement (Kok, 1997). Additionally, the patterns also revealed an implication of prefrontal sites, which could stem from an under-efficient ocular artifact correction step in our processing chain. Indeed, in order to preserve the cerebral activity as much as possible, we only deleted the 2 out of 32 sources that were the most correlated to respectively the vertical and horizontal EOG channels. Also, given that the MATB-II is a task that is very close to a real work task, it elicits more ocular movements than classical laboratory tasks during which participants are asked to fixate the center of the screen and to limit eye movements and blinks. In any case, if it is indeed ocular activity that our second spatial filter enhanced, it means that this ocular activity allows efficient mental workload estimation. This is not surprising given that blink frequency has been reported to vary depending on task difficulty (Holland and Tarlow, 1972; Tanaka and Yamaoka, 1993). Moreover, it is known that the appearance of an unexpected stimulation leads to a startle eyeblink reflex. This reflex is attenuated during a multiple-task –high workload- compared to a single-task condition –low workload (Neumann, 2002). Therefore, the ocular activity produced in response to an infrequent auditory probe could be an efficient marker of task engagement and mental workload. As Roy et al. (2014) already argued , if ocular activity is helping to discriminate workload levels, why remove it? Hence, it might be interesting for future developments to use a processing chain that either does not include an ocular artifact correction step, or, that does but performs classification by fusing two feature vectors, a clean EEG one and an ocular activity one. Multimodality in terms of origin of the physiological markers (e.g., cerebral or ocular) could therefore be the key to enhance classification accuracy for real-life implementations.

Besides, this study evaluates the relevance of a stimulation paradigm and its dedicated processing chain for an ecological task which is the MATB. Although still in a laboratory setting, this task is very close to that performed by pilots and air traffic controllers. However, it modulates workload only by varying the number of tasks to perform in parallel, that is to say by varying the participants' degree of divided attention. In order to pursue the evaluation of the relevance of this stimulation paradigm, future work should focus on an evaluation of its relevance for several tasks that modulate workload based on different cognitive functions, e.g., working memory load, divided attention, executive functions. To our knowledge, only Berka et al. (2007) assessed the relevance of an EEG marker across several types of tasks. But, they focused on frequency power in the classical EEG bands. Thus, in order to progress toward an efficient estimation in real-life settings, the literature still lacks a thorough comparison of ERP modulations due to workload across several tasks. Also, although the participants of our study told us that they were not annoyed by the auditory infrequent stimulations and generally entirely forgot about it, a more thorough investigation of the real cost of such a paradigm in terms of operator fatigue and efficiency should be carried out. What's more, in order to increase the practicality of EEG measures, the number of electrodes should be diminished. However, this study, along with that of Roy et al. (2015) clearly establishes the relevance of a spatial filtering step in order to enhance the discriminability between the two workload levels. Therefore, future studies should evaluate how to reduce the number of electrodes while keeping enough channels to efficiently apply such a filtering step.

Consequently, this study contributes to the neuroergonomics research topic on mental workload estimation by uncovering three main points. First, the single-stimulus paradigm in which participants are probed by infrequent task-independent and ignored probes allows minimally intrusive workload estimation. Second, a spatial filtering step such as a CCA filtering enables a very accurate AEP-based workload classification. Lastly, the combination of this single-stimulus paradigm with infrequent ignored probes and its dedicated processing chain allows efficient workload estimation for an ecologically valid task such as the MATB.

CONCLUSION

This study has demonstrated as a proof-of-concept that a single-stimulus paradigm based on infrequent ignored auditory probes and its dedicated processing chain could allow a very accurate estimation of mental workload with a classification performance above 80% for every participant. This is also the first study to effectively classify workload based on ERPs elicited by ignored stimuli for a task that is very close to a real-life work situation. It paves the way toward the efficient use of ERPs for MSM and brings us closer to the implementation of user adaptive systems in ecological settings.

AUTHOR CONTRIBUTIONS

Study conception and design: RR, SB, SC, AC. Acquisition of data: RR. Analysis and interpretation of data: RR, SB. Drafting of manuscript: RR. Critical revision: SB, SC, AC.

ACKNOWLEDGMENTS

Grenoble Neurophysiology facility IRMaGe was partly funded by the French program "Investissement d'Avenir" run by the "Agence Nationale pour la Recherche": Grant "Infrastructure d'Avenir en Biologie Santé" (ANR-11-INBS-0006).

REFERENCES

Allison, B. Z., and Polich, J. (2008). Workload assessment of computer gaming using a single stimulus event-related potential paradigm. *Biol. Psychol.* 77, 277–283. doi: 10.1016/j.biopsycho.2007.10.014

Belouchrani, A., Abed-Meraim, K., Cardoso, J.-F., and Moulines, E. (1997). A blind source separation technique using second-order statistics. *IEEE Trans. Signal Process.* 45, 434–444. doi: 10.1109/78.554307

Berka, C., Levendowski, D. J., Lumicao, M. N., Yau, A., Davis, G., Zivkovic, V. T., et al. (2007). EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. *Aviat. Space Environ. Med.* 78, 231–244.

Björck, A., and Golub, G. H. (1973). Numerical methods for computing angles between linear subspaces. *Math. Comput.* 27, 579–594. doi: 10.2307/2005662

Blankertz, B., Lemm, S., Treder, M., Haufe, S., and Müller, K.-R. (2011). Single-trial analysis and classification of ERP components: a tutorial. *Neuroimage* 56, 814–825. doi: 10.1016/j.neuroimage.2010.06.048

Boonstra, T. W., Powell, T. Y., Mehrkanoon, S., and Breakspear, M. (2013). Effects of mnemonic load on cortical activity during visual working memory: linking ongoing brain activity with evoked responses. *Int. J. Psychophysiol.* 89, 409–418. doi: 10.1016/j.ijpsycho.2013.04.001

Brouwer, A.-M., Hogervorst, M. A., van Erp, J. B. F., Heffelaar, T., Zimmerman, P. H., and Oostenveld, R. (2012). Estimating workload using EEG spectral power and ERPs in the n-back task. *J. Neur. Eng.* 9:045008. doi: 10.1088/1741-2560/9/4/045008

Cain, B. (2007). *A Review of the Mental Workload Literature 1.0*. Technical report, Defense Research and Development, Toronto, ON.

Callaway, E., and Halliday, R. (1982). The effect of attentional effort on visual evoked potential N1 latency. *Psychiatry Res.* 7, 299–308. doi: 10.1016/0165-1781(82)90066-X

Combrisson, E., and Jerbi, K. (2015). Exceeding chance level by chance: the caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy. *J. Neurosci. Methods* 250, 126–136. doi: 10.1016/j.jneumeth.2015.01.010

Comstock, J. R. Jr., and Arnegard, R. J. (1992). *The Multi-Attribute Task Battery for Human Operator Workload and Strategic Behavior Research (NASA TM-104174)*. Hampton, VA: NASA Langley Research Center.

Congedo, M., Gouy-Pailler, C., and Jutten, C. (2008). On the blind source separation of human electroencephalogram by approximate joint diagonalization of second order statistics. *Clin. Neurophysiol.* 119, 2677–2686. doi: 10.1016/j.clinph.2008.09.007

Fournier, L. R., Wilson, G. F., and Swain, C. R. (1999). Electrophysiological, behavioral, and subjective indexes of workload when performing multiple tasks: manipulations of task difficulty and training. *Int. J. Psychophysiol.* 31, 129–145. doi: 10.1016/S0167-8760(98)00049-X

Friedrich, E. V., Scherer, R., Sonnleitner, K., and Neuper, C. (2011). Impact of auditory distraction on user performance in a brain–computer interface driven by different mental tasks. *Clin. Neurophysiol.* 122, 2003–2009. doi: 10.1016/j.clinph.2011.03.019

Fu, S., Fedota, J. R., Greenwood, P. M., and Parasuraman, R. (2010). Dissociation of visual C1 and P1 components as a function of attentional load: an event-related potential study. *Biol. Psychol.* 85, 171–178. doi: 10.1016/j.biopsycho.2010.06.008

Gevins, A., and Smith, M. E. (2007). "Electroencephalography (EEG) in neuroergnomics," in *Neuroergonomics: The Brain at Work*, eds R. Parasuraman and M. Rizzo (New York, NY: Oxford University Press Inc), 15–31.

Gomarus, H. K., Althaus, M., Wijers, A. A., and Minderaa, R. B. (2006). The effects of memory load and stimulus relevance on the EEG during a visual selective memory search task: an ERP and ERD/ERS study. *Clin. Neurophysiol.* 117, 871–884. doi: 10.1016/j.clinph.2005.12.008

Hoffmann, U., Vesin, J., and Ebrahimi, T. (2006). "Spatial filters for the classification of event-related potentials," in *Proceeding of the European Symposium Artificial Neural Networks*, Bruges.

Holland, M. K., and Tarlow, G. (1972). Blinking and mental load. *Psychol. Rep.* 31, 119–127. doi: 10.2466/pr0.1972.31.1.119

Holm, A., Lukander, K., Korpela, J., Sallinen, M., and Müller, K. M. I. (2009). Estimating brain load from the EEG. *ScientificWorld J.* 9, 639–651. doi: 10.1100/tsw.2009.83

Hotelling, H. (1936). Relations between two sets of variates. *Biometrika* 28, 321–377. doi: 10.2307/2333955

Kok, A. (1997). Event-related-potential (ERP) reflections of mental resources: a review and synthesis. *Biol. Psychol.* 45, 19–56. doi: 10.1016/S0301-0511(96)05221-0

Kok, A. (2001). On the utility of P3 amplitude as a measure of processing capacity. *Psychophysiology* 38, 557–577. doi: 10.1017/S0048577201990559

Kramer, A. F., Sirevaag, E. J., and Braune, R. (1987). A psychophysiological assessment of operator workload during simulated flight missions. *Hum. Factors* 29, 145–160.

Kramer, A. F., Trejo, L. J., and Humphrey, D. (1995). Assessment of mental workload with task-irrelevant auditory probes. *Biol. Psychol.* 40, 83–100. doi: 10.1016/0301-0511(95)05108-2

Mertens, R., and Polich, J. (1997). P300 from a single-stimulus paradigm: passive versus active tasks and stimulus modality. *Electroencephalogr. Clin. Neurophysiol.* 104, 488–497. doi: 10.1016/S0168-5597(97)00041-5

Miller, M. W., Rietschel, J. C., McDonald, C. G., and Hatfield, B. D. (2011). A novel approach to the physiological measurement of mental workload. *Int. J. Psychophysiol.* 80, 75–78. doi: 10.1016/j.ijpsycho.2011.02.003

Mühl, C., Jeunet, C., and Lotte, F. (2014). EEG-based workload estimation across affective contexts. *Front. Neurosci.* 8:114. doi: 10.3389/fnins.2014.00114

Natani, K., and Gomer, F. E. (1981). *Electrocortical Activity and Operator Workload: A Comparison of Changes in the Electroencephalogram and in Event-Related Potentials*. McDonnell Douglas Technical Report E2427. Long Beach CA: McDonnell Douglas Corporation.

Neumann, D. L. (2002). Effect of varying levels of mental workload on startle eyeblink modulation. *Ergonomics* 45, 583–602. doi: 10.1080/00140130210148546

Noh, E., and de Sa, V. R. (2014). "Discriminative dimensionality reduction for analyzing EEG data," in *Proceedings of the 36th Annual Meeting of the Cognitive Science Society*, Austin, TX, 1090–1095.

Okita, T. (1979). Event-related potentials and selective attention to auditory stimuli varying in pitch and localization. *Biol. Psychol.* 9, 271–284. doi: 10.1016/0301-0511(79)90027-9

Parasuraman, R., Christensen, J., and Grafton, S. (2012). Neuroergonomics: The brain in action and at work. *Neuroimage* 59, 1–3. doi: 10.1016/j.neuroimage.2011.08.011

Rivet, B., Souloumiac, A., Attina, V., and Gibert, G. (2009). xDAWN algorithm to enhance evoked potentials: application to brain-computer interface. *IEEE Trans. Biomed. Eng.* 56, 2035–2043. doi: 10.1109/TBME.2009.2012869

Roy, R. N., Bonnet, S., Charbonnier, S., Jallon, P., and Campagne, A. (2015). A comparison of ERP spatial filtering methods for optimal mental workload estimation. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 2015, 7254–7257. doi: 10.1109/EMBC.2015.7320066

Roy, R. N., Charbonnier, S., and Bonnet, S. (2014). Eye blink characterization from frontal EEG electrodes using source separation and pattern recognition algorithms. *Biomed. Signal Process. Control* 14, 256–264. doi: 10.1016/j.bspc.2014.08.007

Roy, R. N., Charbonnier, S., Campagne, A., and Bonnet, S. (2016). Efficient mental workload estimation using task-independent EEG features. *J. Neural Eng.* 13:026019. doi: 10.1088/1741-2560/13/2/026019

Schäfer, J., and Strimmer, K. (2005). A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Stat. Appl. Genet. Mol. Biol.* 4, 1175–1189.

Schultheis, H., and Jameson, A. (2004). Assessing cognitive load in adaptive hypermedia systems: physiological and behavioral methods. *Lect. Notes Comput. Sci.* 313, 225–234. doi: 10.1007/978-3-540-27780-4_26

Sirevaag, E., Kramer, A., Wickens, C. D., Reisweber, M., Strayer, D., and Grenell, J. (1993). Assessment of pilot performance and mental workload in rotary wing aircraft. *Ergonomics* 36, 1121–1140. doi: 10.1080/00140139308967983

Spüler, M., Walter, A., Rosenstiel, W., and Bogdan, M. (2014). Spatial filtering based on canonical correlation analysis for classification of evoked or event-related potentials in EEG data. *IEEE Trans. Neural Syst. Rehabil. Eng.* 22, 1097–1103. doi: 10.1109/TNSRE.2013.2290870

Sternberg, S. (1969). Memory-scanning: mental processes revealed by reaction-time experiments. *Am. Sci.* 57, 421–457.

Tanaka, Y., and Yamaoka, K. (1993). Blink activity and task difficulty. *Percept. Mot. Skills* 77, 55–66. doi: 10.2466/pms.1993.77.1.55

Ullsperger, P., Freude, G., and Erdmann, U. (2001). Auditory probe sensitivity to mental workload changes – an event-related potential study. *Int. J. Psychophysiol.* 40, 201–209. doi: 10.1016/S0167-8760(00)00188-4

van Erp, J., Lotte, F., and Tangermann, M. (2012). Brain-computer interfaces: beyond medical applications. *Computer* 45, 26–34. doi: 10.1109/MC.2012.107

Zander, T. O., and Kothe, C. (2011). Towards passive brain–computer interfaces: applying brain–computer interface technology to human–machine systems in general. *J. Neural Eng.* 8:025005. doi: 10.1088/1741-2560/8/2/025005

Zijlstra, F. R. H. (1993). *Efficiency in Work Behavior*. Doctoral dissertation, Technical University, Delft.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Roy, Bonnet, Charbonnier and Campagne. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 11 January 2017
doi: 10.3389/fnhum.2016.00647

Gaussian Process Regression for Predictive But Interpretable Machine Learning Models: An Example of Predicting Mental Workload across Tasks

Matthew S. Caywood[1], Daniel M. Roberts[1,2], Jeffrey B. Colombe[1], Hal S. Greenwald[1] and Monica Z. Weiland[1]*

[1] The MITRE Corporation, McLean, VA, USA, [2] Department of Psychology, George Mason University, Fairfax, VA, USA

OPEN ACCESS

Edited by:
Thorsten O. Zander,
Technical University of Berlin,
Germany

Reviewed by:
Virginia R. De Sa,
Vanderbilt University, USA
Nicolas Langer,
Child Mind Institute, USA
Sheng-Hsiou Hsu,
University of California, San Diego,
USA

***Correspondence:**
Jeffrey B. Colombe
jcolombe@mitre.org

Received: 23 October 2015
Accepted: 05 December 2016
Published: 11 January 2017

Citation:
Caywood MS, Roberts DM, Colombe JB, Greenwald HS and Weiland MZ (2017) Gaussian Process Regression for Predictive But Interpretable Machine Learning Models: An Example of Predicting Mental Workload across Tasks. Front. Hum. Neurosci. 10:647. doi: 10.3389/fnhum.2016.00647

There is increasing interest in real-time brain-computer interfaces (BCIs) for the passive monitoring of human cognitive state, including cognitive workload. Too often, however, effective BCIs based on machine learning techniques may function as "black boxes" that are difficult to analyze or interpret. In an effort toward more interpretable BCIs, we studied a family of N-back working memory tasks using a machine learning model, Gaussian Process Regression (GPR), which was both powerful and amenable to analysis. Participants performed the N-back task with three stimulus variants, auditory-verbal, visual-spatial, and visual-numeric, each at three working memory loads. GPR models were trained and tested on EEG data from all three task variants combined, in an effort to identify a model that could be predictive of mental workload demand regardless of stimulus modality. To provide a comparison for GPR performance, a model was additionally trained using multiple linear regression (MLR). The GPR model was effective when trained on individual participant EEG data, resulting in an average standardized mean squared error (sMSE) between true and predicted N-back levels of 0.44. In comparison, the MLR model using the same data resulted in an average sMSE of 0.55. We additionally demonstrate how GPR can be used to identify which EEG features are relevant for prediction of cognitive workload in an individual participant. A fraction of EEG features accounted for the majority of the model's predictive power; using only the top 25% of features performed nearly as well as using 100% of features. Subsets of features identified by linear models (ANOVA) were not as efficient as subsets identified by GPR. This raises the possibility of BCIs that require fewer model features while capturing all of the information needed to achieve high predictive accuracy.

Keywords: EEG, BCI, Gaussian Process Regression, machine learning, neuroergonomics

INTRODUCTION

Neuroimaging methods, particularly inexpensive and non-invasive techniques such as electroencephalography (EEG) and functional near infrared spectroscopy (fNIRS), are increasingly being used to continuously assess the cognitive state of individuals during task performance, an example of Neuroergonomics (Parasuraman, 2003; Parasuraman and Rizzo, 2006). This information can be used to better understand the demands of the task being performed, assess the limitations of the individual, or be fed back into the system to adjust the task relative to the individual's current state. The use of physiological data to assess operator state has also recently been described as a 'passive' brain-computer interface (BCI) (Zander et al., 2009; Zander and Kothe, 2011), in contrast to traditional 'active' BCIs which utilize physiological data to allow an individual to act on the outside world (Wolpaw and Wolpaw, 2012).

Workload, the demand on the individual's attention and working memory, is a cognitive state of special interest for passive measurement during task performance. Cognitive Load Theory (CLT) (Sweller et al., 1998) for example, suggests that maintaining an optimal level of workload for a given task can assist in learning new material. Further, Coyne et al. (2009) incorporate CLT with Multiple Resource Theory (MRT) (Wickens, 2008), which distinguishes between different modes of mental demand, suggesting that real-time measurement of participant workload could be utilized to optimally redirect mental demand across the several modes of resources available, for example presentation of information in a spatial versus verbal code as delineated by MRT (see Coyne et al., 2009). Physiological measures of workload have been sought in a variety of tasks including N-back (Grimes et al., 2008; Baldwin and Penaranda, 2011; Ayaz et al., 2012; Brouwer et al., 2012), the Sternberg Memory Scanning Task (Wilson and Fisher, 1995; Baldwin and Penaranda, 2011), memory span tests (Baldwin and Penaranda, 2011; Chaouachi et al., 2011), the MAT-B multi-tasking scenario (Wilson and Russell, 2003b; Kothe and Makeig, 2011); and operational simulations (Wilson and Russell, 2003a; Ayaz et al., 2012). The extent of this literature reflects scientific awareness of the limitations of behavioral or subjective workload assessment techniques, including limited sensitivity (Gevins and Smith, 2003; Just et al., 2003), subjective bias, and intrusiveness.

EEG based workload monitoring has been explored using a variety of different machine learning approaches, including step-wise linear discriminant analysis (SWDA) (Wilson and Fisher, 1995; Wilson and Russell, 2003a), artificial neural networks (ANN) (Wilson and Russell, 2003a; Baldwin and Penaranda, 2011), naïve Bayes models (Grimes et al., 2008), and least-angle regression (Kothe and Makeig, 2011).

To estimate workload as defined above based on EEG spectra, we applied a supervised machine learning approach, performing a statistical regression to take processed neurophysiological signals as inputs and to predict the load parameter N from the N-back task as an output. Many previous projects seeking to predict mental workload have used a classifier rather than a regressor approach. For example Baldwin and Penaranda (2011) used three working memory tasks, each with two levels of imposed difficulty, Wilson and Fisher (1995) used a battery of tasks, each with two levels of difficulty, while both Wilson and Russell (2003b) and Kothe and Makeig (2011) used the Multi-Attribute Task Battery with two levels of difficulty. Wilson and Russell (2003a) distinguished between up to seven conditions, using three different simulated ATC tasks with three, three, and one level of difficulty, respectively. However, the condition levels were treated as categorical, with the authors using ANN and stepwise discriminant analysis as classifiers to discriminate between data from each condition. Similarly, Grimes et al. (2008) predicted working memory load within 4 levels of the N-back task (0- through 3-back), but classified the levels as categorical labels, rather than as a continuous construct.

Treating mental workload as a series of categorical states has the effect of forcing estimates of workload to reside in discrete categorical bins without any continuous variation. The N-back task is comprised of discrete task load levels $N = 1$, 2, 3, and considered in isolation, this task is readily amenable to prediction based on a classifier. However, we conceptualize the mental state of workload as potentially lying along a continuum of values that the N-back task visits at discrete levels due solely to the structure of the task, not necessarily due to the inherent structure of working memory and attentional resources. The neurophysiological data is continuous in nature, and in order to preserve any potential information about workload as a continuously varying mental state, we treated the predicted N as a continuous variable even though all the training data for N was discrete. This required the use of a regression method rather than a classification method. One consequence of treating workload as a continuous measure is that the appropriate measure of error to be minimized in supervised training, as well as for operational testing, is continuous rather than discrete. For this reason, we present predictor performance primarily in standardized mean square error (sMSE), discussed more fully in the Section "Materials and Methods."

Our choice of regression on a continuous task load variable was also motivated by a follow-on application of methods described here for estimating cognitive workload in a highly realistic en-route air traffic control (ATC) simulation, in which task difficulty was multivariate, and in each dimension highly granular and ordinal. This required a regressor rather than a classifier. The results presented here are meant to relate workload estimation to the dominant baseline literature on workload, and to generalize those studies to a broad variety of operational contexts including but not limited to ATC.

We employed Gaussian Process Regression (GPR; Rasmussen and Williams, 2005), a type of non-parametric regression, in which a single unknown target variable's status (in this case, the number 'N' back) is estimated as a function of the state of one or more known input variables (in this case, power spectra at each electrode in the EEG montage).

Parametric regression methods, for example multiple linear regression (MLR), replace training data with a user-specified function, such as a line or curve or surface in the geometric

space of inputs and outputs, whose parameters can be fitted to optimize estimation of outputs from inputs over the training data. For parametric methods, after the regression weights have been obtained, the original training data may be discarded. Non-parametric regression methods, by contrast, may keep the original training data to use as a scaffold for constructing a regressor function. Test data is compared to the training data points, with output value of the test point estimated via the distance of the test data input to the training data input. As a result of this weighting, estimates of output values form a locally smooth surface spanning the input data, in a process often referred to simply as smoothing. Non-parametric regression only assumes that data points with similar input values will be close in the output space. For GPR specifically, the form of the local weighting is defined by the covariance function and associated hyperparameters learned during model training.

This non-parametric GPR approach has several benefits with respect to cognitive monitoring. First, GPR makes few assumptions about the shape of the estimator function beyond the assumptions associated with the choice of covariance function. This is beneficial especially in high-dimensional input spaces, as is the case when there are many known variables for each data point, and the shape of the relationship between knowns and unknowns cannot easily be visualized and understood by a researcher.

Second, a GPR model can be constructed to change the width of the local weighting functions separately for each known input dimension during training, providing an indirect measure of that input dimension's relevance. Measuring relevance adds interpretability to the model, and can be used to relate the features used by the model to existing literature, or aid in understanding which of the input variables could be left out of the analysis with little or no reduction in predictive accuracy.

A third benefit of GPR is its robustness to spontaneous failure of sources of input during operational test use of a BCI, such as the loss of good electrical contact by an EEG electrode or other equipment failure. Changes in the set of features available to machine learning methods challenge parametric methods such as linear or quadratic models, which typically have dependencies between features. In contrast, GPR depends more directly on the data and is robust to such changes; it can even be applied to data containing many fewer features than the model received during training.

Finally, a fourth major benefit of GPR for cognitive monitoring is its inherently probabilistic nature, returning both point predictions and confidence intervals around those predictions. Confidence values associated with each prediction may be used to dynamically inform decisions about when to trust a trained model's predictions in operational settings.

While GPR has been used to classify EEG in the context of a BCI task involving imagined hand movement (Zhong et al., 2008; Wang et al., 2009), its use in cognitive state assessment has been limited (although see Chaouachi et al., 2011, 2015).

A reasonable assumption in cognitive neuroscience is that similar regions of the brain are engaged in similar functions across individuals during a specific task. This assumption motivates an approach to research that seeks constant neurophysiologic signatures for cognitive functions that generalize broadly among human participants. The present study has employed a more conservative and directed approach based on another reasonable assumption, which is that brain function involves learning, and that as a result, meaningful idiosyncratic differences may be expected among individuals with different learning histories, or within an individual over learning timescales. As such, we focused our analysis on a same-day, same-individual construct for training and testing our machine learning methods. Further, we did not set out to evaluate the neural basis of working memory and attention during task loading, although we regard this as an important goal for other research. Our goal was simply to evaluate the effectiveness and interpretability of a best-of-class machine learning approach for real-time, passive BCI targeted to cognitive monitoring in its simplest and most direct form.

We present a paradigm for assessment of cognitive workload for an operator, specifically the working memory and attentional demand based on measurable task load. We predict workload within several N-back tasks by training a GPR model, then testing it on held-out data from the same participant and session. The N-back task variants, which were designed to have face validity to an operational ATC task, include the following variations: auditory, numeric, and spatial. Finally, we analyze the GPR model to identify which EEG electrode sites, frequency bands, and derived features are essential to the predictive accuracy of the model, which serves to set a lower bound on the number of features required for accurate prediction.

MATERIALS AND METHODS

Participants

The study included 16 male participants, aged 39–62 years old, selected for operational experience in the target operational domain of ATC. All participated voluntarily, and provided written informed consent after having had the procedures of the study described to them. Personally identifiable information for all participants was anonymized and kept secure by a trusted agent. All participants were salaried employees of the MITRE Corporation, and were compensated by allowing them to apply the time spent participating in the study to their work hours. Human subjects procedures were approved by the MITRE Corporation Institutional Review Board (MIRB), to which the Code of Federal Regulations, Title 45 (Public Welfare), Department of Health and Human Services, Part 46 (Protection of Human Subjects) applies for federally funded research involving human subjects.

Task

To change working memory load in a controlled manner, we used an N-back working memory task in one of three stimulus modes (Auditory, Numeric, Spatial) and three task levels ($N = 1$, 2, 3) for each mode. The N-back task required participants to view a series of stimuli and press the spacebar key when the currently presented stimulus matched the stimulus presented

N stimuli before the current one. The task was implemented in BrainWorkshop (Hoskinson, 2011), modified to synchronize with the EEG system.

The Auditory stimuli were NATO letters ('Alpha,' 'Bravo,' 'Charlie,' etc.) spoken by a computer-generated voice. Numeric stimuli were numbers of 3 or 4 digits, e.g., "505" or "6099," presented in the center of the screen. Spatial stimuli were blue squares presented in one of eight spatial locations on the screen, in a 3 × 3 grid leaving out the center square (**Figure 1**). Within each condition, eight unique stimuli were presented over the course of the block. Within the spatial condition, these eight stimuli were the aforementioned eight spatial positions. Within the Auditory and Numeric blocks, these eight stimuli were eight sounds or images randomly selected from a pool of 26 possible NATO letter sounds or 26 possible Numeric images. Each trial lasted 3 s, with visual stimuli in the Numeric and Spatial conditions remaining onscreen for the first 500 ms of the trial. The stimuli for each trial were selected pseudorandomly from the eight possible stimuli within the block, with an N-back match additionally forced on 1/8 of trials. The combination of the inherent 1/8 probability of random match and the independent forced match probability of 1/8 results in an overall 76.56% (7/8 * 7/8) chance of non-matching stimuli and 23.44% chance of matching stimuli. Participants were instructed to respond to matching stimuli by pressing the spacebar key on a standard computer keyboard, while non-matching trials did not require a response.

The task was performed in blocks of 100 comparison trials of a single modality and task level. Participants were allowed to take short breaks between 100 trial blocks. Three 100-trial blocks of each N-back level were performed for each of three stimulus modes, totaling 900 trials for each participant. Stimulus modes were counterbalanced across participants, while N-back levels were performed in the order 1-Back, 2-Back, 3-Back, within each modality block. Before each block, a resting baseline condition was recorded, however, data from this resting baseline condition was not included in the regression models. Prior to the experimental blocks, participants completed 20 practice trials at each N-back level of a Color variant, in which participants indicated if the color of the current stimulus (a square presented in the center of the display) matched the color of the stimulus presented N stimuli prior.

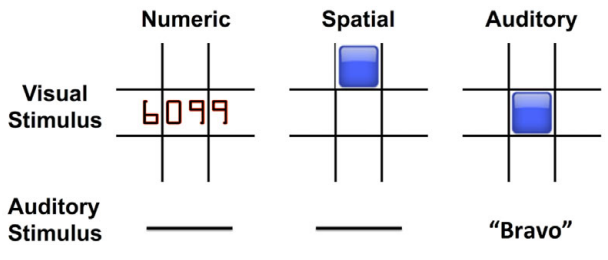


FIGURE 1 | N-back working memory tasks used three stimulus modes: Numeric, Spatial, and Auditory. The visual and auditory stimuli associated with each mode are shown.

Following each block, participants reported their subjective rating of block difficulty (subjective workload) on a 1–7 Likert scale from low to high workload. Subjective workload ratings were collected in order to confirm that the N-back task was subjectively experienced as more demanding as N-back level increased, as well as to investigate any subjective differences in demand between N-back modalities used (Auditory, Numeric, Spatial).

Behavioral Data

Accuracy on the N-back task is evaluated within each block as the number of true positive (TP) responses (correctly responding when the current stimulus matched the stimulus presented "N" back), divided by the sum of the TP responses, false positive (FP) responses (incorrectly responding when the current stimulus did not match the stimulus presented "N" back), and false negative (FN) responses (incorrectly failing to respond when the current stimulus matched the stimulus presented "N" back). This is equivalently described as accuracy = TP/(TP + FP + FN). As the N-back match probability was 23.44%, this places an upper limit of chance performance at 23.44%. For example, responding to all stimuli regardless of N-back match would generate an accuracy of 23.44%, while responding to no stimuli regardless of N-back match would generate an accuracy of 0%.

Behavioral accuracy and subjective workload were assessed via separate two-way repeated-measures ANOVAs, with factors N-back level (1, 2, 3) and task mode (Auditory, Numeric, Spatial). Mauchly's test was used to assess sphericity, with *F*-values adjusted via Greenhouse–Geisser correction where appropriate. Effect size is indicated by generalized eta squared (η^2_G) (Olejnik and Algina, 2003), a measure of effect size appropriate for repeated measures designs (Bakeman, 2005).

EEG Collection

EEG data were collected via a 32-channel actiCAP active electrode system and BrainAmp amplifier at a sampling rate of 500 Hz using Recorder software (Brain Products GmbH), with online reference at electrode FCz and online bandpass filter from 0.1 to 250 Hz.

Processing

Offline data analysis was completed with the EEGLAB toolbox for MATLAB (Delorme and Makeig, 2004) and custom MATLAB scripts. EEG signals were band-pass filtered to 1–50 Hz, down-sampled to 250 Hz, and re-referenced to the average of the left and right mastoid sites (TP9 and TP10).

Trial epochs were extracted from 0 to 3 s post stimulus onset, and labeled according to N-back level, stimulus mode, and behavioral accuracy. Channels and epochs containing paroxysmal artifacts such as gross EMG or cap movement were identified via visual inspection, and were removed from further analysis (Delorme et al., 2007). Between 0 and 4 electrodes were removed per participant (mean of 0.75 electrodes were removed). The remaining epochs were decomposed via independent component analysis (ICA), using the extended InfoMax algorithm as implemented in EEGLAB. For each participant, independent components (IC) representing sources

of artifact including eye blinks, lateral eye movements, and muscle activity were manually identified based on IC topography, frequency spectra, and time-domain activity, and were removed from the data.

Feature Extraction

Band-power features were extracted by transforming each epoch from the time to frequency domain via the Welch method. The Welch method averages the Fast Fourier Transform (FFT) results from several overlapping Hamming windowed segments. A window size of 500 points (2 s) and overlap of 250 points (1 s) were used, along with a 512 point FFT.

For each channel, frequencies were averaged into 6 pre-specified bands, delta: 1–3 Hz, theta: 4–7 Hz, low alpha: 8–10 Hz, high alpha: 11–12 Hz, beta: 13–25 Hz, gamma: 26–40 Hz.

Band power values were then converted to the natural logarithm of their original values to more closely approximate a Gaussian distribution, and each feature was then zero-centered and normalized by its standard deviation on the training set. The same normalization was applied to trials from both training and test sets; both were *z*-scored relative to the mean and standard deviation of the training set. Trials from the test set were *z*-scored relative to the mean and standard deviation of the training set, rather than the test set, to place the test trials on the same scale as the training set. Scaling the test set trials to the test set mean and standard deviation could eliminate meaningful differences that could be present between training and test sets. For example, when using a trained model to derive workload predictions on a new task that is on average more difficult than the training task. In addition, for online prediction applications the mean and standard deviation of the full test set are unknown in advance.

Machine Learning: Gaussian Process Regression

The Information present in EEG band-power features about task level was analyzed using a continuum of methods including ANOVA, MLR and GPR. Additionally, while imposed task level (the number 'N' back) was of primary interest, models were additionally constructed using participants' subjective rating of their mental demand as labels.

For machine learning, a feature vector composed of each of the 6 bands at each of the 32 electrode sites, less any electrodes rejected due to excessive artifact or poor electrode contact, was taken as input. The features were normalized as described in Section "Feature Extraction." The length of the feature vector is the product of the number of bands and number of electrode sites analyzed, and was thus of length 192 (6 * 32) for the 11 participants for whom no electrodes were rejected due to artifact, and six elements (frequency bands) less for each rejected channel for the remaining five participants.

Gaussian Process Regression

A GPR model, a form of Bayesian non-linear regression, was trained using the Gaussian Processes for Machine Learning (GPML) library for MATLAB (Rasmussen and Williams, 2005; Rasmussen and Nickisch, 2010). A GPR model is defined primarily by the selection of a covariance function, which defines how the expected value of the target variable changes as values change across the input space. Here, a squared-exponential covariance function with automatic relevance determination (ARD) was used, in conjunction with a constant zero mean function. ARD refers to the inclusion of a length-scale for each feature within the covariance function, which can be examined after training to determine the relative importance of that feature to prediction. As described by Rasmussen and Williams (2005), the squared exponential covariance function with ARD is defined as:

$$k(x_p, x_q) = \sigma_f^2 * \exp\left(-\frac{1}{2}(x_p - x_q)^\intercal * (diag(\ell)^{-2}) * (x_p - x_q)\right) \quad (1)$$

Where x_p and x_q represent values in the input space, σ_f^2 represents the noise free signal variance, and ℓ is a vector of length-scales (one for each feature).

This covariance function is stationary in the sense that the relationship between values in the input space depends only on their distance, not to their particular location in the space. The squared exponential covariance function was selected *a priori* based on its relative simplicity, the assumption inherent in its use is that data points that are close in the input space will tend to be close in the output space. The constant zero mean function was selected as the data was normalized to have zero mean in the training set. Rasmussen and Williams (2005) present an in-depth presentation of the properties of different covariance and mean functions in the context of GPR.

The covariance and mean functions were used in conjunction with a Gaussian likelihood for prediction via the following equations, all from Rasmussen and Williams (2005):

$$f_*|X, y, X_* \sim N(\bar{f}_*, cov(f_*)) \quad (2)$$

Where f_* is a posterior distribution, X is a matrix of training inputs, y is a vector of training targets, X_* is a matrix of test inputs, $\bar{f}_*$ is the posterior mean, and cov(f_*) is the posterior covariance.

The posterior mean is specified as:

$$\bar{f}_* \triangleq \mathbb{E}[f_*|X, Y, X_*] = K(X_*, X)[K(X, X) + \sigma_n^2 I]^{-1} y \quad (3)$$

The posterior covariance is specified as:

$$cov(f_*) = K(X_*, X_*) - K(X_*, X)[K(X, X) + \sigma_n^2 I]^{-1} K(X, X_*) \quad (4)$$

Where K indicates a covariance matrix, and σ_n^2 is a noise variance term.

The covariance function contains several hyperparameters, which are optimized during model training. Hyperparameters for the covariance function include a length-scale for each feature (ℓ), and a noise free signal variance (σ_f^2). In addition, the covariance function is evaluated using a Gaussian likelihood function, which has a single hyperparameter, the noise variance (σ_n^2). The constant zero mean function has no hyperparameters.

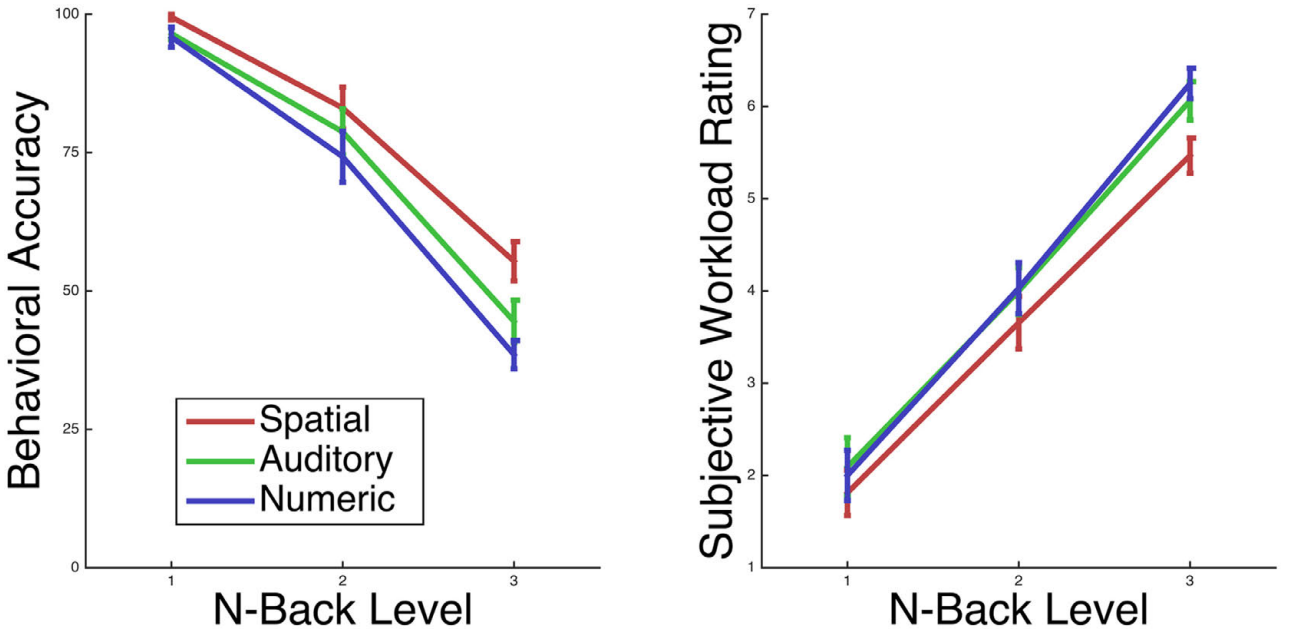


FIGURE 2 | Participant performance and subjective workload on the N-back task changed monotonically with level.

Prior to each model run, these hyperparameters are set to default values, which are subsequently adjusting during model training. Here, for hyperparameters associated with the squared exponential covariance function with ARD, the length-scale for each feature was set to 10, and the signal variance was set to 1. Additionally, for the hyperparameter associated with the Gaussian likelihood function, the likelihood variance was set to 1. These hyperparameters are then optimized within each model run by the GPML library, by minimizing the negative log marginal likelihood on the training set, over 100 function evaluations.

After training the model, new predictions are made via the conditional distribution of target output values, given the test inputs, training inputs, training targets, covariance function, and associated hyperparameters. The mean and variance of the posterior target distribution are used to generate point predictions and confidence intervals, respectively.

Evaluation of Model Performance

Model performance at predicting N-back task level (N) was assessed via fivefold cross-validation with a five trial buffer between training and test sets. Data from each modality and N-back level block (9 blocks total) was split into five partitions, with each partition containing a contiguous block of trials. On any given fold of the fivefold cross-validation procedure, 4 of the 5 partitions (80% of data) were used for training the GPR model, with the remaining partition held out as a test set for assessing model performance. Additionally, any trials from the test set that occurred within five trials of a member of the training set were removed from the test set and not included in measures of model performance. Trials were removed from the test set, and not the training set, to ensure a constant amount of training data (4 of 5 partitions or 80%) across runs. These neighboring trials were removed in order to reduce any short-time scale effects of attention or participant posture on model performance. After identification of the training and test trials from each of the 9 blocks, the data, from these 9 blocks (3 N-back levels and 3 modalities) were pooled for training and testing, labeled by N-back level and subjective workload rating provided by each participant after each block, but not labeled by modality. Data from the three modalities were pooled in an attempt to identify features indicative of working memory load independent of any particular stimulus modality. Measures of prediction quality were obtained for each participant by combining the results from the five model runs. Specifically, for each participant, the true and predicted values from each model run of collected and used to compute a single sMSE and a single Pearson correlation coefficient for that participant. On average, 661.25 trials were included in each training set, and 88.66 trials were included in each test set. Despite the use of fivefold cross-validation, the number of trials in the average test set is less than 1/4 of the trials in the average training set due to the removal of trials from the set test partitions that occurred within five trials of a trial from the training set.

As a parametric regression model for performance comparison to GPR, we used MLR with one linear term per feature plus a constant term. The model training and testing functions were implemented using BCILAB (Kothe and Makeig, 2013). Our BCILAB plugins for Gaussian Processes (a BCILAB wrapper around the GPML library), and for MLR (a BCILAB wrapper around the 'regress' function in MATLAB), are available as open source code.

TABLE 1 | Predictive ability of feature subsets.

	max # features	Correlation *r*	Standardized MSE
Feature subset			
All	192	0.75 ± 0.03	0.44 ± 0.04
GPR-ARD subsets			
50% shortest length scales	96	0.75 ± 0.03	0.44 ± 0.04
25% shortest length scales	48	0.74 ± 0.03	0.46 ± 0.04
ANOVA subsets			
Top 50% ANOVA features	96	0.73 ± 0.03	0.47 ± 0.05
Top 25% ANOVA features	48	0.68 ± 0.03	0.59 ± 0.06
Electrode site subsets			
B-Alert X-10 channels	54	0.53 ± 0.04	0.72 ± 0.04
Emotiv EPOC[a] channels	96	0.73 ± 0.03	0.46 ± 0.04
Parietal channels only	54	0.56 ± 0.04	0.68 ± 0.05
Occipital channels only	30	0.52 ± 0.04	0.72 ± 0.05
Frequency band subsets			
Delta	32	0.21 ± 0.03	0.95 ± 0.02
Theta	32	0.23 ± 0.04	0.94 ± 0.02
Low alpha	32	0.24 ± 0.04	0.92 ± 0.03
High alpha	32	0.25 ± 0.04	0.93 ± 0.02
Beta	32	0.65 ± 0.03	0.57 ± 0.05
Gamma	32	0.74 ± 0.03	0.45 ± 0.04

Feature subsets are categorized by whether they are subsets of electrode sites (over all frequency bands), frequency bands (at all electrode sites), or selected from sites × bands. [a]*For this analysis, Emotiv EPOC sites AF3 and AF4, which were not included in our configuration, were substituted by adjacent sites Fp1 and Fp2.*

Continuous prediction accuracy was quantified using two metrics: standardized mean squared error (sMSE) and Pearson correlation coefficient (*r)*. sMSE is the mean squared error (MSE) of true and predicted values, divided by the variance of the true values. sMSE has a characteristic scale of 0–1 and, due to the standardization on the variance of the true values, is dimensionless, unlike the MSE. Like MSE, sMSE equals 0 for a perfectly accurate prediction. However, due to standardization sMSE equals 1 for a naïve model which always predicts the mean of the ground truth values, and exceeds 1 for predictions that are more erroneous than could be obtained by only predicting the mean of the ground truth values. For machine learning purposes, *r* ranges from 1 (perfect accuracy) to 0 (uncorrelated); however, a naïve model predicting a constant output will show positive *r*. Additionally, although mental workload is argued to be best treated as a continuous, rather than discrete, variable, we have also included discretized versions of the continuous MLR and GPR output. These predictions were included to allow the presented results to be more readily compared with other reports in which discrete classification is performed, and are computed by rounding each continuous prediction to the nearest label in the training set (i.e., a continuous prediction of 2.4 is relabeled as 2), then computing the fraction of predictions which have the correct label.

While predicting the imposed task load is of primary focus, an additional model was trained to predict subjectively experienced workload, using the reports provided by each participant following each task block. This model was computed in the same manner as the previously described model for imposed task load, with the exception of each trial being labeled according to the subjective workload provided by that participant for that block (a value that can range from 1 to 7), rather than the imposed task load (1–3).

Additionally, models were constructed with data from single task variants, in order to investigate the ability of the model to predict the task load within task variants relative to across task variants. Data from each task variant and load was split into five partitions, with a separation of at least five trials between partitions, as previously described for the primary analysis. While the primary analysis combined data across the three task variants for a given fold, the present analysis used data from only a single task variant for training, and a single task variant for testing. For example, the first run of training on the Auditory task and testing on the Auditory task uses the first training fold and first testing fold of exclusively Auditory task data. In contrast, the first run of training on Auditory task and testing on Spatial task uses the first training fold of exclusively Auditory task data, and first testing fold of exclusively Spatial task data. As three task variants were included in the experiment, generating nine combinations of training and test task variants.

Feature Analysis

To illuminate the association between individual participants' EEG features and working memory load prediction, we used two techniques. First, we applied a one-way ANOVA for task level (the number N-back; 1, 2, 3) to individual participants' EEG data. Second, using the trained GPR predictive model, we examined ARD length scales of each feature to identify which played the greatest role in prediction. ARD length scales were also used to evaluate the predictive power of alternate EEG electrode

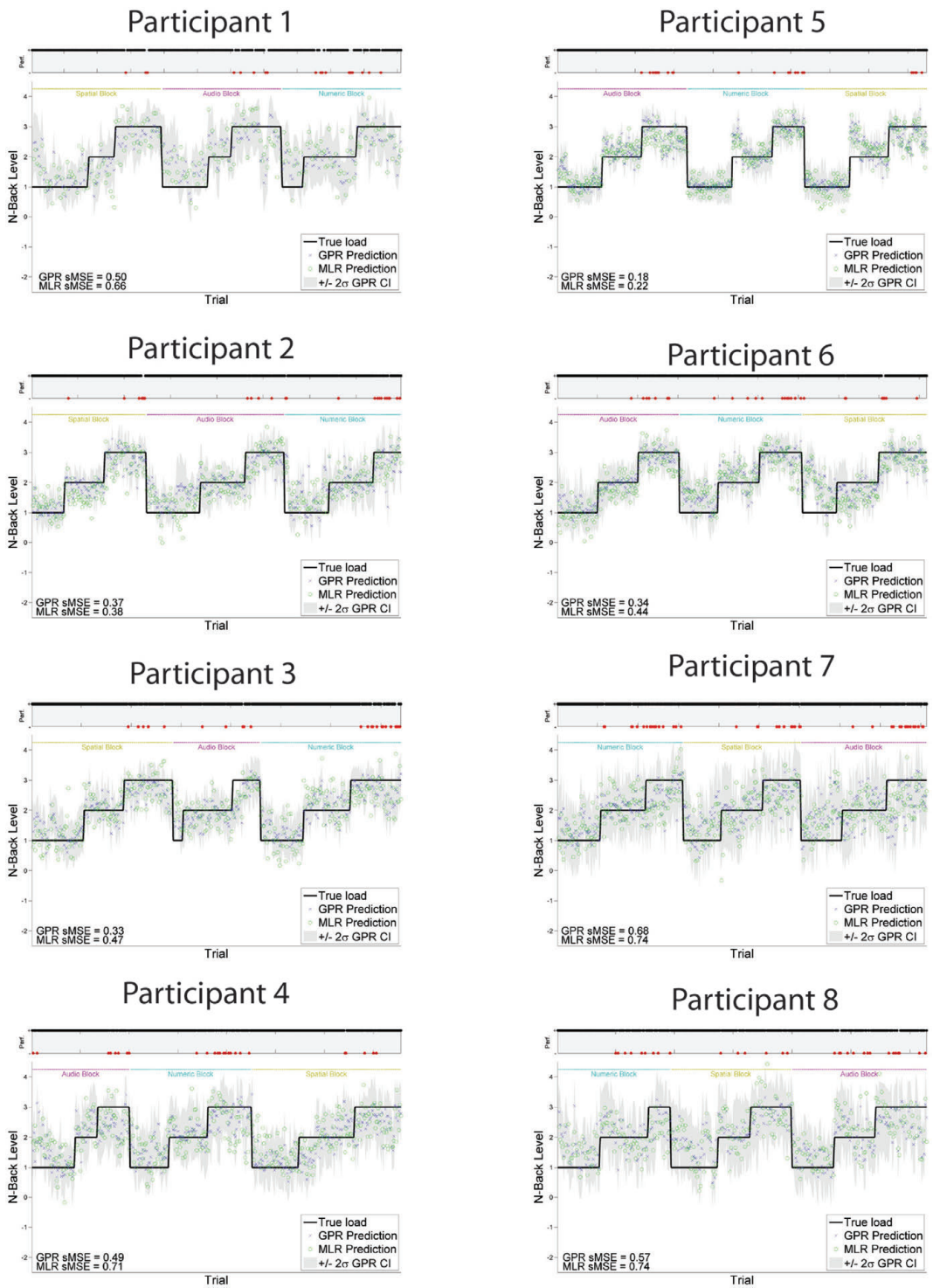


FIGURE 3 | GPR predictions of N-back level for participants 1–8. The figure displays predictions derived from the 5 cross-validation folds in a single graph. Main graph shows ground truth task load (black line). Predicted load is represented as the point prediction for each trial, for both Gaussian Process (GP) and Multiple Linear Regression (MLR) predictors, displayed in blue x's and green o's, respectively. Task block for each section of the experiment (Auditory, Numeric, Spatial) is indicated by colored labels above the predictions. The gray region shows the $\pm$ 2σ confidence interval for each GP point prediction generated by the model. sMSE for both GP and MLR models are included in the lower left of each participant panel. Participant behavioral performance is shown in the line graph at top of each subplot (+ = correct, − = incorrect, with incorrect points also colored in red) in order to visually examine the relation between model prediction and participant behavioral performance on the task.

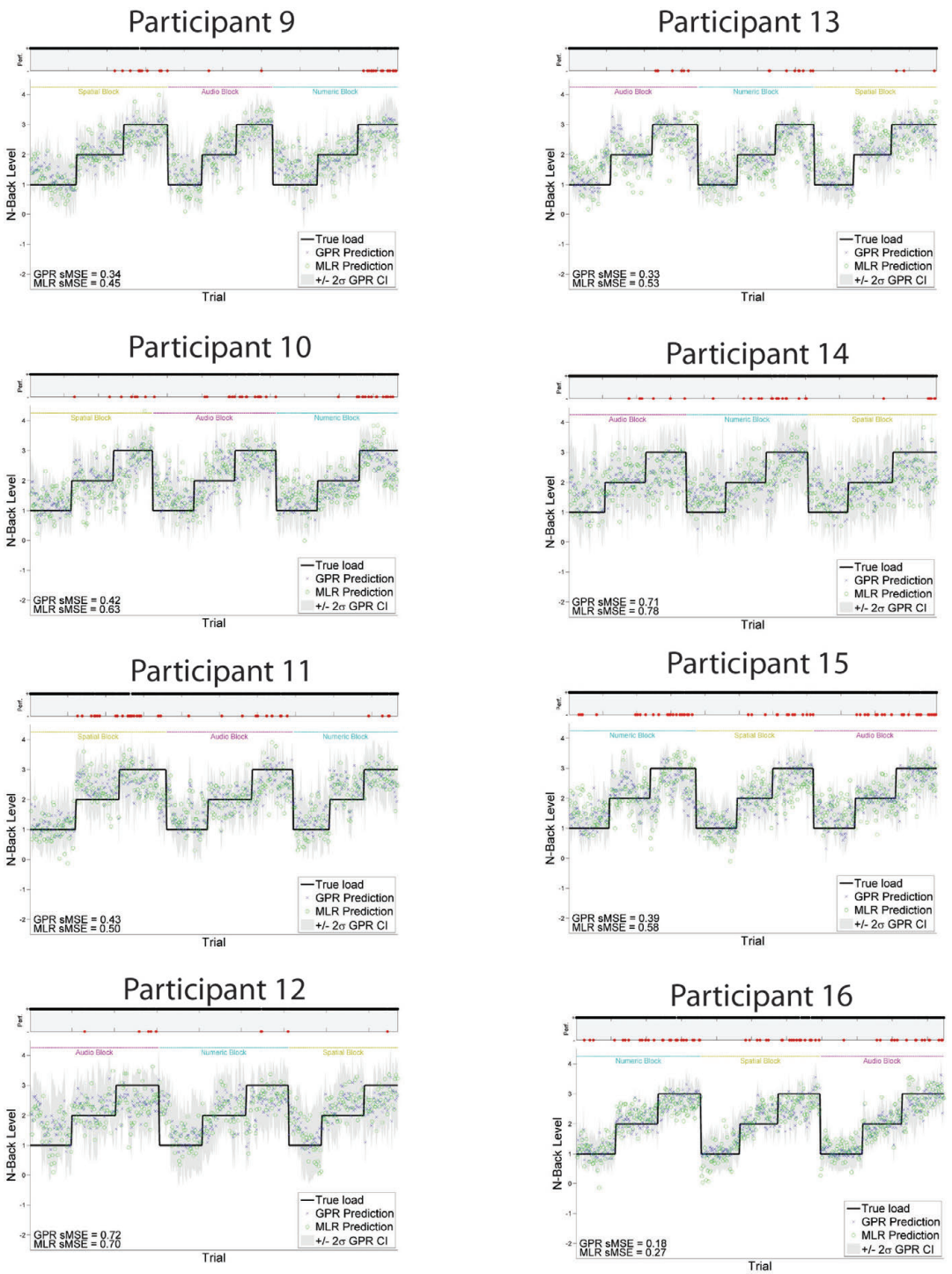


FIGURE 4 | GPR predictions of N-back level for participants 9–16. The figure displays predictions derived from the 5 cross-validation folds in a single graph. Main graph shows ground truth task load (black line). Predicted load is represented as the point prediction for each trial, for both Gaussian Process (GP) and Multiple Linear Regression (MLR) predictors, displayed in blue x's and green o's, respectively. Task block for each section of the experiment (Auditory, Numeric, Spatial) is indicated by colored labels above the predictions. The gray region shows the $\pm 2\sigma$ confidence interval for each GP point prediction generated by the model. sMSE for both GP and MLR models are included in the lower left of each participant panel. Participant behavioral performance is shown in the line graph at top of each subplot (+ = correct, − = incorrect, with incorrect points also colored in red) in order to visually examine the relation between model prediction and participant behavioral performance on the task.

TABLE 2 | Predictive ability of Gaussian Process Regression (GPR) model in comparison to multiple linear regression (MLR) model, predicting either task load or subjective workload using all model features.

Predicted Variable	Prediction method	*r*	sMSE	Classification
Task load	GP	0.75 ± 0.03	0.44 ± 0.04	0.70 ± 0.02
Task load	MLR	0.69 ± 0.03	0.55 ± 0.04	0.63 ± 0.02
Subjective workload	GP	0.76 ± 0.03	0.43 ± 0.04	0.52 ± 0.03
Subjective workload	MLR	0.70 ± 0.02	0.54 ± 0.04	0.44 ± 0.02

Model performance is provided via Pearson's correlation coefficient 'r', standardized mean squared error (sMSE), and categorical classification accuracy. Values displayed are the mean performance across participants ± the standard error of the mean across participants.

montages mapped to other commercial EEG equipment, as is further explained in Section "Results."

RESULTS

Behavioral Data

Participants completed all N-back working memory tasks (Auditory, Numeric and Spatial tasks) at above chance performance within all N-back levels. As N-back level increased, performance significantly decreased. Across all participants and modalities, mean 1-back performance was 97%, mean 2-back performance was 79%, and mean 3-back performance was 46% (**Figure 2**). There was a main effect of level, $F(2,30) = 135.108$, $p < 0.001$, $\eta^2_G = 0.747$, as well as a main effect of mode, $F(2,30) = 14.457$, $p < 0.001$, $\eta^2_G = 0.097$, and a level by mode interaction, $F(4,60) = 3.336$, $p = 0.016$, $\eta^2_G = 0.033$.

As indicated by the reported measure of effect size generalized eta squared (η^2_G), the effect of N-back level on performance was of greater magnitude than the effect of modality of performance.

Participants reported subjective workload levels spanning from 1 to 7, with mean 1-back workload 2.0, mean 2-back workload 3.9, and mean 3-back workload 5.9 (**Figure 2**). For subjective workload, there is both a main effect of level, $F(2,30) = 181.449$, $p < 0.001$, $\eta^2_G = 0.735$, and a main effect of mode, $F(2,30) = 7.773$, $p = 0.002$, $\eta^2_G = 0.042$, while the level by mode interaction was not significant ($p > 0.10$).

Similar to task accuracy, according to the reported measure of effect size, generalized eta squared, the effect of N-back level on subjective workload was of greater magnitude than the effect of modality on subjective workload. Participants performed the Spatial task more accurately, and additionally rated it as lower in subjective workload, in comparison to the Auditory or Numeric tasks. Although each task used only 8 stimuli within each block, it is possible that the consistent use of the same 8 spatial locations across blocks of the spatial task contributed to this performance and subjective workload difference.

Predictive Accuracy of BCI

The GPR with ARD was trained to predict task level for individual participants on a mixture of all three N-back tasks, and tested on the left-out test data using fivefold cross-validation.

Individually trained GPR models were able to predict task level across participants with high accuracy. sMSE mean and standard error across multiple participants was 0.44 ± 0.04, where 0 is perfect prediction and 1 is a model which performs no better than a naive model always predicting the mean of ground truth (**Table 1**). Pearson's *r* correlation was 0.75 ± 0.03, where $r = 1$ is perfect, $r = 0$ is uncorrelated. (All error estimates are given as standard error of the mean.) The GPR predictions of task level for each trial are presented within **Figure 3** (participants 1–8) and **Figure 4** (participants 9–16). The predictions derived from the 5 model folds have been merged into a single dataset for presentation.

The models trained using GPR and all features performed significantly better than models trained using MLR and all features, using the same training and test folds. GPR models had mean sMSE of 0.44 ± 0.04, while MLR models had mean sMSE of 0.55 ± 0.04, $t(1,15) = -6.28$, $p < 0.001$. Similarly, models trained to predict subjective workload ratings performed better using GPR than MLR. The subjective workload model trained using GPR had mean sMSE of 0.43 ± 0.04, while the analogous model trained using MLR had mean sMSE of 0.54 ± 0.04, $t(1,15) = -6.07$, $p < 0.001$. Measures of model quality in terms of Pearson *r* and discretized classification are provided in **Table 2** for comparison with other paradigms. The GPR predictions of subjective workload for each trial are presented within **Figure 5** (participants 1–8) and **Figure 6** (participants 9–16). The predictions derived from the 5 model folds have been merged into a single dataset for presentation. **Table 3** additionally displays the sMSE for each participant, both collected across the 5 runs prior to calculating sMSE, and the mean and standard deviation of sMSE calculated by first computing sMSE within run. The sMSE for each participant collected across the 5 runs prior to calculating sMSE is very similar to the mean of sMSE calculated by first computing sMSE within runs.

Comparing the performance of the GPR model trained on task level to the equivalent model trained on subjective workload, the ability to predict the two label types was not significantly different $t(1,15) = 0.53$, $p = 0.606$.

Feature Analysis in the ANOVA, GPR, and MLR Models

To determine which EEG band-site features were significantly associated with N-back level, we applied a one-way ANOVA for level to individual participants' EEG data. Because the predictive model was also individualized, it was necessary to analyze individual data rather than group effects as is commonly done in cognitive neuroscience.

Using the GPR predictive model, we examined the set of features to identify which played the greatest role in prediction. When ARD is used in training a GPR, the resulting length scale

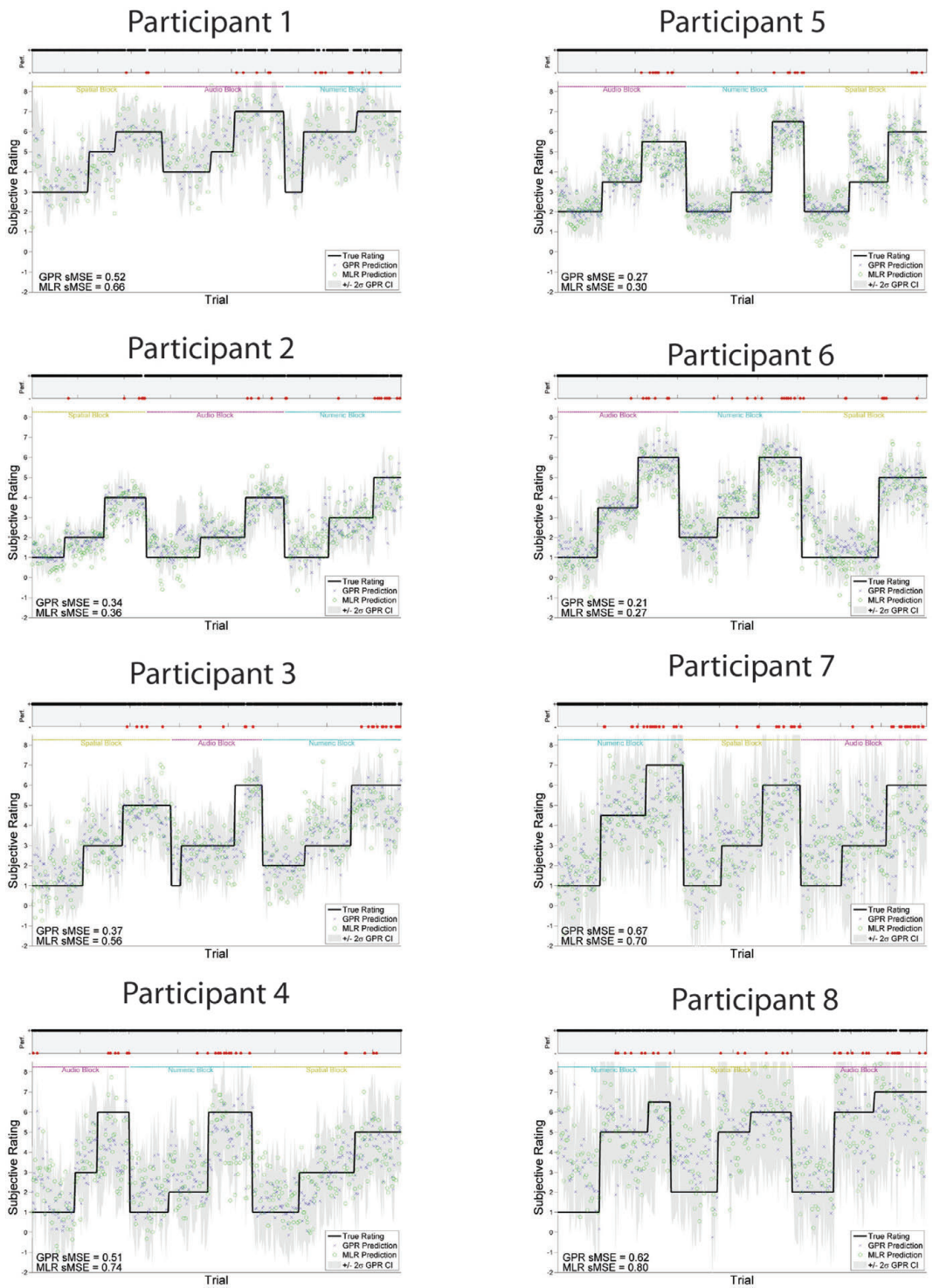


FIGURE 5 | GPR predictions of subjective workload for participants 1–8. The figure displays predictions derived from the 5 cross-validation folds in a single graph. Main graph shows ground truth subjective workload (black line) provided by the participant at the end of the block. Predicted subjective workload is represented as the point prediction for each trial, for both Gaussian Process (GP) and Multiple Linear Regression (MLR) predictors, displayed in blue x's and green o's, respectively. Task block for each section of the experiment (Auditory, Numeric, Spatial) is indicated by colored labels above the predictions. The gray region shows the $\pm$ 2σ confidence interval for each GP point prediction generated by the model. sMSE for both GP and MLR models are included in the lower left of each participant panel. Participant behavioral performance is shown in the line graph at top of each subplot (+ = correct, − = incorrect, with incorrect points also colored in red) in order to visually examine the relation between model prediction and participant behavioral performance on the task.

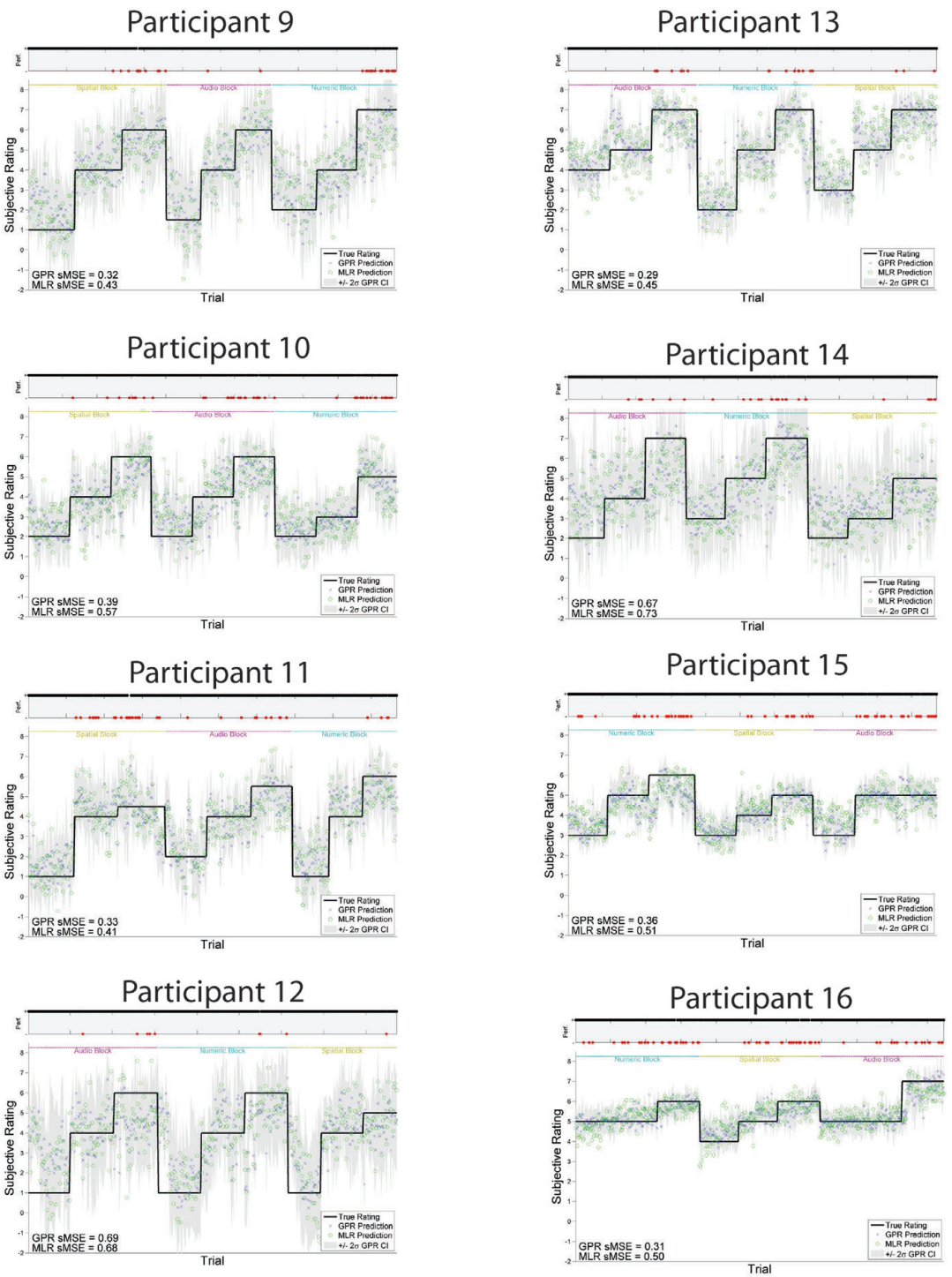


FIGURE 6 | GPR predictions of subjective workload for participants 9–16. The figure displays predictions derived from the 5 cross-validation folds in a single graph. Main graph shows ground truth subjective workload (black line) provided by the participant at the end of the block. Predicted subjective workload is represented as the point prediction for each trial, for both Gaussian Process (GP) and Multiple Linear Regression (MLR) predictors, displayed in blue x's and green o's, respectively. Task block for each section of the experiment (Auditory, Numeric, Spatial) is indicated by colored labels above the predictions. The gray region shows the $\pm 2\sigma$ confidence interval for each GP point prediction generated by the model. sMSE for both GP and MLR models are included in the lower left of each participant panel. Participant behavioral performance is shown in the line graph at top of each subplot ($+$ = correct, $-$ = incorrect, with incorrect points also colored in red) in order to visually examine the relation between model prediction and participant behavioral performance on the task.

TABLE 3 | Individual participant GPR model performance.

Participant	Total sMSE	sMSE Mean Over 5 Runs	sMSE standard deviation over 5 runs
1	0.50	0.47	0.24
2	0.37	0.34	0.11
3	0.33	0.30	0.14
4	0.49	0.45	0.27
5	0.18	0.18	0.10
6	0.34	0.32	0.15
7	0.68	0.66	0.21
8	0.57	0.54	0.17
9	0.34	0.32	0.17
10	0.42	0.40	0.15
11	0.43	0.41	0.19
12	0.72	0.65	0.45
13	0.33	0.31	0.23
14	0.71	0.67	0.28
15	0.39	0.37	0.22
16	0.18	0.18	0.04

The total sMSE combines the results of each run into vectors of true and predicted values, which weights each test example equally. The sMSE mean over 5 runs averages the sMSE results derived from each test run. The sMSE standard deviation is the standard deviation of the sMSE results derived from each test run. The total sMSE and sMSE mean are not equal because each test run can have different numbers of test trials, due to the removal of any test trials that occurred within 5 trials of a training trial.

of each feature indicates the relative sensitivity of the model to changes in that feature's value (MacKay, 2003; Rasmussen and Williams, 2005). A model is more sensitive to features with short length scales and least sensitive (most invariant) to features with long length scales.

Figures 7 and **8** show one-way ANOVA *F*-values compared to GPR length scales for each channel × band power feature, for each of the 16 participants. The values displayed are the average of the values for that participant, over the 5 runs of cross-validation performed. Although there is substantial between-participant variability, gamma band features at occipital and temporal sites are commonly utilized by the GPR models for prediction.

Unlike the features with significant level effects in ANOVA, the (most sensitive) features with the shortest length scales are not generally clustered into individual frequency bands, with the exception of the gamma band, where several channels are uniformly short in length scale. This lack of spatial patterning was also typical across participants.

As the MLR predictions were derived from a multivariate regression, multicollinearity between features can make interpretation of the resulting regression coefficients difficult or misleading (Haufe et al., 2014). Weights from the MLR models were therefore transformed into activation patterns via Equation (6) from Haufe et al. (2014). Specifically, the activations are derived by:

$$A = \sum_x W \sum_{\hat{S}}^{-1} \quad (5)$$

Where $\sum_x$ is the covariance of the data, W is the multivariate regression weights, and $\sum_{\hat{S}}^{-1}$ is the inverse covariance matrix of the latent factors, in this case simply the N-back level labels. **Figure 9** displays the activation patterns from the MLR models predicting task load.

Feature Selection and Prediction Accuracy

For each participant, we compared the predictive ability of several feature subsets. The feature subset "All" (i.e., all electrodes × all bands) was the upper bound on predictive accuracy for this data set (**Table 1**).

To illuminate which frequency bands are important to the task, we considered the predictive accuracy of feature subsets corresponding to single frequency bands (e.g., the 32 features corresponding to the beta band at all electrodes). While all bands contributed to predictive accuracy, the largest contribution came from features in the beta and gamma frequency range. This suggests that information that might be discounted by standard EEG analysis can be highly informative in the context of a BCI predicting workload.

How important was it for the BCI to include all 32 electrodes for this task? We considered feature subsets with a smaller number of EEG electrode sites than were actually measured (but all frequency bands). The model's accuracy for several subsets of electrode sites, averaged over all 16 participants, is also shown (**Table 1**). We compared the montage of our laboratory EEG headset to the montage of two EEG headsets including fewer electrodes, one focused on rapid deployment (B-Alert X10) and one on affordability for home use by consumers (Emotiv EPOC). In this task, the 16 channels present in the Emotiv EPOC device, primarily near equatorial sites such as F7, F8, P3, P4, P7, and P8, capture much of the model's predictive ability. However, the montage of channels present in the B-Alert X-10, more along midline sites such as Fz, Cz, and POz are less effective, generating similar performance as achieved by only looking at a single region's channels (e.g., parietal channels or occipital channels).

One operationally relevant scenario is that a full laboratory electrode cap might be used to calibrate a model for a participant before switching to a simpler EEG device for operational use. We tested this concept by training the GPR model on the full feature set, leaving out features using ARD or ANOVA F-values, then testing the newly reduced model's predictive power. With this paradigm, we found that selecting a reduced feature model using GPR length-scales was more resilient than reducing models using ANOVA features. For both the top 25% and top 50% of features, selection based on training data GPR length-scale generated a model with lower test set sMSE in comparison to selection based on training data ANOVA *F*-value, as evaluated with paired samples *t*-tests; [$t(1,15) = -5.62$, $p < 0.001$ for the top 25% of features, $t(1,15) = -2.68$, $p = 0.017$] for the top 50% of features, see **Table 1**.

A similar method allowed us to measure the absolute minimum number of features required for prediction of task level. Each individual's feature length scales were sorted from shortest to longest, and the GPR model was tested on subsets of increasing size, from 1 to 100% of total features, in increments of

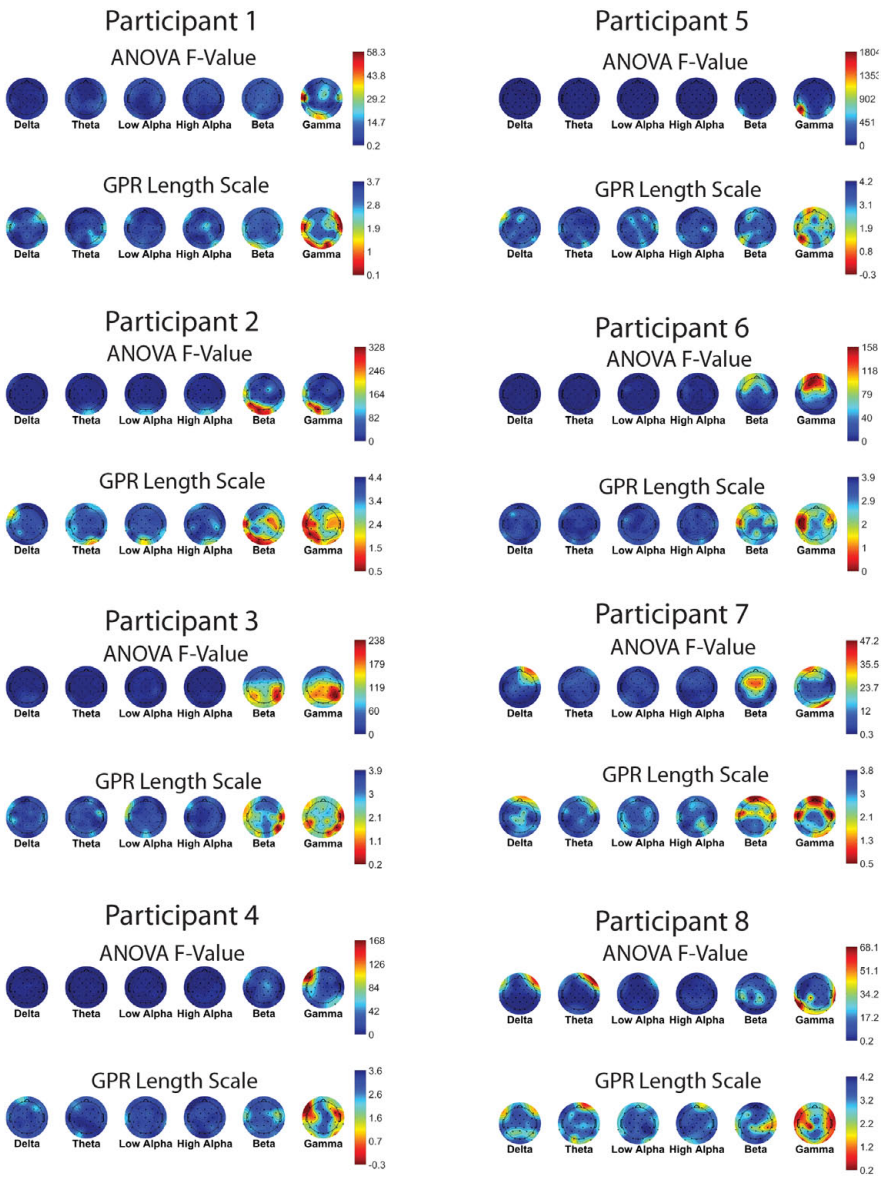


FIGURE 7 | EEG features (bands and sites) for participants 1–8. Top of each participant plot: ANOVA *F*-values for memory load. Red (higher values) corresponds to larger *F*-value for feature. Bottom of each participant plot: log scaled Gaussian Process model length scales. Red (lower values) corresponds to higher sensitivity for feature. The directionality of the colormap is swapped between ANOVA and GPR in order for warmer (redder) colors to always indicate greater feature relevance.

1% (**Figure 10**). As features are added beyond the minimum level required for the model to function, the trend is for classifier error to decrease monotonically until it plateaus near the minimum sMSE of the full model. Approximately 20% of the total number of features are sufficient for prediction quality near the full model.

Prediction within and across Task Variants

Predictions obtained using single task variants for training and testing are obtained in **Table 4**. When the training and test data are from the same task variant (the diagonal of the table), sMSE is approximately the same or lower than what was obtained by combining training and test data across pooled modalities. However, when training and test data are obtained from differing task variants, prediction is no better than what could be obtained by naively predicting the mean of the target distribution.

DISCUSSION

We used GPR to train a model capable of accurately predicting N-back working memory load or workload. When data from

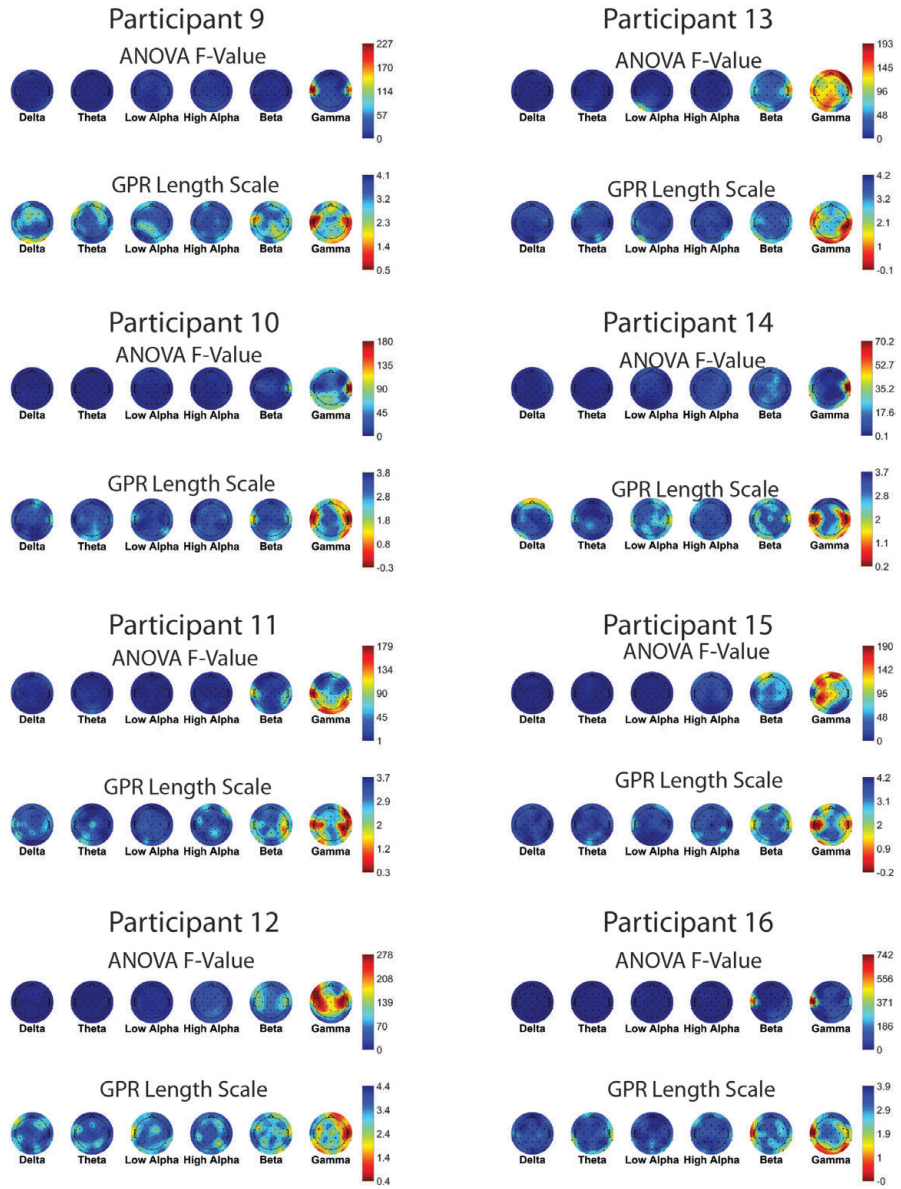


FIGURE 8 | EEG features (bands and sites) for participants 9–16. Top of each participant plot: ANOVA F-values for memory load. Red (higher values) corresponds to larger *F*-value for feature. Bottom of each participant plot: log scaled Gaussian Process model length scales. Red (lower values) corresponds to higher sensitivity for feature. The directionality of the colormap is swapped between ANOVA and GPR in order for warmer (redder) colors to always indicate greater feature relevance.

all three task variants were pooled for training and testing, above chance predictions were obtained. This result is consistent with a meta-analysis of functional magnetic resonance imaging (fMRI) studies using N-back variations which showed a frontoparietal network which, although affected by the nature of the information retained, is generally active across all N-back variants (Owen et al., 2005). However, if data was trained exclusively on a single task variant, then prediction on alterative task variants was no better than a naïve model which always predicts the mean task load. Training exclusively on a single task variant may overfit to that particular variant, impairing prediction when the test variant differs. It is possible that improved cross-variant prediction could be obtained by modification of the GPR model to account for greater uncertainty in predicting a new task variant.

For the pooled data, predictive accuracy was high overall (sMSE $= 0.44$, $r = 0.75$), although the GPR model was less able to predict (or extrapolate to) extreme values, tending to smooth extreme values to middling values. This limitation is typical of interpolation based regressors such as GPR, especially given the limited number of data points (∼800) relative to the high dimensionality of the data (up to 192 per participant, dependant

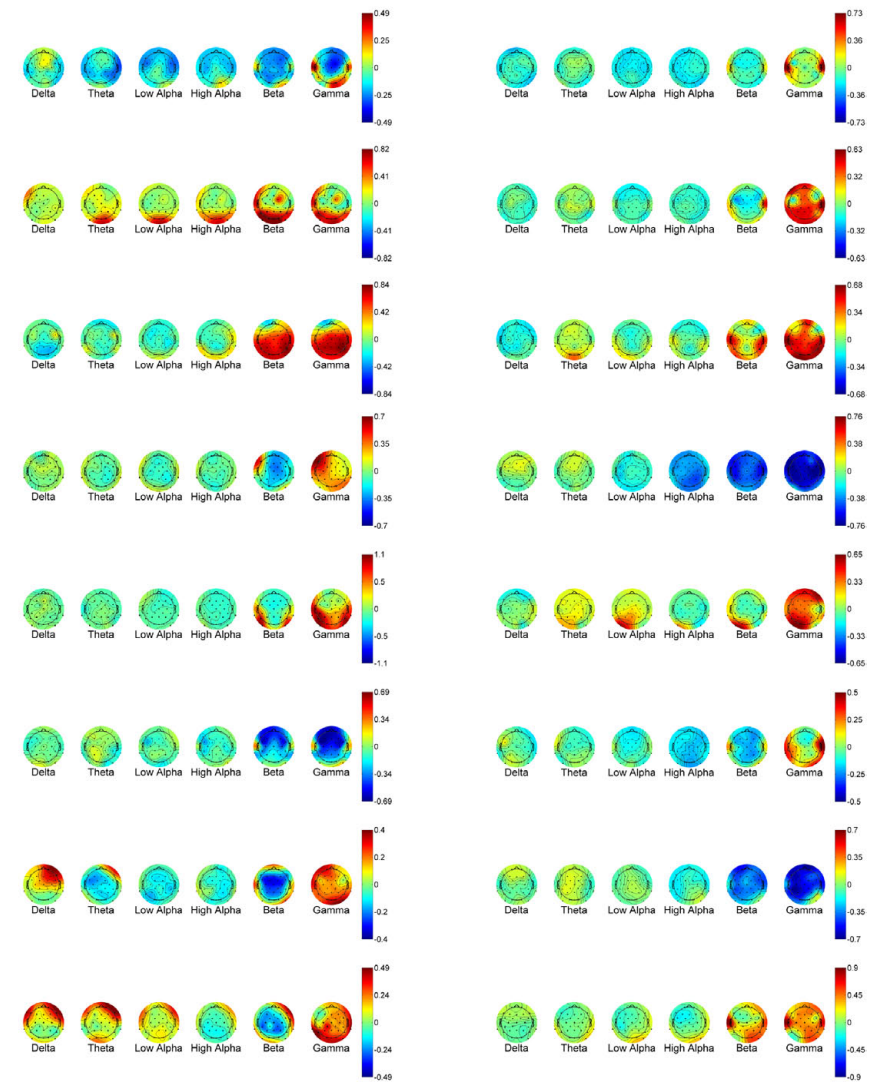


FIGURE 9 | Activations derived from the Multiple Linear Regression models predicting task load, at each band and electrode site for all participants. Red values indicate more positive activations; blue values indicate more negative activations.

on whether any channels were removed due to excessive artifact). Extrapolation might be improved by the use of alternative covariance functions incorporating linear terms.

GPR performed significantly better than the baseline performance established by a simpler parametric technique, MLR. This was the case for models trained and tested on N-back task load, as well as for models trained and tested on subjective workload ratings provided by the participants after each block of the task. Model performance between N-back task load and subjective workload was similar, likely due to the strong relation between N-back task load and subjectively reported workload as reported in the behavioral results.

Applying feature subset selection to the model revealed that feature subsets selected based on techniques such as ANOVA are significantly less efficient at prediction than the subsets identified by GPR with ARD. For example, using the top 25% of features derived from GPR generates model performance approximately equivalent to the top 50% of features derived from an ANOVA model. Similarly, models using GPR consistently outperformed models utilizing MLR, a simpler but less flexible approach.

Periods of data containing obvious muscular artifacts were manually rejected from the dataset prior to training the machine learning model. Despite this, features in the gamma band of the EEG were most sensitive to variations in N-back level. Several works have cautioned against the use of higher frequency band power features such as beta and gamma for workload estimation (Gerjets et al., 2014; Brouwer et al., 2015), due to EMG contamination from differential motor activity in different blocks

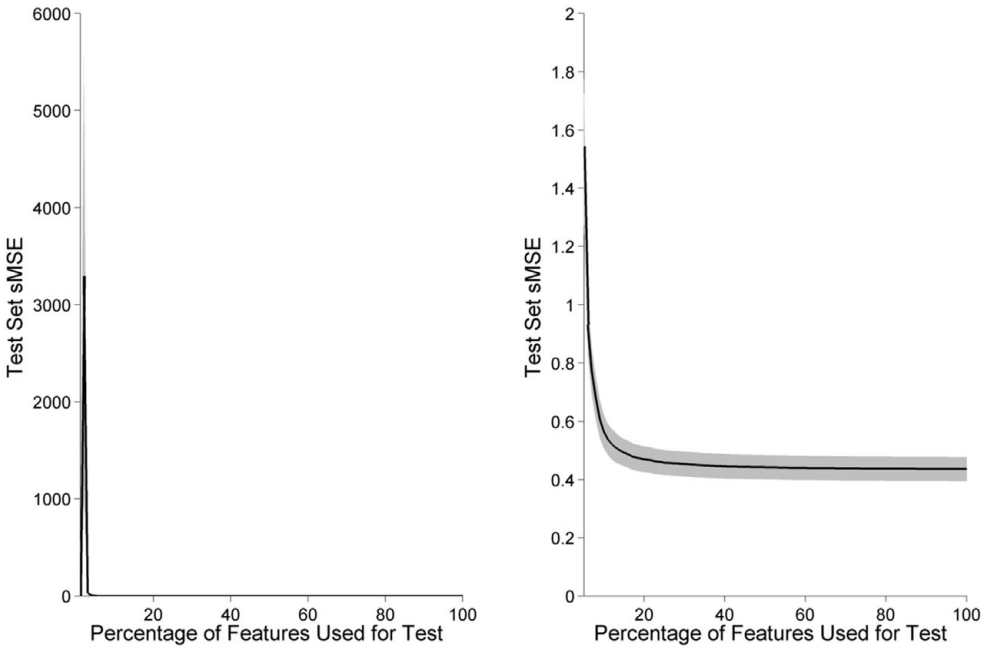


FIGURE 10 | Model accuracy increases as features are added in order of informativeness. Black line indicates model performance as the percentage of features are increased, averaged across participants. Gray shading indicates the standard error of the mean, across participants. Below a minimum number of features, model error increases due to bias from insufficient dimensionality. After approximately the top 20% of features have been included in the model, adding additional features provides little improvement to model performance. The left panel displays 1–100% of features, while for clarity the right panel displays only 5–100% of features, where the GPR model is able to function.

TABLE 4 | Predictive ability of Gaussian Process Regression (GPR) models trained and tested on trials from single working memory task variants.

	Test task		
Training task	**Auditory**	**Numeric**	**Spatial**
Auditory	0.45 ± 0.06	1.15 ± 0.13	0.97 ± 0.09
Numeric	1.07 ± 0.12	0.33 ± 0.04	1.15 ± 0.13
Spatial	1.13 ± 0.14	1.11 ± 0.13	0.35 ± 0.05

The predicted variable for all models was task load. Model performance is provided as the mean and standardized error of the mean of standardized mean square error (sMSE).

of the task. Here, the N-back task was utilized, in which mental workload between task levels is varied by task instruction rather than alteration of the perceptual or motor demands of the task. While perceptual demands do vary between the modality variants of the task utilized (auditory, numeric, spatial), as our predictor was trained and tested on a random selection of data from each of three modalities at each N-back level, the N-back level groupings do not contain systematic differences in perceptual or motor demands that would aid prediction.

As gamma band power is susceptible to contamination from muscular artifact (Muthukumaraswamy, 2013), the source of these features cannot be assumed to be of neural origin. However, the consistency of the gamma band features across participants despite the constant motor demands of the N-back suggests that, if predictions result in part due to EMG activity, more direct measures of EMG activity may prove useful for mental workload classification. It is possible that the diagnosticity of the gamma band feature is due not to a confound in how participants respond between levels of the task, but subtle postural changes on the part of participants as mental demand increases. In this sense, gamma band features may be considered an artifact when EEG is used to measure electrical activity of exclusively neural origin, but in our analysis may also represent a feature that is truly diagnostic of mental demand, and not simply an experimental design confound.

It is often desirable to minimize the number of electrode sites required for a BCI. In laboratory settings with standard electrode caps, using fewer sites can reduce experimental preparatory time or allow experimenters to allocate more time to ensuring a low-impedance connection at key sites, improving data quality. In custom-designed electrode caps, fewer sites may also reduce the size, weight, and power (SWaP) and cost of BCI systems intended to operate in real-time. Taking advantage of our non-parametric GPR model, we were able to demonstrate a method for determining subsets of channels that capture the full predictive accuracy of the entire electrode cap – and even determine the minimum number of channels, or even EEG features, required for accuracy. We observed that the 16 channels present in a commercial off-the-shelf device, mostly lateral sites near the head's equator, capture a very large fraction of the predictive ability of the full 32 channel laboratory cap. Devices with such electrode montages might be used in future experiments, provided their EEG signal quality is acceptable.

Despite what we believe to be an overall contribution to the field, several limitations of the current report should be noted. The present paradigm used 80% of the available data for training on each cross-validation fold. This amount of training data may not be practical to acquire before a real-time device could be utilized. Additionally, our models were trained and tested within each participant. A more optimal model would be participant independent. It possible that these issues could be partially mitigated by adapting data or Gaussian Process hyperparameters that were learned from previous participants to reduce the training time required for new participants. Additionally, the present work contains data from 16 participants, all of whom are male and middle-aged. Future reports should expand workload prediction using larger and more demographically variable participant samples. Finally, while stimulus modality was randomized, participants completed increasingly demanding experimental blocks within each modality. Therefore, fatigue or tiredness could potentially contribute to estimations of mental demand.

CONCLUSION

There is potentially great value in real-time, non-invasive monitoring of cognitive states by 'passive' BCI using methods such as electroencephalography (EEG). Cognitive variables such as workload, which are predictive of operational errors, are potentially valuable targets for real-time monitoring. Information about these variables may be useful in a variety of downstream applications, including providing situational awareness for human operators, alerting operators about high-workload situations, testing and training operators, redesign of interfaces, and redesign of working practices to optimize operator performance.

In this paper, we used EEG to monitor cognitive workload during a simple working memory task (N-back) in multiple sensory and cognitive modalities (Auditory, Numeric, and Spatial). Calibration from training data was demonstrated to be effective using GPR, out performing a more basic model utilizing MLR. GPR also provided the ability to assess the relative predictive value of each input variable (EEG electrode sites, and frequency bands at each site, together summarized as EEG 'features') in predicting the workload variable of interest. The GPR approach was superior to conventional analysis of variance (ANOVA) methods in determining which reduced subsets of EEG features from the training set would be most predictive about the cognitive variable of interest in the test set. This type of analysis may inform engineering efforts to produce EEG systems with few electrodes placed at the most highly informative sites on the scalp for the desired evaluations.

The current approach can be placed within a class of methods that seek to use techniques from machine learning to not only make predictions, but glean useful information about the neural or behavioral processes under study. In another example, Noh and de Sa (2014) have reported that a machine learning model trained on a subset of EEG data can be used to select features for traditional hypothesis testing on an independent test set. As this method derives candidate features for discriminating between conditions from the independent training set, it avoids the issue of multiple comparisons encountered when performing traditional hypothesis testing on several potential features within a single set of data.

In addition, in contrast to more traditional statistical methods such as MLR, the GPR approach provides confidence intervals around each prediction. Information regarding the confidence of a predictor may be useful in operational domains in order to determine when to trust the outputs of the predictive model. For example, a test point that contains data that is far outside what was observed within the training set would be predicted with a large confidence interval.

AUTHOR CONTRIBUTIONS

MC: Co-led research effort, collected and analyzed data, wrote software code, wrote paper. DR: Collected majority of data, wrote software code, analyzed data, contributed to writing of paper. JC: Participated in data collection, edited paper. HG: Participated in data collection, wrote stimulus/task software. MW: Initiated project plan, co-led research effort, participated in data collection.

ACKNOWLEDGMENTS

This research was supported by the MITRE Innovation Program. Approved for public release, distribution unlimited, MITRE case number 15-2950.

REFERENCES

Ayaz, H., Shewokis, P. A., Bunce, S., Izzetoglu, K., Willems, B., and Onaral, B. (2012). Optical brain monitoring for operator training and mental workload assessment. *Neuroimage* 59, 36–47. doi: 10.1016/j.neuroimage.2011.06.023

Bakeman, R. (2005). Recommended effect size statistics for repeated measures designs. *Behav. Res. Methods* 37, 379–384. doi: 10.3758/BF03192707

Baldwin, C. L., and Penaranda, B. N. (2011). Adaptive training using an artificial neural network and EEG metrics for within-and cross-task workload classification. *Neuroimage* 59, 48–56. doi: 10.1016/j.neuroimage.2011.07.047

Brouwer, A.-M., Hogervorst, M. A., van Erp, J. B. F., Heffelaar, T., Zimmerman, P. H., and Oostenveld, R. (2012). Estimating workload using EEG spectral power and ERPs in the n-back task. *J. Neural Eng.* 9, 045008. doi: 10.1088/1741-2560/9/4/045008

Brouwer, A.-M., Zander, T. O., van Erp, J. B. F., Korteling, J. E., and Bronkhorst, A. W. (2015). Using neurophysiological signals that reflect cognitive or affective state: six recommendations to avoid common pitfalls. *Front. Neurosci.* 9:136. doi: 10.3389/fnins.2015.00136

Chaouachi, M., Jraidi, I., and Frasson, C. (2011). *"Modeling Mental Workload Using EEG Features for Intelligent Systems," in User Modeling, Adaption and Personalization*. Berlin: Springer, 50–61.

Chaouachi, M., Jraidi, I., and Frasson, C. (2015). "MENTOR: A Physiologically Controlled Tutoring System," in *User Modeling, Adaptation and Personalization Lecture Notes in Computer Science*, eds F. Ricci, K. Bontcheva, O. Conlan, and S. Lawless (Berlin: Springer International Publishing), 56–67.

Coyne, J. T., Baldwin, C., Cole, A., Sibley, C., and Roberts, D. M. (2009). "Applying real time physiological measures of cognitive load to improve training," in *Foundations of Augmented Cognition. Neuroergonomics and Operational Neuroscience*, eds D. D. Schmorrow I, V. Estabrooke, and M. Grootjen (Berlin: Springer), 469–478.

Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21. doi: 10.1016/j.jneumeth.2003.10.009

Delorme, A., Sejnowski, T., and Makeig, S. (2007). Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. *Neuroimage* 34, 1443–1449. doi: 10.1016/j.neuroimage.2006.11.004

Gerjets, P., Walter, C., Rosenstiel, W., Bogdan, M., and Zander, T. O. (2014). Cognitive state monitoring and the design of adaptive instruction in digital environments: lessons learned from cognitive workload assessment using a passive brain-computer interface approach. *Neuroprosthetics* 8:385. doi: 10.3389/fnins.2014.00385

Gevins, A., and Smith, M. (2003). Neurophysiological measures of cognitive workload during human-computer interaction. *Theor. Issues Ergon. Sci.* 4, 113–131. doi: 10.1080/14639220210159717

Grimes, D., Tan, D. S., Hudson, S. E., Shenoy, P., and Rao, R. P. (2008). "Feasibility and pragmatics of classifying working memory load with an electroencephalograph," in *Proceeding of the Twenty-Sixth Annual SIGCHI Conference on Human Factors in Computing Systems*, Florence, 835–844.

Haufe, S., Meinecke, F., Görgen, K., Dähne, S., Haynes, J.-D., Blankertz, B., et al. (2014). On the interpretation of weight vectors of linear models in multivariate neuroimaging. *Neuroimage* 87, 96–110. doi: 10.1016/j.neuroimage.2013.10.067

Hoskinson, P. (2011). *BrainWorkshop*. Available at: http://brainworkshop.sourceforge.net/

Just, M. A., Carpenter, P., and Miyake, A. (2003). Neuroindices of cognitive workload: neuroimaging, pupillometric and event-related potential studies of brain work. *Theor. Issues Ergon. Sci.* 4, 56–88. doi: 10.1080/14639220210159735

Kothe, C. A., and Makeig, S. (2011). Estimation of task workload from EEG data: new and current tools and perspectives. *Conf. IEEE Eng. Med. Biol. Soc.* 2011, 6547–6551. doi: 10.1109/IEMBS.2011.6091615

Kothe, C. A., and Makeig, S. (2013). BCILAB: a platform for brain–computer interface development. *J. Neural Eng.* 10, 056014. doi: 10.1088/1741-2560/10/5/056014

MacKay, D. J. C. (2003). *Information Theory, Inference and Learning Algorithms*. Cambridge: Cambridge University Press.

Muthukumaraswamy, S. D. (2013). High-frequency brain activity and muscle artifacts in MEG/EEG: a review and recommendations. *Front. Hum. Neurosci.* 7:2013. doi: 10.3389/fnhum.2013.00138

Noh, E., and de Sa, V. R. (2014). "Discriminative dimensionality reduction for analyzing eeg data," in *Proceedings of the 36th Annual Meeting of the Cognitive Science Society*, (Austin, TX: Cognitive Science Society), 1090–1095.

Olejnik, S., and Algina, J. (2003). Generalized eta and omega squared statistics: measures of effect size for some common research designs. *Psychol. Methods* 8, 434–447. doi: 10.1037/1082-989X.8.4.434

Owen, A. M., McMillan, K. M., Laird, A. R., and Bullmore, E. (2005). N-back working memory paradigm: a meta-analysis of normative functional neuroimaging studies. *Hum. Brain Mapp.* 25, 46–59. doi: 10.1002/hbm.20131

Parasuraman, R. (2003). Neuroergonomics: research and practice. *Theor. Issues Ergon. Sci.* 4, 5–20. doi: 10.1080/14639220210199753

Parasuraman, R., and Rizzo, M. (eds) (2006). *Neuroergonomics: The Brain at Work*, 1st Edn. Oxford: Oxford University Press.

Rasmussen, C. E., and Nickisch, H. (2010). Gaussian processes for machine learning (GPML) toolbox. *J. Mach. Learn. Res.* 11, 3011–3015.

Rasmussen, C. E., and Williams, C. K. I. (2005). *Gaussian Processes for Machine Learning*. Cambridge, MA: The MIT Press.

Sweller, J., van Merrienboer, J. J. G., and Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educ. Psychol. Rev.* 10, 251–296. doi: 10.1023/A:1022193728205

Wang, B., Wan, F., Un Mak, P., In Mak, P., and Vai, M. I. (2009). "EEG signals classification for brain computer interfaces based on gaussian process classifier," in *Proceedings of the 7th International Conference on Information, Communications and Signal Processing*, Macau.

Wickens, C. D. (2008). Multiple resources and mental workload. *Hum. Factors* 50, 449–455. doi: 10.1518/001872008X288394

Wilson, G. F., and Fisher, F. (1995). Cognitive task classification based upon topographic EEG data. *Biol. Psychol.* 40, 239–250. doi: 10.1016/0301-0511(95)05102-3

Wilson, G. F., and Russell, C. A. (2003a). Operator functional state classification using multiple psychophysiological features in an air traffic control task. *Hum. Factors* 45, 381–389. doi: 10.1518/hfes.45.3.381.27252

Wilson, G. F., and Russell, C. A. (2003b). Real-time assessment of mental workload using psychophysiological measures and artificial neural networks. *Hum. Factors* 45, 635–643. doi: 10.1518/hfes.45.4.635.27088

Wolpaw, J., and Wolpaw, E. W. (2012). *Brain-Computer Interfaces: Principles and Practice*. Oxford: Oxford University Press.

Zander, T. O., and Kothe, C. (2011). Towards passive brain–computer interfaces: applying brain–computer interface technology to human–machine systems in general. *J. Neural Eng.* 8, 025005. doi: 10.1088/1741-2560/8/2/025005

Zander, T. O., Kothe, C., Welke, S., and Roetting, M. (2009). "Utilizing secondary input from passive brain-computer interfaces for enhancing human-machine interaction," in *Foundations of Augmented Cognition. Neuroergonomics and Operational Neuroscience Lecture Notes in Computer Science*, eds D. D. Schmorrow I, V. Estabrooke, and M. Grootjen (Berlin: Springer), 759–771.

Zhong, M., Lotte, F., Girolami, M., and Lécuyer, A. (2008). Classifying EEG for brain computer interfaces using Gaussian processes. *Pattern Recognit. Lett.* 29, 354–359. doi: 10.1016/j.jneumeth.2015.01.010

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 The MITRE Corporation. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 18 May 2016
doi: 10.3389/fnhum.2016.00223

Evaluation of an Adaptive Game that Uses EEG Measures Validated during the Design Process as Inputs to a Biocybernetic Loop

Kate C. Ewing, Stephen H. Fairclough and Kiel Gilleade*

School of Natural Sciences and Psychology, Liverpool John Moores University, Liverpool, UK

Biocybernetic adaptation is a form of physiological computing whereby real-time data streaming from the brain and body is used by a negative control loop to adapt the user interface. This article describes the development of an adaptive game system that is designed to maximize player engagement by utilizing changes in real-time electroencephalography (EEG) to adjust the level of game demand. The research consists of four main stages: (1) the development of a conceptual framework upon which to model the interaction between person and system; (2) the validation of the psychophysiological inference underpinning the loop; (3) the construction of a working prototype; and (4) an evaluation of the adaptive game. Two studies are reported. The first demonstrates the sensitivity of EEG power in the (frontal) theta and (parietal) alpha bands to changing levels of game demand. These variables were then reformulated within the working biocybernetic control loop designed to maximize player engagement. The second study evaluated the performance of an adaptive game of Tetris with respect to system behavior and user experience. Important issues for the design and evaluation of closed-loop interfaces are discussed.

Keywords: psychophysiology, EEG, gaming, physiological computing, adaptive interface, effort, engagement

OPEN ACCESS

Edited by:
Mikhail Lebedev,
Duke University, USA

Reviewed by:
Dimitrios Kourtis,
Central European University, Hungary
Viktor Müller,
Max Planck Institute for Human Development, Germany
Jean-Arthur Micoulaud Franchi,
Centre National de la Recherche Scientifique, France

****Correspondence:***
Stephen H. Fairclough
s.fairclough@ljmu.ac.uk

Received: *14 December 2015*
Accepted: *29 April 2016*
Published: *18 May 2016*

Citation:
Ewing KC, Fairclough SH and Gilleade K (2016) Evaluation of an Adaptive Game that Uses EEG Measures Validated during the Design Process as Inputs to a Biocybernetic Loop. Front. Hum. Neurosci. 10:223. doi: 10.3389/fnhum.2016.00223

INTRODUCTION

Biocybernetic control describes how the implicit measurement of physiological signals from the brain or body can be transformed into a control input for real-time software adaptation. This category of physiological computing system (Fairclough, 2009) has also been described as a passive brain-computer interface (Zander and Kothe, 2011) because the user simply responds to events at the interface without any requirement for volitional control. The purpose of biocybernetic adaptation is to create a seamless and tacit form of human-computer interaction where software adaptation is timely and intuitive from the perspective of the user.

The biocybernetic model has been applied to a range of domains, such as: adaptive automation (Bailey et al., 2006), detection of negative emotions (Kapoor et al., 2007), adaptive robotics (Liu et al., 2007) and support for social behavior (Chanel and Mühl, 2015). An early example of a working biocybernetic control loop was developed by NASA in the 1990s where the real-time analysis of electroencephalography (EEG) signals was converted into an input variable for the control of the level of system automation during simulated aviation tasks (Pope et al., 1995; Freeman et al., 1999; Prinzel et al., 2000; Scerbo et al., 2003). This control loop was designed to sustain operator engagement within an optimal

zone that avoided complacency and inattention by selectively disabling system automation in order to oblige the operator to engage with a manual interface. This example of biocybernetic control set a blueprint for a data processing protocol wherein electrocortical activity interacts with a computerized system within a negative feedback loop. This model of closed-loop control detects deviations from an optimal state of brain activity and uses these variations to cue changes at the human-computer interface in order to "pull" the psychological state of the user in a desired direction.

The design of a biocybernetic closed-loop incorporates a number of distinct processing stages: (1) data collection from sensors, (2) filtering of raw data coupled with artifact correction techniques, (3) data analysis for the extraction of meaningful metrics that permit a valid inference of the user state, (4) conversion of the metrics in order to instigate adaptation at the user interface, i.e., defining criteria/triggers for adaptation or by categorizing data using machine learning algorithms (Baldwin and Penaranda, 2012; Novak et al., 2012); and (5) adaptation of the user interface in a manner designed to promote a desirable user state.

All biocybernetic closed-loop systems are rooted in a psychophysiological inference; for example: inferring increased arousal from increases in skin conductance level, inferring negative affect from activation of the corrugator supercilli. The validity of this inference is fundamental to the integrity of a working loop, but the process of establishing validity is complex (Cacioppo and Tassinary, 1990). The loop is designed to utilize software adaptation in order to influence a key psychological concept or dimension in the user, e.g., engagement, mental workload, attention. If the fundamental link between input measures, the psychological concept targeted by those measures and the adaptive logic of the loop is weak or tenuous, then the effectiveness of the closed-loop system will be compromised (Fairclough, 2007, 2009). Because the loop works in real-time, it is important that measures are: (a) sufficiently sensitive to changes in the relevant psychological dimension; and (b) specific to that dimension, i.e., not confounded with other psychological variables. Consequently it is important to construct biocybernetic loops on the basis of measures that have either been scientifically validated according to research literature or tested and validated in the context of the target task or application.

This article describes the development of an adaptive computer game where the software responds in real-time in order to enhance the experience of the player by making the game appropriately challenging. Optimizing task difficulty is one of several methods of adapting gaming experiences using biocybernetics, others include enhancing emotional engagement and reducing player frustration (Gilleade et al., 2005). This closed-loop approach employs the same logic that underpins the integration of biofeedback mechanics into gaming applications (Nacke et al., 2011) and the design of adaptive games dedicated to the creation of a specific emotion (Dekker and Champion, 2007). One goal for an adaptive game is to deliver a level of difficulty tailored to the skills of the player via closed-loop control such that the game is personalized to the skills and abilities of each player. This article will describe the development of an adaptive game of Tetris designed to sustain player engagement (see also Chanel et al., 2011) and also an experimental study intended to validate the psychophysiological inference underpinning the system that was conducted prior to the creation of the working prototype.

Game construction began with the formulation of a conceptual framework upon which to model the responses of the adaptive game. Our framework was based upon the Motivational Intensity Model (MIM: Wright, 2008) which describes the relationship between effort investment and task demand; a model that has been corroborated via a number of experimental studies (e.g., Wright and Kirby, 2001; Richter et al., 2008; Richter and Gendolla, 2009). One prediction of this model is that effort rises proportionally with increases in task difficulty until demand is so great that the human deems task success unlikely and withdraws effort, the result of which is a shark-fin shaped effort curve (**Figure 1**). The MIM was adapted to provide a conceptual framework for defining a desirable state of player engagement that could serve as the target for the biocybernetic loop. The adaptation took account of research upon the gaming experience to define an ideal "zone" state for the player. For instance, Csikszentmihalyi (1990) described the ideal or optimal level of engagement as "flow"; a state where engagement with a task is full to the point that time seems to slip away. According to Nacke and Lindley (2008) flow is characterized by an absence of undesirable mental states (i.e., boredom) and entails a positive emotional experience. Similar states, such as being in the zone or total immersion have been described by Chen (2004) and Ryan et al. (2006) respectively. The observation has also been made that situations of high effort promote skill development and an opportunity to demonstrate mastery or competence that leads to a positive gaming experience (Ryan et al., 2006). Thus, the MIM was adapted to represent four broad categories of player state; boredom, engagement, zone and overload (**Figure 1**). The conceptual distinction between these four categories was used to define adaptive goals for the biocybernetic loop, namely:

- To avoid boredom by increasing game demand whenever boredom was detected
- To reduce demand when overload was detected

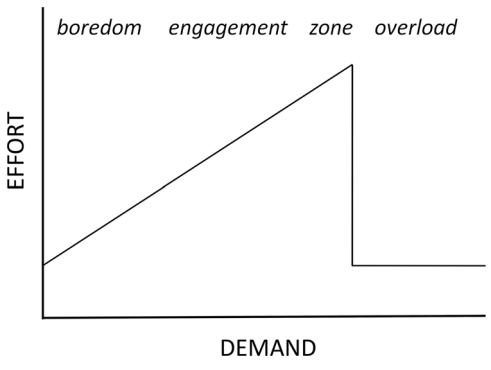


FIGURE 1 | Motivational Intensity Model (MIM) adapted by the addition of four categories of user state.

- To make no adjustment when the player occupied the target states of engagement and zone

In order for the control loop to work within this framework (**Figure 1**) the model must be operationalized using psychophysiological measures. The MIM has been extensively corroborated by cardiovascular indices of mental effort (e.g., Wright and Dill, 1993; Wright and Kirby, 2001; Richter et al., 2008; Richter and Gendolla, 2009), however cardiovascular measures have a number of limitations as inputs to a biocybernetic loop including an inability to diagnose and monitor individual psychological dimensions of effort, e.g., reactivity in blood pressure is simultaneously sensitive to motivation, cognitive effort and physical effort (Cacioppo et al., 2000). By contrast EEG provides a wide choice of metrics that permit a multidimensional monitoring of engagement, including spontaneous oscillations, evoked and event-related potentials (EPs and ERPs), different frequency bands, scalp locations and power values. Multivariate combinations of EEG measures have demonstrated impressive levels of accuracy at discriminating user workload (e.g., Gevins et al., 1998; Prinzel et al., 2003; Scerbo et al., 2003; Chanel et al., 2008; Christensen et al., 2012). Of particular interest are EEG oscillations in the alpha (7.5–13 Hz) and theta (4–7 Hz) bands, which are reliable measures of cortical activation and mental effort (e.g., Gevins et al., 1998; Klimesch, 1999; Wilson, 2002, 2003). In an earlier study (Fairclough et al., 2013), measures of power in the alpha and theta bands were sensitive to manipulations of cognitive demand and motivational incentives using the *N*-back working memory task, however, the capacity of these metrics to index demand and motivation in the context of a computer game remained unknown.

STUDY ONE: VALIDATION OF INPUT MEASURES

Introduction

An experimental study was conducted to evaluate the sensitivity and reliability of the EEG alpha and theta bands to variations in game demand and motivation during the play upon the popular game Tetris. The study aimed to establish: (a) the most suitable EEG measures to use as inputs to a real-time biocybernetic loop; and (b) an appropriate framework for the operationalization of the MIM with respect to measures of spontaneous EEG. The study employed a within subjects design and involved game based manipulations of motivation and demand: three levels of game demand were tested (low, high, excessive) along with two incentive conditions whereby a game-based incentive was present in one condition and absent in the other. It was expected that changes in oscillatory EEG activity in the alpha and theta bands would capture: (1) situations of low effort (i.e., due to boredom or overload); (2) instances of effort increasing in line with demand (when players were engaged with the game) and most significantly; and (3) when players were in the "zone" (when maximal effort was apparent; **Figure 1**). It was also anticipated that the addition of an incentive would increase effort investment provided that game success was likely (Wright, 2008).

Method

Participants

Twenty participants (11 females) took part in the experiment. Participants were aged between 19 and 36 years, and had a mean age of 23.2 years ($SD = 4.02$). All participants were volunteers who gave their written informed consent prior to data collection in accordance with the Declaration of Helsinki.

Game Demand

Cognitive demand was manipulated using an adapted version of the Tetris game. The game requires participants to rotate and move falling pieces in order to build rows of blocks at the bottom of a game board. Falling pieces were one of seven possible colored shapes; each comprised of four squares arranged in different configurations. Pieces were selected to fall in random order. In order to allow gameplay for a fixed duration of 180 s the conventional Tetris game-board was adapted to prevent game-death (when pieces stack to the top of the board to signal game-over). The adaptation consisted of shifting the game-board upwards so that the highest stacked piece was maintained at the center of the game board, and was unable to rise above this level (**Figure 2**).

The speed and quantity of the falling pieces were systematically manipulated to create three levels of game demand (low, high or excessive). In the low demand condition, an average of 22.1 pieces fell with a drop speed of 2.5 board squares s^{-1} An average of 66.2 pieces fell with a drop speed of 6.7 board squares s^{-1} in the high demand condition. In the excessive demand condition, an average of 217 pieces fell at a drop speed of 20 board squares s^{-1}. These parameters were determined on the basis of a small pilot study ($N = 7$).

Incentives

Games were presented in one of two incentive conditions (incentive + performance feedback vs. no incentive + no performance feedback). Each participant completed both incentive conditions (i.e., within-subjects). In the incentive + feedback condition, game coins could be earned for completing rows of Tetris pieces. Coins were accrued in proportion to the number of rows cleared relative to the maximum possible row clearance, such that a maximum of 70 coins could be earned (representing 100% possible clearance). Between zero and seven coins were accumulated every 10 s depending on the proportion of maximum cleared rows achieved at the time of accrual, i.e., at the end of each game, best performance = 70 coins and worst performance = 0 coins (**Figure 2**). Sounds were presented with each award of coins: "kerching" with an award (if current total was less than 35 coins) or "coin jackpot" (if total was over 35 coins). In the no-incentive (+ no feedback) condition, the display related to the coin incentive was absent and no sound effects related to the award of coins were played. For both incentive conditions sound effects occurred when rotating the

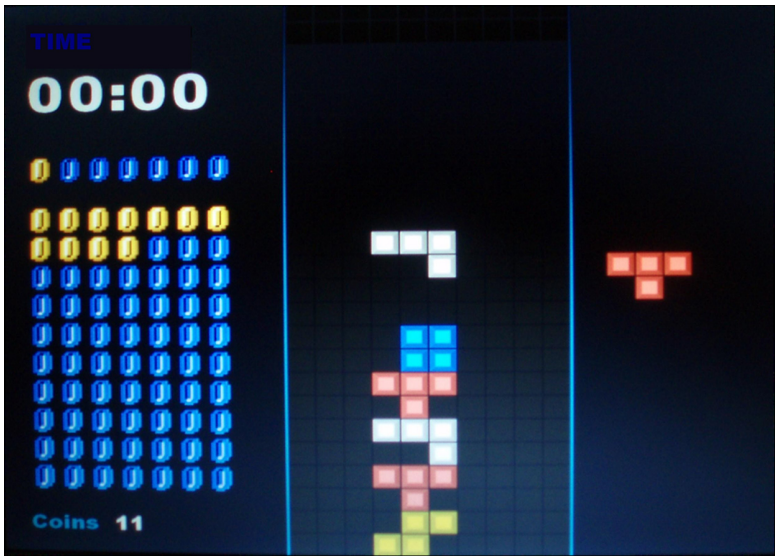


FIGURE 2 | Game-board during incentive + feedback condition: coins are displayed pictorially in a 7 × 10 matrix on the left of the screen and turn from dark blue to gold to indicate coin achievement. A separate row of coins above indicates the number of coins awaiting award at the next 10 s time-point (one coin in this example). The coin score (bottom) and remaining game time are presented in numerals on the left of the screen.

pieces (small "pop") and shifting the pieces left or right (small "snap").

Experimental Design

The experiment consisted of six 180 s games (2 incentive blocks × 3 levels of demand per incentive block). Incentive blocks were delivered in a counterbalanced order and each level of demand presented in random order within each incentive block. *Post hoc T*-tests questionnaires were completed after each game. Throughout each game EEG was measured along with task performance; the total duration of the experimental session was approximately 40 min. Participants practiced by playing each of the six game versions once prior to the experiment and the fitting of EEG equipment. The procedure for the experiment and data collection protocol was approved by the Liverpool John Moores University (LJMU) University Research Ethics Committee and the experiment was conducted in accordance with the recommendations of the LJMU University Research Ethics Committee.

Subjective Questionnaires

Subjective workload was assessed using the NASA Task Load Index (TLX; Hart and Staveland, 1988) which consists of six scales (subjective effort, mental demand, temporal demand, physical demand, perception of performance and frustration). Subjective levels of motivation were assessed using the Dundee State Stress Questionnaire (DSSQ) v1.2 motivation scale, which includes eight items relating to motivation, task enjoyment, desire for success, task value, mental effort, agreeableness on completion, concern over poor performance and eagerness to do well (Matthews et al., 1999). Participants completed one version of each questionnaire immediately after each of the six experimental conditions.

EEG Recording and Analysis

EEG was recorded monopolarly from 64 Ag–AgCl pin-type active electrodes mounted in a BioSemi stretch-lycra head cap. Electrodes were positioned using the international 10–20 system and recorded activity from the following sites: frontal pole (FPz, FP1 and FP2), anterior-frontal (AFz, AF3, AF4, AF7 and AF8), frontal (Fz, F1, F2, F3, F4, F5, F6, F7 and F8), fronto-central (FCz, FC1, FC2, FC3, FC4, FC5 and FC6), central (Cz, C1, C2, C3, C4, C5 and C6), temporal (FT7, FT8, T7, T8, TP7 and TP8), parieto-central (CPz, CP1, CP2, CP3, CP4, CP5 and CP6), parietal (Pz, P1, P2, P3, P4, P5, P6, P7, P8, P9 and P10), occipito-parietal (POz, PO3, PO4, PO7 and PO8) and occipital/inion (Oz, O1, O2 and Iz). Two reference electrodes, the "common mode sense" (CMS) and "driven right leg" (DRL) were used; these function via a feedback loop to drive the participant's voltage (acquired via CMS) as close as possible to zero. AC differential amplifiers performed continuous digitization at 16,384 Hz which was then down-sampled online to 256 Hz. No filters were applied online to allow visual inspection of noise. Offline filtering was performed using a notch filter of 50 Hz and high and low pass filters of 0.05 and 40 Hz respectively. The data were visually inspected for artifacts from external electromagnetic sources. Automatic correction of blink artifacts and horizontal and vertical saccades was performed using detection through predefined topographies. Muscle activity over 100 μV was also excluded. Fast fourier transforms (FFTs) were computed over 50% overlapped windows of 2 s (512 points). The total power in μV^2 was obtained for lower alpha frequency band (7.5–10 Hz), upper alpha frequency band (10.5–13 Hz) and theta frequency band (4–7 Hz; Klimesch, 1999). For the analysis of spectral power in the alpha bands data from the electrodes most spatially representative of the regions of interest were used i.e., frontal (F3, F4); temporal (T7, T8); central (C3, C4); parietal

(P3, P4); occipital (O1, O2). This selection permitted analysis of distributed signals whilst minimizing type one error. The theta band used in this study consisted of a 1 Hz window taken around the frequency of peak modulation within the 4–7 Hz theta range for each participant. This was in order to individualize measurements and maximize their validity. As the majority of participants tend not to produce a clear peak frequency within the theta band, and because there tends to be a large inter-individual variability in the magnitude of the theta response to demand, individualization of the measure was deemed necessary (Gevins et al., 1998). The method involved (for each participant) plotting the spectral power values that lay within the 4–7 Hz theta band for each demand condition on a graph where frequency was represented on the *x*-axis and spectral power on the *y*-axis. The graph for each participant was then visually inspected to discern the theta frequency possessing the greatest demand related modulation of power. Many participants did not display a unique frequency with the greatest power modulation, but instead a small window of similar frequencies that displayed greater modulation than the other theta frequencies; for this reason a 1 Hz window was selected for each participant. Power spectra values for both alpha and theta bands were log transformed (using the natural log) to normalize distribution. A single 180 s continuously recorded data segment was analyzed for each experimental condition.

Statistical Analysis

A priori hypotheses concerning effects for demand were tested using repeated measures analyses of variance (ANOVA). Multivariate analyses are reported using the Pillai's Trace statistic and where multivariate tests failed to reach significance, due to a small sample size ($N = 20$) significant univariate analyses are reported. Greenhouse-Geisser corrections were applied for violations of sphericity as indicated by Mauchly's test. Alpha levels for *a priori* tests were set at 0.05. Significant omnibus effects have been followed up with *post hoc* tests where the alpha levels were corrected to minimize Type one errors using the Bonferroni adjustment.

Results

Performance

A 2 × 3 (incentive × demand) repeated measures ANOVA was performed on game performance scores (i.e., the percentage of successful line completions), which revealed an omnibus effect for demand ($F_{(2,18)} = 504.8$, $p < 0.01$, $\eta^2 = 0.98$). There were no main or interaction effects for the incentive. *Post hoc* tests revealed that performance was significantly reduced at excessive compared to high demand ($p < 0.01$) and low demand ($p < 0.01$). Performance scores were also significantly lower at high compared to low demand ($p < 0.01$), descriptive statistics are presented in **Table 1**.

TABLE 1 | Mean scores and standard deviation (in brackets) for Tetris performance (the percentage of rows completed; $N = 20$).

Demand	Low		High		Excessive	
Incentive	Inc.	No inc.	Inc.	No inc.	Inc.	No inc.
Performance	**70.67** (12.67)	**63.44** (13.43)	**58.59** (26.9)	**49.63** (28.55)	**2.98** (3.26)	**2.41** (2.35)

Subjective Self-Report Data

A 2 × 3 (incentive × demand) MANOVA on scores for the six scales of the NASA TLX revealed significant main effects for demand ($F_{(12,220)} = 22.64$, $p < 0.01$, $\eta^2 = 0.55$) and incentive ($F_{(6,109)} = 2.85$, $p < 0.05$, $\eta^2 = 0.14$). Ratings of mental, physical and temporal demand increased significantly with each increment in demand (all $p < 0.05$). Effort ratings increased from low to high demand ($p < 0.01$) and showed a marginally significant increase at excessive vs. high demand ($p = 0.05$). Perceptions of performance quality were reduced at excessive vs. high and low demand (both $p < 0.01$) while frustration was elevated at excessive vs. high and low demand (both $p < 0.01$). Ratings of mental demand, physical demand and effort all increased with incentive ($p < 0.05$). However there was no effect for incentive upon the ratings of temporal demand, frustration and perception of performance quality; descriptive statistics are provided in **Table 2**.

Scores on items from the DSSQ Motivation subscale had a high internal consistency (Cronbach's alpha = 0.88) so were collapsed into one index of subjective motivation. A demand (3) × incentive (2) repeated measures ANOVA revealed significant main effects for demand ($F_{(2,18)} = 29.42$, $p < 0.01$, $\eta^2 = 0.77$) and incentive ($F_{(1,19)} = 15.16$, $p < 0.05$, $\eta^2 = 0.44$). *Post hoc* *T*-tests indicated enhanced motivation at high demand (high vs. low: $p < 0.01$; high vs. excessive: $p = 0.01$). Motivation was also elevated when the incentive was present for all demand conditions ($p = 0.01$; **Table 2**).

EEG Theta Power

A 2 × 3 repeated measures MANOVA was conducted on theta power data from five frontal (F, FC) and AF sites (AFz, Fz, FCz, F1, F2). This analysis produced a main effect for demand ($F_{(2,18)} = 21.89$, $p < 0.01$, $\eta^2 = 0.71$) and site ($F_{(4,16)} = 38.73$, $p < 0.01$, $\eta^2 = 0.91$). A quadratic trend for demand was significant ($F_{(1,19)} = 19.71$, $p < 0.01$, $\eta^2 = 0.51$) indicating maximum power at high demand. There was no effect of incentive on frontal theta power.

To locate the effects for demand paired sample *T*-tests were conducted on data that had been collapsed across the levels of site and incentive. Theta power was significantly elevated at high vs. low and excessive demand ($p < 0.01$). There was also a marginally significant increase of theta power during excessive compared to low demand ($p = 0.05$).

EEG Alpha Power (7.5–13 Hz)

To discern effects of the manipulations upon spectral power in the alpha band, repeated measures (2 × 3 × 5 × 2) ANOVAs with factors of incentive (incentive, no incentive) × demand (low, high, excessive) × site (frontal (F3, F4), parietal (P3, P4), occipital (O1, O2), central (C3, C4), temporal (T7, T8)) × hemisphere (left, right)

TABLE 2 | Mean and standard deviation (brackets) scores for the six NASA TLX Scales (mental demand, physical demand, temporal demand, frustration, effort and perception of performance) and the DSSQ motivation scale.

Demand	Low		High		Excessive	
Incentive	Inc.	No inc.	Inc.	No inc.	Inc.	No inc.
Mental demand	3.77 (2.29)	3.00 (2.17)	6.50 (1.87)	5.32 (2.03)	7.73 (1.95)	7.00 (2.77)
Physical demand	3.00 (2.25)	2.09 (1.51)	5.50 (2.8)	4.05 (2.58)	6.73 (2.79)	7.00 (2.77)
Temporal demand	2.00 (0.87)	1.95 (1.48)	6.09 (1.45)	6.00 (1.8)	9.27 (1.13)	8.72 (1.92)
Frustration	4.18 (2.58)	3.95 (2.42)	4.18 (2.33)	4.36 (2.53)	8.18 (2.02)	7.64 (2.4)
Effort	4.41 (2.4)	2.95 (1.93)	7.23 (1.78)	5.68 (2.24)	8.41 (1.53)	6.77 (2.61)
Perception of performance	7.27 (2.54)	7.55 (2.05)	7.14 (2.41)	5.73 (2.64)	1.50 (0.99)	2.23 (1.78)
Motivation	6.14 (0.91)	4.86 (1.88)	7.60 (1.04)	6.20 (1.38)	5.73 (1.68)	5.40 (1.31)

Inc., incentive; No inc., no incentive; N = 20.

were performed separately on lower and upper alpha band power.

The omnibus analyses for lower alpha band power (7.5–10 Hz) produced main effects for site ($F_{(4,16)} = 41.05$, $p < 0.01$, $\eta^2 = 0.91$) and hemisphere ($F_{(1,19)} = 4.92$, $p < 0.04$, $\eta^2 = 0.21$). Trend analysis showed a linear trend for hemisphere with reduced lower band power in right hemisphere (statistic as for effect). Interactions were also present in the analysis of lower alpha power for incentive × hemisphere ($F_{(1,19)} = 5.73$, $p < 0.03$, $\eta^2 = 0.23$) and demand × site ($F_{(4,82)} = 4.01$, $p < 0.01$, $\eta^2 = 0.17$). *Post hoc* tests indicated the incentive × hemisphere interaction was related to greater reduction of alpha power in right hemisphere during the incentive condition ($p = 0.02$). The demand × site interaction was linked to a reduction of lower alpha power at occipital sites during high compared to excessive demand ($p = 0.03$); lower alpha was also suppressed at high compared to low demand at temporal sites ($p < 0.01$). Summary statistics for the *post hoc* tests are presented in **Table 3**.

The omnibus ANOVA for upper alpha band (10.5–13 Hz) produced main effects for incentive ($F_{(1,19)} = 6.41$, $p < 0.03$, $\eta^2 = 0.25$), demand ($F_{(2,18)} = 6.62$, $p < 0.01$, $\eta^2 = 0.42$) and site ($F_{(4,16)} = 25.22$, $p < 0.01$, $\eta^2 = 0.86$). There were significant linear trends indicating that upper alpha power decreased as demand increased ($F_{(1,19)} = 13.63$, $p < 0.01$, $\eta^2 = 0.42$) and when the incentive was offered (statistic as for effect). Interactions were also present for incentive × hemisphere ($F_{(1,19)} = 6.81$, $p < 0.02$, $\eta^2 = 026$) and demand × site ($F_{(4,81)} = 8.69$, $p < 0.01$, $\eta^2 = 0.31$).

Post hoc T-tests revealed a reduction of upper alpha power when game coins were present ($p = 0.02$). Upper alpha was also suppressed at excessive compared to high and low demand ($p < 0.01$) and at high compared to low demand ($p = 0.02$) indicating a concomitant drop in upper alpha power as game demand increased.

Analysis of the demand × site interaction revealed a stepwise reduction of upper alpha power as demand increased at parietal, frontal and central sites. However, this demand effect was not apparent at occipital and temporal sites. *Post hoc* tests indicated that the hemisphere × incentive interaction was related primarily to a reduction in power during the incentive condition compared to the no-incentive condition in the right hemisphere ($p < 0.01$). The *t*-values and effect sizes for these *post hoc* tests are displayed in **Table 4**.

TABLE 3 | Differences in power between levels of Tetris demand by region for lower alpha band (*N* = 20).

	Lower alpha band power			
Site		*t*	*p*	η^2
Occipital	excess > high	2.29	0.03	0.22
Temporal	high > low	2.92	0.01	0.31

Discussion

This study was performed to assess the suitability of oscillatory EEG metrics for the real time monitoring of effort and cognitive demand during Tetris play. The results indicated frontal theta was robustly sensitive to objective game demand but that alpha activity only responded to demand at specific sites. For both frontal theta power and subjective motivation there were significant quadratic trends with maxima at high demand indicating that this level stimulated the highest subjective motivation and effort investment, as predicted by the MIM (**Figures 1**, **3**). Upper alpha band (10.5–13 Hz) indicated a linear increase in cortical activation as the challenge of the game increased (**Figure 4**), which corresponded with the trend in subjective workload (**Table 2**). There was no main effect for either manipulation upon the lower alpha band (7.5–10 Hz) however, an interaction with site revealed sensitivity to demand

TABLE 4 | Differences in power between levels of Tetris demand by region for upper alpha band (*N* = 20).

	Upper alpha band power			
Site		*t*	*p*	η^2
Central	low > excess	5.14	<0.01	0.58
	low > high	4.72	<0.01	0.54
	high > excess	3.95	0.01	0.45
Parietal	low > excess	4.24	<0.01	0.49
	low > high	3.55	0.02	0.40
	high > excess	3.18	<0.01	0.35
Frontal	low > excess	4.18	<0.01	0.48
	high > excess	3.31	<0.01	0.37
	low > high	2.46	0.02	0.24

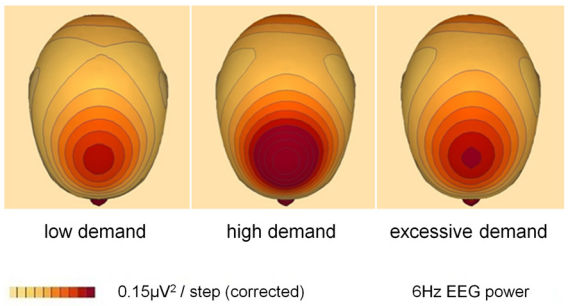


FIGURE 3 | Grand average topographic distribution of spectral power at the frequency of peak power for low, high and excessive demand (*N* = 20: incentive and no-incentive conditions are collapsed). Peak frequency = 6 Hz (the frequency at which a clear peak in EEG power was evident within the 4–7 Hz range); this was identified by visual inspection of the grand average frequency-power spectral plot. Images were constructed using spherical spline interpolation.

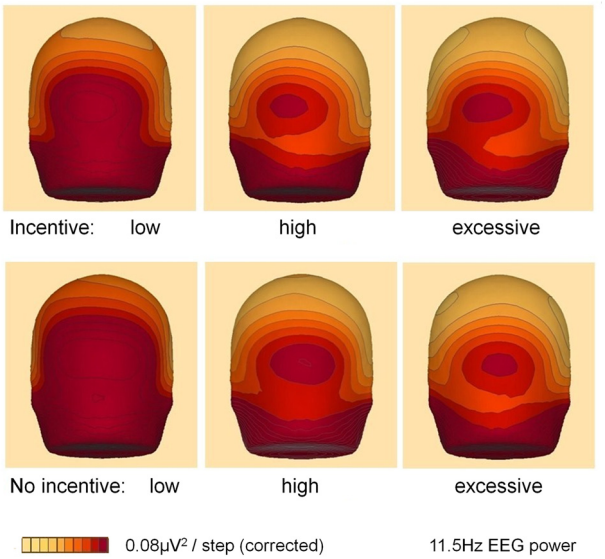


FIGURE 4 | Grand average spectral electroencephalography (EEG) power at 11.5 Hz (*N* = 20) for low, high and excessive cognitive demand on the Tetris game with and without a game based incentive (spherical spline interpolated; image displays rear of scalp).

over temporal and occipital areas of the scalp. The sensitivity of upper alpha activity to game demand was specific to frontal, central and parietal sites. In addition, upper alpha was the only frequency band to respond to the incentive coins (greater power reduction when game coins were present over the right hemisphere).

Augmentation of frontal theta has been widely reported in association with sustained attention, increased cognitive control and working memory (Gevins et al., 1998; Klimesch, 1999; Jensen and Tesche, 2002; Gevins and Smith, 2003; Sauseng et al., 2005; Cavanagh and Frank, 2014; Hsieh and Ranganath, 2014; Clayton et al., 2015). However, the decline of frontal theta power under conditions of excessive demand (**Figure 3**) has not previously been observed. The reproduction of this pattern in Tetris players provided an indication of the ecological validity of this metric and the ability of frontal theta to retain sensitivity to demand when generalized to spatial cognition in a gaming context. The capacity of frontal theta to act as a "generic" index of mental effort makes it an appropriate input to a closed-loop system since games typically use different elements of cognition at different stages of play. In addition, frontal theta demonstrated a degree of face validity owing to the similar pattern of modulation between EEG activity in this band and subjective motivation. The large effect sizes attest to the sensitivity of this measure and its capacity to discriminate between three or more categories of demand as well as detect the "tipping point" where effort is withdrawn due to overload (**Figure 1**).

Alpha power in the upper band, which is associated with task-specific cognitive processes (Klimesch, 1999), was suppressed as demand increased from low to high to excessive levels (**Figure 4**); a finding supported by a significant body of literature on cortical activation (e.g., Pfurtscheller, 1992; Gevins et al., 1998; Fournier et al., 1999; Klimesch, 1999). However, this main effect did not extend to lower band power (an index of cortical arousal and alertness), instead an interaction between demand and site showed that sensitivity of lower alpha band was limited to occipital and temporal areas. The lessening of power in the upper alpha band, and hence the level of cortical activation, was maximal during excessive demand despite a reduction of frontal theta power at this level. This suggests that upper alpha reflected the objective level of task demand upon spatial cognition (e.g., the processing of high numbers of fast moving stimuli in the form of falling Tetris blocks) whereas frontal theta represented the level of effort mobilization in the face of excessive demand (i.e., a withdrawal of effort). These findings suggested that a two-dimensional space could be created akin to the MIM (**Figure 1**) wherein demand is represented by upper alpha power and frontal theta power is used as an index of mental effort.

The sensitivity of the alpha band was found to vary across recording sites. Upper band effects occurred at frontal, central and parietal sites which provides some agreement with other studies linking these cortical areas with mental rotation—a key cognitive component of Tetris play (e.g., Inoue et al., 1998; Yoshino et al., 2000). Conversely, the effects of demand in the lower alpha band were restricted to temporal and occipital electrodes. In addition, the lower band revealed stronger activation in right hemisphere, which is traditionally associated with spatial tasks (Hellige, 1993), whereas the upper band indicated bilateral sensitivity to game demand. This regional variation indicates the importance of targeting the right cortical sites in order to maximize the sensitivity of the EEG metrics to the chosen psychological variables.

The results from the study identified two EEG measures as suitable inputs to a biocybernetic loop designed to control an adaptive game of Tetris. Frontal theta was selected to index mental effort due to its sensitivity to this variable, its reliability and its specificity (i.e., theta did not respond to

the incentive+feedback manipulation), The sensor location Fz, which generally lies at the center of the scalp area associated with frontal theta augmentation was selected as the recording site. Power in the upper alpha band (10.5–13 Hz) was selected to index the level of task cognition; this variable was sensitive to the objective difficulty of the task and demonstrated a linear pattern over the three levels of demand in accordance with subjective workload ratings. There is also strong literature based support for the involvement of upper alpha band with task related cognition, including mental rotation (for a review see Klimesch, 1999). The right parietal site P4 was the chosen sensor input for the sampling of upper alpha oscillations. A parietal site was selected because in the first study parietal sites P3 and P4 detected sensitivity to game demand; central sites were also responsive but there were concerns that these would be subject to confounds from motor activity associated with game play. Although sensitivity was recorded at frontal sites this was smaller in magnitude than the parietal response to demand (**Table 4**). The choice of recording site was also constrained to the set of sites analyzed in study one i.e., frontal (F3, F4), temporal (T7, T8), central (C3, C4), parietal (P3, P4) and occipital (O1, O2) to preserve the validity of the psychophysiological inference regarding game-related cognition. Although there was no interaction of demand with hemisphere in the first study to guide this selection of site, the right hemisphere electrode P4 was selected on the basis of a robust association of right hemisphere with spatial cognition (Klimesch, 1999).

To summarize, the selection of the two EEG inputs to the biocybernetic loop made it possible to operationalize the adapted MIM (**Figure 1**) i.e., frontal theta was used to represent effort and parietal upper alpha to represent game demand (**Figure 5**). According to this conceptual model the desirable states of "zone" and "engagement" are associated with high effort while undesirable states are defined by low effort combined with high demand (overload) or combined with low demand (boredom).

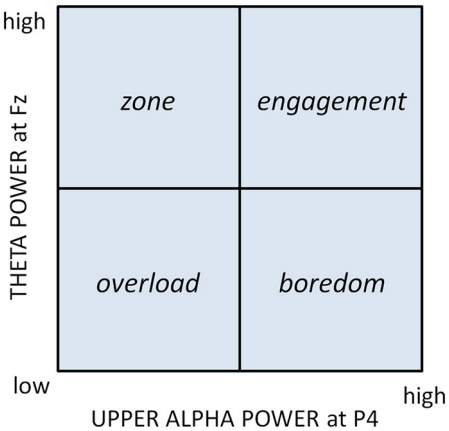


FIGURE 5 | Two dimensional representation of the user state using EEG measures (cortical activation is inversely proportional to alpha band power).

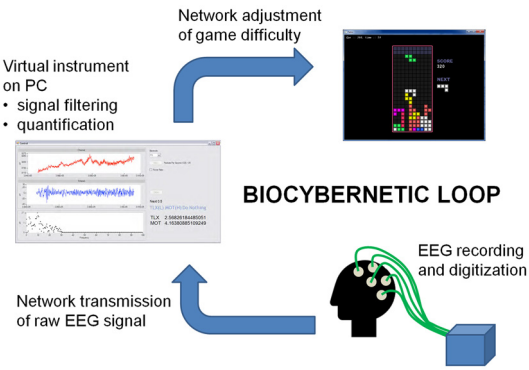


FIGURE 6 | Components of the biocybernetic loop.

DEVELOPMENT OF THE REAL-TIME ADAPTIVE GAMING SYSTEM

The working biocybernetic loop was created from a network that involved the connection of two PCs; one PC that ran the adaptive Tetris Software and a second PC that hosted a virtual instrument (VI) constructed with LabVIEW. Raw EEG data were transmitted to the VI to be filtered and averaged prior to transformation into estimates of motivation and workload by a state classification algorithm. These estimates were defined in terms of the four states of boredom, engagement, zone and overload (**Figure 5**). If the state fell within the undesirable categories of boredom or overload, a signal would be transmitted to the adaptive Tetris Software in order to adjust the level of game demand. The components of this loop are illustrated in **Figure 6**.

EEG data was recorded monopolarly from two Ag-AgCl pin-type active electrodes mounted in a BioSemi head cap at the locations Fz and P4 (sites determined by the 10–20 system). AC differential amplifiers amplified signals at source with continuous digitization at 16,384 Hz and online down sampling to 512 Hz. No filters were applied online to allow visual inspection of noise. The EEG signal was filtered using a Kasier Finite Impulse Response (FIR) of 2–30 Hz then subjected to a FFT in real time using a 2 s Hanning window. Theta activity between 4–8 Hz was obtained from the midfrontal electrode Fz and activity in the upper alpha band (10.5–13 Hz) was derived from right parietal site P4. The FFT calculated power spectra for each frequency band to generate total power values for each measure. These values were then converted to estimates of workload (upper alpha) and motivation (frontal theta).

For the operational model to trigger adaptations of game demand in real-time, it was necessary to select criteria for adaptation so that the four regions of the user state model could be defined (**Figure 5**). To maximize the effectiveness of adaptation, it was desirable to calibrate the criteria or trigger levels to individual players to counteract individual variability in the magnitude of EEG responses to game demand (Gevins et al., 1998).

The criteria for triggering adaptations of the Tetris interface were developed based upon patterns of theta and upper alpha oscillations that were observed relative to a baseline reading. Our participants were required to watch a relaxing video clip (Piferi et al., 2000) in order to establish baseline EEG levels of frontal theta and (parietal) upper alpha for each participant. Baselined derivatives of theta and alpha were captured in 5-s windows during subsequent game play. For example, if frontal theta activity increased or decreased from baseline by 100% in any 5-s window whilst parietal alpha increased or decreased by 100% then system adaptation may be triggered. In practice, frontal theta and parietal alpha were assessed every 5 s as the participant played the adaptive version of Tetris. If the system detected that frontal theta had decreased by 100% or more (from baseline) whilst parietal alpha had increased by 100 or more (from baseline), the player was assessed to be in a state of boredom (**Figure 5**). If the decrease of frontal theta was accompanied by a decrease of parietal alpha, the player was deemed to be in a state of overload.

A straightforward strategy for the adaptation of the game interface was used, i.e., reducing or increasing the drop speed of the falling Tetris blocks to manipulate game difficulty. Speed was increased if the player was deemed to be in a state of boredom and decreased if overload was detected (**Figure 5**). If neither of those states were detected by the system, the drop speed of the Tetris blocks was maintained. This assessment took place in 5-s epochs, hence the drop speed of the game increased or decreased over a period of play depending on the relative frequency of "boredom" or "overload" epochs that occurred within that period.

A series of pilot tests were conducted to determine an appropriate magnitude of the drop speed changes and whether or not to incorporate feedback of drop speed into the interface. The outcomes from these tests indicated that small adjustments without any overt feedback of drop speed were the most acceptable version of the Tetris interface from a user perspective. This design corresponded to a covert adaptive strategy where the adaptive process is expected to produce a gradual impact rather than an immediate impact on player state. This strategy was adopted in order to focus the attention of the players on the game as opposed to the ongoing activity of the biocybernetic loop.

STUDY TWO: EVALUATION OF THE BIOCYBERNETIC LOOP

Introduction

A study was conducted to evaluate the adaptive Tetris game with respect to two questions: (1) does adaptation improve player experience compared to a manual adjustment of game demand; and (2) how does varying the reactivity of the biocybernetic loop (i.e., liberal vs. conservative trigger levels) impact upon player experience and the behavior of the closed-loop. The first question contrasts a covert, automated process of adjustment with a scenario where adjustments of game demand are both overt and manually instigated by the player. The second question pertains to the design of the trigger events for adaptation and how psychophysiological criteria can impact upon the process of system adaptation and the player experience.

Method

Design

Three types of biocybernetic loop were compared: (a) a conservative system that produced an upward or downward adjustment of game demand (i.e., drop speed) when changes in frontal theta and parietal alpha substantially deviated from baseline (greater than 200%); (b) a liberal system that adjusted game demand in response to smaller deviations from baseline EEG activity (100%); and (c) a moderate system that responded to moderate changes in EEG (150%). It was anticipated that the conservative system would be the least reactive and would respond slowly and only to extreme examples of boredom and overload. By contrast, the liberal system was expected to make frequent adjustments and be the most responsive to instances of boredom/overload. For the fourth system, which operated under manual control, participants were required to speak aloud an instruction to increase ("higher") or decrease ("lower") the speed of the falling blocks. These adjustments were made in real-time by an experimenter sitting behind a screen in the laboratory. Ten participants played each of the four Tetris games (conservative closed-loop, liberal closed-loop, moderate closed-loop, manual) for 5 min. The order of presentation of each system was counterbalanced and participants were given a 5 min rest break between each game. Every game began on the slowest speed setting. If the blocks reached the top of the board and "game death" occurred, the game would restart with an empty board on the slowest speed setting. The procedure for the experiment and data collection protocol was approved by the LJMU University Research Ethics Committee and the experiment was conducted in accordance with the recommendations of this same committee.

Participants

Ten volunteers (6 females) participated in the evaluation session. A repeated measures design was used where each participant encountered each of the four versions of the system (conservative/liberal/moderate/manual). All participants were volunteers who gave their written informed consent prior to data collection in accordance with the Declaration of Helsinki.

Subjective Measures

Player experience was analyzed using subjective measures of mood and game immersion. The mood adjective checklist (UMACL; Matthews et al., 1990) assesses three components of mood: energetical arousal (EA: tired-alert), tense arousal (TA: relaxed-tense) and hedonic tone (HT: happy-sad). The UMACL was administered before and after each game to allow calculation of the change scores (post- minus pre-game) for each mood component. Participants also completed the Immersive Experience Questionnaire (IEQ) designed to capture the immersive quality of the gaming experience (Jennett et al., 2008); this scale was administered after each game.

Measures of System Behavior

Data were obtained in order to quantify the behavior of each version of the system. This enabled the three versions of the adaptive closed-loop to be contrasted with one another and an understanding to be acquired of how they differed from the manual control system. Three aspects of system behavior were measured for each system version:

- The mean frequency of increases and decreases in game demand
- The mean frequency of game deaths/resets (when blocks reached the top of the board and the game required resetting)
- The average demand level of each game (game difficulty could vary between 1 and 10 in accordance with the drop speed of falling blocks)

Results

An ANOVA analysis was conducted on each of the three measures of system behavior (see **Table 5** below for the descriptive statistics). Each measure (mean freq. of increases/decreases in demand, mean freq. of game deaths (resets), average game difficulty) was subjected to a one-way ANOVA to assess statistical significance. The number of adjustments to increase task demand was significantly higher for the conservative system compared to the other three systems; unsurprisingly, all three biocybernetic systems exhibited a higher rate of upward adjustment compared to the manual system ($F_{(3,7)} = 79.40$, $p < 0.01$). The analysis of downward adjustment (to decrease game demand) revealed that automated decreases of demand occurred more frequency during games played with the moderate and liberal versions of the biocybernetic loop ($F_{(3,7)} = 18.4$, $p < 0.01$). The analysis of reset frequency indicated that resets were most common in the conservative system, however, this increase failed to reach statistical significance. The analysis of mean demand level indicated that difficulty was significantly lower for the liberal system compared to all other systems ($F_{(3,7)} = 12.3$, $p < 0.01$).

The impact of system adaptation on the user experience was assessed using two types of subjective questionnaire; the IEQ and the UMACL mood adjective checklist. The UMACL was administered before and after each game session in order for us to calculate a change score that quantified the changes in the three components of mood: EA (alert-tired), TA (tense-relax) and HT (happy-sad). All three components were subjected to a one-way ANOVA; mean values are displayed in **Table 6**.

TABLE 5 | Mean values for measures of system adaptation across the four systems ($N = 10$).

System	Increase demand	Decrease demand	Mean reset	Mean difficulty level
Conservative	63.6	43.2	1.3	3.8
Moderate	41.7	56.6	0.6	2.4
Liberal	28.6	62.4	0.4	1.9
Manual	9.4	1.7	0.5	3.3

Highest difficulty level = 10; lowest difficulty = 1.

TABLE 6 | Mean values for subjective data: EA, energetic arousal (change score); TA, tense arousal (change score); HT, hedonic tone (change score); IEQ, immersion ($N = 10$).

System	EA	TA	HT	IEQ
Conservative	4.3	3.1	0.0	64.7
Moderate	2.0	2.4	−1.9	65.5
Liberal	0.2	1.1	−2.0	66.1
Manual	1.1	0.7	−1.6	73.9

The mean values for the changes in mood indicated some consistent trends, namely that participants found the game to be alerting and conducive to tension and negative affect. An ANOVA analysis of all three mood components revealed a significant effect for EA (energetic arousal) only ($F_{(3,7)} = 5.48$, i.e., $p < 0.05$), i.e., participants found the experience of playing the conservative version of the biocybernetic game to be more alerting compared to the liberal version ($p < 0.05$). The analysis of responses to the immersion questionnaire was insignificant, but a trend was observed that participants found the manual version of the game to be the most immersive.

Discussion

This evaluation study demonstrated how the reactivity of the biocybernetic loop affected the performance of the system and the experience of players.

The analysis of system behavior revealed that the conservative system provided the greatest level of challenge, i.e., it produced the highest average level of demand and made the largest number of adjustments to increase game demand. This skew towards increased adjustment of demand was mirrored by the liberal system, which tended to adjust difficulty in the opposite direction, such that the liberal version produced the lowest number of game deaths and lowest average level of demand. The moderate system produced a pattern of upward and downward adjustments that represented a midpoint between that of the conservative and liberal systems. As anticipated, the number of adjustments made manually by participants was lower than the numbers produced by the biocybernetic loop as they tended to simply increase the level of difficulty to their preferred level early in the game without making any subsequent adjustments. The mean level of difficulty during play on the manual system (included as benchmark to compare with the adaptive protocols) provided an indication of the optimal level of demand for the group (3.3). By contrast the conservative system generally pushed the players to a higher level of demand (3.8), resulting in the greatest number of game deaths; the moderate and liberal systems tended to set difficulty at a lower level than the manual system on average. Therefore, the three adaptive systems and their respective triggers tended to either over- or undershoot the mean level of demand that was preferred by our participants.

It was noteworthy that the conservative system produced a large number of upward adjustments in game demand (63.6) suggesting that this system was detecting boredom via the EEG (i.e., a 200% decrease in theta and increase in alpha relative to

baseline; **Figure 5**). Boredom may have resulted from games starting on the slowest drop speed setting and the return to the slowest speed when the game was reset (**Table 5**). By contrast, the liberal system produced more downwards than upwards adjustments even though games started at the easiest level, meaning on some occasions where the trigger criteria for a downward adjustment was fulfilled the interface was unable to slow the speed because the player was already at the lowest level. The liberal system was also detecting more player overload than the other two adaptive systems (more downward adjustments), i.e., the trigger criteria of a 100% decrease in theta and alpha power relative to baseline were fulfilled the most frequently (**Figure 5**). This is surprising in view of the low levels of demand delivered by the liberal system. One explanation may be that the EEG indicators of overload used were incorrect and that simultaneous decreases in alpha and theta power of around 100% indicate low effort (reduced frontal theta) combined with low levels of sensory processing (reduced parietal alpha) instead of overload (Klimesch, 1999). It may be that deviations of 200% or more are required to indicate overload where excessive demand leads to a high level of alpha power suppression, as occurred when demand was excessive in the first study (**Figure 4**). This underlines the importance of not only selecting the best combination of input measures for the biocybernetic loop, but also of defining accurate trigger criteria in terms of the relative magnitudes of the input variables.

It was expected that player experience would be affected by the different outcomes in system behavior between the four versions of the system. However, there were few statistically significant effects on mood and immersion. Alertness was enhanced under the conservative system relative to the liberal system but there were no other significant effects. Of the four systems analyzed the conservative version of the loop produced the most desirable overall impact on player mood, i.e., it evinced the greatest increase in arousal and least negative affect which may be because participants were too challenged to dwell upon their emotional state. This may be because the conservative system was the most successful at detecting boredom and alleviating it with increases in demand. Conversely ratings of immersion in the game were greatest for the manual system, which may reflect the impact of taking momentary breaks from the game to voluntarily control difficulty with a verbal instruction. Even very short breaks from a task are known to increase vigilance performance (Ariga and Lleras, 2011) and opportunities for control can increase the intrinsic motivation for a task (Fisher, 1978). Alternatively, it may be argued the level of demand during the manual control condition was optimal for enhancing immersion. The observation that play on the game increased negative affect under all but the conservative system was unanticipated and may reflect the impact of the lower levels of challenge experienced by participants.

Based upon the results, it would appear that the criteria used to define the three versions of the biocybernetic loop were too similar to evince much difference in player experience. It may be speculated that if players were provided with more time to experience play upon each version of the system they may have been better able to differentiate their respective experiences.

The biocybernetic loop employed a straightforward linear process of calibration to the individual instead of machine learning algorithms. The rationale for this approach was that our psychophysiological measures, EEG theta and upper alpha frequency power, had been validated prior to construction of the loop—and we wished to preserve the transparency of both measures and criteria when testing the working loop. However, there may have been scope to use machine learning during calibration such that more precise linear models may have been generated especially for each participant.

The results highlighted a number of questions surrounding the evaluation of working biocybernetic systems, particularly with respect to the benchmarking of system performance. In this study, a manual system was selected as the benchmark for comparative purposes on the assumption that participants would tailor gameplay to their personal preference. However, this comparison was asymmetrical because the locus of system control for a manual system resided with the user while control was automated within the biocybernetic loop. This is a significant factor when comparing player experience across automated and manual systems since the opportunity for control over a task (as provided by the manual system) is known to affect the level of engagement with that task (Fisher, 1978; Wright and Kirby, 2001). Comparisons with other autonomous systems may therefore be more informative. For example, benchmarking against a system that adapts game demand in a random fashion without an objective rationale, or by using a "yoked" system where the game responds to the physiology of another individual (Bailey et al., 2006). Either of these options may have provided a more parsimonious comparison with the three versions of the working biocybernetic loop.

GENERAL DISCUSSION

This article has described the process of creating a working biocybernetic loop whereby hypotheses derived from experimental work on EEG were first validated in a gaming context in order to select the input measures for the loop. Predictions regarding the modulation of EEG frontal theta and alpha power by variations in the level of cognitive demand and effort were validated during Tetris play; subsequently an adaptive game of Tetris was built that used a biocybernetic loop with the EEG measures tested during the validation stage. Our development process for this prototype exemplifies the principle of designing interactive technologies based upon a theory-driven process of psychophysiological inference (Fairclough, 2009).

The evaluation of autonomous, closed-loop control systems raises important issues for the development of biocybernetic adaptation. The relationship between criteria or categories of psychophysiological activity and the triggering of adaptive responses at the interface requires careful design. The derivation of valid input measures and effective categorization of psychophysiological data in real-time is one stage of this process. Once a method of categorizing the states of the user has been defined (**Figure 5**), these classes must be mapped onto

appropriate responses at the interface. This mapping reflects more than a simple linkage between state x and response y; decisions must be made regarding the frequency and likelihood of those responses as well as the temporal characteristics and relative magnitude of the adaptations. As was demonstrated in the evaluation study, once a working biocybernetic loop has been constructed, responses may be adjusted to optimize the user experience, a process that inevitably involves exploring the interaction between the user and the adaptive response. The behavior of the biocybernetic loop and the interaction between user psychophysiology and adaptive control is an object of study in itself.

Together these two studies provide a potential blueprint for the development and evaluation of a biocybernetic loop. However, further research is required to incorporate psychophysiological theory into the design of physiological computing systems and to develop an effective methodology for system evaluation.

AUTHOR CONTRIBUTIONS

The design and experimental protocol used for Study 1 was developed by KCE and SHF. The Tetris game used in Study 1 was built by KG. Data collection and analysis for Study 1 was performed by KCE. The design and experimental protocol for Study 2 was developed by KG and SHF. Build of the biocybernetic loop and adaptive game of Tetris for Study 2 was performed by KG as was the data collection and analysis for this study. The manuscript was written and edited by KCE (primary author), SHF and KG.

FUNDING

This research was funded by a grant from the REFLECT Project; the European Union's Future and Emerging Technologies Scheme, 7th Framework Programme (FP7). Grant number 215893. Project website: http://reflect.pst.ifi.lmu.de/.

REFERENCES

Ariga, A., and Lleras, A. (2011). Brief and rare mental "breaks" keep you focused: deactivation and reactivation of task goals pre-empt vigilance decrements. *Cognition* 118, 439–443. doi: 10.1016/j.cognition.2010.12.007

Bailey, N. R., Scerbo, M. W., Freeman, F. G., Mikulka, P. J., and Scott, L. A. (2006). Comparison of a brain-based adaptive system and a manual adaptable system for invoking automation. *Hum. Factors* 48, 693–709. doi: 10.1518/001872006779166280

Baldwin, C. L., and Penaranda, B. N. (2012). Adaptive training an artificial neural network and EEG metrics for within- and cross-task workload classification. *Neuroimage* 59, 48–56. doi: 10.1016/j.neuroimage.2011.07.047

Cacioppo, J. T., and Tassinary, L. G. (1990). Inferring psychological significance from physiological signals. *Am. Psychol.* 45, 16–28. doi: 10.1037/0003-066x.45.1.16

Cacioppo, J. T., Tassinary, L. G., and Berntson, G. G. (2000). "Psychophysiological science," in *Handbook of Psychophysiology*, 2nd Edn. eds J. T. Cacioppo, L. G. Tassinary and G. G. Berntson (Cambridge: Cambridge University Press), 3–26.

Cavanagh, J. F., and Frank, M. J. (2014). Frontal theta as a mechanism for cognitive control. *Trends Cogn. Sci.* 18, 414–421. doi: 10.1016/j.tics.2014.04.012

Chanel, G., and Mühl, C. (2015). Connecting brains and bodies: applying physiological computing to support social interaction. *Interact. Comput.* 27, 534–550. doi: 10.1093/iwc/iwv013

Chanel, G., Rebetez, C., Betrancourt, M., and Pun, T. (2008). "Boredom, engagement and anxiety as indicators for adaptation to difficulty in games," in *MindTrek '08 Proceedings of the 12th International Conference on Entertainment and Media in the Ubiquitous Era.* (New York, NY, USA), 13–17.

Chanel, G., Rebetez, C., Betrancourt, M., and Pun, T. (2011). Emotion assessment from physiological signals for adaptation of game difficulty. *IEEE Trans. Syst. Man Cybern. Part A Syst. Hum.* 41, 1052–1063. doi: 10.1109/tsmca.2011.2116000

Chen, J. (2004). Flow in games (and everything else). *Commun. ACM* 50, 31–34. doi: 10.1145/1232743.1232769

Christensen, J. C., Estepp, J. R., Wilson, G. F., and Russell, C. A. (2012). The effects of day-to-day variability of physiological data on operator functional state classification. *Neuroimage* 59, 57–63. doi: 10.1016/j.neuroimage.2011.07.091

Clayton, M. S., Yeung, N., and Cohen Kadosh, R. (2015). The roles of cortical oscillations in sustained attention. *Trends Cogn. Sci.* 19, 188–195. doi: 10.1016/j.tics.2015.02.004

Csikszentmihalyi, M. (1990). *Flow: The Psychology of Optimal Experience.* New York, NY: Harper Perennial.

Dekker, A., and Champion, E. (2007). Please biofeed the zombies: enhancing the gameplay and display of a horror game using biofeedback. *DiGRA* 550–558.

Fairclough, S. H. (2007). Psychophysiological inference and physiological computer games," in *Paper Presented at the ACE Workshop-Brainplay'07: Brain-Computer Interfaces and Games* (Salzburg, Austria).

Fairclough, S. H. (2009). Fundamentals of physiological computing. *Int. Comput.* 21, 133–145. doi: 10.1016/j.intcom.2008.10.011

Fairclough, S. H., Gilleade, K., Ewing, K. C., and Roberts, J. (2013). Capturing user engagement via psychophysiology: measures and mechanisms for biocybernetic adaptation. *Int. J. Auton. Adapt. Commun. Syst.* 6, 63–79. doi: 10.1504/ijaacs.2013.050694

Fisher, C. D. (1978). The effects of personal control, competence and extrinsic reward systems on intrinsic motivation. *Organ. Behav. Hum. Perform.* 21, 273–288. doi: 10.1016/0030-5073(78)90054-5

Fournier, L. R., Wilson, G. F., and Swain, C. R. (1999). Electrophysiological, behavioral and subjective indexes of workload when performing multiple tasks: manipulations of task difficulty and training. *Int. J. Psychophysiol.* 31, 129–145. doi: 10.1016/s0167-8760(98)00049-x

Freeman, F. G., Mikulka, P. J., Prinzel, L. J., and Scerbo, M. W. (1999). Evaluation of an adaptive automation system using three EEG indices with a visual tracking task. *Biol. Psychol.* 50, 61–76. doi: 10.1016/s0301-0511(99)00002-2

Gevins, A., and Smith, M. E. (2003). Neurophysiological measures of cognitive workload during human-computer interaction. *Theoretical Issues in Ergonomic Science* 4, 113–131. doi: 10.1080/14639220210159717

Gevins, A., Smith, M. E., Leong, H., McEvoy, L., Whitfield, S., Du, R., et al. (1998). Monitoring working memory load during computer-based tasks with EEG pattern recognition models. *Hum. Factors* 40, 79–91. doi: 10.1518/001872098779480578

Gilleade, K. M., Dix, A., and Allanson, J. (2005). Affective videogames and modes of affective gaming: assist me, challenge me, emote me, in *Proceedings of DIGRA'2005*, (Vancouver, BC: Canada). http://www.hcibook.com/alan/papers/DIGRA2005/.

Hart, S. G., and Staveland, L. E. (1988). "Development of NASA-TLX (task load index): results of empirical and theoretical research," in *Human Mental Workload*, eds P. A. Hancock and N. Meshkati (Amsterdam: Elsevier Science), 139–183.

Hellige, J. B. (1993). *Hemispheric Asymmetry.* Cambridge, MA: Harvard University Press.

Hsieh, L. T., and Ranganath, C. (2014). Frontal midline theta oscillations during working memory maintenance and episodic encoding and retrieval. *Neuroimage* 85, 721–729. doi: 10.1016/j.neuroimage.2013.08.003

Inoue, M., Yoshino, A., Suzuki, A., Ogasawara, T., and Nomura, S. (1998). Topographic study of human event-related potentials using a task requiring mental rotation,. *Neurosci. Lett.* 253, 107–110. doi: 10.1016/s0304-3940(98)00620-x

Jennett, C., Cox, A. L., Cairns, P., Dhoparee, S., Epps, A., Tijs, T., et al. (2008). Measuring and defining the experience of immersion in games. *Int. J. Hum. Comput. Stud.* 66, 641–661. doi: 10.1016/j.ijhcs.2008.04.004

Jensen, O., and Tesche, C. D. (2002). Frontal theta activity increases in humans with memory load in a working memory task. *Eur. J. Neurosci.* 15, 1395–1399. doi: 10.1046/j.1460-9568.2002.01975.x

Kapoor, A., Burleson, W., and Picard, R. W. (2007). Automatic prediction of frustration. *Int. J. Hum. Comput. Stud.* 65, 724–736. doi: 10.1016/j.ijhcs.2007.02.003

Klimesch, W. (1999). EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain Res. Rev.* 29, 169–195. doi: 10.1016/s0165-0173(98)00056-3

Liu, C., Conn, K., Sarkar, N., and Stone, W. (2007). "Affect recognition in robot assisted rehabilitation of children with autistic spectrum disorder," in *IEEE International Conference on Robotics and Automation* (Rome, Italy), 1755–1760.

Matthews, G., Jones, D. M., and Chamberlain, A. G. (1990). Refining the measurement of mood: the UWIST mood adjective checklist. *Br. J. Psychol.* 81, 17–42. doi: 10.1111/j.2044-8295.1990.tb02343.x

Matthews, G., Joyner, L., Gilliland, K., Campbell, S. E., Huggins, J., and Falconer, S. (1999). "Validation of a comprehensive stress state questionnaire: towards a state 'Big Three'?" in *Personality Psychology in Europe* (Vol. 7), eds I. Mervielde, I. J. Deary, F. De Fruyt and F. Ostendorf (Tilburg: Tilburg University Press), 335–350.

Nacke, L. E., Kalyn, M., Lough, C., and Mandryk, R. L. (2011). "Biofeedback game design: using direct and indirect physiological control to enhance game interaction," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC: ACM), 103–112.

Nacke, L., and Lindley, C. A. (2008). "Flow and immersion in first-person shooters: measuring the player's gameplay experience," in *Proceedings of the Conference on Future Play: Research, Play, Share* (New York, NY: ACM).

Novak, D., Mihelj, M., and Munih, M. (2012). A survey of methods for data fusion and system adaptation using autonomic nervous system responses in physiological computing. *Int. Comput.* 24, 153–172. doi: 10.1016/j.intcom.2012.04.003

Pfurtscheller, G. (1992). Event-related synchronization (ERS): an electrophysiological correlate of cortical areas at rest. *Electroencephalogr. Clin. Neurophysiol.* 83, 62–69. doi: 10.1016/0013-4694(92)90133-3

Piferi, R., Kline, K. A., Younger, J., and Lawler, K. A. (2000). An alternative approach for achieving cardiovascular baseline: viewing an aquatic video. *Int. J. Psychophysiol.* 37, 207–217. doi: 10.1016/s0167-8760(00)00102-1

Pope, A. T., Bogart, E. H., and Bartolome, D. S. (1995). Biocybernetic system evaluates indices of operator engagement in automated task. *Biol. Psychol.* 40, 187–195. doi: 10.1016/0301-0511(95)05116-3

Prinzel, L. J., Freeman, F. G., Scerbo, M. W., Mikulka, P. J., and Pope, A. T. (2000). A closed-loop system for examining psychophysiological measures for adaptive task allocation. *Int. J. Aviat. Psychol.* 10, 393–410. doi: 10.1207/s15327108ijap1004_6

Prinzel, L. J., Parasuraman, R., Freeman, F. G., Scerbo, M. W., Mikulka, P. J., and Pope, A. T. (2003). *Three Experiments Examining the use of Electroencephalogram, Event-Related Potentials and Heart-Rate Variability for Real-Time Human-Centerd Adaptive Automation Design. NASA/TP-2003-212442*. Langley, VA: NASA Langley Research Center.

Richter, M., and Gendolla, G. H. E. (2009). The heart contracts to reward: monetary incentives and preejection period. *Psychophysiology* 46, 451–457. doi: 10.1111/j.1469-8986.2009.00795.x

Richter, M., Friedrich, A., and Gendolla, G. H. (2008). Task difficulty effects on cardiac activity. *Psychophysiology* 45, 869–875. doi: 10.1111/j.1469-8986.2008.00688.x

Ryan, R. M., Rigby, C. S., and Przybylski, A. (2006). The motivational pull of video games: a self-determination approach. *Motiv. Emot.* 30, 344–360. doi: 10.1007/s11031-006-9051-8

Sauseng, P., Klimesch, W., Doppelmayr, M., Pechersorfer, T., Freunberger, R., and Hanslmayr, S. (2005). EEG alpha synchronization and functional coupling during top-down processing in a working memory task. *Hum. Brain Mapp.* 26, 148–155. doi: 10.1002/hbm.20150

Scerbo, M. W., Freeman, F. G., and Mikulka, P. J. (2003). A brain-based system for adaptive automation. *Theor. Issues Ergon. Sci.* 4, 200–219. doi: 10.1080/1463922021000020891

Wilson, G. F. (2002). An analysis of mental workload in pilots during flight using multiple psychophysiological measures. *Int. J. Aviat. Psychol.* 12, 3–18. doi: 10.1207/s15327108ijap1201_2

Wilson, G. F. (2003). "Pilot workload, operator functional state and adaptive aiding," in *Operator Functional State*, eds G. R. J. Hockey, W. K. Gaillard and O. Burov (Amsterdam: IOS Press), 194–203.

Wright, R. A. (2008). Refining the prediction of effort: Brehm's distinction between potential motivation and motivation intensity. *Soc. Personal. Psychol. Compass* 2, 682–701. doi: 10.1111/j.1751-9004.2008.00093.x

Wright, R. A., and Dill, J. D. (1993). Blood pressure responses and incentive appraisals as a function of perceived ability and objective task demand. *Psychophysiology* 30, 152–160. doi: 10.1111/j.1469-8986.1993.tb01728.x

Wright, R. A., and Kirby, L. D. (2001). "Effort determination of cardiovascular response: an integrative analysis with applications in social psychology," in *Advances in Experimental Social Psychology* (Vol. 33), ed. M. Zanna (San Diego, CA: Academic Press), 255–3071.

Yoshino, A., Inoue, M., and Suzuki, A. (2000). A topographic electrophysiologic study of mental rotation. *Brain Res. Cogn. Brain Res.* 9, 121–124. doi: 10.1016/s0926-6410(99)00046-4

Zander, T. O., and Kothe, C. (2011). Towards passive brain-computer interfaces: applying brain-computer interface technology to human-machine systems in general. *J. Neural Eng.* 8:025005. doi: 10.1088/1741-2560/8/2/025005

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Ewing, Fairclough and Gilleade. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution and reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 02 May 2016
doi: 10.3389/fnhum.2016.00185

Neural Mechanisms of Inhibitory Response in a Battlefield Scenario: A Simultaneous fMRI-EEG Study

Li-Wei Ko[1,2,3], Yi-Cheng Shih[1,2], Rupesh Kumar Chikara[2,3], Ya-Ting Chuang[1,2] and Erik C. Chang[4]**

[1] Institute of Bioinformatics and Systems Biology, National Chiao Tung University, Hsinchu, Taiwan, [2] Brain Research Center, National Chiao-Tung University, Hsinchu, Taiwan, [3] Department of Biological Science and Technology, National Chiao Tung University, Hsinchu, Taiwan, [4] Institute of Cognitive Neuroscience, National Central University, Taoyuan, Taiwan

The stop-signal paradigm has been widely adopted as a way to parametrically quantify the response inhibition process. To evaluate inhibitory function in realistic environmental settings, the current study compared stop-signal responses in two different scenarios: one uses simple visual symbols as go and stop signals, and the other translates the typical design into a battlefield scenario (BFS) where a sniper-scope view was the background, a terrorist image was the go signal, a hostage image was the stop signal, and the task instructions were to shoot at terrorists only when hostages were not present but to refrain from shooting if hostages appeared. The BFS created a threatening environment and allowed the evaluation of how participants' inhibitory control manifest in this realistic stop-signal task. In order to investigate the participants' brain activities with both high spatial and temporal resolution, simultaneous functional magnetic resonance imaging (fMRI) and electroencephalography (EEG) recordings were acquired. The results demonstrated that both scenarios induced increased activity in the right inferior frontal gyrus (rIFG) and presupplementary motor area (preSMA), which have been linked to response inhibition. Notably, in right temporoparietal junction (rTPJ) we found both higher blood-oxygen-level dependent (BOLD) activation and synchronization of theta-alpha activities (4–12 Hz) in the BFS than in the traditional scenario after the stop signal. The higher activation of rTPJ in the BFS may be related to morality judgments or attentional reorienting. These results provided new insights into the complex brain networks involved in inhibitory control within naturalistic environments.

Keywords: electroencephalography (EEG), function magnetic resonance imaging (fMRI), inhibitory control, theta-alpha band, right temporoparietal junction (rTPJ)

OPEN ACCESS

Edited by:
Klaus Gramann, Berlin Institute of Technology, Germany

Reviewed by:
Marcus Heldmann, University of Lübeck, Germany
Scott Edward Kerick, US Army Research Laboratory, USA

****Correspondence:***
Li-Wei Ko
lwko@mail.nctu.edu.tw;
Erik C. Chang
auda@ncu.edu.tw

Received: *16 December 2015*
Accepted: *11 April 2016*
Published: *02 May 2016*

Citation:
Ko L-W, Shih Y-C, Chikara RK, Chuang Y-T and Chang EC (2016) Neural Mechanisms of Inhibitory Response in a Battlefield Scenario: A Simultaneous fMRI-EEG Study. Front. Hum. Neurosci. 10:185. doi: 10.3389/fnhum.2016.00185

INTRODUCTION

Inhibitory control is a crucial aspect of cognitive control processes. It allows one to stop ongoing action when it is deemed inappropriate (Aron, 2007). Bari and Robbins (2013) suggested to divide inhibitory control into two categories: cognitive inhibition and behavioral inhibition. Cognitive inhibition can be defined as "the stopping or overriding of a mental process, in whole or in part, with or without intention" (MaCleod, 2007), and is usually measured by the interference task (Kipp, 2005; Leroux et al., 2006). In contrast, behavioral inhibition,

which is the focus of the current study, refers to the suppression of actualizing behavioral outcome, and can be measured by the stop signal task (SST) or go/no-go task (GNGT). Both SST and GNGT use frequent go trials which require participants to perform an action (e.g., press a key button) and infrequent stop (no-go) trials which requires participants to inhibit preparative action (e.g., not to press a key button) upon receiving an additional SST or a different target stimulus (GNGT).

Previous studies have adopted either GNGT or SST to explore the neuroanatomical loci and temporal characteristics of associated brain activities with functional magnetic resonance imaging (fMRI) and electroencephalography (EEG), respectively. In the neuroanatomical domain, many studies found that the prefrontal gyrus (PFG) is important for executive control (for a comprehensive review, see Miller and Cohen, 2001). Consistent activation for response conflict, novelty, working memory (number of elements and delay) and perceptual difficulty has been observed in the inferior frontal gyrus (IFG), dorsal anterior cingulate gyrus (ACG), dorsolateral prefrontal gyrus (DLPFG), but not other frontal regions, regardless of the specific contrast task (Duncan and Owen, 2000). Aron et al. (2004) concluded that the right IFG (rIFG) was more closely related to inhibitory control because damage of the rIFG crucially affected performance in executive cognitive control paradigm, apparently by disrupting inhibition.

However, a number of studies have also proposed that the rIFG is recruited across different task conditions that require sustained attention (Shallice et al., 2008a,b; Simmonds et al., 2008). Hampshire et al. (2010) also suggests that the rIFG serves a general role in attentional control, which rapidly adapts in order to respond to relevant and salient stimuli related to inhibitory control in GNGT and SST. Hence, the suppression of an already initiated response likely depends on rIFG, yet exactly how the inhibitory function is manifested in the motor system remained to be investigated. On the other hand, Aron and Poldrack (2006) had shown that the subthalamic nucleus (STN), which is a part of the basal ganglia, may play a role to suppress the "direct" fronto-striatal pathway that is activated by response initiation and also involved the pre-supplemetary motor area (preSMA). The findings by Mostofsky et al. (2003) suggest that the preSMA appears necessary for inhibiting unwanted movements (stop or no-go condition). Based on previous studies (Aron and Poldrack, 2006; Nachev et al., 2008; Verbruggen and Logan, 2008), Duann et al. (2009) had applied Granger causality analysis in an fMRI study on stop-signal task to explore the functional connectivity of IFG and preSMA. Their study found that preSMA and primary motor gyrus (PMG) have functional interconnectivity via the basal ganalia circuitry to mediate response inhibition, whereas IFG connects with preSMA to modulate the basal ganglia circuitry. According to Duann et al. (2009), the PMG is mediated by IFG and preSMA via basal ganalia circuitry and the functional connectivity between IFG and preSMA is "bi-directional" in SST. Recently, IFG has been hypothesized to serve various functions including resolution of stimulus conflict, attentional orienting or the monitoring of behavior. Consequently, results from some studies have suggested that preSMA is more directly related to response inhibition than IFG, given its involvement in motor control (Bari and Robbins, 2013; Obeso et al., 2013; Aron et al., 2014).

Although the imaging studies are informative about the neuroanatomical loci of response inhibition in the brain, equallythe brain, equally important is how the inhibitory process evolved across time upon its inception. Huster et al. (2013) reviewed EEG studies on the response inhibition under GNGT and SST. Most empirical reports mainly examined event-related potentials (ERP), and it is commonly observed that both stop and no-go conditions evoked two different ERP components which are usually absent in the go condition: a fronto-central negativity occurring around 200–300 ms after stimulus onset (stop or no-go stimulus), followed by a positive potential with a delay of approximately 150 ms exhibiting a fronto-central to centro-parietal topography. These two components have often been conjointly referred to as the N2/P3 complex. Nevertheless, N200s and P300s were also evoked in a broad range of paradigms, including but not limited to response inhibition (e.g., SST, GNGT, Stroop task, Flanker task and Simon task; Kopp et al., 1996; Liotti et al., 2000; Falkenstein et al., 2002; Nieuwenhuis et al., 2004; Ramautar et al., 2006; Johnstone et al., 2007; Bruchmann et al., 2010).

To more specifically determine the temporal marker(s) for response inhibition, an alternative way of analyzing EEG data is through time-frequency analysis for uncovering the oscillatory components involved in inhibitory response (e.g., Herrmann et al., 2005). Basar et al. (1999) demonstrated that EEG can be investigated in the frequency domain and oscillations of specific frequencies are related to specific cognitive functions, such as alpha band (8–12 Hz) fluctuations during both sustained and directed attention (Mathewson et al., 2014). While ERP analysis generally compares latencies or magnitudes of components elicited by different conditions (e.g., go condition vs. stop or no-go conditions), in time-frequency analysis the oscillations of frequency bands associated with different conditions are usually compared. Recently, a number of studies have applied time-frequency analysis in response inhibition tasks. The most common findings from these time-frequency analyses are a burst in frontal-midline theta power for no-go and stop signal conditions as compared to the go condition between 200 and 600 ms after the no-go or stop signal presentation, which falling well into the time range of N2/P3 complex (Huster et al., 2013). In addition, Schmiedt-Fehr and Basar-Eroglu (2011) also reported activity in the delta power for the same time window using a GNGT. These time-frequency components seem to more specifically associated with response inhibition.

Most studies explored the "inhibitory network" by using stimuli with simple configuration in SST or GNGT (e.g., circle as the go stimulus, and an "X" as the stop (no-go) signal; Chang et al., 2014; Lavallee et al., 2014) to investigate the properties of the inhibitory network. How this inhibitory network for typical SST generalizes to response inhibition in more realistic scenarios remains to be investigated. The generalizability issue is not new in cognitive experiments, and not many studies have explored how well cognitive phenomena established in simple

scenes can be generalized to more complex and realistic ones. Lapenta et al. (2014) used transcranial direct current stimulation (tDCS) to explore inhibitory control of EEG under food craving using realistic food picture as go signal in GNGT. They first induced the participant's food craving by a brief movie showing scenes of food and then required the participant to complete a visual analog scale for appearance, smell and taste of the exposed food. All participants were then required to perform GNGT twice: one was performed with active tDCS at F4 and F3 (10–20 EEG coordinate system) and the other with sham tDCS. Their results indicated that tDCS reduced magnitude of frontal N2 component but enhanced the P3a component, as compared with the sham condition. Regenbogen et al. (2010) used real and virtual computer game scenarios to compare the pattern of brain activation between gamers and non-gamers. They analyzed fMRI data by contrasting different combination of conditions, including Violent vs. Nonviolent scenarios under real and virtual modality for gamers and non-gamers, respectively. The activity pattern of non-gamers under the contrast Violent vs. Nonviolent is more complex than gamers in both real and virtual scenes. More importantly, when the neural activities of real modality were compared with virtual modality between gamers and non-gamers, they found non-gamers have more activated brain regions when contrasting Violent vs. Nonviolent conditions, and when contrasting Real vs. Virtual scenes. Based on the findings above, it seems that real and virtual scenes may recruit the brain in distinct ways.

Given the paucity in literature exploring response inhibition by combining methods with high spatial and temporal resolutions, by applying time-frequency analyses, and by contrasting performance under simple vs. naturalistic scenarios, the current study aims to compare behavioral performance and neural mechanisms of inhibitory response under simple and realistic scenarios with simultaneous recording of EEG and fMRI. Scenes from a well-known shooting game "Count Strike" were adopted as the visual background in the battlefield scenario (BFS), where image of a "terrorist" holding a gun served as the go sign, and a "hostage" image as the stop signal. Besides higher extent of visual complexity, this scenario is supposed to induce stressful feelings in the participants. As a control condition, the conventional SST in which simple symbols represent go and stop signals, namely a symbol scenario (SBS), was also adopted. In order to investigate both the rapid brain dynamics and precise spatial loci of the inhibitory process, simultaneous fMRI and EEG recordings were carried out to acquire signals of brain activation from sources with high spatial and temporal resolutions, respectively. Comparing to independent recording of fMRI and EEG, simultaneous fMRI-EEG can confirm that the characterization of functional activations and frequency oscillations of brain networks are under the same experimental condition, and thus more likely the same neural networks (Mulert, 2013). The current study examined significant differences in fMRI and EEG responses associated with successful-stop (SS) vs. successful-go (SG) trials to identify inhibition-related brain activations/dynamics, and SS vs. fail-stop (FS) trials to identify error-related brain activations/dynamics (Li et al., 2006; Boehler et al., 2010; Swick et al., 2011). Based on the literature of inhibitory control reviewed above, we predict that preSMA will show fMRI activation and modulations in theta-alpha band power under both scenarios of SST. However, for the comparison between SBS and BFS, it remains an empirical question whether additional neural networks related to cognitive processing of emotional or social information, such as amygdala or middle temporal gyrus would be involved.

MATERIALS AND METHODS

Participants

All participants (n = 35; mean age = 23.39; SD = 1.86) were right-handed, had normal or corrected-to-normal vision, and none reported history of neurological or psychiatric disorders. Each participant provided written informed consent approved by the Research Ethics Committee of the National Taiwan University prior to participation. Data from three participants were excluded from analyses due to low performance in SST (SG ratio is lower than 2SD below the group mean). Among the remaining participants, simultaneous fMRI-EEG data were successfully acquired from 11 participants, and 21 participants only have fMRI data. Therefore, the fMRI results were based on 32 datasets, whereas the EEG results were based on 11 datasets. Although there were only 11 participants for the EEG analysis, given that each participants made responses to 105 trials, the total amount of epochs is 1155. These epochs are distributed into the four conditions (SG = 705, SS = 170, FG = 92, FS = 188). We consider this amount of epochs are sufficient for our EEG analyses.

Experimental Design

The experiment implemented the stop-signal task under two different scenarios (**Figure 1**), where one consisted of simple symbol (i.e., SBS) and the other battlefield images (i.e., BFS). Every participant was asked to respond to a go stimulus (a circle for SBS and a ruffian for BFS). They hold their response (stop stimulus), when appeared (a cross for SBS and a hostage for BFS), when it was presented after the go stimulus. The critical stop signal delay (cSSD), which is approximately 50% probability of SS, was measured by using a staircase tracking procedure before they performed formal experimental trials in the scanner. The staircase tracking procedure worked in the following way: SSD started at 150 ms and if the participant successful-stopped their response, SSD would increase by 50 ms; on the contrary, SSD would reduce 50 ms and the lower bound of SSD was 150 ms. The formal task used five different SSDs (cSSD, cSSD $\pm$ 40 ms, and cSSD $\pm$ 80 ms) and each SSD had equal number of trials. The participant performed four runs in the fMRI experiment and each run was equally divided into one half for BFS and the other for SBS, where the order of scenario was completely counterbalanced across runs. Each block of scenario in a run had 105 trials of which 25% were stop trials while the rest were go trials. Each go trial began with a fixation cross lasting for a random duration (0.5–6.5 s), followed by a go-signal lasting for 1 s or until response. In a stop trial, the stop-signal is presented

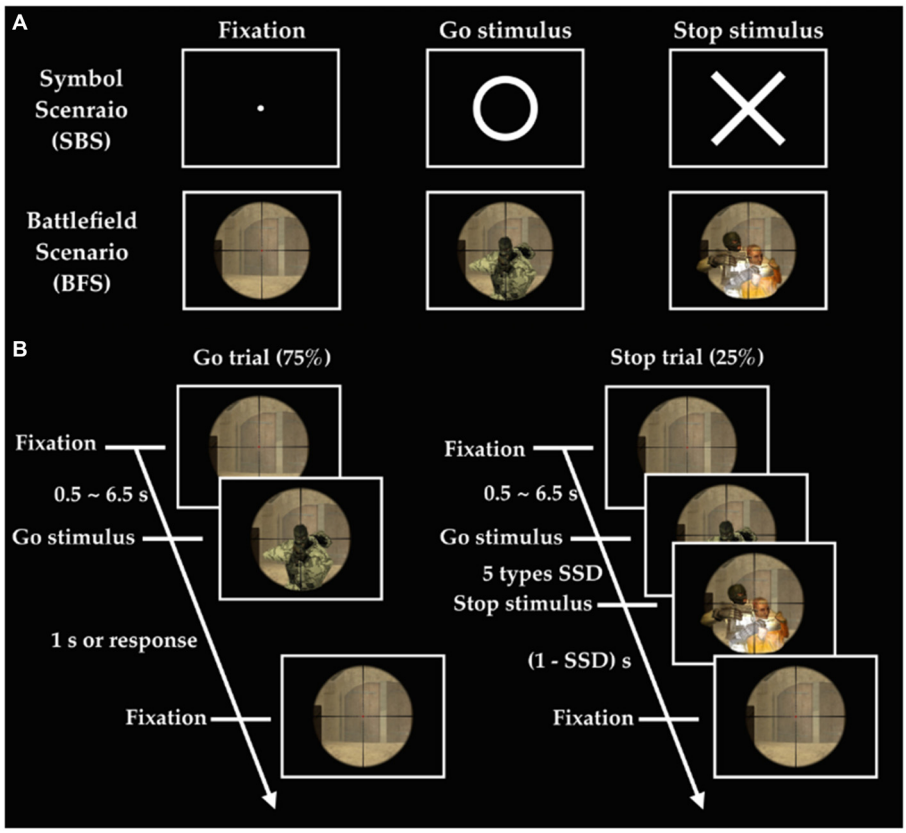


FIGURE 1 | Experimental design. (A) Stimuli in two scenarios; **(B)** order of events in the battlefield scenario (BFS). SSD indicates stop signal delay, and there are five different SSDs including SSD (cSSD), cSSD ± 40 ms, and cSSD ± 80 ms.

N milliseconds after the go-signal, where *N* was defined by the SSD assigned to that trial.

fMRI Signal Acquisition and Preprocessing

Participants performed the task in a Siemens 3T MAGNETOM Skyra scanner located in the Taiwan Mind and Brain Imaging Center at National Chengchi University, Taipei. Structural T1-weighted images were acquired using the MPRAGE sequence (TR: 2530 ms; TE: 3.03 ms; flip angle: 7°; matrix size: 224 × 256; field of view: 224 × 256 mm; in-plane resolution: 1 × 1 mm; slice thickness: 1 mm; 192 slices). Functional brain images were acquired using a gradient echo-planar imagine sequence (TR: 2000 ms; TE: 25 ms; flip angle: 90°; matrix size: 64 × 64; field of view: 220 × 220 mm; voxel size: 3.438 × 3.438 × 4.0 mm^3; 292 volumes per run). The preprocessing stream as well as statistical analyses was completed using the Analysis of Functional Neuroimages (AFNI) software (Cox, 1996). The preprocessing stream included image reconstruction, slice-time correction (time-shifting the time series using Fourier interpolation), and motion-correction (linear least-squared alignment via affine transformation with three translational and three rotational parameters). Activation outside the brain was removed using edge detection techniques. After the preprocessing, each participant's anatomical image was transformed into the standard space of the Montreal Neurological Institute (MNI) 152 brain template using an automated feature-matching algorithm (Collins et al., 1994). Each participant's functional data was first aligned to their own anatomical image and then transformed into the standardized MNI space.

EEG Signal Acquisition and Preprocessing

An MR-compatible 34-channel amplifier (BrainAmp MR; Brain Products) and a MR-compatible EEG cap (BrainCap-MRI 32-Channel-Standard) with a head volume coil were applied in this study. EEG was recorded in the MR scanner room simultaneously with fMRI acquisition. The EEG cap had 31 electrodes for brainwave recording and one for electrocardiography (ECG) recording. Electrode-skin impedance was kept smaller than 10 kOhms by using abrasive electrolyte-gel (ABRALYT HiCl). Data were transferred through fiber-optic cables to an IBM-compatible laptop and recorded by the BrainVision Program (BrainVision Recorder, Brain Products) synchronized with the BOLD signals via triggers from the MR scanner. The EEG signals were recorded with a passband of 1–250 Hz, digitized at 5000 Hz with 32-bit of resolution (equivalent to 0.5 μV; dynamic range: 16.38 mV). The EEG data were band-pass

(1–50 Hz) filtered, re-referenced to the average of channel TP9 and TP10. The MR gradient artifacts in the EEG data were corrected. The MR-denoised EEG data were then down-sampled to 500 Hz, and the cardioballistic signals from the ECG recording were used to adjust EEG signals via peak-detection algorithms in the BrainVision Analyzer software. Severe artifacts of EEG signal induced by muscle activities, environmental noise, eye movements, and blinking were manually removed to minimize their impacts on the subsequent analysis.

Behavioral Data Analysis

We calculated SG and SS ratio of both scenarios to verify if each participant's performance met the criterion. Behavioral characteristics of performance in the stop-signal task, including the go reaction time (Go-RT) and cSSD were analyzed with student's *t* test (BFS vs. SBS). Furthermore, the stop-signal reaction time (SSRT) based on the horse-race model of stopping (Logan et al., 1984) was computed to represent one's inhibitory ability. Since the stopping mechanism itself cannot be directly measured, the SSRT was calculated by subtracting SSD from the Go-RT. The inhibition function was computed as the number of SS trials divided by the number of all stop trials, and subjected to a two-way within-subject ANOVA to assess the effect of Scenarios (BFS vs. SBS), SSD (cSSD, cSSD ± 40 ms, and cSSD ± 80 ms), and their interaction.

fMRI Data Analysis

The fMRI analysis was also completed in AFNI. Stimulus types and participant's response conjointly determined four conditions for each scenario, including SG, SS, FS and fail go. The first-level statistical analysis for each participant was carried out in a general linear model (GLM) by convolving the onset of go stimulus in the SG, SS, and FS conditions, respectively, with a canonical hemodynamic response function (the BLOCK function in 3dDeconvolve of AFNI). Here the effects of interest are inhibitory control and error detection. The active brain areas for inhibitory control was defined by the contrast between SG and SS; on the other hand, the active brain regions for error detection was defined by the contrast between SS and FS. The scenario effect of inhibitory control and error detection were examined by comparing the "difference of difference", namely $(SS - SG)_{BFS} - (SS - SG)_{SBS}$ and $(FS - SS)_{BFS} - (FS - SS)_{SBS}$, respectively. In the second level analysis, the between-scenario differences were analyzed with a linear mixed-effect model (3dMEMA), and the whole-brain type I error was controlled at a cluster threshold (alpha) of 0.05 via Monte Carlo simulation (3dClustSim).

To more sensitively detect activations associated with inhibitory control and error detection, we also carried out region of interest (ROI) analysis by both adopting ROIs related to stop-signal task in the literature (literature-based ROIs) and by selecting ROIs surviving the whole-brain analysis from the inhibitory control and error detection contrasts, respectively, regardless of scenarios (empirical-based ROIs). For the empirical-based ROIs, the leave-one-subject-out (LOSO) method (Esterman et al., 2010) was applied to extract the GLM coefficients, and the differences between scenarios were statistically assessed. It turned out the literature-based ROIs did not yield any significant difference between scenarios and will not be further described. On the other hand, six ROIs empirically identified from the whole brain analysis of inhibitory control and error detection, respectively, regardless of scenarios were analyzed to verify the between scenario difference. Empirical-based ROIs for inhibitory control included rIFG, left insula, preSMA, left inferior parietal gyrus (IPG), right middle occipital gyrus (rMOG) and left MOG. ROIs for error detection included right middle frontal gyrus (rMFG), left IFG, right IPG, right superior temporal gyrus (STG), right inferior occipital gyrus (IOG) and left MOG. The empirical MNI coordinates of inhibitory control and error detection were listed in the Supplementary Materials Table 1, while the literature-based ROIs were listed in the Supplementary Materials Table 2.

EEG Data Analysis

The EEG analysis was completed in EEGLab. Independent Component Analysis (ICA; Makeig et al., 1996; Delorme and Makeig, 2004) was used to separate out temporally independent time course of the activation of which dipole source location (Oostenveld and Oostendorp, 2002) was localized in the brain of each participant for group analysis (cross-subject analysis). We removed artifact components manually and then performed component clustering based on *k*-means ($k = 5$) criteria and dipole-fitting coordinates to identify the most representative clusters. The value of k was determined both by considering potential number of sources associated with the stop-signal task, and the number of ROIs identified in the fMRI results. One of the five resultant clusters was excluded because less than 70% of participants have it. Therefore, four clusters (preSMA, rMFG, and bilateral MOGs) and their dipole locations were identified (see **Figure 2**) to investigate brain dynamics following the go events and the subsequent stop events. Note that the preSMA and bilateral MOG clusters were in anatomical proximity of the inhibitory control ROIs of fMRI results, and the rMFG and the left MOG cluster was close to the error detection ROIs of fMRI results.

Each epoch was separately transformed into the time-frequency domain using the event-related spectral perturbation (ERSP) routine (Delorme and Makeig, 2004). Three conditions, namely SG, SS and FS, were identified as the effect of interest. The baseline was defined as the signals between −0.5 and 0 s before Go-stimulus for comparing response magnitudes of corresponding epochs. A two-way Scenario × Condition ANOVA was conducted on the baseline data to verify whether they are equivalent across scenarios and conditions. We have explored not only the power spectrum of each condition, but also the power spectrum of inhibitory control and error detection, respectively, in each scenario which was also done in the fMRI analyses.

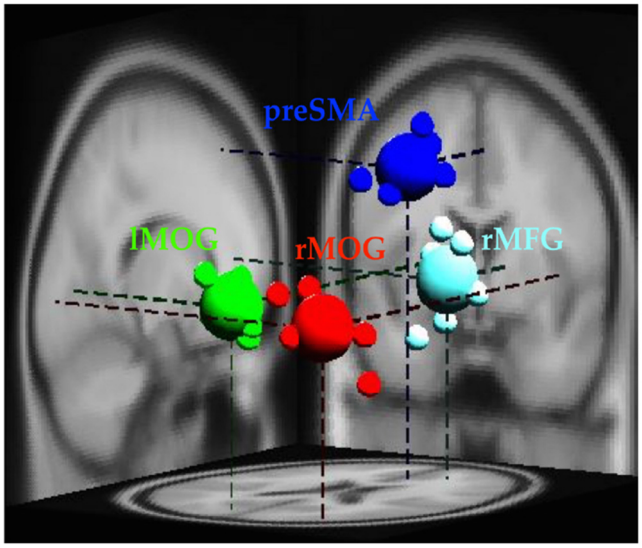


FIGURE 2 | Clusters of dipole locations for the analysis of EEG dynamics. PreSMA and rMFG are for of inhibitory control and error detection, whereas lMOG and rMOG are used for processing visual stimul. Small spheres indicate individual participant's dipole location, while large spheres indicate diploe locations of each cluster. lMOG, Left middle occipital gyrus; rMOG, Right middle occipital gyrus; preSMA, Pre-supplementary motor area; rMFG, Right middle frontal gyrus.

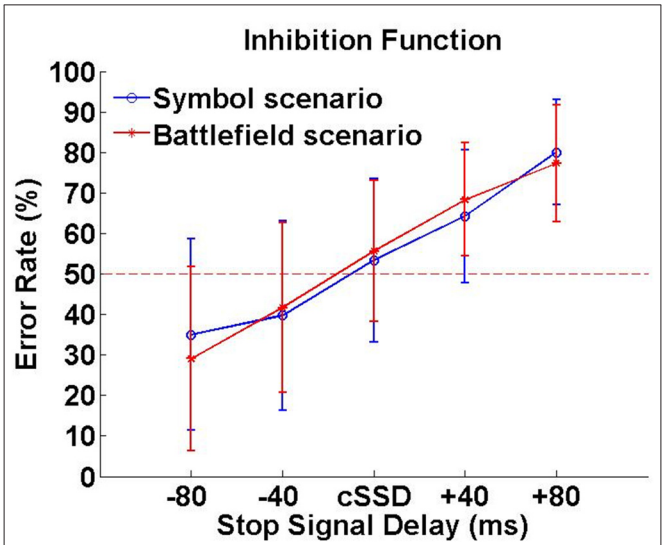


FIGURE 3 | Inhibition function. Error rates (%) were calculated by dividing number unsuccessful stop trials with all stop trials under each SSD.

RESULTS

Behavioral Results

In SBS, the Go-RT, cSSD, SSRT, SG ratio and SS ratio of SBS were 425 ± 62 ms, 188 ± 50 ms, 240 ± 60 ms, 94.0 ± 7.6% and 45.6 ± 16.5%, respectively. In BFS, the Go-RT, cSSD, SSRT, SG ratio and SS ratio of SBS (BFS) were 422 ± 58 ms, 195 ± 68 ms, 230 ± 53 ms, 93.0 ± 10.3% and 45.6 ± 13.6%, respectively. When compared between scenarios, none of these behavioral outcomes reached significance (all *p*s > 0.05; **Figure 3**). In addition, the averaged inhibition function approached 50% at cSSD and error rate level increased with the length of SSD (**Figure 3**).

Imaging Results

Inhibitory Control

Whole Brain Analysis

Tables 1A,B summarized brain regions that were more activated in the SS than in the SG condition, namely the inhibitory control component, under SBS and BFS, respectively. **Figure 4** also shows these activations under the two scenarios conjointly so that overlapping brain regions are explicit. In SBS, the MOG and a few different frontal areas were activated in this contrast (see **Table 1A** and **Figure 4** right panel). On the other hand, the brain areas activated by the BFS (see **Table 1B** and **Figure 4** left panel) was similar to those in SBS (see the purple regions colored in purple in the **Figure 4** middle panel). Moreover, when directly contrasting the two scenarios under inhibitory control, the only significant loci (BFS > SBS) fell within the right temporal-parietal junction (rTPJ; MNI: $x = 48$, $y = -74$, $z = 11$; cluster size = 39).

ROI Analysis

Pairwise *t* tests between the BFS and SBS in the six empirically defined ROIs for inhibitory control revealed significantly higher activation in BFS than in SBS at the left IPG ($t_{(32)} = 2.4$, $p = 0.02$) and rMOG ($t_{(32)} = 2.5$, $p = 0.02$).

Error Detection

Whole Brain Analysis

Tables 1C,D summarized brain regions more activated in SS than fail stop under SBS and BFS, respectively. **Figure 5** also shows these activations in both volumetric (left and right panels) and surface (middle panel) views as described in section "Inhibitory Control" for the inhibitory control. In SBS, the MOG, the bilateral IFG, the rMFG and right postcentral gyrus were activated in this contrast (see **Table 1C** and **Figure 5** right panel). On the other hand, the rMFG, left IFG, the right IPG, the fusiform gyrus, the right precuneus and the rMOG were activated by the BFS (**Table 1D** and **Figure 5** left panel). There was very few overlapping brain regions (purple regions in the middle panel of **Figure 4**). When directly contrasting the two scenarios under error detection, no region showed significant difference.

ROI Analysis

Paired *t* tests between the BFS and SBS in the six ROIs mentioned above revealed only significantly higher activation in BFS than in SBS at right IOG ($t_{(32)} = 2.7$, $p = 0.01$).

EEG Results

Figure 2 shows the four clusters (rMFG, preSMA, and bilateral MOGs) and their dipoles that fulfilled the cluster selection

TABLE 1 | Brain regions more activated in (A) SS compared with SG under SBS, (B) SS compared with SG under BFS, (C) SS compared with FS under SBS, (D) SS compared with FS under BFS.

Side	Region	BA	MNI coordinate (mm)			Cluster Size (Voxels)
			X	*Y*	*Z*	
(A) SS > SG under SBS						
R	Middle Frontal Gyrus	10	38	56	19	40
R	Superior Frontal Gyrus	8	2	26	59	159
R	Inferior Frontal Gyrus	47	48	16	1	844
L	Inferior Frontal Gyrus	13	−34	16	13	193
R	Superior Temporal Gyrus	21	50	−28	1	56
L	Inferior Parietal Gyrus	40	−60	−40	41	528
R	Middle Occipital Gyrus	19	44	−86	5	1752
L	Middle Occipital Gyrus	18	−34	−98	7	944
(B) SS > SG under BFS						
R	Superior Frontal Gyrus	6	12	10	71	32
R	Middle Frontal Gyrus	6	50	−2	43	555
L	Inferior Frontal Gyrus	9	−40	−2	37	35
L	Inferior Frontal Gyrus	47	−30	20	1	95
R	Inferior Parietal Gyrus	40	60	−38	55	43
R	Superior Parietal Gyrus	7	32	−64	55	218
L	Superior Parietal Gyrus	7	−28	−68	55	31
R	Middle Occipital Gyrus	18	50	−74	11	1257
L	Inferior Occipital Gyrus	19	−42	−80	−5	694
(C) SS > FS under SBS						
R	Inferior Frontal Gyrus	47	32	22	5	20
R	Inferior Frontal Gyrus	46	56	26	19	52
R	Middle Frontal Gyrus	8	48	2	49	21
L	Inferior Frontal Gyrus	45	−52	32	7	35
R	Postcentral Gyrus	4	66	−28	49	32
L	Middle Occipital Gyrus	18	−28	−82	−5	37
(D) SS > FS under BFS						
R	Middle Frontal Gyrus	6	42	−2	49	21
L	Inferior Frontal Gyrus	45	−58	14	29	20
R	Inferior Parietal Gyrus	40	44	−40	59	89
L	Fusiform Gyrus		−34	−74	−7	216
R	Precuneus	7	26	−58	59	49
R	Middle Occipital Gyrus	19	32	−94	13	501

Voxelwise threshold, p = 0.0001; cluster alpha < 0.01; BA, Brodmann Area; R, Right; L, Left.

criteria (see "EEG Data Analysis" Section). Because the rMFG is considered as a crucial area for sustaining attention rather than stopping action and preSMA is considered as directly related to response inhibition, preSMA and rMFG were subject to the analysis at the time period when sustained attention and response inhibition were supposed to be ongoing. On the other hand, because bilateral MOGs were considered only relevant to visual perception that are relatively minor to the stop-signal task, their power ERSPs were analyzed at the time period of visual processing and described in the supplementary materials (Supplementary Figures 5, 6). For the rMFG and preSMA clusters described in the main text, the focus is on the contrasts for inhibitory control (i.e., SS vs. SG) and for error detection (i.e., FS vs. SS) in each scenario. The significant modulations within the individual conditions (i.e., SS, SG, and FS) can be found in **Figures 6**–**9**, which are mainly described in the following sections.

The baseline power of EEG oscillations were supposed to be equivalent between the SS and SG conditions as well as between the FS and SS conditions in both scenarios, because participants should be under similar state before the presentation of stimulus in each condition. Consistent with this assumption, one-way ANOVAs comparing SS and SG in BFS and SBS found no significant difference, and so did the one way ANOVAs comparing FS and SS. The analyses of baseline power are described in the Supplementary Figures 2, 4.

Inhibitory Control

Figures 6, **7** show the results of time-frequency analyses in preSMA and rMFG, respectively. In the preSMA component (**Figure 6**), the brain dynamics for inhibitory control can be examined by contrasting the SS and SG conditions. In this contrast, the burst of delta and theta band power was observed in BFS, whereas the suppression of alpha and beta band power was observed in SBS.

In the rMFG component (**Figure 7**), the brain dynamics for inhibitory control (SS vs. SG) showed delta, theta and alpha band power desynchronization after response in BFS, whereas

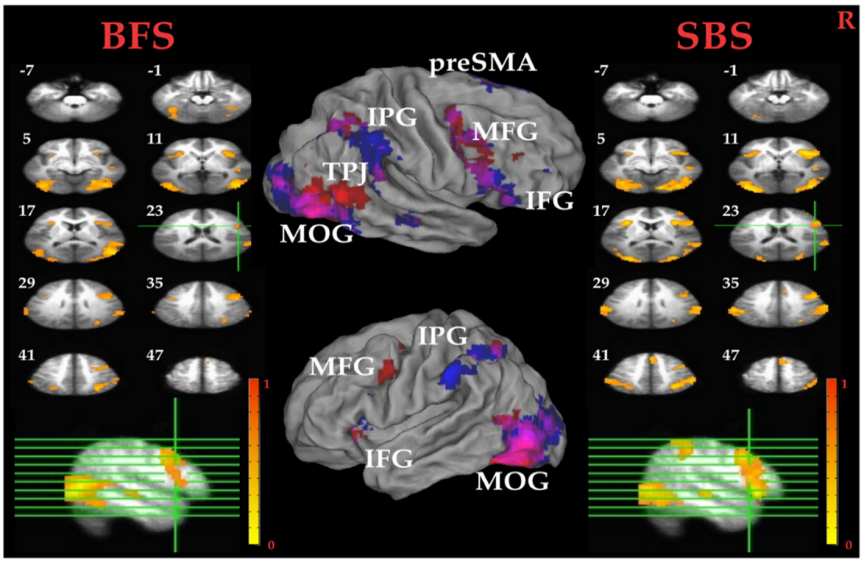


FIGURE 4 | Inhibitory control related brain areas. All results were mapped onto a standard brain surface model in Caret (Van Essen et al., 2001). Left panel: horizontal sections under the BFS; middle panel: visualization of significant activations on the cortical surface for both scenarios (Red: BFS; Blue: symbol scenario [SBS]; Purple: overlap of both scenarios); right panel: horizontal slices under the SBS. The top-left number besides each slice indicate the *z*-axis. Right hemisphere is at the right side of the figure. Voxelwise statistical threshold was set at $p < 0.0001$, and cluster threshold alpha <0.01.

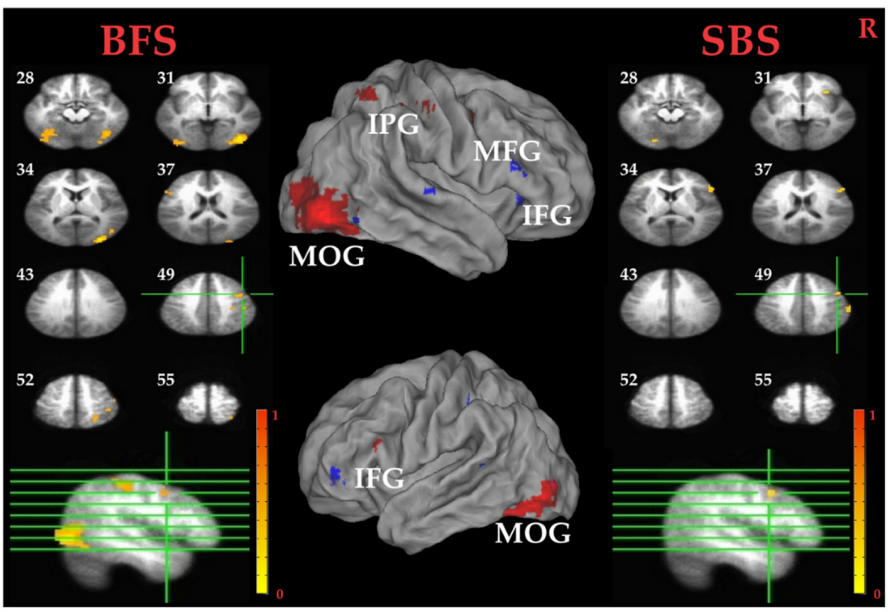


FIGURE 5 | Error detection related brain areas. Left panel: horizontal sections under the BFS; Middle panel: visualization of significant activations on the cortical surface for both scenarios (Red: BFS; Blue: SBS; Purple: overlap of both scenarios); right panel: horizontal slices under the SBS. The top-left number besides each slice indicate the *z*-axis. Right hemisphere is at the right side of the figure. Voxelwise statistical threshold was set at $p < 0.0001$, and cluster threshold alpha <0.01.

beta band power was in synchronization after go stimulus in SBS.

Error Detection

In the preSMA component (**Figure 8**), the brain dynamics for error detection (SS vs. FS conditions) showed the suppression of theta and alpha band power in BFS; on the other hand, all frequency bands power of FS condition displayed much greater magnitude than SS in SBS.

In the rMFG component (**Figure 9**), the brain dynamics for error detection (SS vs. FS) showed that delta and theta band power were in desynchronization after response in the

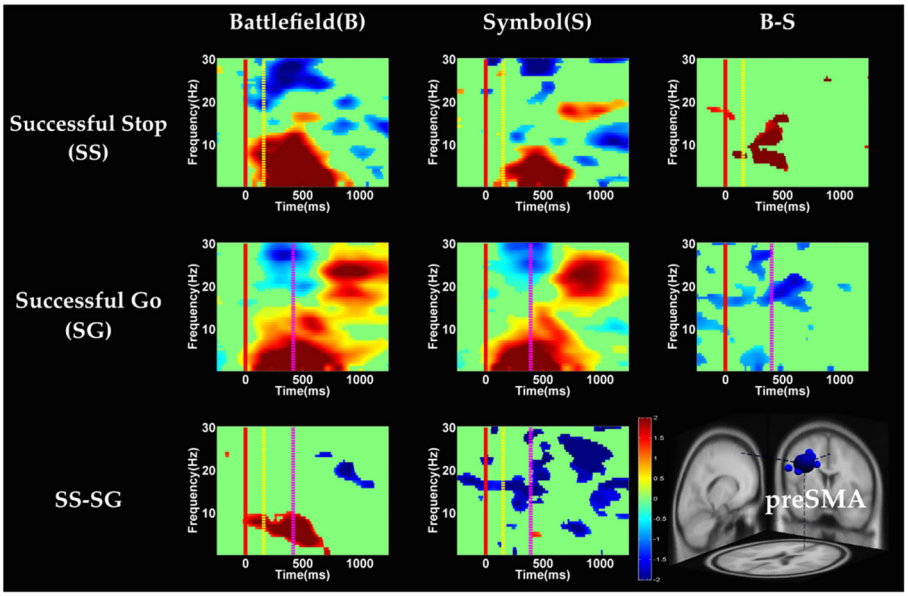


FIGURE 6 | The event-related spectral perturbation (ERSP) images of preSMA cluster under inhibitory control. Red solid line: onset of the go stimulus; yellow dash line: onset of the stop signal; purple dash line: onset of response; color bars indicate the magnitude of the ERSPs; statistical threshold at $p < 0.01$.

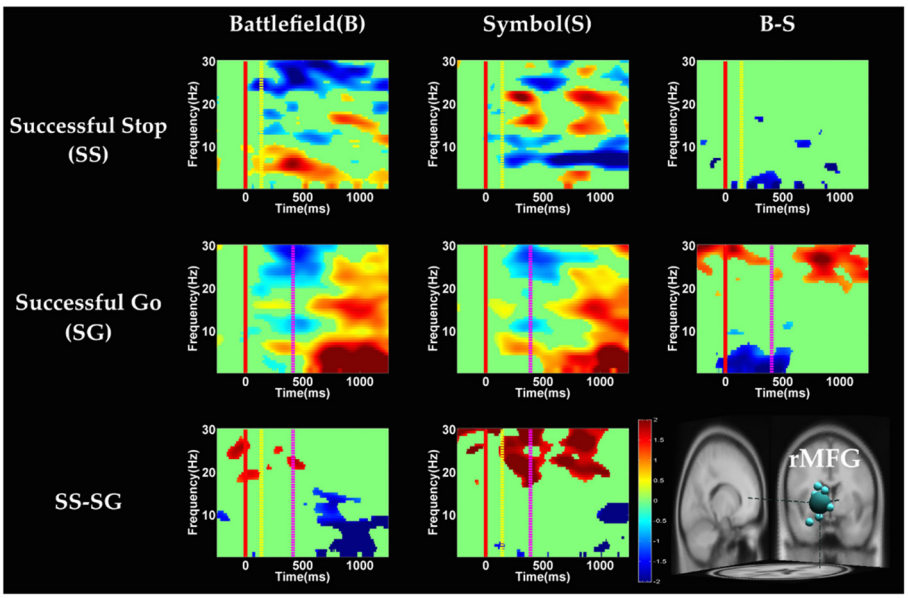


FIGURE 7 | The ERSP images of rMFG cluster under inhibitory control. Red solid line: onset of the go stimulus; yellow dash line: onset of the stop signal; purple dash line: onset of response; color bars indicate the magnitude of the ERSPs; statistical threshold at $p < 0.01$.

BFS, whereas beta band power was in synchronization after go stimulus in SBS.

DISCUSSION

The current study aims to compare inhibitory functions and the associated brain mechanisms underlying realistic and simplified scenarios. Based on the behavioral results, participants successfully performed the stop-signal task under BFS and SBS (the SG ratio was above 90% and SS ratio approached 50% for both scenarios). The SSRT has been suggested to be an indicator of one's inhibitory ability (Band et al., 2003). Since SSRTs of the two scenarios do not differ, performance on response inhibition does not

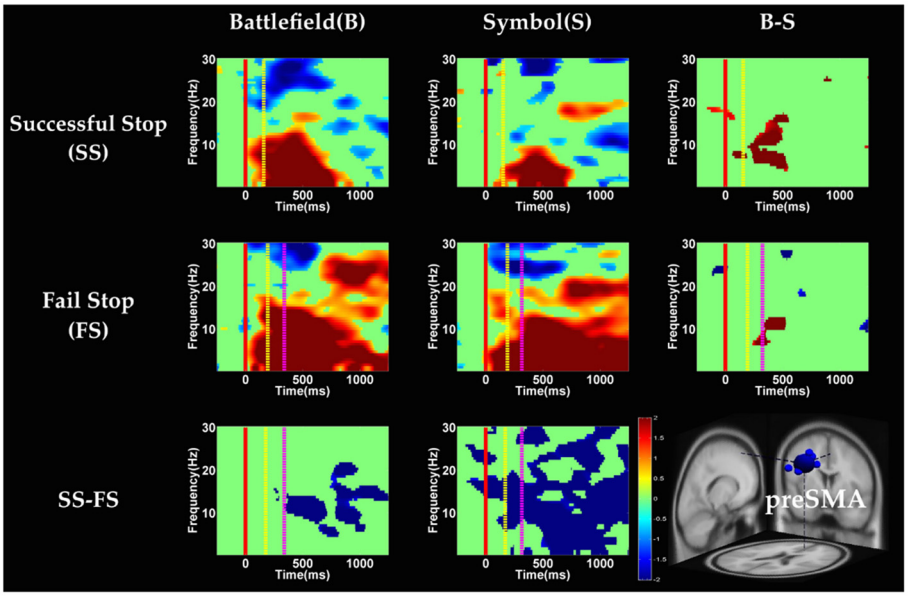


FIGURE 8 | The ERSP images of preSMA cluster under error detection. Red solid line: onset of the go stimulus; yellow dash line: onset of the stop signal; purple dash line: onset of response; color bars indicate the magnitude of the ERSPs; statistical threshold at $p < 0.01$.

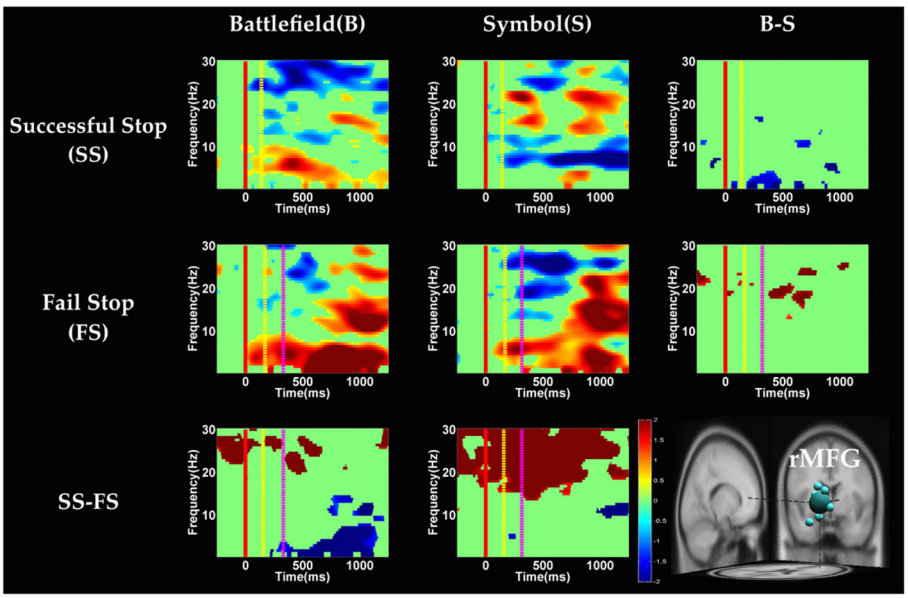


FIGURE 9 | The ERSP images of rMFG cluster under error detection. Red solid line: onset of the go stimulus; yellow dash line: onset of the stop signal; purple dash line: onset of response; color bars indicate the magnitude of the ERSPs; statistical threshold at $p < 0.01$.

seem to be influenced by different scenarios one faces, likely due to highly adaptive nature of human's inhibitory processing. The brain mechanisms for inhibition under the two scenarios can be compared on equivalent bases of behavioral performance.

To summarize, main findings in the fMRI and EEG data are as the following: in the whole-brain analysis of fMRI data, significant difference between the battlefield and SBSs was found only in rTPJ for inhibitory control, and no significant region was found for error detection. In the ROI analysis of fMRI data, significant difference between the two scenarios (BFS > SBS) was found in left IPG and rMOG for inhibitory control, and in right IOG for error detection. As for the EEG results, for inhibitory control in the preSMA, the burst of delta and theta

band power was observed in BFS, whereas the suppression of alpha and beta band power was observed in SBS. In the rMFG, there were delta, theta and alpha band power desynchronization after response in BFS, and beta band power synchronization after go stimulus in SBS. For error detection in the preSMA, there was the suppression of theta and alpha band power in BFS, and broadband synchronization in SBS. In the rMFG, there were delta and theta band power desynchronization after response in the BFS, and there was beta band power synchronization after go stimulus in SBS.

Neural Mechanisms of Inhibitory Control

The fMRI results show that, under the contrast of inhibitory control, the stop-signal task in BFS and SBS activate overlapped brain areas including preSMA, rIFG, bilateral IPG, bilateral MOG. All of these brain areas are either involved in target detection or attention to salient events (Corbetta and Shulman, 2002; Eckert et al., 2009; Menon and Uddin, 2010). Specifically, one of MOG functions is visual form perception and recognition (Grill-Spector and Malach, 2004), the parietal lobe is a crucial locus for spatial attention (Yantis et al., 2002), and the rIFG and preSMA show significant activation for contrast between SS and SG conditions (Boehler et al., 2010; Swick et al., 2011). While participants performed stop-signal task in both scenarios, we expect to observe stronger activations in BFS than SBS because BFS contains more complex visual information and may evoke other cognitive functions involved in the inhibitory network.

With respect to the main goal of the current study, when contrasting the inhibitory control component in both scenarios in the whole-brain analysis, we observe higher activation for the BFS in the rTPJ. The rTPJ has been implicated, together with the rIPG, in detecting behaviorally relevant salient events (Corbetta and Shulman, 2002; Husain and Nachev, 2007). Chang et al. (2013) uses transcranial magnetic stimulation (TMS) to interfere with bilateral TPJ to probe the function in attentional networks, and find that the rTPJ is critically involved in attentional reorienting. In addition, rTPJ is also involved in the "theory-of-mind" (ToM) network which includes the medial PFG, precuneus, right superior temporal sulcus and bilateral TPJ (Saxe and Kanwisher, 2003; Aichhorn et al., 2009). The ToM network increases metabolic activity when one thinks about other people's thoughts. Koster-Hale et al. (2013) use multi-voxel pattern analysis to examine the difference between intentional and accidental harms on other people, and conclude that rTPJ is associated with moral judgments. In the current study, the rTPJ may serve one or a few of the functions mentioned above in BFS because the task involve shooting decision which may aim at innocent hostage.

There is a greater potential negative consequence of failing to stop a shooting response in the presence of an innocent hostage, which may actually decrease response impulsivity but yet still increase the level of activation of inhibitory systems. To verify this speculation with enhanced sensitivity, six brain areas are selected from the whole-brain analysis of inhibitory control that were localized by contrast orthogonal to the scenario effect, including the rIFG, preSMA, left insula, left IPG, rMOG and left MOG. The left IPG and rMOG show a greater activation in BFS. To relate the findings with the roles of these ROIs in previous studies, the left IPG has been implicated in tool manipulation (Ishibashi et al., 2011) or executive function (Kübler et al., 2003), which are both relevant in the current study because participants might have connected the task to firing with a gun (using a tool) to shoot terrorist in BFS. According to Slotnick et al. (2003), the reallocation of visual attention to external stimulus will result in an increase in occipital activation. In ROI results, the stronger activation of rMOG suggests that participants might have focused on terrorist and hostage and ignore the battlefield background. However, as BFS and SBS not only differed in their contextual information but also their visual complexity and emotional implications, the above conjectures need to be considered with caution.

With respect to the temporal dynamics, we first examine common findings in SS and SG conditions of both scenarios. In the preSMA source, there is a burst of each frequency band power except the beta band following the go stimulus, which lasted for 400–600 ms. This phenomenon is consistent with what was found after no-go and stop signal trials in previous studies (Schmiedt-Fehr and Basar-Eroglu, 2011; Huster et al., 2013). Because the preSMA is essential for the conversion from volitional thoughts to actions (Penfield and Welch, 1951; Fried et al., 1991), the beta band power has been generally considered as a marker of explicit responses. The event related desynchronization (ERD) of beta band occurs before and during response and then the event related synchronization (ERS) would follow actual response (Schulz et al., 2014). In the current experiment, the ERD of beta band occurs in SS and SG conditions of two scenarios likely because participants have already prepared to respond when they see the go stimuli; however, the ERS of beta band only occurs in SG condition of both scenarios because participants do not make actual response in the SS condition. Furthermore, the spectral perturbation of SS between BFS and SBS show that the power of theta-alpha band is much greater in BFS. According to Huster et al. (2013), the burst of frontal theta band power is associated with successful inhibition. In the current study, we observed synchronization of theta-alpha band power of SS under BFS and SBS in the preSMA. Furthermore, the theta-alpha band power in preSMA of BFS is higher than SBS, which suggests that the impulsivity in BFS is stronger than in SBS.

One thing worth noticing is that, in the preSMA brain source (**Figure 6**), there is no difference in the baseline power between scenarios, likely because each go stimulus may or may not be followed by a stop signal. This indicates that these two scenarios had the same baseline states when preparing for inhibiting prepotent response in the current trial regardless of stop signal. However, unlike in fMRI analysis, the rTPJ does not show significant differences between scenarios in the EEG analysis.

On the other hand, Swann et al. (2012) demonstrate that the power of 4–15 Hz is suppressed and beta band power would increase in right frontal lobe after the stop signal. Beta band power from the right frontal lobe may serve to compute coherence with preSMA. Therefore, they suggest that right

frontal lobe monitors and detects the stop signal and then transfers the information to preSMA (coherent beta activity). This finding about the role of the right frontal lobe is similar to our results of SS of SBS at the rMFG, but not in the BFS. Perhaps the rMFG is involved in transferring information but not directly in inhibitory control so that different scenarios evoked different spectral perturbation. Finally, the spectral perturbation of two scenarios under bilateral MOG are similar, likely due to their similar roles in processing visual stimuli.

Neural Mechanisms of Error Detection

In the whole-brain analysis, we observe higher activation in the IFG, MFG and MOG for the SS than the FS condition in both scenarios. These brain regions may reflect different cognitive functions in attention during visual processing, decision making, response execution and post-response processing (Iannaccone et al., 2015). Previous fMRI studies have indicated that attention neural network modulates visual cortical activation and facilitation of visual stimulus processing through inhibition of unattended stimulus information (Brefczynski and Deyoe, 1999; Smith et al., 2000; Slotnick et al., 2003). Although higher activation in MOG for error detection can be observed in both scenarios, we expect to observe stronger activations in BFS than SBS because BFS is a more complex situation requiring participants to correct their error and evoke other cognitive functions involved in inhibitory control (see also "Neural Mechanisms of Inhibitory Control" Section).

To verify the above speculation with improved sensitivity, rMFG, left IFG, right STG, right IPG, right IOG and left MOG were identified as ROIs from a contrast (SS—FS) orthogonal to the scenario effect in the whole-brain analysis. Only the rMOG shows greater activation in BFS than SBS. This result supports the idea that participants need to pay more attention to SS in BFS (Slotnick et al., 2003). Although we did not find significantly different activation of MFG between the two scenarios, the current findings still suggest that these middle and inferior frontal regions may differ in the post-response processes in the error detection (i.e., FS vs. SS). The middle and inferior frontal areas have been implicated in error detection and conflict monitoring (Braver et al., 2001; Menon et al., 2001; Rubia et al., 2003, 2005; Rushworth et al., 2004).

The current study also finds that both scenarios have stronger activation of MFG in error detection. The activation of MFG may reflect stronger performance monitoring after FS. Furthermore, we observed only the activation of fusiform gyrus in BFS (**Table 1D**) due to our stimuli design and may reflect face recognition (George et al., 1999).

With respect to the temporal dynamics, when analyzing the effect of scenarios, although ROI analysis in fMRI results reveal significant differences in the right IOG, we do not observe significant difference between scenarios in rMOG (Supplementary Figure 5). Although we do not find the effect of scenarios, we observe the suppression of theta band after response in error detection in both scenarios. Previous EEG studies have indicated that the oscillation of theta-alpha band is associated with the attention network (Fan et al., 2007). The current study finds that the suppression of theta band may be associated with the activation of right occipital gyrus for greater attention to the visual stimuli during SS when compared with FS. On the other hand, we explore the temporal dynamic in preSMA and rMFG brain source and observe the duration of burst of delta, theta and alpha band power in the FS condition were longer then SS condition in preSMA and rMFG. The prolonged duration of the FS condition may reflect error detection. The findings of previous EEG studies suggest that the burst of theta and alpha band power after response in the frontal lobe reflect error processing (Cavanagh and Frank, 2014; Cohen, 2015; Shou and Ding, 2015). Finally, this study reveals that the EEG oscillation of preSMA brain source is related to not only inhibitory control but also error detection.

CONCLUSION

This study uses BFS to translate stop signal paradigm in simulated threatened situation and demonstrates that when human inhibits their action under threatened situation, the rTPJ is involved in the mediation of inhibitory control. The power of theta-alpha band under threatened situation is greater than normal situation that may be associated with the rising activation level of preSMA. Through over half a century of investigations on cognitive functions, significant amount of knowledge of basic cognitive processes has been acquired using stimuli with extremely simple configuration. From the behavioral performance of the current study we demonstrated that findings discovered with simple stimuli remains valid when carefully and comparably transformed into complex and realistic ones. At the meantime, additional brain regions relevant to the new configuration may be involved dynamically for the more complex stimuli, as can be identified from sources of signals differentially specialized in spatial and temporal resolutions. How these simultaneously recorded sources of signals (e.g., EEG and fMRI) are conjointly related to the valence, complexity, and motivational effects induced by scenes embedding the basic cognitive process remain an intriguing and important issue for future studies.

AUTHOR CONTRIBUTIONS

L-WK initiated the main idea of this study, designed the experiment and advised the EEG data analysis methods on the collecting fMRI-EEG data. Y-CS, RKC and Y-TC were performing the fMRI-EEG data collection and data analysis. ECC provided advises on modifying the experiment design and analyzing the fMRI data.

ACKNOWLEDGMENTS

The current study was supported in part by the Taiwan Ministry of Science and Technology (MOST) under grant numbers 102-2420-H-009-003-MY3 and 103-2410-H-009-019-MY2, and in part by UST-UCSD International Center of Excellence in Advanced Bioengineering sponsored by MOST I-RiCE Program

under grant number MOST 103-2911-I-009-101. Research was also in part sponsored by the Army Research Laboratory and was accomplished under Cooperative Agreement Number W911NF-10-2-0022. The authors are grateful for the technical support of Taiwan Mind and Brain Imaging Center (TMBIC).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fnhum.2016.00185/abstract

REFERENCES

Aichhorn, M., Perner, J., Weiss, B., Kronbichler, M., Staffen, W., and Ladurner, G. (2009). Temporo-parietal junction activity in theory-of-mind tasks: falseness, beliefs, or attention. *J. Cogn. Neurosci.* 21, 1179–1192. doi: 10.1162/jocn.2009.21082

Aron, A. R. (2007). The neural bsasis of inhibition in cognitive control. *Neuroscientist* 13, 214–228. doi: 10.1177/1073858407299288

Aron, A. R., and Poldrack, R. A. (2006). Cortical and subcortical contributions to Stop signal response inhibition: role of the subthalamic nucleus. *J. Neurosci.* 26, 2424–2433. doi: 10.1523/JNEUROSCI.4682-05.2006

Aron, A. R., Robbins, T. W., and Poldrack, R. A. (2004). Inhibition and the right inferior frontal cortex. *Trends Cogn. Sci.* 8, 170–177. doi: 10.1016/j.tics.2004.02.010

Aron, A. R., Robbins, T. W., and Poldrack, R. A. (2014). Inhibition and the right inferior frontal cortex: one decade on. *Trends Cogn. Sci.* 18, 177–185. doi: 10.1016/j.tics.2013.12.003

Band, G. P., van Der Molen, M. W., and Logan, G. D. (2003). Horse-race model simulations of the stop-signal procedure. *Acta Psychol. (Amst)* 112, 105–142. doi: 10.1016/s0001-6918(02)00079-3

Bari, A., and Robbins, T. W. (2013). Inhibition and impulsivity: behavioral and neural basis of response control. *Prog. Neurobiol.* 108, 44–79. doi: 10.1016/j.pneurobio.2013.06.005

Basar, E., Basar-Eroglu, C., Karakas, S., and Schürmann, M. (1999). Oscillatory brain theory: a new trend in neuroscience. *IEEE Eng. Med. Biol. Mag.* 18, 56–66. doi: 10.1109/51.765190

Boehler, C. N., Appelbaum, L. G., Krebs, R. M., Hopf, J. M., and Woldorff, M. G. (2010). Pinning down response inhibition in the brain—conjunction analyses of the Stop-signal task. *Neuroimage* 52, 1621–1632. doi: 10.1016/j.neuroimage.2010.04.276

Braver, T. S., Barch, D. M., Gray, J. R., Molfese, D. L., and Snyder, A. (2001). Anterior cingulate cortex and response conflict: effects of frequency, inhibition and errors. *Cereb. Cortex* 11, 825–836. doi: 10.1093/cercor/11.9.825

Brefczynski, J. A., and Deyoe, E. A. (1999). A physiological correlate of the'spotlight'of visual attention. *Nat. Neurosci.* 2, 370–374. doi: 10.1038/7280

Bruchmann, M., Herper, K., Konrad, C., Pantev, C., and Huster, R. J. (2010). Individualized EEG source reconstruction of Stroop interference with masked color words. *Neuroimage* 49, 1800–1809. doi: 10.1016/j.neuroimage.2009.09.032

Cavanagh, J. F., and Frank, M. J. (2014). Frontal theta as a mechanism for cognitive control. *Trends Cogn. Sci.* 18, 414–421. doi: 10.1016/j.tics.2014.04.012

Chang, A., Chen, C.-C., Li, H.-H., and Li, C.-S. R. (2014). Event-related potentials for post-error and post-conflict slowing. *PLoS One* 9:e99909. doi: 10.1371/journal.pone.0099909

Chang, C. F., Hsu, T. Y., Tseng, P., Liang, W. K., Tzeng, O. J., Hung, D. L., et al. (2013). Right temporoparietal junction and attentional reorienting. *Hum. Brain Mapp.* 34, 869–877. doi: 10.1002/hbm.21476

Cohen, M. X. (2015). Comparison of different spatial transformations applied to EEG data: a case study of error processing. *Int. J. Psychophysiol.* 97, 245–257. doi: 10.1016/j.ijpsycho.2014.09.013

Collins, D. L., Neelin, P., Peters, T. M., and Evans, A. C. (1994). Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. *J. Comput. Assist. Tomogr.* 18, 192–205. doi: 10.1097/00004728-199403000-00005

Corbetta, M., and Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nat. Rev. Neurosci.* 3, 201–215. doi: 10.1038/nrn755

Cox, R. W. (1996). AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput. Biomed. Res.* 29, 162–173. doi: 10.1006/cbmr.1996.0014

Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21. doi: 10.1016/j.jneumeth.2003.10.009

Duann, J. R., Ide, J. S., Luo, X., and Li, C. S. (2009). Functional connectivity delineates distinct roles of the inferior frontal cortex and presupplementary motor area in stop signal inhibition. *J. Neurosci.* 29, 10171–10179. doi: 10.1523/JNEUROSCI.1300-09.2009

Duncan, J., and Owen, A. M. (2000). Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends Neurosci.* 23, 475–483. doi: 10.1016/s0166-2236(00)01633-7

Eckert, M. A., Menon, V., Walczak, A., Ahlstrom, J., Denslow, S., Horwitz, A., et al. (2009). At the heart of the ventral attention system: the right anterior insula. *Hum. Brain Mapp.* 30, 2530–2541. doi: 10.1002/hbm.20688

Esterman, M., Tamber-Rosenau, B. J., Chiu, Y. C., and Yantis, S. (2010). Avoiding non-independence in fMRI data analysis: leave one subject out. *Neuroimage* 50, 572–576. doi: 10.1016/j.neuroimage.2009.10.092

Falkenstein, M., Hoormann, J., and Hohnsbein, J. (2002). Inhibition-related ERP components: variation with modality, age and time-on-task. *J. Psychophysiol.* 16, 167–175. doi: 10.1027//0269-8803.16.3.167

Fan, J., Byrne, J., Worden, M. S., Guise, K. G., McCandliss, B. D., Fossella, J., et al. (2007). The relation of brain oscillations to attentional networks. *J. Neurosci.* 27, 6197–6206. doi: 10.1523/JNEUROSCI.1833-07.2007

Fried, I., Katz, A., McCarthy, G., Sass, K. J., Williamson, P., Spencer, S. S., et al. (1991). Functional organization of human supplementary motor cortex studied by electrical stimulation. *J. Neurosci.* 11, 3656–3666.

George, N., Dolan, R. J., Fink, G. R., Baylis, G. C., Russell, C., and Driver, J. (1999). Contrast polarity and face recognition in the human fusiform gyrus. *Nat. Neurosci.* 2, 574–580. doi: 10.1038/9230

Grill-Spector, K., and Malach, R. (2004). The human visual cortex. *Annu. Rev. Neurosci.* 27, 649–677. doi: 10.1146/annurev.neuro.27.070203.144220

Hampshire, A., Chamberlain, S. R., Monti, M. M., Duncan, J., and Owen, A. M. (2010). The role of the right inferior frontal gyrus: inhibition and attentional control. *Neuroimage* 50, 1313–1319. doi: 10.1016/j.neuroimage.2009.12.109

Herrmann, C. S., Grigutsch, M., and Busch, N. A. (2005). "EEG oscillations and wavelet analysis," in *Event-Related Potentials: A Methods Handbook*, ed. T. C. Handy (Cambridge, MA: MIT Press), 229–259.

Husain, M., and Nachev, P. (2007). Space and the parietal cortex. *Trends Cogn. Sci.* 11, 30–36. doi: 10.1016/j.tics.2006.10.011

Huster, R. J., Enriquez-Geppert, S., Lavallee, C. F., Falkenstein, M., and Herrmann, C. S. (2013). Electroencephalography of response inhibition tasks: functional networks and cognitive contributions. *Int. J. Psychophysiol.* 87, 217–233. doi: 10.1016/j.ijpsycho.2012.08.001

Iannaccone, R., Hauser, T. U., Staempfli, P., Walitza, S., Brandeis, D., and Brem, S. (2015). Conflict monitoring and error processing: new insights from simultaneous EEG-fMRI. *Neuroimage* 105, 395–407. doi: 10.1016/j.neuroimage.2014.10.028

Ishibashi, R., Lambon Ralph, M. A., Saito, S., and Pobric, G. (2011). Different roles of lateral anterior temporal lobe and inferior parietal lobule in coding function and manipulation tool knowledge: evidence from an rTMS study. *Neuropsychologia* 49, 1128–1135. doi: 10.1016/j.neuropsychologia.2011.01.004

Johnstone, S. J., Dimoska, A., Smith, J. L., Barry, R. J., Pleffer, C. B., Chiswick, D., et al. (2007). The development of stop-signal and Go/Nogo response inhibition in children aged 7–12 years: performance and event-related

potential indices. *Int. J. Psychophysiol.* 63, 25–38. doi: 10.1016/j.ijpsycho.2006.07.001

Kipp, K. (2005). A developmental perspective on the measurement of cognitive deficits in attention-deficit/hyperactivity disorder. *Biol. Psychiatry* 57, 1256–1260. doi: 10.1016/j.biopsych.2005.03.012

Kopp, B., Rist, F., and Mattler, U. (1996). N200 in the flanker task as a neurobehavioral tool for investigating executive control. *Psychophysiology* 33, 282–294. doi: 10.1111/j.1469-8986.1996.tb00425.x

Koster-Hale, J., Saxe, R., Dungan, J., and Young, L. L. (2013). Decoding moral judgments from neural representations of intentions. *Proc. Natl. Acad. Sci. U S A* 110, 5648–5653. doi: 10.1073/pnas.1207992110

Kübler, A., Murphy, K., Kaufman, J., Stein, E. A., and Garavan, H. (2003). Co-ordination within and between verbal and visuospatial working memory: network modulation and anterior frontal recruitment. *Neuroimage* 20, 1298–1308. doi: 10.1016/s1053-8119(03)00400-2

Lapenta, O. M., Sierve, K. D., de Macedo, E. C., Fregni, F., and Boggio, P. S. (2014). Transcranial direct current stimulation modulates ERP-indexed inhibitory control and reduces food consumption. *Appetite* 83, 42–48. doi: 10.1016/j.appet.2014.08.005

Lavallee, C. F., Herrmann, C. S., Weerda, R., and Huster, R. J. (2014). Stimulus-response mappings shape inhibition processes: a combined EEG-fMRI study of contextual stopping. *PLoS One* 9:e96159. doi: 10.1371/journal.pone.0096159

Leroux, G., Joliot, M., Dubal, S., Mazoyer, B., Tzourio-Mazoyer, N., and Houde, O. (2006). Cognitive inhibition of number/length interference in a Piaget-like task in young adults: evidence from ERPs and fMRI. *Hum. Brain Mapp.* 27, 498–509. doi: 10.1002/hbm.20194

Li, C. S., Huang, C., Constable, R. T., and Sinha, R. (2006). Imaging response inhibition in a stop-signal task: neural correlates independent of signal monitoring and post-response processing. *J. Neurosci.* 26, 186–192. doi: 10.1523/JNEUROSCI.3741-05.2006

Liotti, M., Woldorff, M. G., Perez, R., and Mayberg, H. S. (2000). An ERP study of the temporal course of the Stroop color-word interference effect. *Neuropsychologia* 38, 701–711. doi: 10.1016/s0028-3932(99)00106-2

Logan, G. D., Cowan, W. B., and Davis, K. A. (1984). On the ability to inhibit simple and choice reaction time responses: a model and a method. *J. Exp. Psychol. Hum. Percept. Perform.* 10, 276–291. doi: 10.1037/0096-1523.10.2.276

MaCleod, C. M. (2007). "The concept of inhibition in cognition," in *Inhibition in Cognition*, eds D. S. Gorfein and C. M. MacLeod (Washington, DC, US: American Psychological Association), 3–23.

Makeig, S., Bell, A. J., Jung, T.-P., and Sejnowski, T. J. (1996). Independent component analysis of electroencephalographic data. *Adv. Neural Inf. Process. Syst.* 8, 145–151.

Mathewson, K. E., Beck, D. M., Ro, T., Maclin, E. L., Low, K. A., Fabiani, M., et al. (2014). Dynamics of alpha control: preparatory suppression of posterior alpha oscillations by frontal modulators revealed with combined EEG and event-related optical signal. *J. Cogn. Neurosci.* 26, 2400–2415. doi: 10.1162/jocn_a_00637

Menon, V., Adleman, N. E., White, C. D., Glover, G. H., and Reiss, A. L. (2001). Error–related brain activation during a Go/NoGo response inhibition task. *Hum. Brain Mapp.* 12, 131–143. doi: 10.1002/1097-0193(200103)12:3<131::aid-hbm1010>3.0.co;2-c

Menon, V., and Uddin, L. Q. (2010). Saliency, switching, attention and control: a network model of insula function. *Brain Struct. Funct.* 214, 655–667. doi: 10.1007/s00429-010-0262-0

Miller, E. K., and Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.* 24, 167–202. doi: 10.1146/annurev.neuro.24.1.167

Mostofsky, S. H., Schafer, J. G. B., Abrams, M. T., Goldberg, M. C., Flower, A. A., Boyce, A., et al. (2003). fMRI evidence that the neural basis of response inhibition is task-dependent. *Brain Res. Cogn. Brain Res.* 17, 419–430. doi: 10.1016/s0926-6410(03)00144-7

Mulert, C. (2013). Simultaneous EEG and fMRI: towards the characterization of structure and dynamics of brain networks. *Dialogues Clin. Neurosci.* 15, 381–386.

Nachev, P., Kennard, C., and Husain, M. (2008). Functional role of the supplementary and pre-supplementary motor areas. *Nat. Rev. Neurosci.* 9, 856–869. doi: 10.1038/nrn2478

Nieuwenhuis, S., Yeung, N., and Cohen, J. D. (2004). Stimulus modality, perceptual overlap and the go/no-go N2. *Psychophysiology* 41, 157–160. doi: 10.1046/j.1469-8986.2003.00128.x

Obeso, I., Cho, S. S., Antonelli, F., Houle, S., Jahanshahi, M., Ko, J. H., et al. (2013). Stimulation of the pre-SMA influences cerebral blood flow in frontal areas involved with inhibitory control of action. *Brain Stimul.* 6, 769–776. doi: 10.1016/j.brs.2013.02.002

Oostenveld, R., and Oostendorp, T. F. (2002). Validating the boundary element method for forward and inverse EEG computations in the presence of a hole in the skull. *Hum. Brain Mapp.* 17, 179–192. doi: 10.1002/hbm.10061

Penfield, W., and Welch, K. (1951). The supplementary motor area of the cerebral cortex: a clinical and experimental study. *AMA Arch. Neurol. Psychiatry* 66, 289–317. doi: 10.1001/archneurpsyc.1951.02320090038004

Ramautar, J. R., Kok, A., and Ridderinkhof, K. R. (2006). Effects of stop-signal modality on the N2/P3 complex elicited in the stop-signal paradigm. *Biol. Psychol.* 72, 96–109. doi: 10.1016/j.biopsycho.2005.08.001

Regenbogen, C., Herrmann, M., and Fehr, T. (2010). The neural processing of voluntary completed, real and virtual violent and nonviolent computer game scenarios displaying predefined actions in gamers and nongamers. *Soc. Neurosci.* 5, 221–240. doi: 10.1080/17470910903315989

Rubia, K., Smith, A. B., Brammer, M. J., and Taylor, E. (2003). Right inferior prefrontal cortex mediates response inhibition while mesial prefrontal cortex is responsible for error detection. *Neuroimage* 20, 351–358. doi: 10.1016/s1053-8119(03)00275-1

Rubia, K., Smith, A. B., Brammer, M. J., Toone, B., and Taylor, E. (2005). Abnormal brain activation during inhibition and error detection in medication-naive adolescents with ADHD. *Am. J. Psychiatry* 162, 1067–1075. doi: 10.1176/appi.ajp.162.6.1067

Rushworth, M. F., Walton, M. E., Kennerley, S. W., and Bannerman, D. M. (2004). Action sets and decisions in the medial frontal cortex. *Trends Cogn. Sci.* 8, 410–417. doi: 10.1016/j.tics.2004.07.009

Saxe, R., and Kanwisher, N. (2003). People thinking about thinking people The role of the temporo-parietal junction in "theory of mind". *Neuroimage* 19, 1835–1842. doi: 10.1016/s1053-8119(03)00230-1

Schmiedt-Fehr, C., and Basar-Eroglu, C. (2011). Event-related delta and theta brain oscillations reflect age-related changes in both a general and a specific neuronal inhibitory mechanism. *Clin. Neurophysiol.* 122, 1156–1167. doi: 10.1016/j.clinph.2010.10.045

Schulz, H., Ubelacker, T., Keil, J., Müller, N., and Weisz, N. (2014). Now I am ready-now i am not: the influence of pre-TMS oscillations and corticomuscular coherence on motor-evoked potentials. *Cereb. Cortex* 24, 1708–1719. doi: 10.1093/cercor/bht024

Shallice, T., Stuss, D. T., Alexander, M. P., Picton, T. W., and Derkzen, D. (2008a). The multiple dimensions of sustained attention. *Cortex* 44, 794–805. doi: 10.1016/j.cortex.2007.04.002

Shallice, T., Stuss, D. T., Picton, T. W., Alexander, M. P., and Gillingham, S. (2008b). Mapping task switching in frontal cortex through neuropsychological group studies. *Front. Neurosci.* 2, 79–85. doi: 10.3389/neuro.01.013.2008

Shou, G., and Ding, L. (2015). Detection of EEG spatial-spectral-temporal signatures of errors: a comparative study of ICA-based and channel-based methods. *Brain Topogr.* 28, 47–61. doi: 10.1007/s10548-014-0397-z

Simmonds, D. J., Pekar, J. J., and Mostofsky, S. H. (2008). Meta-analysis of Go/No-go tasks demonstrating that fMRI activation associated with response inhibition is task-dependent. *Neuropsychologia* 46, 224–232. doi: 10.1016/j.neuropsychologia.2007.07.015

Slotnick, S. D., Schwarzbach, J., and Yantis, S. (2003). Attentional inhibition of visual processing in human striate and extrastriate cortex. *Neuroimage* 19, 1602–1611. doi: 10.1016/s1053-8119(03)00187-3

Smith, A. T., Singh, K. D., and Greenlee, M. W. (2000). Attentional suppression of activity in the human visual cortex. *Neuroreport* 11, 271–278. doi: 10.1097/00001756-200002070-00010

Swann, N. C., Cai, W., Conner, C. R., Pieters, T. A., Claffey, M. P., George, J. S., et al. (2012). Roles for the pre-supplementary motor area and the right inferior frontal gyrus in stopping action: electrophysiological responses and functional and structural connectivity. *Neuroimage* 59, 2860–2870. doi: 10.1016/j.neuroimage.2011.09.049

Swick, D., Ashley, V., and Turken, U. (2011). Are the neural correlates of stopping and not going identical? Quantitative meta-analysis of two response

inhibition tasks. *Neuroimage* 56, 1655–1665. doi: 10.1016/j.neuroimage.2011.02.070

Van Essen, D. C., Drury, H. A., Dickson, J., Harwell, J., Hanlon, D., and Anderson, C. H. (2001). An integrated software suite for surface-based analyses of cerebral cortex. *J. Am. Med. Inform. Assoc.* 8, 443–459. doi: 10.1136/jamia.2001.0080443

Verbruggen, F., and Logan, G. D. (2008). Response inhibition in the stop-signal paradigm. *Trends Cogn. Sci.* 12, 418–424. doi: 10.1016/j.tics.2008.07.005

Yantis, S., Schwarzbach, J., Serences, J. T., Carlson, R. L., Steinmetz, M. A., Pekar, J. J., et al. (2002). Transient neural activity in human parietal cortex during spatial attention shifts. *Nat. Neurosci.* 5, 995–1002. doi: 10.1038/nn921

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Ko, Shih, Chikara, Chuang and Chang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution and reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 13 May 2016
doi: 10.3389/fnhum.2016.00219

Exploring Neuro-Physiological Correlates of Drivers' Mental Fatigue Caused by Sleep Deprivation Using Simultaneous EEG, ECG, and fNIRS Data

Sangtae Ahn[1†], Thien Nguyen[2†], Hyojung Jang[1], Jae G. Kim[2] and Sung C. Jun[1*]*

[1] School of Electrical Engineering and Computer Science, Gwangju Institute of Science and Technology, Gwangju, South Korea, [2] Department of Biomedical Science and Engineering, Gwangju Institute of Science and Technology, Gwangju, South Korea

Investigations of the neuro-physiological correlates of mental loads, or states, have attracted significant attention recently, as it is particularly important to evaluate mental fatigue in drivers operating a motor vehicle. In this research, we collected multimodal EEG/ECG/EOG and fNIRS data simultaneously to develop algorithms to explore neuro-physiological correlates of drivers' mental states. Each subject performed simulated driving under two different conditions (well-rested and sleep-deprived) on different days. During the experiment, we used 68 electrodes for EEG/ECG/EOG and 8 channels for fNIRS recordings. We extracted the prominent features of each modality to distinguish between the well-rested and sleep-deprived conditions, and all multimodal features, except EOG, were combined to quantify mental fatigue during driving. Finally, a novel driving condition level (DCL) was proposed that distinguished clearly between the features of well-rested and sleep-deprived conditions. This proposed DCL measure may be applicable to real-time monitoring of the mental states of vehicle drivers. Further, the combination of methods based on each classifier yielded substantial improvements in the classification accuracy between these two conditions.

Keywords: EEG/ECG/EOG/fNIRS, neuro-physiological correlates, drivers' mental fatigue, sleep deprivation, simulated driving, multimodal integration, driving condition level

OPEN ACCESS

Edited by:
Klaus Gramann, Berlin Institute of Technology, Germany

Reviewed by:
Tzyy-Ping Jung, University of California San Diego, USA
Chin-Teng Lin, National Chiao-Tung University, Taiwan

***Correspondence:**
Jae G. Kim
jaekim@gist.ac.kr;
Sung C. Jun
scjun@gist.ac.kr

[†]These authors have contributed equally to this work.

Received: *05 November 2015*
Accepted: *27 April 2016*
Published: *13 May 2016*

Citation:
Ahn S, Nguyen T, Jang H, Kim JG and Jun SC (2016) Exploring Neuro-Physiological Correlates of Drivers' Mental Fatigue Caused by Sleep Deprivation Using Simultaneous EEG, ECG, and fNIRS Data. Front. Hum. Neurosci. 10:219. doi: 10.3389/fnhum.2016.00219

INTRODUCTION

Neuroergonomics is an emerging field that investigates human mental states and their workloads in order to improve the reliability of human performance, and ensure its stability in various environments (Parasuraman, 2003; Parasuraman and Rizzo, 2008). In neuroergonomics, both the fundamental principles of neuroscience and human factors are considered thoroughly, and neural behaviors have been investigated primarily when people are engaged in tasks in a work environment (Parasuraman and Wilson, 2008). Due to the implications for public safety, a major application of neuroergonomics is the assessment of driver fatigue. In general, driver fatigue is categorized as either mental or physical. Mental fatigue occurs because of gradual and cumulative mental effort (Grandjean, 1979) during driving, or sleep deprivation before driving (Durmer and Dinges, 2005). In contrast, physical fatigue represents reduced muscular strength and coordination. Physical

fatigue may be countered by deliberate action; however, mental fatigue is difficult to resolve. Because of mental fatigue, drivers begin to doze involuntary, which often results in traffic accidents (Horne and Reyner, 1999; Connor et al., 2002; Herman et al., 2014).

One potential method that may be used to reduce traffic accidents is to measure inherent mental fatigue before or during driving, in order to predict a driver's mental condition and determine whether s/he can drive safely. Because driving requires complex cognitive processes and sustained concentration, predicting a driver's mental fatigue before or during driving could be effective in preventing traffic accidents. Thus, we attempted in this work to explore neuro-physiological correlates in two different conditions, one well-rested with a low risk of fatigue, and the other sleep-deprived with a high risk of fatigue.

Among many studies performed to evaluate drivers' fatigue in real-time, computer vision-based systems have been used widely. Bergasa et al. (2006) proposed a noninvasive system to monitor a driver's vigilance using several parameters, including percentage or duration of eye closure, blinking, and the frequency of nodding. By using a fuzzy classifier, the researchers then inferred the level of the drivers' fatigue. However, the reliability of the findings decreased when the drivers wore glasses or the surrounding brightness changed. To address these problems, D'Orazio et al. (2007) designed an experimental paradigm that incorporated conditions in which some subjects had different eye colors, wore glasses, and drove vehicles in light of varying intensities. Using the proposed visual framework, the authors obtained robust results. In addition, various visual cues that characterized eyelid, gaze, and head movements, as well as facial expressions were employed in a probabilistic model developed to predict fatigue (Ji et al., 2004) that yielded even more robust results. Recently, Wang et al. (2014) developed an online, closed-loop lapse detection system featuring a mobile wireless electroencephalograph (EEG), and were able to extract certain EEG signatures associated with fatigue.

To date, EEG has been found to be a promising indicator for investigations of driver fatigue (Lal and Craig, 2001). EEG data have shown that there is a significant increase in theta and delta activity, and a decrease in heart rate (HR) associated with fatigue (Lal and Craig, 2002). Further, in a subsequent study that considered three phases of fatigue (early, medium, and extreme), software was developed to monitor driving fatigue, and was validated with EEG data from 35 subjects engaged in a simulated driving task (Lal et al., 2003). Another study (Lin et al., 2005) estimated drowsiness and driver performance by correlating changes in log power spectra. To detect drowsiness, they constructed an individualized linear regression model to assess EEG dynamics continuously based on an independent component analysis. Because drowsiness is a crucial factor in driving, Lin and his colleagues investigated the effect of continuous arousing auditory feedback on sustained attention in a driving simulator (Lin et al., 2010). They found that spectral powers in alpha and theta bands were suppressed and lasted 30 s or longer after feedback. This finding was introduced to estimate classification accuracy; as a result, they achieved a classification accuracy of approximately 78% using the maximum likelihood classifier (Lin et al., 2013) and applied it to develop an online, closed-loop system for practical lapse detection in real environments (Wang et al., 2014).

Various other methods have been used to explore drivers' mental fatigue, such as a support vector machine (SVM) (Shen et al., 2008; Yeo et al., 2009), Bayesian network (Yang et al., 2010), wavelet analysis (Kar et al., 2010; Li and Chung, 2013), and others. In addition to EEG studies, electrocardiography (ECG) and electrooculography (EOG) have been used to determine neuro-physiological correlates of drivers' mental fatigue. One study (Patel et al., 2011) used neural network analysis and demonstrated that the variability in drivers' HRs differed significantly in alert and fatigued states. They investigated the power spectral density behaviors between the two states during long-term driving and reported that the neural network was 90% accurate in classifying mental state.

Eyelid-related features from EOG data also have been reported to be possible candidates to detect whether or not a driver is drowsy (Hu and Zheng, 2009). In this report, they used vertical and horizontal EOG channels to extract and validate eye blinks according to eyelid movement parameters, such as blink duration, speed, and amplitude. Three conditions (alert, sleepy, and very sleepy) were classified with high reliability using SVM. Simultaneous recording of EEG/ECG (Zhao et al., 2012) and the combination of multimodal features from EEG, EOG, and ECG data (Khushaba et al., 2011) demonstrated significant differences during long-term driving. In this study, the researchers developed an efficient, fuzzy mutual information-based wavelet packet transformation that combined EEG, EOG, and ECG features to detect drivers' drowsiness; this technique yielded a classification accuracy greater than 90%.

An emerging portable and noninvasive brain functional imaging technique, functional near infra-red spectroscopy (fNIRS), has been introduced to monitor cognitive workload or fatigue in simulated environments (Ayaz et al., 2012). fNIRS data from the prefrontal cortex were collected during a complex air-traffic control task that required the subjects to prevent collisions between aircraft in their sectors. As the number of aircraft in their sector increased, a concomitant increase in prefrontal cortex activation was observed, which suggests that fNIRS provides a sensitive index of cognitive workload. fNIRS also demonstrated changes in prefrontal activation during skill acquisition in both basic working memory tasks (McKendrick et al., 2014) and more complex piloting tasks (Harrison et al., 2014; Gateau et al., 2015).

A portable fNIRS device was developed for use in mobile neuroimaging of the prefrontal cortex (Ayaz et al., 2013). In a driving environment, Li et al. (2009) observed changes in cerebral oxygenation during prolonged simulated driving. Forty healthy subjects were divided randomly into two groups (driving vs. non-driving), and the driving group performed a simulated 3 h driving task. The authors found a relative increase in frontal cortex oxygenation in the driving group by comparison to the non-driving group, and oxygenation decreased gradually after the driving task. Considering real driving situations, Yoshino et al. (2013) investigated the changes in cerebral oxygen exchange during actual driving on an expressway. An fNIRS signal was recorded in the subjects' parietal and prefrontal cortices using

an fNIRS device mounted in the vehicle. They found that the areas activated varied depending on the driving task, such as parking, acceleration, driving at constant speed, deceleration, and U-turns. Thus, the use of fNIRS may be an effective approach to evaluate brain activity in various driving environments.

Recently, hybrid approaches that combine two different modalities (Pfurtscheller et al., 2010) to improve performance and reduce classification error have been reported as promising for future brain-computer interfaces (BCI). One example of a hybrid BCI that incorporates both EEG electrical activity and fNIRS hemodynamic changes yielded improved classification performance in sensorimotor rhythm-based BCI systems (Fazli et al., 2012). The researchers calculated classification accuracies in the movements executed and motor imagery by estimating a meta-classifier. After the estimation of both classifiers (EEG and fNIRS), the combination of outputs of each classifier resulted in improved classification accuracy. Khan et al. (2014) decoded four movement directions (left, right, forward, and backward) using the mixed features of EEG and fNIRS, in which EEG features were used to classify left/right, and fNIRS features were used to classify forward/backward. In addition, hybrid BCI may be used as a brain switch that determines whether a certain task is active. Koo et al. (2015) employed a novel experimental paradigm to detect the occurrence of motor imagery in fNIRS data. Threshold-based detection with a feature value of the fNIRS data determined whether or not the action of a motor imagery task was attempted. The combination of EEG and fNIRS is also applicable to language studies (Wallois et al., 2012) and cortical current estimation (Morioka et al., 2014). Hybrid BCIs may provide a good opportunity to increase BCI performance by offering the synergistic effects of multimodal brain imaging techniques.

In this work, we recorded multimodal EEG/ECG/EOG and fNIRS data simultaneously in a driving simulator and combined their features to distinguish drivers with high- and low-risks of fatigue using neuro-physiological correlates and a classification method. Hemodynamic changes in the prefrontal cortex (Li et al., 2009; Ayaz et al., 2012, 2013; Yoshino et al., 2013; Harrison et al., 2014; McKendrick et al., 2014; Gateau et al., 2015) have been used to neuro-physiological correlates, and these activities were reported to play an important role in neuroergonomics, such as mental workload (Mandrick et al., 2013a), cognitive operation (Mandrick et al., 2013b), and emotional function (Doi et al., 2013). Furthermore, it is clear that EEG, ECG, and EOG are also promising indicators that may be used to investigate the neuro-physiological correlates of drivers' mental fatigue (Lal and Craig, 2001). Therefore, combining this hybrid system with prefrontal fNIRS may be a far more informative measure for identifying neuro-physiological correlates under varying driving conditions. To the best of our knowledge, this multimodal approach has been tested rarely to explore neuro-physiological correlates of drivers' mental fatigue.

Thus, the goal of this study was to determine modality-specific features of EEG, EOG, ECG, and fNIRS. These features were then used to distinguish between well-rested and sleep-deprived conditions, and resulted in a classifier that showed whether or not a driver was in an alert mental state. The use of a reasonable combination of these multimodal features may improve classification accuracy and its quantification may yield a real-time strategy to monitor drivers' mental fatigue.

MATERIALS AND METHODS

Experimental Procedure

Eleven healthy subjects (10 males, 1 female, aged 26.6 ± 1.4, range = 24–28) who had valid driver's licenses participated in a custom-built virtual driving simulation task, as depicted in **Figure 1A**. The subjects practiced repeatedly until they were familiar with the simulation system. The purposes of, and instructions for, the experiment were explained in advance, and all of the subjects signed an informed consent. Subjects received approximately $10 per h as compensation for their participation. Each subject performed simulated driving under two conditions (well-rested and sleep-deprived) on different days. Under the well-rested condition, subjects were instructed to sleep at least 7 h before the experiment, as sleeping seven or more hours is known to maintain healthy mental alertness (Kripke et al., 2002). In the sleep-deprived condition, the subjects were instructed to stay up all night in order to produce mental fatigue.

Driving tests in both conditions were performed before 9 a.m. In this experiment, we assumed that subjects would be significantly mentally fatigued after one night of sleep deprivation. To determine the degree of fatigue produced by sleep deprivation, a subjective questionnaire was administered to the subjects before the experiment to score their levels of fatigue, and the scores demonstrated clearly that the sleep-deprived subjects were substantially more fatigued than were the well-rested subjects. The subjects sat in a comfortable driver's seat and drove on an oval track for a minimum of 30 min. The maximum driving speed was set at 100 km/h in both conditions. The steering wheel vibrated whenever the vehicle collided with a crash barrier in order to prevent the drivers from falling asleep completely. A high-definition webcam (Logitech HD Pro C920) was used to record each subject's behavior in real-time. This experiment was approved by the Institutional Review Board at the Gwangju Institute of Science and Technology (20150615-HR-18-02-06).

Data Recording of EEG/ECG/EOG and fNIRS

Sixty-four EEG electrodes were attached to the drivers' scalps according to the 10–20 international position system. Horizontal and vertical EOGs were used and two ECG electrodes were attached to the left/right chest (Biosemi ActiveTwo System). These data were collected at a 512 Hz sampling rate using BCI2000 software (Schalk et al., 2004). Biosemi ActiView software monitored the stability and reliability of the EEG signal. After the experiment, bad channels that contained abnormal noise were identified by visual inspection and excluded from the analysis.

A custom-built fNIRS system (continuous wave, 10 Hz sampling rate) was used to record hemodynamic changes in the brain. This was an updated version of one described in a

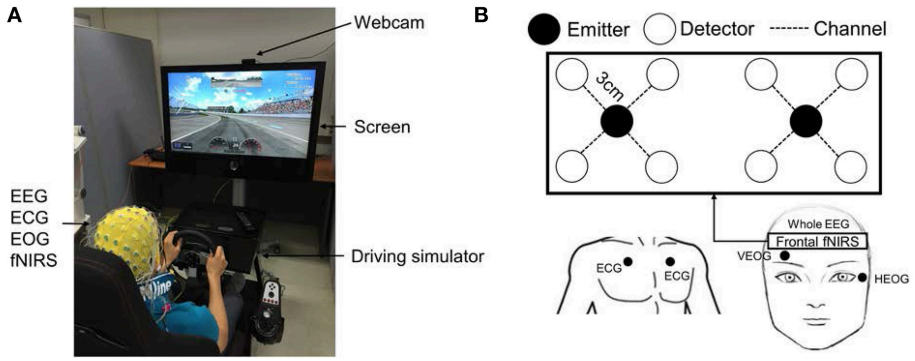


FIGURE 1 | Experimental setup for simulated driving. (A) Driving simulation environment with EEG/ECG/EOG/fNIRS measurements. **(B)** fNIRS setup in the prefrontal cortex. Two emitters and eight detectors (eight channels) were attached to the forehead. The distance between emitter and detector was 3 cm.

previous work (Kim et al., 2015). The system consists of probe and control circuits. The probe includes 2 LEDs (emitters) and 8 photodetectors (detectors). The LEDs emit near infrared (NIR) light at two wavelengths (735 and 850 nm). The emitter and four surrounding detectors were separated by 3 cm, as Homma et al. (1996) suggested that in soft tissues, NIR is able to attain a penetration depth equal to half of the emitter-detector separation. Therefore, with a 3.0 cm emitter-detector separation, our system should have been able to collect brain activity at a depth of 1.5 cm below the scalp. An emitter-detector pair form one fNIRS channel that measures hemodynamic changes midway between the emitter and the detector. Given a suitable geometric arrangement, many detectors may receive light from one emitter. This enabled us to design an 8-channel probe with 2 LEDs and 8 photodetectors. The 8-channel probe was attached to the prefrontal region to investigate the subjects' mental state, as illustrated in **Figure 1B** (Li et al., 2009; Sato et al., 2013). The control circuit receives a signal from the probe, amplifies it, and sends it to the computer via serial communication. Matlab-based software was programmed to record, process, and display the hemodynamic signals. Interference between EEG/EOG/ECG electrodes and fNIRS emitters has been observed and is believed to result from light leakage from the emitters, which may cause deterioration in the quality of electrical data (Koo et al., 2015). This interference was removed by blocking light leakage from the emitters and applying a simple pre-processing technique. Two desktop computers were used to record the EEG/ECG/EOG and fNIRS data simultaneously. Triggers for start and end times were sent to the BCI2000 software to synchronize the multimodal EEG/ECG/EOG and fNIRS data. The computer that recorded EEG/ECG/EOG data sent a start trigger, at which time the second computer began to record fNIRS data. The end time of the experiment was marked in the same way.

Data Analysis

Feature Extraction and Classification from EEG

After the experiment, the data collected were inspected visually and bad channels were rejected. The logistic infomax independent component analysis (Bell and Sejnowski, 1995) was used to remove EOG artifacts and the data were then band-pass filtered from 1 to 50 Hz. We analyzed data from the first 30 min only after the drivers began the task, because fatigue levels between well-rested and sleep-deprived subjects were likely to be quite different during the initial minutes of driving. From the real-time webcam video monitoring data, we observed that even some well-rested subjects became drowsy and quite bored after that length of time.

EEG data (30 min) were divided into 10 s (a trial) to yield a total of 180 trials for each driving condition. A power spectral density was computed for each trial using the EEGLAB library (Delorme and Makeig, 2004), and a relative power level (RPL) was computed in order to reduce session/subject variability (Ahn et al., 2013a,b, 2014; Cho et al., 2015). To calculate the RPL, we considered five spectral band ranges: delta (1–4 Hz), theta (4–8 Hz), alpha (8–13 Hz), beta (13–30 Hz), and gamma (30–50 Hz). Next, each band-power was normalized by the total power, defined as the sum over all band powers, after which we extracted the most informative RPL features between the two driving conditions. On the other hand, to discriminate between the well-rested and sleep-deprived conditions, pre-processed data (180 trials) from each driving condition (well-rested and sleep-deprived) were firstly divided into 2 groups (training and test) and according to time sequence; training and test groups were composed of 126 (70%) and 54 trials (30%), respectively. Then, to avoid temporal dependency between groups, last 6 trials (1 min) for each group were excluded; thus, for each of driving conditions, 120 and 48 trials for training and test were obtained, respectively. By this grouping, temporal dependency (adjacency) was included within groups, but was excluded between groups. This procedure was repeated 30 times by sliding temporal window of 1 min (6 trials) and then choosing training and test groups. Thereafter, each feature vector of the training and test data using RPL was fed into the classifier. The training group was used to construct a classifier based on Fisher's linear discriminant analysis (FLDA), and the test group was input to a constructed classifier in order to measure classification accuracy. A classifier was generated from the training data and the classification accuracy was estimated from the test data.

Finally, 30 classification accuracies were estimated to obtain an average accuracy.

Feature Extraction and Classification from ECG and EOG

The HRs of each subject were extracted using two ECG channels (left/right chest). During pre-processing, ECG data were band-pass filtered from 0.1 to 30 Hz and were detrended to remove the baseline shift. After detrending, a QRS-complex was observed to be the most prominent repeating peak in the ECG signal. The QRS-wave is used commonly to determine subjects' HRs or predict abnormalities in cardiac function. Specifically, the emergence of an R-peak indicated a subject's HR clearly and was extracted easily by adjusting a deterministic threshold of the ECG magnitude. Next, the number of R-peaks per minute was counted and used to determine HR per minute. HRs from the two ECG channels on the left and right chest were calculated for the entire 30 min and averaged to reduce possible detection error and bias. To classify mental state from the ECG data, we adopted the extraction of RR-peak interval features (de Chazal et al., 2004). After detection of the R-peak in each 10-s trial, the intervals between one R-peak and the next were averaged, and the procedure was repeated for all trials. In this way, 180 R-peak intervals were estimated as a feature set. The EOG signal was used to extract the rate of eye blinking in each 1-min trial, which has been reported to be associated well with a human's mental state (Schleicher et al., 2008): when the eye blinks, a clear, sharp wave is observed. After baseline drift removal was applied, a peak detection algorithm (Pettersson et al., 2013) was used with a given threshold of signal magnitude. Finally, the number of peaks per minute, which represented the eye-blinking rate, was used as the EOG feature.

Feature Extraction and Classification from fNIRS

We adopted the modified Beer-Lambert's law (mBLL) to retrieve relative concentration changes from the light intensities of the 8 detectors (Cope et al., 1988; Kocsis et al., 2006). The change in optical density at two wavelengths (735 and 850 nm) is related to changes in oxy-hemoglobin concentration (HbO) and deoxy-hemoglobin concentration (HbR). Data with abnormal noise were removed by visual inspection, and the remaining data were filtered with a 0.01 Hz high-pass filter to remove baseline drifts. Light intensities for 30 s after the initiation of the experiment were averaged and set as baseline intensities. HbO and HbR were estimated with the following equations:

$$\triangle HbO = \frac{log\frac{I_b^{\lambda_1}}{I_t^{\lambda_1}}\varepsilon_{HbR}^{\lambda_2} - log\frac{I_b^{\lambda_2}}{I_t^{\lambda_2}}\varepsilon_{HbR}^{\lambda_1}}{d \cdot DPF\left[\varepsilon_{HbO}^{\lambda_1}\varepsilon_{HbR}^{\lambda_2} - \varepsilon_{HbO}^{\lambda_2}\varepsilon_{HbR}^{\lambda_1}\right]}, \quad (1)$$

$$\triangle HbR = \frac{log\frac{I_b^{\lambda_2}}{I_t^{\lambda_2}}\varepsilon_{HbO}^{\lambda_1} - log\frac{I_b^{\lambda_1}}{I_t^{\lambda_1}}\varepsilon_{HbO}^{\lambda_2}}{d \cdot DPF\left[\varepsilon_{HbO}^{\lambda_1} \cdot \varepsilon_{HbR}^{\lambda_2} - \varepsilon_{HbO}^{\lambda_2} \cdot \varepsilon_{HbR}^{\lambda_1}\right]}, \quad (2)$$

where
I_b^{λ} : *baseline intensity* (λ_1 : 735 *nm* λ_2 : 850 *nm*)
I_t^{λ} : *transient intensity*
d: *emitter* − *detector separation*
$\varepsilon_{Hb}^{\lambda}$: *extinction coefficient*
DPF : *differential path length factor*

In continuous wave fNIRS, the differential path length factor (DPF) is unknown. However, it is similar for both wavelengths and is included conventionally in the unit of hemodynamic changes as a scaling factor. Thus, HbO and HbR have the same unit of mM/DPF, and the extinction coefficients are specific for HbO and HbR at each wavelength. Matcher et al. (1995) measured extinction coefficients of hemoglobin at different wavelengths as follows:

at wavelength λ_1= 735 nm,

$$\varepsilon_{HbO}^{\lambda_1} = 0.4646\text{m}M^{-1}cm^{-1} \text{ and } \varepsilon_{HbR}^{\lambda_1} = 1.2959\text{m}M^{-1}cm^{-1},$$

at wavelength λ_2= 850 nm,

$$\varepsilon_{HbO}^{\lambda_2} = 1.1596\text{m}M^{-1}cm^{-1} \text{ and } \varepsilon_{HbR}^{\lambda_2} = 0.7861\text{m}M^{-1}cm^{-1}.$$

Like EEG feature extraction, 10 s of data were defined as one trial, which yielded a total of 180 trials per condition. Next, relative concentration changes were estimated for each trial. To reduce the effects of fluctuations and noise, fNIRS data were smoothed using 10-s temporal windowing with a 50% overlap. Finally, the amplitudes of HbO and HbR were used as informative features for classification of the two driving conditions.

RESULTS

Relative Power Level from EEG

We investigated RPL values over five spectral bands—delta, theta, alpha, beta, and gamma—and found that the RPL values for delta, theta, and gamma did not differ statistically between the two driving conditions. However, alpha and beta RPL values differed clearly in the two conditions, as shown in **Figure 2A**. Grand-averaged topographies were described for each condition, and alpha RPL in the sleep-deprived condition was activated to a greater degree in the right centro-parietal region. Such an increase in alpha power has been reported in the literature as a notable marker in driving (Lal and Craig, 2001; Simon et al., 2011). A decrease in beta RPL over the fronto-central region was observed in the sleep-deprived condition. This decrease in beta power may indicate a lack of arousal, which is consistent with the results of several studies (Tanaka et al., 2012; Zhao et al., 2012).

Figure 2B shows the distributions of alpha and beta RPLs from subject S5 in a two-dimensional Cartesian coordinate space. We note that most subjects showed similar physiological behaviors, and S5's results were chosen as representative because they yielded the highest classification accuracy in the EEG, as shown in **Table 3**. For the purposes of consistency and comparison, the other results (ECG, fNIRS) from subject S5 are also illustrated in the subsequent sections. Each RPL was averaged spatially over significant regions, such as the centro-parietal for alpha and the fronto-central for beta. Each dot

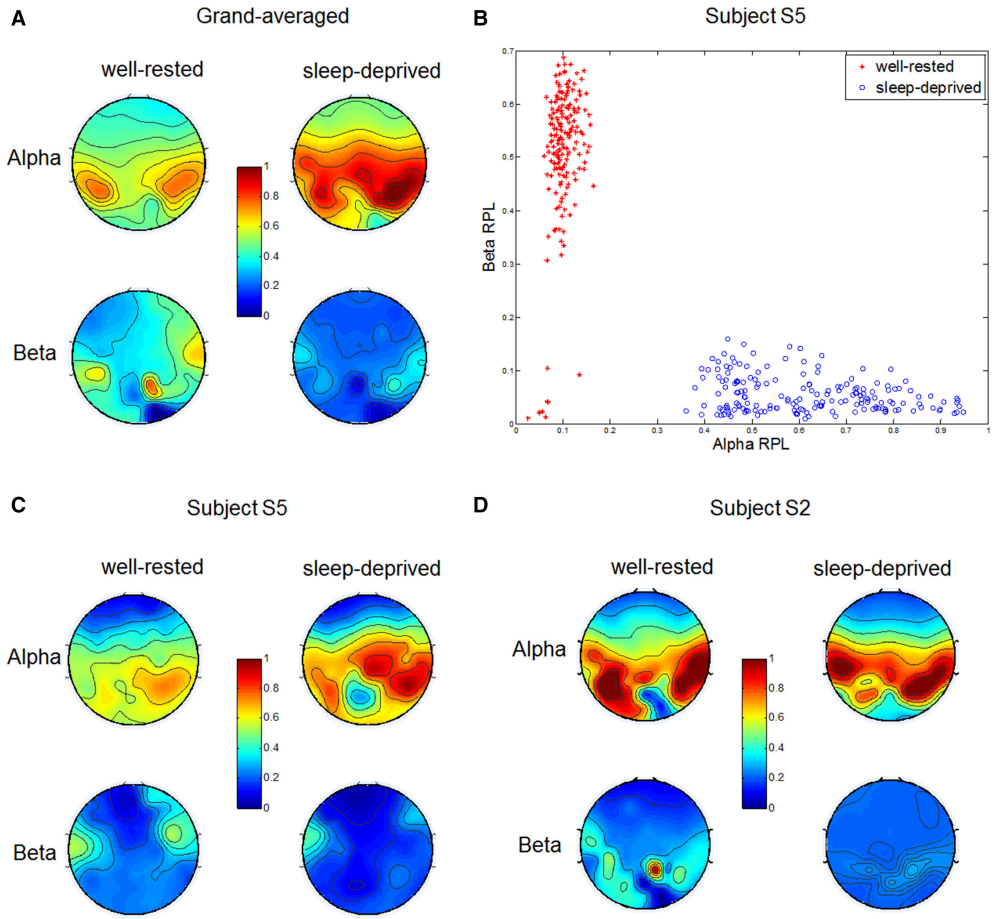


FIGURE 2 | Relative power levels from EEG in two different conditions. (A) Grand-averaged alpha and beta RPLs in well-rested and sleep-deprived conditions. Alpha and beta RPLs differed significantly in the right centro-parietal and frontal regions, respectively. **(B)** Scatter plot of alpha (x-axis) and beta (y-axis) RPLs in two driving conditions (red asterisk: well-rested, blue circle: sleep-deprived) for subject S5. **(C,D)** Indicate alpha and beta RPLs in the well-rested and sleep-deprived conditions, in which subject S5 had the highest and S2 had the lowest classification accuracy, respectively.

represents corresponding alpha (x-coordinate) and beta (y-coordinate) RPLs for one trial in each condition (well-rested and sleep-deprived). In the well-rested condition, most RPL dots were distributed in the upper left area in R^2 space, while they were distributed in the lower right area in the sleep-deprived condition. Thus, these features (centro-parietal alpha RPL and fronto-central beta RPL) in the dataset collected allowed us to achieve a discriminative classification between the two driving conditions quite well.

To investigate inter-subject variability (Ahn and Jun, 2015), we plotted RPL topographies in **Figures 2C,D**, respectively, for two subjects who achieved the highest (S5) and the lowest (S2) EEG classification accuracies (**Table 3**). As depicted in the figures, subject S5 showed a clear alpha RPL increase in the right centro-parietal region and beta RPL decrease in the fronto-central region in the sleep-deprived condition. In contrast, subject S2, who demonstrated the lowest classification accuracy, showed a slight alpha RPL increase and a beta RPL decrease in the sleep-deprived condition. Interestingly, this subject (S2) was likely to have been fatigued already, despite being in the well-rested condition before the experiment. Our investigation of this subject will be described in detail in the Discussion section.

Time Course of Relative Concentration Changes from fNIRS

The time course of the relative concentration changes of HbO and HbR were estimated through mBLL. **Figure 3** depicts the concentration changes at channels 1 and 5 for subject S5. Because all channels were attached to the prefrontal cortex, they all showed similar behaviors over time. Thus, for the purpose of illustration, we chose two representative channels (1 and 5). The concentration changes of HbO increased gradually over time in the well-rested condition and demonstrated the highest level at channels 1 and 5 between 20 and 30 min. Paying attention while driving a vehicle requires high oxygen consumption in the brain, which induces an increase in cerebral blood flow; this increase in cerebral oxygenation, as shown by an increase in HbO and decrease in HbR, indicates that the cerebral blood flow increased during

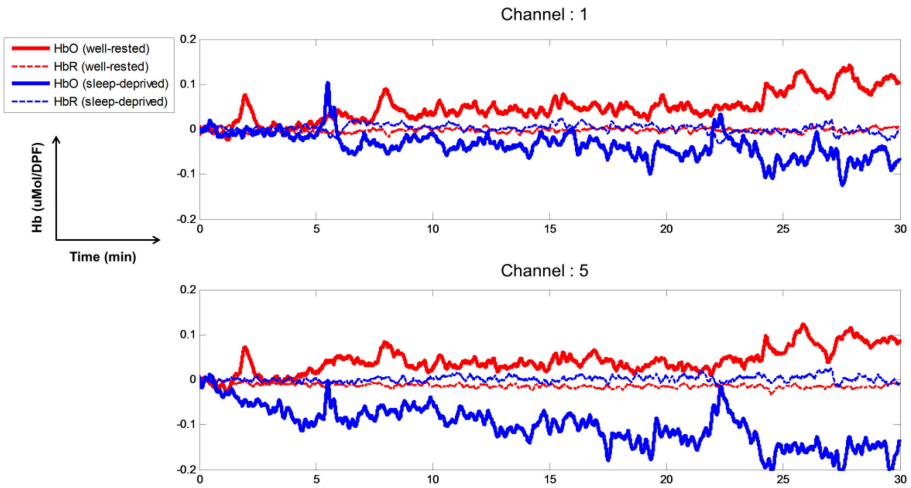


FIGURE 3 | Time course of relative concentration changes of HbO and HbR (channels 1 and 5). Solid, thick red and blue-colored lines indicate HbO for well-rested and sleep-deprived conditions, respectively. Dashed thin lines indicate HbR in the two conditions.

driving under the well-rested condition. On the other hand, concentration changes of HbO decreased slightly (Channel 1) or remained stable (Channel 5) compared to the baseline, and HbR concentration maintained baseline values while driving under the sleep-deprived condition. Less oxygen may be consumed when mentally fatigued, and therefore, brain activity is likely to be suppressed, resulting in less blood flow to the brain.

Reduced Heart Rate and Eye Blinking in the Sleep-Deprived Condition

The mean HRs of all subjects were calculated from ECG signals over the entire 30-min driving period (average of varied HRs), as tabulated in **Table 1** and depicted in **Figure 4A**. As shown in the table and figure, HRs in the sleep-deprived condition were significantly lower than were those in the well-rested condition for all subjects ($p < 0.01$, Wilcoxon signed-rank test). **Figure 4B** shows the HR of subject S5 over time (from initiation to 30 min of driving). HRs in the well-rested condition were higher than those in the sleep-deprived condition; however, the difference in the HR between the two conditions decreased gradually as driving time increased and became quite small at the end of the driving task (∼30 min). We deduced from this time variance in the HR that even a well-rested driver began to feel fatigued after some duration of driving and was considerably fatigued by the end of the task.

Because of the EOG signal, we expected that subjects in the sleep-deprived condition would demonstrate a relatively higher rate of eye blinking than those in the well-rested condition. Instead, we observed (not shown here) that some subjects showed higher rates of eye blinking in the well-rested than in the sleep-deprived condition, although the differences were not statistically significant. From the video data, we found that these subjects closed and opened their eyes frequently to overcome drowsiness, and this action on their part may have affected seriously the rate of eye blinking derived from the EOG data.

Driving Condition Level and Relative Driving Condition Level

We attempted to demonstrate that it may be possible to evaluate neuro-physiological correlates of drivers' mental fatigue using the significant features found in EEG, ECG, and fNIRS data. To that end, we used the three factors extracted from multimodal signals in the previous sections: RPL (ratio of beta to alpha) from EEG, HbO from fNIRS, and the averaged HR from ECG. Each feature was normalized by scaling between 0 and 1 for equal contribution, as formulated in Equation (3). Each value of the EEG, ECG, and fNIRS was distributed well between those values. Significantly abnormal values—greater than 95% (mean ± 2∗σ)—were considered outliers and were shrunk to their maximum or minimum values in the feature set. The summation of all three normalized factors was proposed as the driving condition level (DCL), as depicted in Equation (4). In addition, we estimated the relative difference in DCL between the well-rested and sleep-deprived conditions (rDCL), which represented the degree of the drivers' fatigue compared to that in the well-rested condition, as defined in Equation (5); a higher rDCL indicates greater fatigue.

$$norm(x) = \frac{x - \min(x)}{\max(x) - \min(x)}, \tag{3}$$

$$\begin{aligned} DCL = {} & norm\left(\frac{beta\ RPL}{alpha\ RPL}\right) + norm(HbO) \\ & + norm(HR), \\ & \left(0 \le norm\left(\frac{beta\ RPL}{alpha\ RPL}\right) norm(HbO) norm(HR) \right. \\ & \left. \le 1,\ 0 \le \mathrm{DCL} \le 3\right) \end{aligned} \tag{4}$$

$$rDCL\,(\%) = 100 - \frac{DCL_{sleep-deprived}}{DCL_{well-rested}} * 100, \tag{5}$$

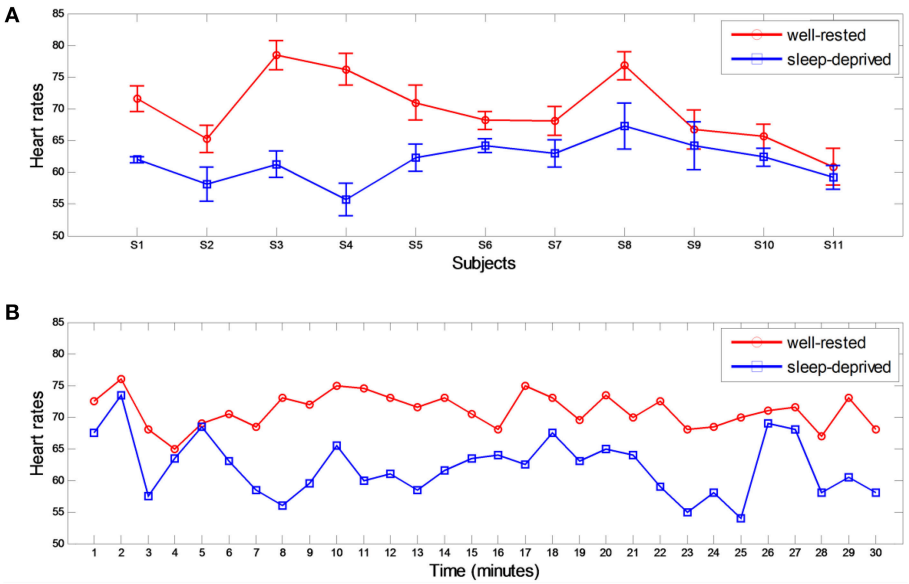


FIGURE 4 | Heart rates from ECG in two different conditions. (A) Averaged heart rates for all subjects in well-rested and sleep-deprived conditions. **(B)** Subject 5's HRs over time. Each HR was estimated every minute.

TABLE 1 | Averaged heart rates (HRs) in well-rested and sleep-deprived conditions.

Subject	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	Mean	*p*-value
Well-rested	71.6 (2.0)	65.3 (2.2)	78.5 (2.3)	76.2 (2.5)	71 (2.8)	68.2 (1.4)	68.1 (2.3)	76.8 (2.2)	66.7 (3.1)	65.7 (1.9)	60.9 (2.9)	69.9 (3.3)	0.0009
Sleep-deprived	62 (0.5)	58.2 (2.7)	61.3 (2.1)	55.7 (2.6)	62.3 (2.2)	64.2 (1.1)	63 (2.2)	67.3 (3.7)	64.2 (3.8)	62.4 (1.4)	59.2 (1.9)	61.8 (3.5)	

Values in parentheses indicate standard deviations for each HR. The Wilcoxon signed-rank test was performed.

Using our proposed definition of DCL (Equation 4), DCL values were estimated in the two conditions for each subject. For all subjects, 30 min of multimodal data were used to estimate the values. DCL ranged from 0 to 3, with a smaller DCL indicating greater fatigue. These DCL values are tabulated in **Table 2**. Subject S8 showed the highest DCL value in the well-rested condition, while subject S4 showed the lowest DCL value in the sleep-deprived condition. We found that the two drivers' conditions (well-rested and sleep-deprived) differed significantly ($p < 0.01$, Wilcoxon signed rank test). In addition, rDCL, which represents the percentage of the fatigue in a drivers' mental condition, was introduced in this work. All rDCLs were consistently greater than 30%, except for those for two of 11 subjects (S7 and S9); thus, when rDCL is tuned more finely with more data, it may be used as a predictor of mental fatigue.

To investigate the drivers' mental fatigue over time, each DCL value (per a minute) was estimated by extracting features of RPL (beta over alpha), HbO, and averaged HR. **Figure 5** shows the DCL values of subject S5, in which the values decreased gradually over time in the well-rested condition, and remained consistent at approximately 1 in the sleep-deprived condition. Sleep-deprived subjects were quite fatigued already at the beginning of the driving task. From the questionnaire, we found a weak correlation ($r^2 = 0.42$) between rDCL values in the sleep-deprived condition and subjects' reported degree of sleepiness (1: rarely sleepy to 5: very sleepy). The average scores for sleepiness over all subjects were 1.4 and 4.1 in the well-rested and sleep-deprived conditions, respectively, while the average hours of sleep reported were 7.36 and 0 h, respectively.

Multimodal Integration to Determine Neuro-Physiological Correlates

In this work, we recorded simultaneous EEG/ECG/EOG and fNIRS signals for multimodal analysis. Multimodal integration is an efficient technique that yields important insights into brain processes (Uludağ and Roebroeck, 2014). Even though the EOG signals did not differ statistically in this work, they were used to eliminate eye movement artifacts in the EEG data. On the other hand, EEG/ECG and fNIRS data yielded clear features that discriminated between the driving conditions, and each feature from the three different modalities differed significantly between the well-rested and sleep-deprived conditions. Based on these features, DCL (summation of these normalized factors) was proposed to determine neuro-physiological correlates of drivers'

TABLE 2 | Driving condition level (DCL) in well-rested and sleep-deprived conditions.

Subject	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	Mean
Well-rested	2.52	2.48	2.31	2.04	2.24	2.75	1.94	2.80	1.94	2.01	2	2.34
Sleep-deprived	1.47	1.17	1.44	1.07	1.11	1.73	1.42	1.88	1.54	1.19	1.21	1.39
rDCL (%)	41.7	52.8	37.7	47.5	50.2	37.1	26.7	32.9	21	40.8	39.5	40.7

Relative DCL (rDCL) indicates percentage of fatigue level by comparison to well-rested condition.

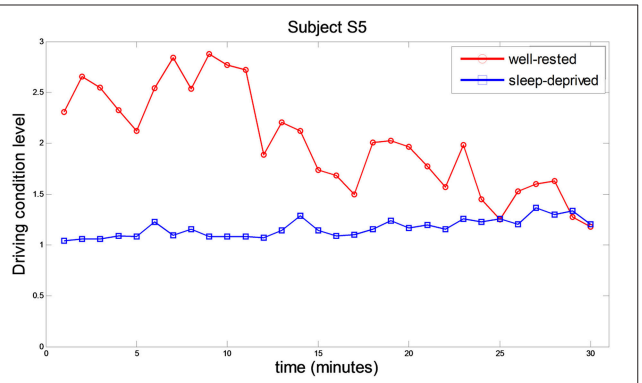


FIGURE 5 | Comparative driving condition level behavior in well-rested and sleep-deprived conditions while subject S5 was driving. Each point represents the averaged DCL value during one minute of driving. Red circles and blue squares represent the DCL in the well-rested and sleep-deprived conditions, respectively.

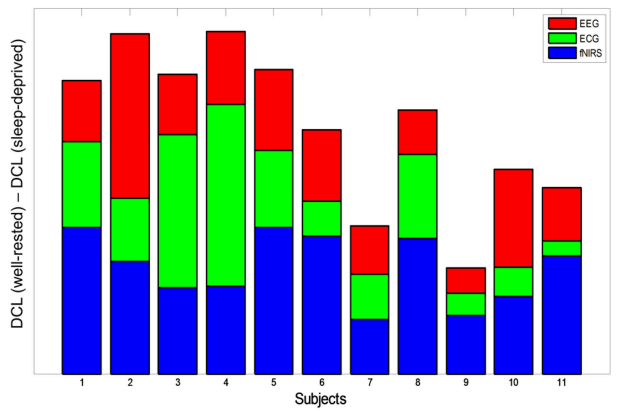


FIGURE 6 | DCL difference (subtraction of sleep-deprived from well-rested) for all subjects. Each color represents modality-specific contribution to DCL difference (red: EEG, green: ECG, blue: fNIRS). Summation of the three colored bars indicates the synergistic effect of multimodal data on classification between the two driving conditions.

mental fatigue in a quantitative manner. As a result, we observed that DCL may offer a reasonable method to discriminate well between the two driving conditions. To illustrate the individual contribution of each modality to the differences in DCL between the two driving conditions, the modality-specific contributions are shown for all subjects in **Figure 6**. Accumulation of the three colored bars indicates DCL differences in the multimodal data (EEG+ECG+fNIRS), which represent the synergistic effect of these data. As shown in this figure, subjects S2 and S4 demonstrated the greatest differences in DCL, while the DCL of subjects S7 and S9 differed the least between the two conditions. Because of the unbiased combination, the averaged contributions of each modality (EEG, ECG, and fNIRS) to the DCL differences were quite similar (0.46, 0.45, and 0.46, respectively).

Comparison of Hybrid Approaches Using EEG/ECG and fNIRS

To investigate the hybrid effect of the classification for the two different driving conditions, we compared various combinations of modalities with respect to the classifiers' outputs. For the combined classifiers in each modality, each classifier's outputs (EEG, ECG, and fNIRS) were regarded as features of the second classifier. Thereafter, the outputs of the second classifier represent the results of the combined classifiers. A flow diagram of this procedure is depicted in **Figure 7**.

Classification accuracies of single modalities (EEG, ECG, fNIRS) and all combinations of modalities (EEG+ECG, EEG+fNIRS, ECG+fNIRS, and EEG+ECG+fNIRS) at the classification level are summarized in **Table 3**. Most of the subjects (8 of 11) achieved improved performance in the combined EEG+ECG+fNIRS, and the average performance of this combination was greater than that of the others. A one-way ANOVA conducted on the seven different approaches indicated that there was a considerably significant difference [$F_{(6, 70)} = 4.38$, $p = 0.0008 < 0.10$]. Further, the EEG+ECG+fNIRS combinations differed significantly from the others ($p < 0.05$, Wilcoxon signed-rank test). Each level of significance is marked with asterisks in **Table 3**. Notably, subject S6 showed the greatest improvement (~30.5%) in the EEG+ECG+fNIRS combination compared to EEG only.

DISCUSSION

EEG Spectral Changes and Driving Conditions

To date, most researchers have investigated driving fatigue using EEG changes, which are promising indicators of this phenomenon (Lal and Craig, 2001), and EEG has the advantages of being portable, noninvasive, inexpensive, and safe to measure during driving. With EEG recording, Lal and Craig (2002) found substantial increases in delta, theta, and alpha activity in the transition to fatigue, which was

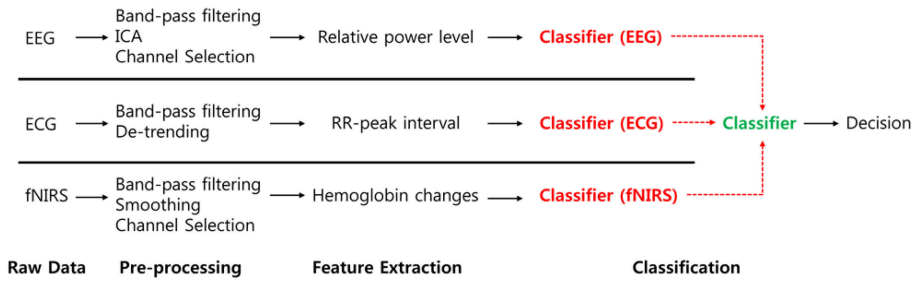


FIGURE 7 | Flow diagram for classifier combination. Each classifier's output is regarded to second classifier's input.

TABLE 3 | Classification accuracies in well-rested and sleep-deprived conditions.

Subject	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	Mean
**EEG	61.3	52.1	58.9	54.4	70.8	55.2	68.9	64.3	61.0	52.2	58.1	59.7
**ECG	79.4	72.0	52.8	76.3	60.5	52.3	63.9	72.2	61.2	58.8	59.7	64.5
*fNIRS	68.4	**84.5**	55.6	65.9	76.8	80.8	**72.9**	61.5	52.6	57.7	58.3	66.8
**EEG+ECG	76.2	60.0	**74.3**	68.1	72.9	68.9	64.3	73.5	71.4	60.4	69.5	69.0
**EEG+fNIRS	83.3	63.3	66.7	65.5	72.7	63.3	63.3	70.0	73.3	60.3	70.0	68.3
**ECG+fNIRS	82.8	64.8	63.3	73.3	73.3	73.3	60.0	76.7	63.3	56.7	73.3	69.2
EEG+ECG+fNIRS	**84.5**	73.7	71.8	**77.9**	**78.1**	**85.7**	67.0	**83.8**	**74.6**	**62.4**	**75.2**	**75.9**

*The highest accuracies among seven approaches are displayed in bold. Asterisks indicate the level of signficance with the EEG+ECG+fNIRS combination (**$p < 0.01$: *$p < 0.05$).*

consistent with existing findings described in a review paper (Sahayadhas et al., 2012). Alpha activity is believed to be the most prominent indicator of driver fatigue. With this reasoning, Simon et al. (2011) verified alpha spindle activity based on a short-time Fourier transformation in real traffic conditions. Statistical analysis of these actual driving data revealed significant increases for all alpha spindle parameters, such as rate, duration, amplitude, and power, between the awake and drowsy states during 20 min of driving. Similarly, in the EEG recordings in this work, we observed a significant increase in alpha based on RPL. To reduce session and subject variability, a normalized alpha RPL was introduced and a significant alpha RPL difference was found in the centro-parietal region.

It is known that attention is also correlated with alpha power suppression. In our experiment, visual attention may be expected to be related closely to a simulated driving task. Such visual attention-related alpha power suppression may be observed normally in the occipital region, as reported in previous studies (Worden et al., 2000; Sauseng et al., 2005; Rihs et al., 2007). However, in our study, notable alpha suppression was observed in the centro-parietal region alone, which is consistent with results in previous studies of fatigue (Lal and Craig, 2001, 2002; Simon et al., 2011; Sahayadhas et al., 2012). For this reason, it is clear that a reduction in power in the alpha band was correlated with fatigue in this experiment. We observed beta RPL changes in the sleep-deprived condition, and beta power may be an additional indicator of mental fatigue. Tanaka et al. (2012) found that beta power densities decreased significantly after tiring cognitive tasks. They calculated EEG power spectra in each band and showed that beta waves decreased significantly in the fronto-central region with increased driving times. It also has been reported that beta rhythm is associated closely with increased alertness and arousal (Okogbaa et al., 1994), which is likely to be applicable to driving situations (Yeo et al., 2009; Yang et al., 2010; Zhao et al., 2012). In this work, we inferred that the lack of arousal and alertness caused by mental fatigue and sleep deprivation may result in a decreased beta rhythm during simulated driving.

Until now, most studies related to the detection of mental fatigue during driving have been experimental, and driving conditions have been divided according to the elapsed duration of driving. For example, data from the first 10 min have been considered to be the normal condition, while those from the last 10 min were specified as the fatigued condition (Li et al., 2009; Simon et al., 2011). Such an approach may not be appropriate, however, as some people may not develop fatigue even after 2–3 h of driving, especially professional drivers. Therefore, in order to discriminate between high- and low-risk conditions explicitly, each subject was both at high and low risk of fatigue before the driving tests, depending on how many hours they slept at night. Because sleep deprivation is well known to affect decision-making, attention, vigilance, human performance, and mental fatigue (Åkerstedt et al., 2004; Alhola and Polo-Kantola, 2007), it is appropriate to refer to sleep deprivation as analogous to fatigue. In addition, in our driving simulator, the steering wheel vibrated whenever the car crashed into a barrier to prevent drivers from actually falling asleep. Preventing the subjects from falling

asleep may have suppressed activation in the delta and theta bands because these waves are associated closely with deep sleep (Maquet et al., 1997) and REM sleep (Jouvet, 1969), respectively.

Observations of Fatigue in the Well-Rested Condition

Before the experiment, all subjects with well-rested condition were instructed to sleep over 7 h to ensure that they were mentally alert and physically refreshed. However, several subjects experienced fatigue in the driving task nonetheless, due to various internal or external environmental factors, even though they reported that they had slept well the previous night. Subjects S7 and S9 had the lowest DCL values in the well-rested condition, as shown in **Table 2** and **Figure 6**. According to their questionnaires, these two subjects recorded a score of 2 in the sleepiness section (1: rarely sleepy to 5: very sleepy) prior to the experiment, yet they often nodded off during driving. Their behavior was recorded on the HD-Webcam, and showed clearly that they were drowsy. Furthermore, they reported scores of 4 and 5, respectively, on the sleepiness scale after the experiment. **Figure 8** represents the DCLs for these two subjects over time. As shown, their DCLs in the well-rested condition were similar to those in the sleep-deprived condition. The DCLs fluctuated in the well-rested condition, as shown by repeated increases and decreases. Moreover, these two subjects achieved low classification accuracies in a single modality, as shown in **Table 3**, although their accuracy improved when measured with mixed features. Clearly, we believe that these two subjects were likely to have been fatigued despite their assignment to the well-rested condition. In this work, we used the HD-webcam only to monitor the subjects, and did not measure or analyze behavioral data. Analyzing subjects' behavior in real-time, such as head or eye movements, may offer supporting evidence that some subjects slept well but experienced mental fatigue nonetheless. We will collect such behavioral data in the online mental fatigue monitoring system, which is currently under investigation.

Temporal Mismatch between EEG and fNIRS

Recently, many studies have tried to combine EEG and fNIRS to improve classification accuracy or increase the degrees of freedom in BCI systems (Fazli et al., 2012; Khan et al., 2014; Putze et al., 2014; Koo et al., 2015; Yin et al., 2015). However, the fNIRS system measures hemodynamic change, which is a delayed response compared to neuronal electrical activity, and it also has a relatively low temporal resolution (<10 Hz), both of which are critical drawbacks in fNIRS measurements. Because of their low temporal resolution, it is sometimes difficult to combine fNIRS data with other brain imaging data. Nevertheless, one of the most significant merits of the fNIRS system is its ability to measure oxygen consumption related to blood flow in the brain, similar to that in functional magnetic resonance imaging (fMRI), and fNIRS has been nicknamed the portable fMRI for this reason.

Considering the advantages and drawbacks of the fNIRS system, we calculated the features during each 1-min period throughout the dataset in this work. Each minute in the 30 min of data yielded a DCL value, which was used to discriminate between the well-rested and sleep-deprived conditions. In addition, oxygen consumption in the prefrontal cortex may represent cognitive workload or fatigue (Ayaz et al., 2012, 2013; Harrison et al., 2014; McKendrick et al., 2014). Therefore, it is likely that fNIRS may be a significant indicator of mental fatigue, and we are sure that employing multimodal data is quite useful in monitoring mental fatigue. Also, the prefrontal cortex is related closely to mental workload (Mandrick et al., 2013a)

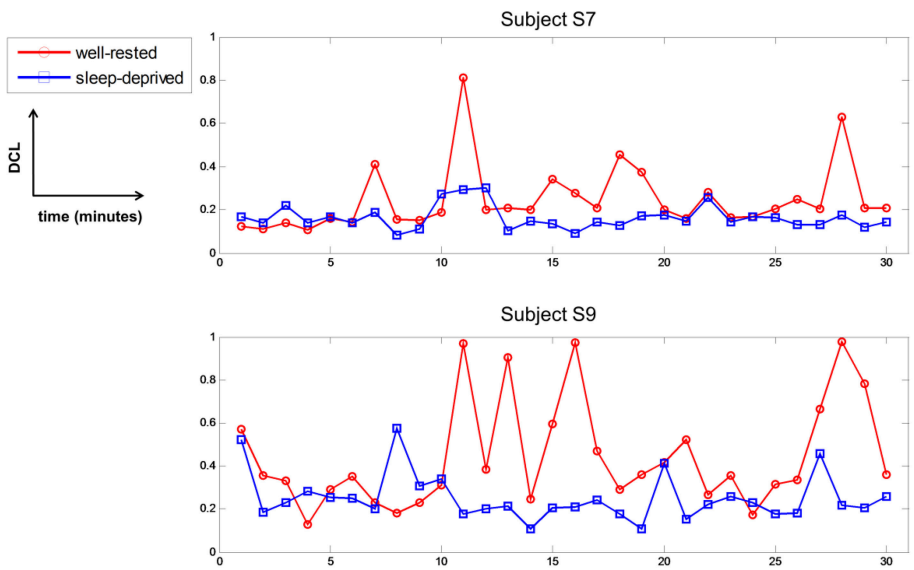


FIGURE 8 | Driving condition level of subject S7 and S9 in two conditions per 1 min over time. Red-circle and blue-square represent the driving condition level in well-rested and sleep-deprived conditions, respectively.

and the performance of cognitive tasks (Mandrick et al., 2013b). For these reasons, we introduced the prefrontal fNIRS in this work. However, whole head fNIRS would be beneficial and will be considered in our future work.

Limitations and Future Work

We proposed here an indicator of drivers' mental fatigue (Equation 4) that was able to discriminate the drivers' mental conditions well. To include equal contributions of EEG, ECG, and fNIRS features in the indicator, we normalized and summed all three. Further, all three were weighted equally to calculate the indicator, although this might not be an optimal method of quantification. Therefore, we calculated a new indicator using a weighting factor of $[DCL = a * norm\left(\frac{beta\ RPL}{alpha\ RPL}\right) + b * norm(HbO) + c * norm(HR)]$. Each weighting factor was defined between 0 and 1 with increments of 0.10; thus, we were able to search for the highest DCL difference in the two conditions (subtraction of sleep-deprived from well-rested). As a result, they had values comparable to those with equal weights and we observed no significant improvement. Even though we were able to apply elegant optimization methods, a somewhat better indicator may be achieved.

The primary reason to monitor drivers' mental fatigue is to prevent car accidents by providing drivers with a rapid and reliable alarm. To achieve this purpose, both making predictions before driving and monitoring a driver's condition in real-time are potential approaches. Our proposed rDCL (Equation 5) can predict a driver's fatigue prior to driving if training data can be obtained when the driver is in an alert state. Compared with DCL in the well-rested condition, the driver's condition prior to driving may be pre-checked by estimating the DCL. Although we used the entire dataset obtained during 30 min in the well-rested condition in this work, data of a shorter duration could be used for baseline. We are investigating the minimum duration needed to predict drivers' fatigue now. Another approach to accident prevention is to monitor drivers' fatigue in real-time.

We performed only an offline analysis in this work, and normalized DCL values were estimated each minute. For online monitoring, however, normalization of each modality's features is quite difficult to implement based on current methods available. One possible approach to solve the normalization issue is to record resting data before driving and use them as baseline data. Alternatively, an adaptive normalization method that updates feature values in real-time is a candidate. We are investigating the most appropriate normalization method for a subsequent online mental fatigue monitoring system.

In this work, we attempted to analyze multimodal data with simultaneous recordings of EEG/ECG/EOG and fNIRS. One of the advantages of multimodal signal integration is that each imaging method provides a physiologically and physically filtered view of one or more brain processes of interest. Thus far, the EEG-fMRI combination has been investigated widely, especially in epilepsy research, to help localize specific regions (Rosenkranz and Lemieux, 2010) by improving spatial and temporal resolution. Now, the EEG-fNIRS combination may be an alternative imaging method with merits that include low cost and simple implementation.

In this study, we custom-built an fNIRS that was already validated in previous study (Kim et al., 2015), and thus enabled us to design the experiment well. Although, a lengthy preparation time was required to attach the detectors and emitters, and test the quality of the light intensity for fNIRS measurement, this EEG-fNIRS integration may be quite beneficial in developing a monitoring system, as reported in the existing literature (Fazli et al., 2012; Wallois et al., 2012; Khan et al., 2014; Morioka et al., 2014; Putze et al., 2014; Koo et al., 2015; Yin et al., 2015). Another concern in analyzing multimodal data is how their differences (physical values, temporal resolution) are considered in an integrated frame. An in-depth investigation is needed to enhance the synergistic effect of multimodal data recording. Similarly, in this study, we were unable to guarantee that the EEG, ECG, and fNIRS, or their combined features, are related linearly to the fatigue level, although in the multimodal results, each modality influenced the fatigue level to some degree, as shown in **Figure 8**.

CONCLUSIONS

The purpose of this study was to use simultaneous EEG/ECG/EOG and fNIRS recordings to determine neuro-physiological correlates that can be used to discriminate sleep deprivation-induced mental fatigue in drivers by comparison to those who are well-rested. To achieve our purpose, we introduced two driving conditions (well-rested and sleep-deprived), and were able to extract significant features from their EEG, ECG, and fNIRS data. However, no significant feature was found in the EOG due to high variability in the subjects' data. The features observed allowed us to determine the mental condition of each driver, and also yielded good discriminative results between two driving conditions. Further, we investigated the synergistic effects of multimodal data to compare the various combinations at the classification level with a single modality. In conclusion, our proposed combined approach of simultaneous EEG/ECG and fNIRS data may be a promising tool with which to monitor drivers' mental fatigue.

AUTHOR CONTRIBUTIONS

SA, TN, HJ, JGK, and SCJ designed the experimental paradigm, and TN and JGK custom-built the fNIRS system. Data collection was conducted by SA, TN, and HJ, and analysis was performed by SA; the entire manuscript was written by SA and SCJ.

ACKNOWLEDGMENTS

This work was supported by Hyundai Next Generation Vehicle (NGV), the National Research Foundation of Korea (#2013R1A1A2009029), and the GIST Research Institute (GRI) in 2016.

REFERENCES

Ahn, M., Ahn, S., Hong, J. H., Cho, H., Kim, K., Kim, B. S., et al. (2013a). Gamma band activity associated with BCI performance: simultaneous MEG/EEG study. *Front. Hum. Neurosci.* 7:848. doi: 10.3389/fnhum.2013.00848

Ahn, M., Cho, H., Ahn, S., and Jun, S. C. (2013b). High theta and low alpha powers may be indicative of BCI-Illiteracy in motor imagery. *PLoS ONE* 8:e80886. doi: 10.1371/journal.pone.0080886

Ahn, M., and Jun, S. C. (2015). Performance variation in motor imagery brain–computer interface: a brief review. *J. Neurosci. Methods* 243, 103–110. doi: 10.1016/j.jneumeth.2015.01.033

Ahn, S., Ahn, M., Cho, H., and Jun, S. C. (2014). Achieving a hybrid brain-computer interface with tactile selective attention and motor imagery. *J. Neural Eng.* 11:066004. doi: 10.1088/1741-2560/11/6/066004

Alhola, P., and Polo-Kantola, P. (2007). Sleep deprivation: impact on cognitive performance. *Neuropsychiatr. Dis. Treat.* 3, 553–567.

Ayaz, H., Onaral, B., Izzetoglu, K., Shewokis, P. A., McKendrick, R., and Parasuraman, R. (2013). Continuous monitoring of brain dynamics with functional near infrared spectroscopy as a tool for neuroergonomic research: empirical examples and a technological development. *Front. Hum. Neurosci.* 7:871. doi: 10.3389/fnhum.2013.00871

Ayaz, H., Shewokis, P. A., Bunce, S., Izzetoglu, K., Willems, B., and Onaral, B. (2012). Optical brain monitoring for operator training and mental workload assessment. *Neuroimage* 59, 36–47. doi: 10.1016/j.neuroimage.2011.06.023

Bell, A. J., and Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Comput.* 7, 1129–1159. doi: 10.1162/neco.1995.7.6.1129

Bergasa, L. M., Nuevo, J., Sotelo, M. A., Barea, R., and Lopez, M. E. (2006). Real-time system for monitoring driver vigilance. *IEEE Trans. Intell. Transp. Syst.* 7, 63–77. doi: 10.1109/TITS.2006.869598

Cho, H., Ahn, M., Kim, K., and Jun, S. C. (2015). Increasing session-to-session transfer in a brain–computer interface with on-site background noise acquisition. *J. Neural Eng.* 12:066009. doi: 10.1088/1741-2560/12/6/066009

Connor, J., Norton, R., Ameratunga, S., Robinson, E., Civil, I., Dunn, R., et al. (2002). Driver sleepiness and risk of serious injury to car occupants: population based case control study. *BMJ* 324:1125. doi: 10.1136/bmj.324.7346.1125

Cope, M., Delpy, D. T., Reynolds, E. O., Wray, S., Wyatt, J., and van der Zee, P. (1988). Methods of quantitating cerebral near infrared spectroscopy data. *Adv. Exp. Med. Biol.* 222, 183–189.

D'Orazio, T., Leo, M., Guaragnella, C., and Distante, A. (2007). A visual approach for driver inattention detection. *Pattern Recognit.* 40, 2341–2355. doi: 10.1016/j.patcog.2007.01.018

de Chazal, P., O'Dwyer, M., and Reilly, R. B. (2004). Automatic classification of heartbeats using ECG morphology and heartbeat interval features. *IEEE Trans. Biomed. Eng.* 51, 1196–1206. doi: 10.1109/TBME.2004.827359

Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21. doi: 10.1016/j.jneumeth.2003.10.009

Doi, H., Nishitani, S., and Shinohara, K. (2013). NIRS as a tool for assaying emotional function in the prefrontal cortex. *Front. Hum. Neurosci.* 7:770. doi: 10.3389/fnhum.2013.00770

Durmer, J. S., and Dinges, D. F. (2005). Neurocognitive consequences of sleep deprivation. *Semin. Neurol.* 25, 117–129. doi: 10.1055/s-2005-867080

Fazli, S., Mehnert, J., Steinbrink, J., Curio, G., Villringer, A., Müller, K.-R., et al. (2012). Enhanced performance by a hybrid NIRS-EEG brain computer interface. *Neuroimage* 59, 519–529. doi: 10.1016/j.neuroimage.2011.07.084

Gateau, T., Durantin, G., Lancelot, F., Scannella, S., and Dehais, F. (2015). Real-time state estimation in a flight simulator using fNIRS. *PLoS ONE* 10:e0121279. doi: 10.1371/journal.pone.0121279

Grandjean, E. (1979). Fatigue in industry. *Br. J. Ind. Med.* 36, 175–186.

Harrison, J., Izzetoglu, K., Ayaz, H., Willems, B., Hah, S., Ahlstrom, U., et al. (2014). Cognitive workload and learning assessment during the implementation of a next-generation air traffic control technology using functional near-infrared spectroscopy. *IEEE Trans. Hum. Mach. Syst.* 44, 429–440. doi: 10.1109/THMS.2014.2319822

Herman, J., Kafoa, B., Wainiqolo, I., Robinson, E., McCaig, E., Connor, J., et al. (2014). Driver sleepiness and risk of motor vehicle crash injuries: a population-based case control study in Fiji (TRIP 12). *Injury* 45, 586–591. doi: 10.1016/j.injury.2013.06.007

Homma, S., Fukunaga, T., and Kagaya, A. (1996). Influence of adipose tissue thickness on near infrared spectroscopic signal in the measurement of human muscle. *J. Biomed. Opt.* 1, 418–424. doi: 10.1117/12.252417

Horne, J., and Reyner, L. (1999). Vehicle accidents related to sleep: a review. *Occup. Environ. Med.* 56, 289–294.

Hu, S., and Zheng, G. (2009). Driver drowsiness detection with eyelid related parameters by support vector machine. *Expert Syst. Appl.* 36, 7651–7658. doi: 10.1016/j.eswa.2008.09.030

Ji, Q., Zhu, Z., and Lan, P. (2004). Real-time nonintrusive monitoring and prediction of driver fatigue. *IEEE Trans. Veh. Technol.* 53, 1052–1068. doi: 10.1109/TVT.2004.830974

Jouvet, M. (1969). Biogenic amines and the states of sleep. *Science* 163, 32–41. doi: 10.1126/science.163.3862.32

Kar, S., Bhagat, M., and Routray, A. (2010). EEG signal analysis for the assessment and quantification of driver's fatigue. *Transp. Res. Part F Traffic Psychol. Behav.* 13, 297–306. doi: 10.1016/j.trf.2010.06.006

Khan, M. J., Hong, M. J., and Hong, K.-S. (2014). Decoding of four movement directions using hybrid NIRS-EEG brain-computer interface. *Front. Hum. Neurosci.* 8:244. doi: 10.3389/fnhum.2014.00244

Khushaba, R. N., Kodagoda, S., Lal, S., and Dissanayake, G. (2011). Driver drowsiness classification using fuzzy wavelet-packet-based feature-extraction algorithm. *IEEE Trans. Biomed. Eng.* 58, 121–131. doi: 10.1109/TBME.2010.2077291

Kim, E., Lee, S., Phillips, Z., and Kim, J. G. (2015). A discrepancy of penile hemodynamics during visual sexual stimulation observed by near-infrared spectroscopy. *BMC Urol.* 15, 11. doi: 10.1186/s12894-015-0005-x

Kocsis, L., Herman, P., and Eke, A. (2006). The modified Beer-Lambert law revisited. *Phys. Med. Biol.* 51, N91–N98. doi: 10.1088/0031-9155/51/5/N02

Koo, B., Lee, H.-G., Nam, Y., Kang, H., Koh, C. S., Shin, H.-C., et al. (2015). A hybrid NIRS-EEG system for self-paced brain computer interface with online motor imagery. *J. Neurosci. Methods* 244, 26–32. doi: 10.1016/j.jneumeth.2014.04.016

Kripke, D. F., Garfinkel, L., Wingard, D. L., Klauber, M. R., and Marler, M. R. (2002). Mortality associated with sleep duration and insomnia. *Arch. Gen. Psychiatry* 59, 131–136. doi: 10.1001/archpsyc.59.2.131

Lal, S. K. L., and Craig, A. (2001). A critical review of the psychophysiology of driver fatigue. *Biol. Psychol.* 55, 173–194. doi: 10.1016/S0301-0511(00)00085-5

Lal, S. K. L., and Craig, A. (2002). Driver fatigue: electroencephalography and psychological assessment. *Psychophysiology* 39, 313–321. doi: 10.1017/S0048577201393095

Lal, S. K. L., Craig, A., Boord, P., Kirkup, L., and Nguyen, H. (2003). Development of an algorithm for an EEG-based driver fatigue countermeasure. *J. Safety Res.* 34, 321–328. doi: 10.1016/S0022-4375(03)00027-6

Li, G., and Chung, W.-Y. (2013). Detection of driver drowsiness using wavelet analysis of heart rate variability and a support vector machine classifier. *Sensors* 13, 16494–16511. doi: 10.3390/s131216494

Li, Z., Zhang, M., Zhang, X., Dai, S., Yu, X., and Wang, Y. (2009). Assessment of cerebral oxygenation during prolonged simulated driving using near infrared spectroscopy: its implications for fatigue development. *Eur. J. Appl. Physiol.* 107, 281–287. doi: 10.1007/s00421-009-1122-6

Lin, C.-T., Huang, K.-C., Chao, C.-F., Chen, J.-A., Chiu, T.-W., Ko, L.-W., et al. (2010). Tonic and phasic EEG and behavioral changes induced by arousing feedback. *Neuroimage* 52, 633–642. doi: 10.1016/j.neuroimage.2010.04.250

Lin, C.-T., Huang, K.-C., Chuang, C.-H., Ko, L.-W., and Jung, T.-P. (2013). Can arousing feedback rectify lapses in driving? Prediction from EEG power spectra. *J. Neural Eng.* 10:056024. doi: 10.1088/1741-2560/10/5/056024

Lin, C.-T., Wu, R.-C., Liang, S.-F., Chao, W.-H., Chen, Y.-J., and Jung, T.-P. (2005). EEG-based drowsiness estimation for safety driving using independent component analysis. *IEEE Trans. Circuits Syst. Regul. Pap.* 52, 2726–2738. doi: 10.1109/TCSI.2005.857555

Mandrick, K., Derosiere, G., Dray, G., Coulon, D., Micallef, J.-P., and Perrey, S. (2013a). Prefrontal cortex activity during motor tasks with additional mental load requiring attentional demand: a near-infrared spectroscopy study. *Neurosci. Res.* 76, 156–162. doi: 10.1016/j.neures.2013.04.006

Mandrick, K., Derosiere, G., Dray, G., Coulon, D., Micallef, J.-P., and Perrey, S. (2013b). Utilizing slope method as an alternative data analysis for functional

near-infrared spectroscopy-derived cerebral hemodynamic responses. *Int. J. Ind. Ergon.* 43, 335–341. doi: 10.1016/j.ergon.2013.05.003

Maquet, P., Degueldre, C., Delfiore, G., Aerts, J., Péters, J.-M., Luxen, A., et al. (1997). Functional neuroanatomy of human slow wave sleep. *J. Neurosci.* 17, 2807–2812.

Matcher, S. J., Elwell, C. E., Cooper, C. E., Cope, M., and Delpy, D. T. (1995). Performance comparison of several published tissue near-infrared spectroscopy algorithms. *Anal. Biochem.* 227, 54–68. doi: 10.1006/abio.1995.1252

McKendrick, R., Ayaz, H., Olmstead, R., and Parasuraman, R. (2014). Enhancing dual-task performance with verbal and spatial working memory training: continuous monitoring of cerebral hemodynamics with NIRS. *Neuroimage* 85 (Pt 3), 1014–1026. doi: 10.1016/j.neuroimage.2013.05.103

Morioka, H., Kanemura, A., Morimoto, S., Yoshioka, T., Oba, S., Kawanabe, M., et al. (2014). Decoding spatial attention by using cortical currents estimated from electroencephalography with near-infrared spectroscopy prior information. *Neuroimage* 90, 128–139. doi: 10.1016/j.neuroimage.2013.12.035

Okogbaa, O. G., Shell, R. L., and Filipusic, D. (1994). On the investigation of the neurophysiological correlates of knowledge worker mental fatigue using the EEG signal. *Appl. Ergon.* 25, 355–365.

Parasuraman, R. (2003). Neuroergonomics: research and practice. *Theor. Issues Ergon. Sci.* 4, 5–20. doi: 10.1080/14639220210199753

Parasuraman, R., and Rizzo, M. (eds.). (2008). *Neuroergonomics: The Brain at Work.* New York, NY: Oxford University Press.

Parasuraman, R., and Wilson, G. F. (2008). Putting the brain to work: neuroergonomics past, present, and future. *Hum. Factors* 50, 468–474. doi: 10.1518/001872008X288349

Patel, M., Lal, S. K. L., Kavanagh, D., and Rossiter, P. (2011). Applying neural network analysis on heart rate variability data to assess driver fatigue. *Expert Syst. Appl.* 38, 7235–7242. doi: 10.1016/j.eswa.2010.12.028

Pettersson, K., Jagadeesan, S., Lukander, K., Henelius, A., Hæggström, E., and Müller, K. (2013). Algorithm for automatic analysis of electro-oculographic data. *Biomed. Eng. OnLine* 12:110. doi: 10.1186/1475-925X-12-110

Pfurtscheller, G., Allison, B. Z., Brunner, C., Bauernfeind, G., Solis-Escalante, T., Scherer, R., et al. (2010). The hybrid BCI. *Front. Neurosci.* 4:42. doi: 10.3389/fnpro.2010.00003

Putze, F., Hesslinger, S., Tse, C.-Y., Huang, Y., Herff, C., Guan, C., et al. (2014). Hybrid fNIRS-EEG based classification of auditory and visual perception processes. *Neuroprosthetics* 8:373. doi: 10.3389/fnins.2014.00373

Rihs, T. A., Michel, C. M., and Thut, G. (2007). Mechanisms of selective inhibition in visual spatial attention are indexed by α-band EEG synchronization. *Eur. J. Neurosci.* 25, 603–610. doi: 10.1111/j.1460-9568.2007.05278.x

Rosenkranz, K., and Lemieux, L. (2010). Present and future of simultaneous EEG-fMRI. *Magn. Reson. Mater. Phys. Biol. Med.* 23, 309–316. doi: 10.1007/s10334-009-0196-9

Sahayadhas, A., Sundaraj, K., and Murugappan, M. (2012). Detecting driver drowsiness based on sensors: a review. *Sensors* 12, 16937–16953. doi: 10.3390/s121216937

Sato, H., Yahata, N., Funane, T., Takizawa, R., Katura, T., Atsumori, H., et al. (2013). A NIRS-fMRI investigation of prefrontal cortex activity during a working memory task. *Neuroimage* 83, 158–173. doi: 10.1016/j.neuroimage.2013.06.043

Sauseng, P., Klimesch, W., Stadler, W., Schabus, M., Doppelmayr, M., Hanslmayr, S., et al. (2005). A shift of visual spatial attention is selectively associated with human EEG alpha activity. *Eur. J. Neurosci.* 22, 2917–2926. doi: 10.1111/j.1460-9568.2005.04482.x

Schalk, G., McFarland, D. J., Hinterberger, T., Birbaumer, N., and Wolpaw, J. R. (2004). BCI2000: a general-purpose brain-computer interface (BCI) system. *IEEE Trans. Biomed. Eng.* 51, 1034–1043. doi: 10.1109/TBME.2004.827072

Schleicher, R., Galley, N., Briest, S., and Galley, L. (2008). Blinks and saccades as indicators of fatigue in sleepiness warnings: looking tired? *Ergonomics* 51, 982–1010. doi: 10.1080/00140130701817062

Shen, K.-Q., Li, X.-P., Ong, C.-J., Shao, S.-Y., and Wilder-Smith, E. P. V. (2008). EEG-based mental fatigue measurement using multi-class support vector machines with confidence estimate. *Clin. Neurophysiol.* 119, 1524–1533. doi: 10.1016/j.clinph.2008.03.012

Simon, M., Schmidt, E. A., Kincses, W. E., Fritzsche, M., Bruns, A., Aufmuth, C., et al. (2011). EEG alpha spindle measures as indicators of driver fatigue under real traffic conditions. *Clin Neurophysiol* 122, 1168–1178. doi: 10.1016/j.clinph.2010.10.044

Tanaka, M., Shigihara, Y., Ishii, A., Funakura, M., Kanai, E., and Watanabe, Y. (2012). Effect of mental fatigue on the central nervous system: an electroencephalography study. *Behav. Brain Funct.* 8:48. doi: 10.1186/1744-9081-8-48

Åkerstedt, T., Knutsson, A., Westerholm, P., Theorell, T., Alfredsson, L., and Kecklund, G. (2004). Mental fatigue, work and sleep. *J. Psychosom. Res.* 57, 427–433. doi: 10.1016/j.jpsychores.2003.12.001

Uludağ, K., and Roebroeck, A. (2014). General overview on the merits of multimodal neuroimaging data fusion. *Neuroimage* 102, (Pt 1), 3–10. doi: 10.1016/j.neuroimage.2014.05.018

Wallois, F., Mahmoudzadeh, M., Patil, A., and Grebe, R. (2012). Usefulness of simultaneous EEG-NIRS recording in language studies. *Brain Lang.* 121, 110–123. doi: 10.1016/j.bandl.2011.03.010

Wang, Y.-T., Huang, K.-C., Wei, C.-S., Huang, T.-Y., Ko, L.-W., Lin, C.-T., et al. (2014). Developing an EEG-based on-line closed-loop lapse detection and mitigation system. *Neuroprosthetics* 8:321. doi: 10.3389/fnins.2014.00321

Worden, M. S., Foxe, J. J., Wang, N., and Simpson, G. V. (2000). Anticipatory biasing of visuospatial attention indexed by retinotopically specific α-bank electroencephalography increases over occipital cortex. *J. Neurosci.* 20:RC63.

Yang, G., Lin, Y., and Bhattacharya, P. (2010). A driver fatigue recognition model based on information fusion and dynamic Bayesian network. *Inf. Sci.* 180, 1942–1954. doi: 10.1016/j.ins.2010.01.011

Yeo, M. V. M., Li, X., Shen, K., and Wilder-Smith, E. P. V. (2009). Can SVM be used for automatic EEG detection of drowsiness during car driving? *Saf. Sci.* 47, 115–124. doi: 10.1016/j.ssci.2008.01.007

Yin, X., Xu, B., Jiang, C., Fu, Y., Wang, Z., Li, H., et al. (2015). A hybrid BCI based on EEG and fNIRS signals improves the performance of decoding motor imagery of both force and speed of hand clenching. *J. Neural Eng.* 12:036004. doi: 10.1088/1741-2560/12/3/036004

Yoshino, K., Oka, N., Yamamoto, K., Takahashi, H., and Kato, T. (2013). Functional brain imaging using near-infrared spectroscopy during actual driving on an expressway. *Front. Hum. Neurosci.* 7:882. doi: 10.3389/fnhum.2013.00882

Zhao, C., Zhao, M., Liu, J., and Zheng, C. (2012). Electroencephalogram and electrocardiograph assessment of mental fatigue in a driving simulator. *Accid. Anal. Prev.* 45, 83–90. doi: 10.1016/j.aap.2011.11.019

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Ahn, Nguyen, Jang, Kim and Jun. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 01 March 2016
doi: 10.3389/fnhum.2016.00073

Steering Demands Diminish the Early-P3, Late-P3 and RON Components of the Event-Related Potential of Task-Irrelevant Environmental Sounds

Menja Scheer[1], *Heinrich H. Bülthoff*[1,2]* *and Lewis L. Chuang*[1]*

[1] *Department of Perception, Cognition and Action, Max Planck Institute for Biological Cybernetics, Tübingen, Germany,*
[2] *Department of Cognitive and Brain Engineering, Korea University, Seoul, South Korea*

The current study investigates the demands that steering places on mental resources. Instead of a conventional dual-task paradigm, participants of this study were only required to perform a steering task while task-irrelevant auditory distractor probes (environmental sounds and beep tones) were intermittently presented. The event-related potentials (ERPs), which were generated by these probes, were analyzed for their sensitivity to the steering task's demands. The steering task required participants to counteract unpredictable roll disturbances and difficulty was manipulated either by adjusting the bandwidth of the roll disturbance or by varying the complexity of the control dynamics. A mass univariate analysis revealed that steering selectively diminishes the amplitudes of early P3, late P3, and the re-orientation negativity (RON) to task-irrelevant environmental sounds but not to beep tones. Our findings are in line with a three-stage distraction model, which interprets these ERPs to reflect the post-sensory detection of the task-irrelevant stimulus, engagement, and re-orientation back to the steering task. This interpretation is consistent with our manipulations for steering difficulty. More participants showed diminished amplitudes for these ERPs in the "hard" steering condition relative to the "easy" condition. To sum up, the current work identifies the spatiotemporal ERP components of task-irrelevant auditory probes that are sensitive to steering demands on mental resources. This provides a non-intrusive method for evaluating mental workload in novel steering environments.

Keywords: steering, mental workload, distraction, MMN, early P3, late P3, RON

OPEN ACCESS

Edited by:
Hasan Ayaz,
Drexel University, USA

Reviewed by:
Kimmo Alho,
University of Helsinki, Finland
Erich Schröger,
University of Leipzig, Germany

***Correspondence:**
Heinrich H. Bülthoff
heinrich.buelthoff@tuebingen.mpg.de;
Lewis L. Chuang
lewis.chuang@tuebingen.mpg.de

Received: *22 December 2015*
Accepted: *15 February 2016*
Published: *01 March 2016*

Citation:
Scheer M, Bülthoff HH and Chuang LL (2016) Steering Demands Diminish the Early-P3, Late-P3 and RON Components of the Event-Related Potential of Task-Irrelevant Environmental Sounds.
Front. Hum. Neurosci. 10:73.
doi: 10.3389/fnhum.2016.00073

INTRODUCTION

Safety concerns have strongly motivated research in determining the demands, or workload, that users experience while performing closed-loop steering tasks, particular in the context of driving a car or piloting an aircraft (for a general review about workload, see Kramer, 1991; Wickens, 2008; Young et al., 2015). Even if competence can be maintained in spite of high mental workload, such scenarios leave little spare capacity for handling unexpected occurrences. There is no doubt that steering places high requirements on visual and motoric resources (Land and Lee, 1994; Salvucci and Gray, 2004).

Besides this, some aspects of steering have also been shown to require mental resources (Wickens et al., 1983, 1984). This has been typically demonstrated with the use of dual-task paradigms that induce a competition for mental resources between the primary steering task and an appropriately chosen secondary task (McLeod, 1977; Wickens and Gopher, 1977). The purpose of this article is to evaluate the demands that steering places on mental resources without requiring the user to perform a secondary task. To do so, we investigate how steering demands modify the event-related potentials (ERPs) to task-irrelevant auditory probes. The steering task is further manipulated for two aspects of steering that are known to influence handling difficulty, namely the bandwidth of disturbance and the complexity of (vehicle) control dynamics.

Workload can be defined as the ratio between the demands of a task and the resources of the human operator. Its concept originates from the idea that human operators possess, at any given time, a limited reserve of mental resources (Kramer, 1991; Wickens, 2008). By introducing a competition for this limited reserve, for example by requiring participants to perform two tasks simultaneously, researchers are able to investigate how difficulty manipulations in a primary task can create a demand for resources that are drawn away from an accompanying secondary task. Changes in resource demands are indexed by secondary task performance. A comparison of performance measures on competing tasks typically demonstrate that participants are capable of varying the relative prioritization of competing tasks (Wickens and Gopher, 1977), but only when the tasks overlap in their resource requirements (McLeod, 1977). The "Multiple Resource Theory" provides a framework that allows researchers and practitioners to define the resource requirements of different tasks and in doing so, predict possible conflicts (Wickens and Yeh, 1983; Wickens, 2002, 2008). Within this framework, a steering task places obvious demands on visual perception and motoric responses. By using electroencephalography (EEG) to measure the ERP to secondary task stimuli, Wickens and colleagues were able to demonstrate the demands of various aspects of steering on mental resources as well.

To date, ERP studies have broadly demonstrated that steering demands tend to reduce the amplitude of the P300, an ERP component that is generated by the target stimuli of a secondary task (e.g., Wickens et al., 1977; Isreal et al., 1980; Wickens and Yeh, 1983). Dual-task studies that investigate steering demands typically require participants to detect and explicitly respond to infrequently presented "oddball" targets as a secondary task. "Oddballs" elicit a prominent P300 component in the EEG signal. The P300 is a positive deflection between 250–400 ms and its amplitude has been used to index the level of experienced workload (Kok, 1997). The finding that steering demands diminish P300 amplitudes in an accompanying "oddball" detection task is commonly interpreted as follows. The primary steering task places prioritized demands on mental resources, resulting in the reduced availability of mental resources that would otherwise be recruited for the detection of secondary "oddball" targets (Wickens et al., 1977; Isreal et al., 1980; Wickens and Yeh, 1983). Hence, the reduced availability of mental resources is reflected in the reduced amplitudes of P300 that are elicited by the detected "oddballs". This serves as a proxy for evaluating the demands for mental resources, given different manipulations of steering difficulty. Some steering parameters exert a uniform cost on P300 amplitudes regardless of their manipulated difficulty levels, while increasing the difficulty levels of other parameters can induce decreased P300s to secondary "oddball" targets. For example, increasing the number of simultaneously tracked dimensions (Wickens et al., 1977; Kramer et al., 1983; Sirevaag et al., 1989), tracking speed (Kida et al., 2004), and the frequency bandwidth of the tracked target (Isreal et al., 1980) do not result in a decrease of P300 amplitudes. In contrast, increasing the complexity of control dynamics (e.g., from a first-order to a second-order integrator; Wickens et al., 1983, 1984; Sirevaag et al., 1989) or the unpredictability of the tracked target (Kida et al., 2004) result in corresponding decreases in P300 amplitudes. Other ERP components have also been analyzed for their sensitivity to changes in steering demands, albeit with mixed results. Kida et al. (2004) reported a decrease in the amplitude of the N140 component to the somatosensory targets of a secondary oddball task, which did not vary with the predictability of the steering task.

Until now, ERP studies of steering demands have mainly been performed in the presence of a secondary task that contains the stimuli for eliciting the ERP. It is generally believed that ERP probes are only effective for evaluating the resource demands of tasks that they are in explicit conflict with. Indeed, Wickens et al. (1983) have shown that the influence of steering demands on P300 amplitudes is removed when the ERP probes were task-irrelevant. Unfortunately, dual-task paradigms present several limitations in understanding steering demands. First, requiring an overt response to a secondary task interferes with the performance of the primary steering task (Wickens et al., 1983). In this regard, the secondary task is not a passive consumer of residual mental resources but is, rather, in direct competition with the primary task for shared resources. Second, the researcher has little control over how participants might choose to divide their resources between primary and secondary task, regardless of explicit instructions. Finally, estimated workload from ERP measurements could be due to the interaction of the primary and the secondary task demands, instead of the primary task alone. These reasons, amongst others, have motivated the development of non-intrusive methods for estimating primary task demands that do not necessitate a secondary task.

In contrast to Wickens et al.'s (1983) findings, ERPs to task-irrelevant stimuli can sometimes be demonstrated to vary with the demands of a task that is performed in isolation. This has been shown with the use of ERP probe stimuli that are more likely to recruit larger momentary shifts of resources than simple beep tones, such as complex environmental sounds (Courchesne et al., 1975; Ullsperger et al., 2001; Polich, 2003). Such stimuli are task-irrelevant and reliably elicit a positive ERP component termed the novelty-P3 (P3a)—that has a similar time-course to the P300 but with a frontal instead of a parietal distribution

(Polich, 2007). Given their task-irrelevant nature, it is more reasonable to assume that their elicited ERP components reflect residual resources that are not consumed by the demands of the investigated task. Task-irrelevant probes have been used to estimate the demands of a variety of tasks including arithmetic and visual monitoring (Ullsperger et al., 2001), working memory task (i.e., *n*-back task; SanMiguel et al., 2008), Tetris® (Miller et al., 2011; Dyke et al., 2015), first-person-shooter (Allison and Polich, 2008) and car racing games (Burns and Fairclough, 2015). It has not always been necessary to employ novel environmental sounds in order to generate ERPs for the evaluation of task demands—simple beep tones have proven to be sufficient in some instances (Burns and Fairclough, 2015). Nonetheless, there are also other examples whereby simple beep tones do not generate ERPs (i.e., P3a) that are sensitive to task demands (e.g., Ullsperger et al., 2001; Muller-Gass et al., 2007). Environmental sounds have the added value of generating larger novelty-P3s that are further separable for an early and late P3 component, which are claimed to be functionally distinct (Alho et al., 1998; Yago et al., 2003; McDonald et al., 2010). Early P3 is claimed to reflect post-sensory detection of unexpected events that contradict the observer's representation of the external world, while late P3 is claimed to reflect attentional processing of the unexpected event. Besides novelty-P3, other ERP components of task-irrelevant probes (i.e., N1/MMN; Ullsperger et al., 2001; Dyke et al., 2015; P2 and N2; Allison and Polich, 2008; late positive potential or LPP; Miller et al., 2011) have also been claimed to be diminished by increased task demands, albeit less consistently.

Taken together, ERP probes can be regarded as distractors that demand resources either through explicit competition with the primary task (Isreal et al., 1980; Wickens et al., 1983, 1984) or by implicitly drawing upon residual resources that are unconsumed by the primary task (SanMiguel et al., 2008; Miller et al., 2011; Burns and Fairclough, 2015; Dyke et al., 2015). Previous work that assessed steering demands might have required ERP probes to be task-relevant because the employed probes (i.e., beep tones) did not recruit sufficient resources to indicate the influence of steering demands.

ERP components that are elicited by distracting stimuli have been suggested to reflect three stages of distraction (Schröger and Wolff, 1998; Escera and Corral, 2007; Wetzel and Schröger, 2014). Based on the specific ERP components that are decreased with an increase of the task demands, inferences about the stages of distraction that are influenced can be drawn. The first stage of distraction is the detection that the model of the environment was violated. When engaged in a task, participants can be expected to be primarily focused on this task. At the same time, the regularities of the acoustic environment are encoded and used to form a predictive model of the surroundings. Whenever a current event violates this predictive model, the distraction process is initiated. This first stage of distraction is reflected in the elicited ERP by the mismatch negativity (MMN). The MMN is an early, negative ERP component that is apparent in the difference wave between the distractor- and the standard stimuli, for example in an oddball paradigm. Thus, the presence of a MMN indicates early sensory detection of an unexpected change in the environment. The second stage is the, voluntary or involuntary, orientation of attention towards the distracting event. Depending on the level of readily available resources and the eliciting event, resources might be directed towards the distracting event in order to process it. This stage is reflected by the occurrence of the novelty-P3 component. The third stage describes a disengagement of resources from the distracting event and a re-orientation back to the task at hand. Disengagement from the distractor stimuli is reflected by the re-orientation negativity (RON), a late negative component.

The current study investigates the influence of steering demands on ERP components that are generated by task-irrelevant auditory distractor stimuli. In the viewing baseline condition, we expect distractor stimuli to elicit ERP components that correspond to the three-stage distraction model, regardless of whether they are infrequently presented beep tones or infrequently presented environment sounds. However, we expect these ERP components to be larger when generated by environment sounds. Furthermore, we expect these ERP components to decrease when participants are required to perform a steering task, but only when they are generated by environmental distractors. We employ a data-driven approach (i.e., mass univariate analyses; Groppe et al., 2011) to ensure the validity of any correspondence between distractor ERP components and steering demands. This approach allows us to define each affected component in terms of its spatial and temporal characteristics, as opposed to restricting our analyses to an *a priori* selection of components (cf., Miller et al., 2011; Dyke et al., 2015). ERP components that are found to be sensitive to steering demands are subsequently submitted for permutation tests to evaluate their suitability for discriminating between manipulated levels of steering difficulty. We manipulate steering difficulty by either increasing the frequency bandwidth of the disturbance that is experienced during steering (cf., Isreal et al., 1980), or by varying the complexity of the control dynamics (cf., Wickens et al., 1983). We expect more participants to demonstrate a significant reduction in these targeted ERP components in the "hard" condition compared to the "easy" condition.

MATERIALS AND METHODS

Participants

We tested 24 right-handed volunteers (seven women, mean age = 27.9 years, *SD* = 5.2). All participants reported normal or corrected-to-normal vision, no hearing impairment and no history of neurological diseases. The experimental procedure was approved by the MPG Ethics Council and all participants gave written informed consent.

Stimuli and Apparatus

The experiment was set up in a dimly-lit, low noise environment. It consisted of a primary steering task and the presentation of task-irrelevant, auditory stimuli. The steering task was

presented via a central display (1027 × 581 mm, resolution 1920 × 1080 px), approximately 180 cm away from the seated participants. Auditory stimuli were presented to both ears via headphones (MDR-CD380, Sony), that where driven by a soundcard (sampling frequency: 96 kHz; DELTA1010LT, M-Audio). A secondary heads-down display informed the participants of their most recent steering performance and the current experimental status. Data collection was performed, using customized software, written in Matlab Simulink. The software version of the NASA-TLX questionnaire (Hart and Staveland, 1988) was presented on a separate notebook.

Two lines (length: 16° visual angle, thickness: 2 px) were presented on a blue background. These lines were a white horizontal non-moving reference line and a second black line that rotated around the joint center of both lines. A right-handed sidestick (Extreme 3D Pro, Logitech) with a spring constant of 0.6 N/degree was used as input device.

During the entire experiment, participants were probed with task-irrelevant stimuli with a random inter-stimulus interval (mean = 1.20 s, *SD* = 62 ms). Infrequently presented environmental sound distractors (prob. of presentation: $p = 0.1$) were intermixed with frequent, standard ($p = 0.8$) and infrequent distractor ($p = 0.1$) beep-tones. Two easily discriminable beep-tones were used (i.e., 300 and 700 Hz) and their probability ($p = 0.1$ and $p = 0.8$) was counter-balanced across participants. The environmental sounds consisted of a set of 30 recognizable complex sounds (e.g., human laughter) that were selected from a database obtained from the New York State Psychiatric Institute (Fabiani et al., 1996). The environmental sounds were presented in quasi-random order without replacement. Environmental sounds, as well as standard and distractor beep-tones, had a mean duration of 336 ms (*SD* = 62.5 ms) and a mean intensity of 60 dB SPL (*SD* =0.31 dB). Both, environmental and beep sounds were always preceded by at least one standard beep.

Task

Participants performed a steering task in which they were required to continuously counteract a quasi-random roll motion of a rotating line. This unpredictable roll motion was defined by the forcing function $f_t(t)$ (see Equation (1) and **Table 1**). Participants were instructed to minimize the displacement $e(t)$ of the rotating line (black in **Figure 1**) relative to the reference line (white in **Figure 1**), with lateral deflections of the sidestick.

Task-irrelevant sounds were presented that our participants were instructed to disregard. The experiment consisted of steering as well as of viewing trials. The viewing trials presented the same visual feedback in all sessions and served as a baseline. In this condition, participants viewed the steering task that was prerecorded. By comparing the steering trials against these viewing trials that both presented the same visualization, we could determine how the demands of the steering task influenced the measured ERPs, independent of the visualization.

Two aspects of the steering task were used to influence the level of workload in the task: (1) the frequency bandwidth of the roll disturbance and (2) the complexity of the internal control dynamics. In every steering trial, one of these aspects was manipulated, leading to two levels of steering task difficulty, namely "easy" and "hard" for each of the two manipulations. The second aspect was kept constant and will be referred to as "standard", in the following. The objective was to create two levels of workload for independent manipulations (cf. Isreal et al., 1980; Wickens et al., 1984). Details of these manipulations of engagement are given in the following.

Manipulation of the Bandwidth of Roll Disturbance

The roll disturbance was designed as a sum of ten sine waves that could be manipulated for the number and intensity of roll reversals by adjusting the frequency bandwidth, such that the "easy" condition presented less power in the higher frequencies, compared to the "hard" condition. The "standard" condition was designed to be an intermediate of these two conditions.

In all conditions, the forcing function was formalized as the sum of ten sine waves that were non-harmonically related, as described in (1):

$$f_t(t) = \sum_{j=1}^{10} A(j) \cdot \sin\left(\omega(j) \cdot t + \phi(j)\right) \tag{1}$$

The amplitude $A(j)$, frequency $\omega(j)$ and phase $\phi(j)$ of these 10 sine waves, for the "standard", the "easy" and the "hard" condition, are given in the **Table 1**.

TABLE 1 | Amplitude $A(j)$, frequency $\omega(j)$ and phase $\phi(j)$ of the ten sine waves, contained in the forcing function, for the "standard", "easy" and "hard" condition.

	Standard			Easy			Hard		
j	$A(j)$ in degree	$\omega(j)$ in rad/s	$\phi(j)$ in rad	$A(j)$ in degree	$\omega(j)$ in rad/s	$\phi(j)$ in rad	$A(j)$ in degree	$\omega(j)$ in rad/s	$\phi(j)$ in rad
1	1.34	0.39	2.69	1.36	0.39	3.27	1.33	0.39	2.42
2	1.03	0.83	5.74	0.93	0.83	5.95	1.10	0.83	2.20
3	0.51	1.76	5.72	0.40	1.76	3.95	0.63	1.76	2.35
4	0.26	2.85	5.92	0.19	2.85	3.93	0.34	2.85	4.59
5	0.16	3.90	1.66	0.12	3.90	2.26	0.21	3.90	4.57
6	0.09	5.45	1.53	0.07	5.45	0.59	0.13	5.45	5.67
7	0.06	7.76	1.90	0.05	7.76	1.65	0.08	7.76	0.74
8	0.04	10.50	4.74	0.04	10.50	3.80	0.05	10.50	0.71
9	0.04	13.11	4.06	0.03	13.11	0.15	0.04	13.11	0.21
10	0.03	17.33	4.53	0.03	17.33	4.83	0.03	17.33	3.39

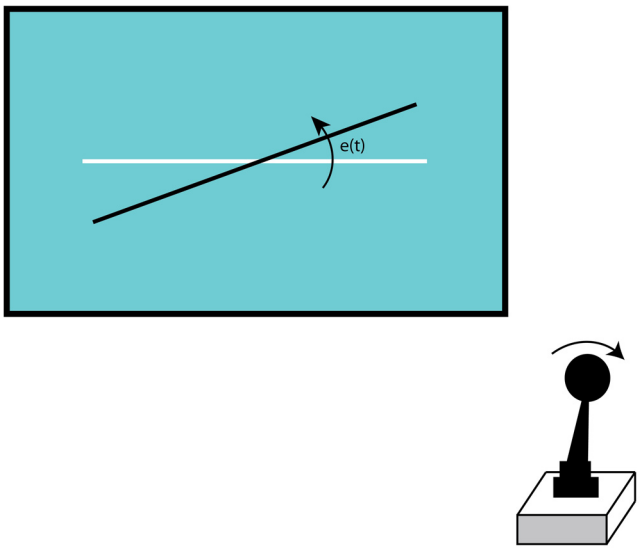


FIGURE 1 | The steering task required the participants to counteract the quasi-random displacement *e(t)* of the rotating line (black) to the non-moving reference line (white), with lateral sidestick deflections.

The forcing function in the "standard" condition had a variance of 1.61 degree^2, adapted from Nieuwenhuizen et al. (2013). In the "easy" condition a variance of 1.47 degree^2 and in the "hard" condition a variance of 1.78 degree^2 was applied.

To sum up, the "hard" condition presented larger amplitudes in the higher frequencies that resulted in more instances of roll-reversals than the "standard" and "easy" condition.

Manipulation of the Control Dynamics

By manipulating the control dynamics, the motion of the rotating line, relative to the sidestick input of the participants, was manipulated. The control dynamics can be formally described as the transfer function *H(s)*.

In the "standard" condition the transfer function had the form of:

$$H_{\text{standard}}(s) = \frac{2.75}{s\,(s+\omega_b)} \tag{2}$$

This represents a hybrid controller that reacts to the sidestick input with a weighted mixture of velocity and acceleration control. In other words, depending on the frequency of the sidestick input of the participant, either the velocity or the acceleration of the rotating line was influenced. To manipulate the internal control dynamics for difficulty levels, we removed either the velocity or the acceleration component, resulting in either a pure velocity controller with the following form for the "easy" condition:

$$H_{\text{easy}}(s) = \frac{1.5}{s} \tag{3}$$

or a pure acceleration controller with the following form for the "hard" condition:

$$H_{\text{hard}}(s) = \frac{5}{s^2} \tag{4}$$

These transfer functions were adopted from Zollner et al. (2010).

Controlling the acceleration has been shown to be more demanding than controlling the velocity (e.g., Wickens et al., 1984; Sirevaag et al., 1989). When the velocity is controlled, the angle of the sidestick translates to the velocity of the controlled line. In this case, keeping the sidestick in the center results in no motion of the controlled line. When the acceleration is controlled instead, keeping the sidestick in the center results in no further acceleration, but the controlled line will maintain its current velocity. Thus, participants have to anticipate the future consequence of their input commands when using a pure acceleration controller.

Design and Procedure

The experiment consisted of two sessions on 2 separate days, one that contained the manipulation of the bandwidth of the roll disturbance and one that contained the manipulation of the complexity of the control dynamics. Session order was counterbalanced across participants. Each of the two sessions consisted of four blocks that contained three trials each. The four blocks differed in terms of the implemented difficulty ("easy" or "hard"). Each block contained two steering and one viewing trial, where the order of the trials was randomized for every participant. Each of the trials lasted 4 min 26 s and trials were separated by 20 s of rest. During EEG preparation, participants were trained on every difficulty level and for each manipulation for at least one trial. Over the whole course of the experiment, after each trial, participants were presented with their performance (normalized root-mean-square error, nRMSerror) to keep them motivated. At the end of each block, participants were asked to rate their perceived workload in the NASA-TLX questionnaire for each level of difficulty, separately.

EEG Signal Processing

The EEG was recorded with 26 active g.tec Ag/AgCl electrodes (g.LADYbird, g.tec), mounted in an elastic cap (g.GAMMAcap, g.tec). The electrooculogram (EOG) was recorded from four additional electrodes: at the outer canthi of the left and right eye, and above and below the left eye. All recorded signals were re-referenced off-line to the linked mastoids. The ground electrode was placed at FPz. The signals were amplified in the range between 0 and 2.4 kHz and digitized with a sampling rate of 256 Hz (g.USBamp, g.tec).

Further processing and analysis of the ERP signal was performed with Matlab and the open source Matlab toolboxes EEGLAB (Delorme and Makeig, 2004) and ERPLAB (Lopez-Calderon and Luck, 2014). In the off-line preprocessing, the data was high pass filtered at 1 Hz and low pass filtered at 15 Hz. Second-order Butterworth filters were used for both filters. From the filtered data, epochs from −200 to 1000 ms, relative to the onset of the presented sounds, were extracted. Epochs that showed blink or eye movement characteristics, in any of the electrodes, were rejected. The remaining epochs were averaged for each auditory stimulus

type (environmental distractor, beep distractor, standard beep tone) and baseline corrected with reference to the pre-stimulus interval. The statistical analysis of the ERPs was based on the difference wave between ERPs that were elicited by distractors (the beep and environmental distractors, separately) and standards. This difference wave has been also referred to as distraction potential (DP; Escera et al., 2003).

Statistical Analysis of the ERPs

We adopted a 2-stage approach for analyzing the ERPs elicited by the environmental and beep distractors. First, we employed mass univariate analyses to: (i) determine the ERP components that were elicited by the distractors; (ii) determine the ERP components that differed between the environmental and beep distractors; and (iii) identify and define the spatiotemporal characteristics of ERP components that were significantly reduced during steering, relative to the viewing baseline condition. To perform the mass univariate analyses, measured brain potentials were compared between the relevant conditions at all time points (between 100–900 ms after the presentation of the auditory stimuli) and all measured electrodes (26 electrodes distributed over the scalp). Two-tailed *t*-tests were performed between the compared conditions to yield *t*-values for every time-point of each electrode. The false discovery rate (FDR) was controlled using the Benjamini and Yekutieli (2001) procedure with a FDR level of 5%. This particular FDR procedure guarantees that the true FDR will approximate the nominal FDR level of 5%, regardless of the dependency structure of the multiple tests (a tutorial review of the mass univariate analysis is provided by Groppe et al., 2011). This revealed ERP time points and their corresponding electrodes that were significantly different between the conditions.

Second, the ERP components that were identified to be sensitive to steering demands were submitted to permutation tests for each individual participant, in order to determine if these components were influenced by our difficulty manipulations for either disturbance bandwidth or control dynamics. A description of these single-subject permutation tests and their interpretation is provided by Maris and Oostenveld (2007). In brief, four key steps are performed for each participant: First, the selected electrode's mean amplitude over the time-range of interest was computed for every trial. Second, these mean amplitudes were submitted to a one-tailed, paired-samples *t*-test to yield a test *t*-value. Third, a null-distribution of *t*-values was generated. All trials were pooled and randomly distributed (without replacement) to two subsets. A paired *t*-test was performed between these two sub-sets to generate a single *t*-value. This was repeated 10,000 times to generate a null distribution. Fourth, the test *t*-value was compared to this generated null-distribution to determine its *z*-value. An alpha-level of 0.05 was adopted to determine if the tested participant showed a significant difference for the difficulty manipulations. This procedure was repeated for each participant and each ERP component of interest.

RESULTS

Steering Performance and Perceived Workload

Steering performance and the perceived workload were analyzed for our manipulations of steering demands. This was performed independently for our manipulations of disturbance bandwidth and control dynamics complexity with the use of a paired-samples *t*-test. This was to validate that our participants responded appropriately to our difficulty manipulations for "easy" and "hard". An alpha-level of 0.05 was adopted for significance testing. The Cohen's *d* is reported for the effect size. Overall, we found medium to large effects in our manipulations of difficulty for both performance and perceived workload.

Steering performance was evaluated based on the root-mean squared deviation of the rotating line from the reference line (i.e., RMSerror). The mean RMSerror was significantly higher in the "hard" than in the "easy" condition for manipulations of the disturbance bandwidth ($t_{(23)} = -6.6$, $p < 0.001$, $d = -1.4$) and control dynamics ($t_{(23)} = -2.2$, $p = 0.04$, $d = -0.4$).

Perceived workload was based on the participants' responses in the NASA-TLX questionnaire. The resulting workload score is the weighted sum of six subscales that were perceived by the participants as contributing to the overall workload in the following proportions: Effort: 24.5%, Mental Demand: 23.1%, Temporal Demand: 17.7%, Performance: 14.3%, Physical Demand: 13.4%, and Frustration: 7.0%. The "hard" condition was rated as being significantly more demanding than the "easy" condition for both manipulations (disturbance bandwidth: $t_{(23)} = -3.4$, $p = 0.00$, $d = -0.7$; control dynamics: $t_{(23)} = -3.6$, $p < 0.001$, $d = -0.7$). **Figure 2** illustrates the distribution of the six subscales over the two manipulations and two levels of difficulty.

ERP Results

This section is divided into three parts that describe the three analyzed aspects of the elicited ERP components. First, we present the comparison of the two distractor stimuli. Second, we present the results of the comparison between the viewing and steering trials. Third, we present the results of the comparison between the two applied manipulations of steering demands.

Comparison of the Two Distractor Stimuli

To begin, we separately identified ERP components that were elicited by the environmental and beep distractors. Therefore, we identified, with mass univariate analysis, the time-periods for which ERP amplitudes were significantly different from the pre-stimulus time interval. **Figure 3** illustrates the grand averaged waveforms and indicates significant ERP components with black bars. The environmental sounds elicited, in the steering and the viewing condition, a MMN, an early and late P3, a RON, a late positive potential (LPP) and a late negativity (LN). The beep distractors elicited a MMN, a P3a that was not further discriminable for early and late P3 sub-components, a RON, and (only in the steering condition) a LN.

Subsequently, we contrasted the ERPs that were elicited by the environmental and beep distractors. This was performed

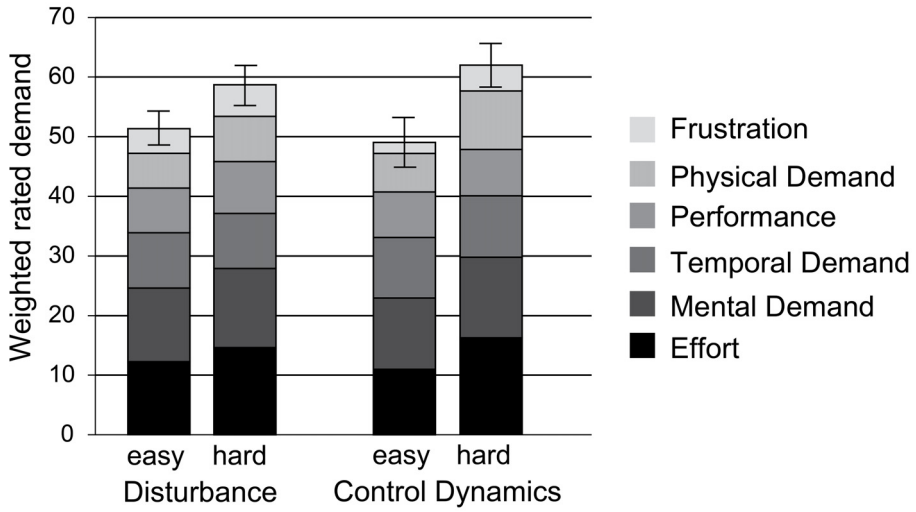


FIGURE 2 | Weighted sum of the six subscales of the NASA-TLX that were perceived by the participants as contributing to the overall workload. The error bars represent the 95% confidence interval.

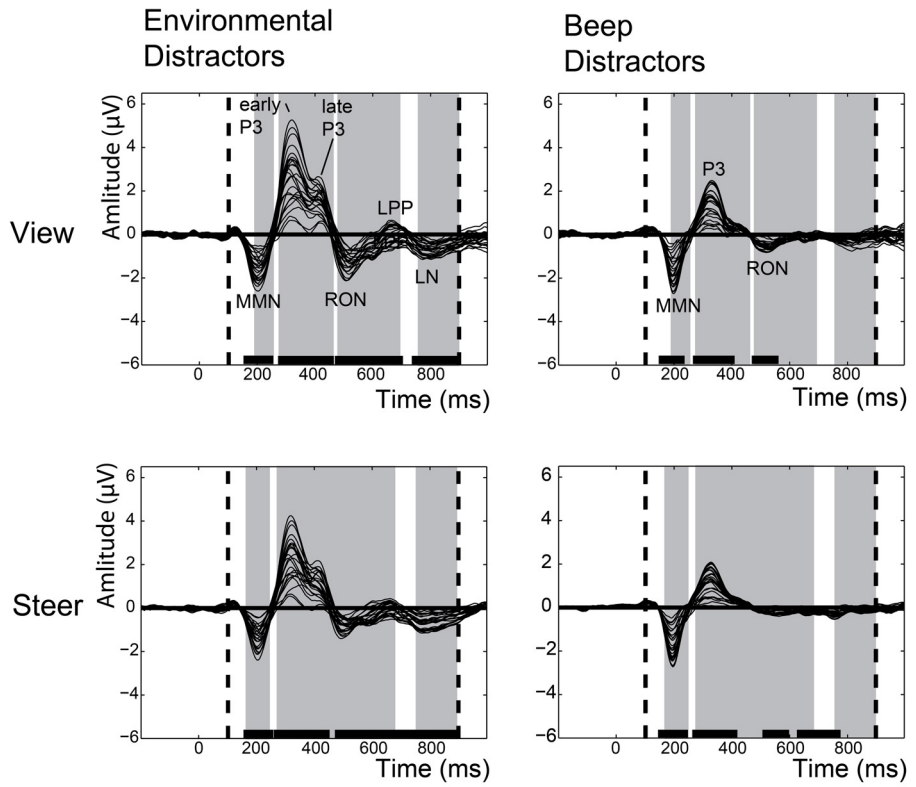


FIGURE 3 | Grand averaged waveform of the event-related potentials (ERPs) that were elicited by the environmental distractors (left column) and the beep distractors (right column), separately for the viewing (top row) and steering (bottom row). The grand averaged waveform shows the difference wave between the ERPs elicited by the environmental/beep distractors and the standard beep-tones. Every line represents one electrode. The dashed vertical lines represent the time window of interest (100–900 ms). The black bars specify the time range when the ERP amplitudes were significantly different from the pre-stimulus time-interval. The gray areas highlight the time-periods where the ERPs of the beep and environmental distractor differed significantly from each other.

separately for the steering and the viewing trials with the use of mass univariate analyses. **Figure 3** highlights (in gray) the time-periods where the ERPs of the beep and environmental distractor differ significantly. This reveals that environmental distractors generate larger P3, RON, LPP and LN components than the beep distractors. The beep distractor generated an MMN that peaked earlier than the environmental distractor.

General Demands of the Steering Task

Here, we determined the influence of steering demands on the elicited ERP components. In the grand averaged waveform (see **Figure 4**), the influence of the steering demands can be mainly observed in the ERPs that were elicited by the environmental distractor stimuli and to a lesser degree, in the beep distractors. As expected, for the ERPs that were elicited by the standard beeps the steering demands did not have a visible influence.

Using a mass univariate analysis, we determined the electrodes and time points for which ERPs were significantly decreased during the steering trials, relative to the viewing trials. This was performed separately for the ERPs that were elicited by the environmental distractors and those elicited by the beep distractors. The ERPs elicited by the beep distractors were not significantly influenced by steering demands for any electrode at any time point. In contrast, the ERPs elicited by the environmental distractors were selectively decreased by steering demands at specific time-points and electrodes. **Figure 5** provides a raster diagram to indicate the time-points and electrodes where ERPs of the environmental distractors were sensitive to steering demands. The scalp topographies for significant ERP components are provided together with the significant electrodes, indicated as white filled circles. Altogether, we find that steering demands diminish an early and late sub-component of the novelty-P3, and the RON. These ERP components have a frontocentral distribution.

Steering demands significantly decrease the early P3 generated by the environmental distractor in the time window between 280–330 ms in the frontocentral electrodes (AF3, AF4, F3, F4, FC5, FC1, FC2, C3, T7, Fz, Cz). The late P3 was significantly decreased between 330–430 ms in the central electrodes (FC1, FC2, FC6, C3, C4, CP1, CP2, P3, CP6, Cz, CPz, Pz). Interestingly, steering demands influence late P3 amplitudes at electrodes that do not correspond with the frontal electrodes, which exhibit the largest late P3 amplitudes. The RON was significantly decreased in the time window of 500–550 ms over the left electrodes (AF3, F3, FC1, FC2, FC5, FC6, C3, CP1, CP2, CP5, P3, PO3, Fz, Cz, CPz, Oz).

Following this, we employed permutation tests to analyze the influence of steering demands on the early P3, late P3, and RON of individual participants, when elicited by environmental distractors. Single trials of the two steering conditions ("easy" and "hard") were independently compared to the baseline viewing condition. For each participant, we submitted the recorded data from the electrodes and time points of the targeted ERP components to the permutation test. This was performed independently for the two different manipulations of steering difficulty, namely disturbance bandwidth and control dynamics complexity. **Figure 6** plots the number of participants that produced significantly larger ERP amplitudes in the viewing compared to the "easy" or "hard" steering trials for the targeted ERP components.

The single-subject analysis produced results that were consistent across both manipulations (i.e., disturbance bandwidth and control dynamics complexity) and all three analyzed components (early P3, late P3 and RON). More participants showed a significant reduction in the three targeted ERP components for the "hard" condition than the "easy" condition, relative to the "viewing" baseline. **Figure 6** also indicates differences across individuals, in terms of how they varied in response to the difficulty manipulations. White bars represent participants whose selected ERP components were diminished in both the "easy" and "hard" conditions. The dark gray bars represent participants whose ERP components were only diminished by the "hard" condition but not by the "easy" condition. The light gray bars represent participants whose ERP

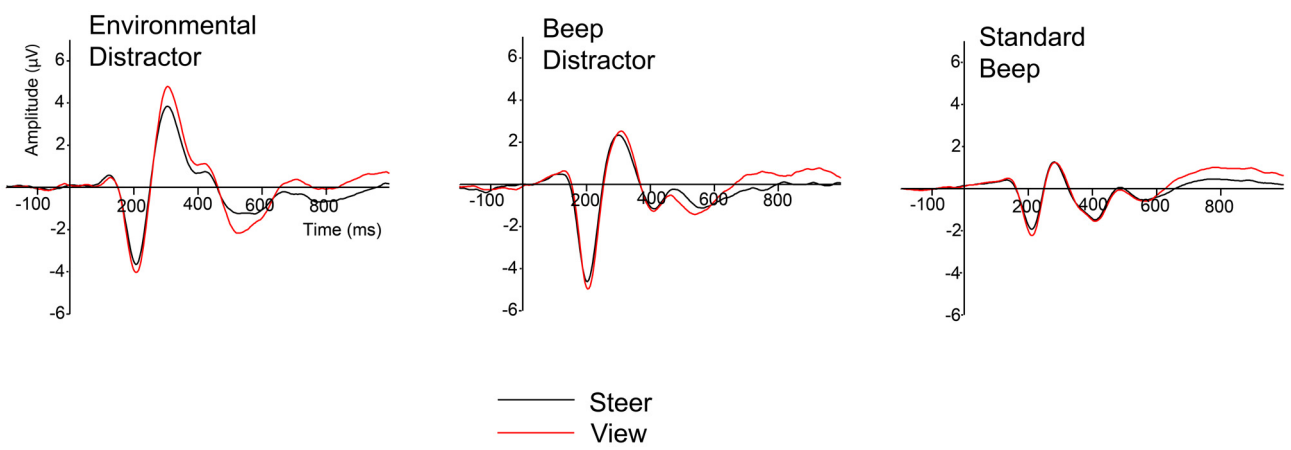


FIGURE 4 | Grand averaged waveforms after the stimulus presentation of the environmental distractors, beep distractors and standard beeps for the viewing (red) and steering (black) trials.

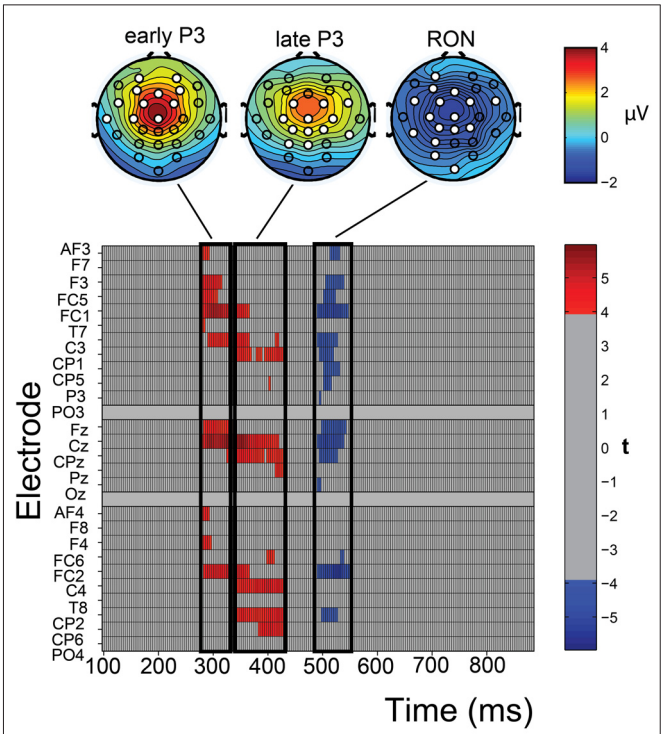


FIGURE 5 | The raster diagram (bottom) shows the comparison results of the environmental distractor ERPs across viewing and steering trials. A mass univariate analysis analyzed every time point (256 Hz) between 100–900 ms for all 26 electrodes. Red/blue rectangles represent time points and electrodes where the difference between the ERPs in the viewing and steering trials was significantly positive/negative. Scalp topographies are provided (top) for the three significant time-intervals where significant differences were found. Scalp potential amplitudes are illustrated as heat maps and significant electrodes that differentiated between the viewing and steering conditions are marked white.

components were only diminished by the "easy" condition but not by the "hard" condition. Overall, the results are in line with our expectations. More participants whose ERPs were unaffected by the "easy" condition were, nonetheless, affected by the "hard" condition than *vice versa*.

Influence of the Steering Manipulations

Permutation tests were conducted to identify the number of participants who reliably exhibited lower amplitudes for the targeted ERP components (i.e., early P3, late P3, and RON) in the "hard" trials relative to the "easy" trials. **Figure 7** represents these results as gray bars. The same analysis was performed based only on the peak-amplitude electrode and corresponding time-window (i.e., ±20 ms around the grand average peak). This is the approach that is employed by comparable research (cf., Miller et al., 2011; Dyke et al., 2015). **Figure 7** represents these results as black bars. A comparison shows that a mass univariate analysis approach identified ERP components that were more sensitive to the current steering manipulations. Finally, more participants responded in the expected direction for the targeted ERP components when the complexity of the control dynamics was manipulated for difficulty than when the bandwidth of disturbance was manipulated.

DISCUSSION

The current study was designed to investigate if the demands of a steering task would attenuate the amplitudes of ERPs to task-irrelevant stimuli. It is in this regard that the current work sets itself apart from previous work that evaluated steering demands by measuring the ERPs to the task-relevant stimuli of a concurrent secondary task (e.g., Wickens et al., 1983, 1984; Sirevaag et al., 1989). The main findings of the current study are that steering demands can significantly reduce the amplitudes of three ERP components (i.e., early P3, late P3, and RON) of task-irrelevant auditory probes. However, this requires the probes to be complex environmental sounds and not simple beep-tones. Two aspects of the steering task (i.e., disturbance bandwidth and control dynamics complexity) were manipulated for steering demands and the found ERP components were significantly diminished in more participants during the difficult conditions relative to the easy conditions for both manipulations. The current results agree with a three-stage distraction model, whereby the ERP probes can be regarded as distractor stimuli that consume mental resources involuntarily (Schröger and Wolff, 1998; Escera and Corral, 2007; Wetzel and Schröger, 2014). Therefore, we will discuss our results within this simple framework. The discussion will be organized as follows. First, we shall discuss the differences between complex environmental sounds and simple beep tones in order to understand why the former elicit ERPs that are sensitive to steering demands while the latter do not. Second, we will discuss the implications of each ERP component that was found to respond to steering demands. Third, we will discuss the observed differences in the ERPs between manipulating either the disturbance bandwidth or the control dynamics complexity.

Comparison of Complex Environmental Sounds and Beep-Tones Distractor Stimuli

Both types of task-irrelevant distractor sounds elicited a characteristic waveform that contained ERP components, which were significantly different from the baseline (see **Figure 3**). In temporal order, they are the MMN, the novelty-P3, and the RON. Respectively, they are claimed to represent the three subsequent stages of how users respond to distraction (Schröger and Wolff, 1998; Escera and Corral, 2007; Wetzel and Schröger, 2014): (1) detection of the unexpected stimulus; (2) orientation towards the stimulus; and (3) disengagement from the distractor to re-orient back to the steering task. In other words, infrequently presented sounds are preferentially processed by the brain in spite of being task-irrelevant, whether they are complex environmental sounds or beep-tones. Two other ERP components (i.e., LPP and LN) were also elicited, but were not sensitive to steering demands.

Environmental sounds elicited ERPs that differed from the beep tones in two ways. First of all, they elicited larger ERPs. Second, their ERPs contained components that were sensitive

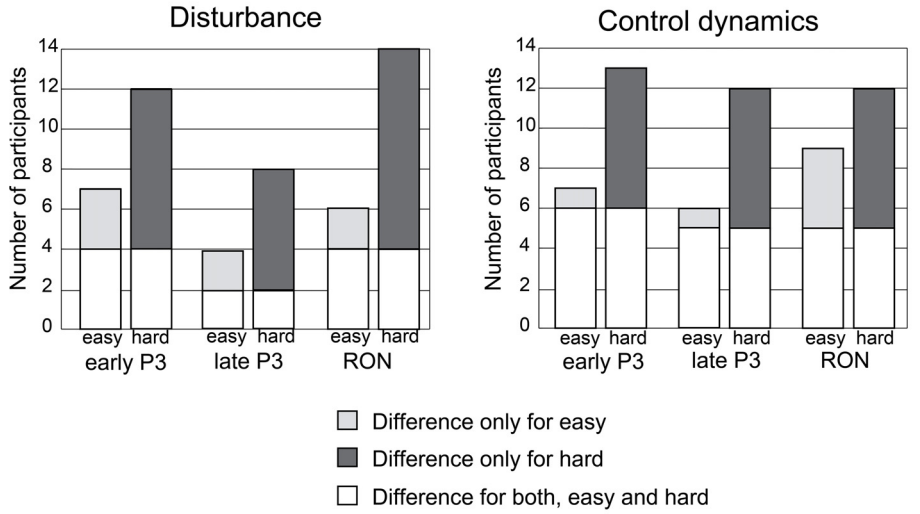


FIGURE 6 | Permutation tests were performed to evaluate steering manipulations of the disturbance (left) and control dynamics (right). Bar plots indicate the number of participants that exhibit a significant difference in their early P3, late P3 and RON for the steering condition ("easy", "hard") relative to the viewing baseline. White bars indicate participants who showed a reliable difference for both "easy" and "hard" conditions. Light/Dark gray bars indicate participants who showed a reliable difference for only the "easy"/"hard" condition.

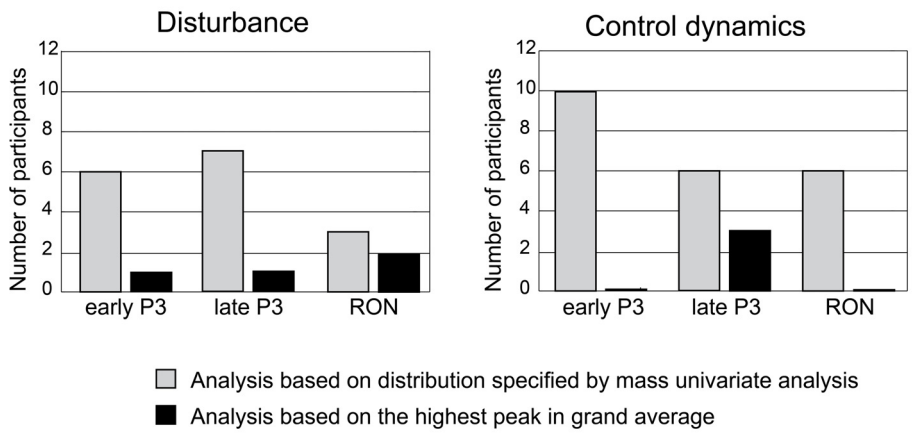


FIGURE 7 | Permutation tests were performed to evaluate steering manipulations of the disturbance (left) and control dynamics (right). Bar plots indicate the number of participants that exhibit a significant difference in their early P3, late P3 and RON between the "easy" and "hard" conditions. Light gray bars indicate participants who showed a reliable difference between the "easy" and "hard" conditions when the analysis was based on the electrodes and time points indicated by mass univariate analysis. Black bars indicate participants who showed a reliable difference between the "easy" and "hard" conditions when the analysis was based on the peak in the grand averaged waveform.

to steering demands. These two aspects are related. To begin, it can be argued that the larger novelty-P3 and RON amplitudes (see gray areas in **Figure 3**) indicate that environmental sounds recruit more corresponding mental resources than the beep sounds (Kok, 1990, 1997). This difference is apparent in the baseline viewing condition during which the participants' mental resources were unoccupied and readily available. Involuntary resource recruitment is attenuated when participants are required to perform a steering task (i.e., in the steering trials), but only for the novelty-P3 and the RON of the environmental distractors (see **Figure 5**). This is because the steering task reduced the amount of available resources to a lower level than task-irrelevant environmental distractors would typically recruit. In view of this, we believe that our use of task-irrelevant

environmental distractors is a more direct assessment of the resource demands of the steering task, when compared to dual-task paradigms that increase the resource demands of task-relevant stimuli that actively compete for resources with the steering task (Wickens et al., 1983, 1984; Sirevaag et al., 1989).

What are the properties of environmental sounds that allow them to recruit more mental resources and hence, generate larger ERPs even when they are task-irrelevant? Previous work suggests that distractor stimuli tend to recruit more resources if they are personally meaningful and/or exhibit high dissimilarity from their context. The personal meaning and dissimilarity from the context are respectively referred to as being stimuli specific and aspecific (Eimer et al., 1996; Hughes, 2014). Specific aspects are parameters that are inherent to the stimulus, which represent its meaning to the observer (Hughes, 2014). For example, one's personal ringtone is more distracting, as reflected by larger elicited ERPs, than another person's ringtone (Roye et al., 2007). In the current study, the environmental distractors represented familiar objects (e.g., dogs, cats, babies), which have more personal meaning than the beep-tone distractors. Thus, they can be expected to recruit more resources. Aspecific aspects of the eliciting stimulus recruit resources involuntarily due to its embedded presentation context. For example, a task-irrelevant female voice has been shown to be less distracting, as reflected by a decrease of performance in a visual recall task, when presented in a series of female voices than when presented in a series of male voices (Hughes et al., 2013). In the current experiment, we presented the environmental sounds as well as the beep sounds against a context of frequent beep tones. Arguably, environmental sounds that are a complex combination of multiple frequencies are more dissimilar to this context than their beep tone counterparts. This raised the likelihood that the environmental sounds would recruit more resources than their beep tone counterpart.

To sum up, task-irrelevant stimuli are more likely to be sensitive to task demands if they are personally meaningful and differ sufficiently from their embedded context. Some studies have been reported that have been successful in using task-irrelevant beep tones to evaluate task demands. However, these studies investigated complex tasks—that is, first person shooter (Allison and Polich, 2008) and racing games (Burns and Fairclough, 2015)—that, presumably, induced higher task engagement and varied in their resource demands at levels that beep tones were sensitive to. We expect the ERPs of task-irrelevant environmental sounds to be even more sensitive than beep tones to the resource demands of such complex tasks.

Influence of Steering Demands on the Measured ERP Components

The current study is the first to employ task-irrelevant ERP probes in a task that allows for the systematic manipulation of different steering demands. Such task-irrelevant probes, in particular environmental sounds, continue to elicit ERPs with components that we have identified to be selectively diminished by steering demands: early P3, late P3 and RON (see **Figures 3**, **5**). As noted before, these components correspond to the mid and late stages of a three-stage distraction model (Schröger and Wolff, 1998; Escera and Corral, 2007; Wetzel and Schröger, 2014). From the perspective of this model, steering demands did not inhibit our participants' capacity for detecting unexpected occurrences. Instead, steering demands significantly diminished the extent to which available mental resources could be directed towards the processing of distractor stimuli. In turn, this hinders an efficient re-orientation away from the distractor stimuli. Altogether, these findings demonstrate that steering places demands on mental resources that would otherwise be directed towards an instinctive evaluation of unexpected events. These resources are based on attentional processes, but at a cognitive rather than a perceptual level. It is interesting to note that our participants were able to articulate this in that they rated the "hard" condition as being more demanding than the "easy" condition in terms of mental rather than physical effort (see **Figure 2**). This supports our research motivation in understanding the demands of a steering task beyond its perceptual and response requirements.

The ability to maintain an appropriate level for "distraction" is a fundamental capability of our attentional system and a critical aspect of effective vehicle handling. On the one hand, the capacity to be distracted by unexpected events is necessary when these events reflect potential dangers in the environment. For example, the phenomenon of "attentional tunneling" refers to scenarios when high-performance pilots miss unexpected hazards given their increased engagement with vehicle handling. Such undesirable instances have even been observed in novel cockpit environments that are designed to promote engagement with vehicle handling, for example when synthetic vision displays with intuitive flight guidance were employed for fixed-wing control (Wickens and Alexander, 2009). On the other hand, distraction presents a danger when it interrupts and prevents one to carry out a safety-critical task. In the United States, driver distraction raises the risk of a light-vehicle near-crash/crash to approximately three times of the baseline level (Klauer et al., 2006; Regan et al., 2011). Task-irrelevant or task-relevant probes can be judiciously employed in steering environments depending on whether the goal is to investigate either involuntary or voluntary distraction. A perspective that considers steering environments in terms of the driver's engagement with the steering task and potential distractions (both voluntary and involuntary) is more likely to yield practical insights and operational recommendations than one that simply evaluates driving workload.

In this study, we show that both, early and late P3 components, were influenced by steering demands. These components are discriminable from each other in terms of their spatial and temporal characteristics. Functionally, the early P3 reflects a sensitivity towards violations of one's model of the environment at a post-sensory stage (Ceponiene et al., 2004). The late P3 relates to the attending of the unexpected event itself, presumably for the purpose of updating one's model

of the environment when deemed necessary (Escera et al., 1998; Yago et al., 2003; SanMiguel et al., 2008). Earlier studies have provided mixed evidence on the relationship of workload and these components. Difficulty manipulations in a complex Tetris® gaming environment have been found to only diminish early P3 amplitude (Dyke et al., 2015), while other studies, in particular those that target memory load, identified the late P3 as the only P3 sub-component that is influenced by workload (Escera et al., 1998; SanMiguel et al., 2008). Until the subtle interactions between workload and these P3 sub-components are better understood, we recommend employing approaches such as mass univariate analyses to determine the role of either sub-components in new task paradigms (e.g., steering), so as to reduce the risk of false positives.

Characterizing the relevant sub-components in terms of their spatial and temporal distributions provides an additional benefit. It allowed us to discriminate between manipulations of steering demands that would not be noticeable by only analyzing the peak, given inter- and intra-individual differences (cf., Munka and Berti, 2006; Miller et al., 2011; Dyke et al., 2015). In the current work, we show that more participants discriminated for the "easy" and "hard" steering trials compared to when the analysis was based on the highest peak in the grand average (see **Figure 7**). Mass univariate analysis also offers an additional benefit in that it more accurately defines the spatial location of the effect of interest. In the case of late P3, we find that the electrodes that are sensitive to steering demands have a more parietal distribution than the peak amplitude electrode. This agrees with the work of Yago et al. (2003) who also defined a discriminable parietal aspect of late P3 that is claimed to be involved with working memory updating and is believed to originate from the posterior and superior parietal lobes.

Besides early and late P3, we found that steering demands significantly decreased RON amplitude. RON is believed to reflect the re-orientation of attention from the distractor stimulus (Schröger and Wolff, 1998; Escera and Corral, 2007; Wetzel and Schröger, 2014). In this sense, it can be regarded as a disengagement of resources from processing distractor stimuli. Our results are comparable to those reported by Berti and Schröger (2003) who also found that increasing workload in the primary forced-choice task reduced RON amplitudes to a distracting task-irrelevant feature. In their experiment, participants were required to discriminate between sounds with "short" and "long" durations. Infrequent changes in the task-irrelevant pitch of the sounds produced RONs with an approximate latency of 500 ms. In their experiment, workload was manipulated either by allowing participants to respond immediately or by requiring them to respond upon the presentation of the next stimuli. The latter was considered to be more difficult as it involved a stimulus-response conflict. The amplitude of RON was found to be diminished in the difficult condition. Our current results indicate that a similar RON component can be diminished by increased task demands, even when the task is presented in a separate modality from the distractor. One reason for this could be that fewer resources were available to begin with, that could be effectively engaged by the distractor stimuli. Another reason could be that mental resources are more likely to be engaged with processing distractor stimuli for longer periods of time when sub-optimal levels of resources are allocated for their processing. In this case, the disengagement from the distractor stimuli could be expected to be less efficient. Whichever the reason, it is important to realize that RON reflects resource (*re*-)allocation processes at a post-sensory stage and that its amplitude does not simply decrease with increased workload. In fact, RON amplitudes have been found to be larger for the 1-back working memory task than its 0-back counterpart (SanMiguel et al., 2008). In this example, the 1-back task required participants to reference information of the primary task from recent history and larger RONs could have reflected a disengagement of resources from the distractor stimulus in addition to the re-allocation of resources to task-relevant information. We believe that our manipulation of steering demands resulted in decreased RON amplitudes because it only reflected the disengagement of resources from task-irrelevant distractor stimuli. If this is true, a dual-task paradigm that entails resource competition between a steering task and a task-relevant probe should result in larger RON amplitudes when steering demands are increased.

The Steering Demands of Manipulating Disturbance and Control Dynamics

In the current study, we manipulated two aspects of steering that are known to influence steering demands—that is, disturbance bandwidth and control dynamics complexity. Both manipulations of steering difficulty had an influence on the identified ERP components in the expected direction (**Figures 6**, **7**). Comparatively, this influence was evident in more participants when the complexity of control dynamics was manipulated. This result is in agreement with previous work that has shown a greater sensitivity of secondary task ERPs to the manipulation of control dynamics in the primary task (Isreal et al., 1980; Wickens et al., 1983, 1984; Sirevaag et al., 1989).

While encouraging, these results should be treated with caution. Our analyses reveal that our manipulations for steering demands do not influence the identified ERP components in all of our participants. In fact, some participants responded to steering demands only in the "easy" but not the "hard" condition, albeit to a lesser extent than *vice versa* (**Figure 6**). We believe that this reflects two aspects of inter-participant variance that are difficult to control for with the use of task-irrelevant ERP probes. First, the amount of resources that are involuntarily recruited for the processing of task-irrelevant probes. Second, steering competence and engagement with the steering task.

Participants can be expected to differ in terms of how meaningful they perceive different environmental sounds. Such differences could vary the extent to which these task-irrelevant distractors attract resources for their processing. If "insufficient" resources are recruited, changes in the level of available resources due to manipulations in steering demands can be expected to

go undetected. To mediate this, future studies could consider employing environmental distractors that are not as easily recognizable. It has been shown that larger frontal and parietal novelty-P3s are elicited by environment sounds that are not as easily recognizable, compared to their more recognizable counterparts (Opitz et al., 1999). Moreover, it has been shown that the novelty-P3's amplitude decreases with the repetition of familiar sounds but not unfamiliar sounds, presumably because participants are more effective in ignoring them (Cycowicz and Friedman, 1998, 2007).

Participants can be expected to vary in terms of steering proficiency. Therefore, some participants may only start to exhibit reduced levels of available resources under highly demanding scenarios. In fact, this is reflected in our results (see **Figure 6**). The current experiment employed fixed levels of steering difficulty. Subsequent studies could calibrate levels of steering difficulty for individual participants so that their performance discriminates sufficiently between "easy" and "hard" conditions. This would be similar to the use of adaptive methods in psychophysics to calibrate stimuli settings to individual differences in perception (Kingdom and Prins, 2010).

In spite of these limitations, our current findings are consistent with previous findings. The ERP components, which we have identified as being sensitive to steering demands, are more likely to differentiate for "easy" and "hard" conditions when disturbance bandwidth was manipulated than when control dynamics complexity was manipulated (cf., Isreal et al., 1980). This difference between the two manipulations is more prominent for early P3 and RON than it is for late P3. This suggests that increasing the complexity of the control dynamics limits how resources are directed towards and away from distractor stimuli.

Conclusion and Outlook

To conclude, we have shown that the demands of a steering task influence how the brain responds to task-irrelevant stimuli. Specifically, steering demands diminish the amplitudes of the early P3, late P3, and RON that are elicited by task-irrelevant auditory distractors, which are personally meaningful and distinct from the background. A three-stage distraction model would suggest that steering demands decreases one's sensitivity and likelihood to attend to unexpected events (early/late P3), as well as one's capacity to re-orient back to the steering task at hand (RON). In particular, we found this to be true for steering manipulations that increased the complexity of control dynamics.

The three-stage model of distraction, and its associated ERP components, is a simplification. It assumes a serial chain of information processing of the distractor stimulus and is agnostic to how the stages could be selectively influenced by factors that do not pertain to the distractor stimulus itself. Thus, its explanatory power is limited. Our finding, that environment sound distractors are more "distracting" than deviant beep tones (and result in larger MMN, P3a, and RON), is in line with the predictions of the three-stage distraction model. However, the three-stage distraction model does not explain why steering demands selectively influence P3a and RON amplitudes but not MMN. In fact, there is accumulating evidence to suggest that dissociations exist between the three stages of distraction. Factors such as the predictability of the distractor, which is not dependent on the distractor *per se* but on the homogeneity of the sequence of stimuli that precedes it, can influence MMN and P3a but not RON (Horváth et al., 2008). Converse dissociations have been reported whereby increasing the predictability of an auditory distractor with a visual cue can decrease P3a and RON amplitudes but leave MMN intact (e.g., Sussman et al., 2003). Hence, more complex accounts have since been proposed that not only consider how distractor stimuli are processed but also how their processing might interact with the perceived regularity of the auditory scene (for example, see Bendixen, 2014). For now, it is sufficient to note that the demands of a steering task are reflected in how it modulates the distractibility of task-irrelevant environment sounds, as reflected in the early/late P3 and RON that they elicit. Besides electrophysiological responses, future experiments should be designed to investigate the behavioral consequences of distraction on steering performance (c.f., Parmentier, 2014). This could elucidate differences between distractor stimuli that passively reflect steering engagement and those that pose an involuntary conflict with the cognitive processes that underlie steering itself.

Task-irrelevant stimuli can be expected to be more easily integrated into real-world operations than the use of ERP probes that require an explicit response. In this regard, our current findings raise the opportunity of estimating steering demands across a wider range of scenarios than was previously considered to be practical. Furthermore, the use of task-irrelevant and task-relevant distractor stimuli can reveal complementary aspects of how mental resources are managed during steering. In this regard, they can be effectively employed to understand the demands of steering and users' level of engagement with the steering task and their environment.

AUTHOR CONTRIBUTIONS

HHB and LLC: conception of the work; LLC and MS: design and interpretation of the data; LLC and MS: acquisition and analysis of the data; HHB, LLC and MS drafting and revision of the work; HHB, LLC and MS: final approval of the version to be published and agreement to be accountable for all aspects of the work.

FUNDING

This research was supported by the German Research Foundation (DFG) within project C03 of SFB/Transregio 161 as well as by the Max Planck Society.

ACKNOWLEDGMENTS

Special thanks to Christiane Glatz, Nina Flad, Eva Symeonidou, Alessandro Nesti, Tonja Machulla and two anonymous reviewers for their insightful comments and suggestions.

REFERENCES

Alho, K., Winkler, I., Escera, C., Huotilainen, M., Virtanen, J., Jääskeläinen, I. P., et al. (1998). Processing of novel sounds and frequency changes in the human auditory cortex: magnetoencephalographic recordings. *Psychophysiology* 35, 211–224. doi: 10.1111/1469-8986.3520211

Allison, B. Z., and Polich, J. (2008). Workload assessment of computer gaming using a single-stimulus event-related potential paradigm. *Biol. Psychol.* 77, 277–283. doi: 10.1016/j.biopsycho.2007.10.014

Bendixen, A. (2014). Predictability effects in auditory scene analysis: a review. *Front. Neurosci.* 8:60. doi: 10.3389/fnins.2014.00060

Benjamini, Y., and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Ann. Stat.* 29, 1165–1188. doi: 10.1214/aos/1013699998

Berti, S., and Schröger, E. (2003). Working memory controls involuntary attention switching: evidence from an auditory distraction paradigm. *Eur. J. Neurosci.* 17, 1119–1122. doi: 10.1046/j.1460-9568.2003.02527.x

Burns, C. G., and Fairclough, S. H. (2015). Use of auditory event-related potentials to measure immersion during a computer game. *Int. J. Hum. Comput. Stud.* 73, 107–114. doi: 10.1016/j.ijhcs.2014.09.002

Ceponiene, R., Lepistö, T., Soininen, M., Aronen, E., Alku, P., and Näätänen, R. (2004). Event-related potentials associated with sound discrimination versus novelty detection in children. *Psychophysiology* 41, 130–141. doi: 10.1111/j.1469-8986.2003.00138.x

Courchesne, E., Hillyard, S., and Galambos, R. (1975). Stimulus novelty, task relevance and the visual evoked potential in man. *Electroencephalogr. Clin. Neurophysiol.* 39, 131–143. doi: 10.1016/0013-4694(75)90003-6

Cycowicz, Y. M., and Friedman, D. (1998). Effect of sound familiarity on the event-related potentials elicited by novel environmental sounds. *Brain Cogn.* 36, 30–51. doi: 10.1006/brcg.1997.0955

Cycowicz, Y. M., and Friedman, D. (2007). Visual novel stimuli in an ERP novelty oddball paradigm: effects of familiarity on repetition and recognition memory. *Psychophysiology* 44, 11–29. doi: 10.1111/j.1469-8986.2006.00481.x

Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21. doi: 10.1016/j.jneumeth.2003.10.009

Dyke, F. B., Leiker, A. M., Grand, K. F., Godwin, M. M., Thompson, A. G., Rietschel, J. C., et al. (2015). The efficacy of auditory probes in indexing cognitive workload is dependent on stimulus complexity. *Int. J. Psychophysiol.* 95, 56–62. doi: 10.1016/j.ijpsycho.2014.12.008

Eimer, M., Nattkemper, D., Schröger, E., and Prinz, W. (1996). "Involuntary attention," in *Handbook of Cognition: Perception and Action*, Vol. 3, eds O. Neumann, and A. F. Sanders (London: Academic Press), 155–184.

Escera, C., Alho, K., Winkler, I., and Näätänen, R. (1998). Neural mechanisms of involuntary attention to acoustic novelty and change. *J. Cogn. Neurosci.* 10, 590–604. doi: 10.1162/089892998562997

Escera, C., and Corral, M. J. (2007). Role of mismatch negativity and novelty-P3 in involuntary auditory attention. *J. Psychophysiol.* 21, 251–264. doi: 10.1027/0269-8803.21.34.251

Escera, C., Yago, E., Corral, M.-J., Corbera, S., and Nuñez, M. I. (2003). Attention capture by auditory significant stimuli: semantic analysis follows attention switching. *Eur. J. Neurosci.* 18, 2408–2412. doi: 10.1046/j.1460-9568.2003.02937.x

Fabiani, M., Kazmerski, V. A., Cycowicz, Y. M., and Friedmann, D. (1996). Naming norms for brief environmental sounds: effects of age and dementia. *Psychophysiology* 33, 462–475. doi: 10.1111/j.1469-8986.1996.tb01072.x

Groppe, D. M., Urbach, T. P., and Kutas, M. (2011). Mass univariate analysis of event-related brain potentials/fields I: a critical tutorial review. *Psychophysiology* 48, 1711–1725. doi: 10.1111/j.1469-8986.2011.01273.x

Hart, S. G., and Staveland, L. E. (1988). Development of NASA-TLX (task load index): results of empirical and theoretical research. *Adv. Psychol.* 52, 139–183. doi: 10.1016/S0166-4115(08)62386-9

Horváth, J., Winkler, I., and Bendixen, A. (2008). Do N1/MMN, P3a, and RON form a strongly coupled chain reflecting the three stages of auditory distraction? *Biol. Psychol.* 79, 139–147. doi: 10.1016/j.biopsycho.2008.04.001

Hughes, R. W. (2014). Auditory distraction: a duplex-mechanism account. *Psych J.* 3, 30–41. doi: 10.1002/pchj.44

Hughes, R. W., Hurlstone, M. J., Marsh, J. E., Vachon, F., and Jones, D. M. (2013). Cognitive control of auditory distraction: impact of task difficulty, foreknowledge, and working memory capacity supports duplex-mechanism account. *J. Exp. Psychol. Hum. Percept. Perform.* 39, 539–553. doi: 10.1037/a0029064

Isreal, J. B., Chesney, G. L., Wickens, C. D., and Donchin, E. (1980). P300 and tracking difficulty: evidence for multiple resources in dual-task performance. *Psychophysiology* 17, 259–273. doi: 10.1111/j.1469-8986.1980.tb00146.x

Kida, T., Nishihira, Y., Hatta, A., Wasaka, T., Tazoe, T., Sakajiri, Y., et al. (2004). Resource allocation and somatosensory P300 amplitude during dual task: effects of tracking speed and predictability of tracking direction. *Clin. Neurophysiol.* 115, 2616–2628. doi: 10.1016/j.clinph.2004.06.013

Kingdom, F. A. A., and Prins, N. (2010). *Psychophysics: A Practical Introduction.* London: Academic Press.

Klauer, S. G., Dingus, T. A., Neale, V. L., Sudweeks, J. D., and Ramsey, D. J. (2006). *The Impact of Driver Inattention on Near-Crash/Crash Risk: An Analysis using the 100 Car Naturalistic Driving Study Data, DOT HS-810 594, U.S. Department of Transportation, National Highway Traffic Safety Administration.* Washington, DC: National Highway Traffic Safety Administration.

Kok, A. (1990). Internal and external control: a two-factor model of amplitude change of event-related potentials. *Acta Psychol. (Amst)* 74, 213–236. doi: 10.1016/0001-6918(90)90006-2

Kok, A. (1997). Event-related-potential (ERP) reflections of mental resources: a review and synthesis. *Biol. Psychol.* 45, 19–56. doi: 10.1016/s0301-0511(96)05221-0

Kramer, A. F. (1991). "Physiological metrics of mental workload: a review of recent progress," in *Multiple Task Performanc*, ed. D. L. Damos (London: Taylor and Francis), 279–328.

Kramer, A. F., Wickens, C. D., and Donchin, E. (1983). An analysis of the processing requirements of a complex perceptual-motor task. *Hum. Factors* 25, 597–621.

Land, M. F., and Lee, D. N. (1994). Where we look when we steer. *Nature* 369, 742–744. doi: 10.1038/369742a0

Lopez-Calderon, J., and Luck, S. J. (2014). ERPLAB: an open-source toolbox for the analysis of event-related potentials. *Front. Hum. Neurosci.* 8:213. doi: 10.3389/fnhum.2014.00213

Maris, E., and Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* 164, 177–190. doi: 10.1016/j.jneumeth.2007.03.024

McDonald, C. G., Gabbay, F. H., Rietschel, J. C., and Duncan, C. C. (2010). Evidence for a new late positive ERP component in an attended novelty oddball task. *Psychophysiology* 47, 809–813. doi: 10.1111/j.1469-8986.2010.00986.x

McLeod, P. (1977). A dual task response modality effect: support for multiprocessor models of attention. *Q. J. Exp. Psychol.* 29, 651–667. doi: 10.1080/14640747708400639

Miller, M. W., Rietschel, J. C., McDonald, C. G., and Hatfield, B. D. (2011). A novel approach to the physiological measurement of mental workload. *Int. J. Psychophysiol.* 80, 75–78. doi: 10.1016/j.ijpsycho.2011.02.003

Muller-Gass, A., Macdonald, M., Schröger, E., Sculthorpe, L., and Campbell, K. (2007). Evidence for the auditory P3a reflecting an automatic process: elicitation during highly-focused continuous visual attention. *Brain Res.* 1170, 71–78. doi: 10.1016/j.brainres.2007.07.023

Munka, L., and Berti, S. (2006). Examining task-dependencies of different attentional processes as reflected in the P3a and reorienting negativity components of the human event-related brain potential. *Neurosci. Lett.* 396, 177–181. doi: 10.1016/j.neulet.2005.11.035

Nieuwenhuizen, F. M., Mulder, M., van Paassen, M. M., and Bülthoff, H. H. (2013). Influences of simulator motion system characteristics on pilot control behavior. *J. Guid. Control Dyn.* 36, 667–676. doi: 10.2514/1.59257

Opitz, B., Mecklinger, A., Friederici, A. D., and von Cramon, D. Y. (1999). The functional neuroanatomy of novelty processing: integrating ERP and fMRI results. *Cereb. Cortex* 9, 379–391. doi: 10.1093/cercor/9.4.379

Parmentier, F. B. (2014). The cognitive determinants of behavioral distraction by deviant auditory stimuli: a review. *Psychol. Res.* 78, 321–338. doi: 10.1007/s00426-013-0534-4

Polich, J. (2003). "Overview of P3a and P3b," in *Detection of Change: Event-Related Potential and fMRI Findings*, ed. J. Polich (Boston: Kluwer Academic Press), 83–98.

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clin. Neurophysiol.* 118, 2128–2148. doi: 10.1016/j.clinph.2007.04.019

Regan, M. A., Hallett, C., and Gordon, C. P. (2011). Driver distraction and driver inattention: definition, relationship and taxonomy. *Accid. Anal. Prev.* 43, 1771–1781. doi: 10.1016/j.aap.2011.04.008

Roye, A., Jacobsen, T., and Schröger, E. (2007). Personal significance is encoded automatically by the human brain: an event-related potential study with ringtones. *Eur. J. Neurosci.* 26, 784–790. doi: 10.1111/j.1460-9568.2007.05685.x

Salvucci, D. D., and Gray, R. (2004). A two-point visual control model of steering. *Perception* 33, 1233–1248. doi: 10.1068/p5343

SanMiguel, I., Corral, M.-J., and Escera, C. (2008). When loading working memory reduces distraction: behavioral and electrophysiological evidence from an auditory-visual distraction paradigm. *J. Cogn. Neurosci.* 20, 1131–1145. doi: 10.1162/jocn.2008.20078

Schröger, E., and Wolff, C. (1998). Attentional orienting and reorienting is indicated by human event-related brain potentials. *Neuroreport* 9, 3355–3358. doi: 10.1097/00001756-199810260-00003

Sirevaag, E. J., Kramer, A. F., Coles, M. G. H., and Donchin, E. (1989). Resource reciprocity: an event-related potentials analysis. *Acta Psychol. (Amst)* 70, 77–97. doi: 10.1016/0001-6918(89)90061-9

Sussman, E., Winkler, I., and Schröger, E. (2003). Top-down control over involuntary attention switching in the auditory modality. *Psychon. Bull. Rev.* 10, 630–637. doi: 10.3758/bf03196525

Ullsperger, P., Freude, G., and Erdmann, U. (2001). Auditory probe sensitivity to mental workload changes–an event-related potential study. *Int. J. Psychophysiol.* 40, 201–209. doi: 10.1016/s0167-8760(00)00188-4

Wetzel, N., and Schröger, E. (2014). On the development of auditory distraction: a review. *Psych J.* 3, 72–91. doi: 10.1002/pchj.49

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theor. Issues Ergon. Sci.* 3, 159–177. doi: 10.1080/14639220210123806

Wickens, C. D. (2008). Multiple resources and mental workload. *Hum. Factors* 50, 449–455. doi: 10.1518/001872008x288394

Wickens, C. D., and Alexander, A. L. (2009). Attentional tunneling and task management in synthetic vision displays. *Int. J. Aviat. Psychol.* 19, 182–199. doi: 10.1080/10508410902766549

Wickens, C. D., and Gopher, D. (1977). Control theory measures of tracking as indices of attention allocation strategies. *Hum. Factors J. Hum. Factors Ergon. Soc.* 19, 349–365.

Wickens, C. D., Isreal, J., and Donchin, E. (1977). The event related cortical potential as an index of task workload. *Hum. Factors Ergon. Soc. Annu. Meet. Proc.* 21, 282–286). doi: 10.1177/107118137702100404

Wickens, C. D., Kramer, A. F., and Donchin, E. (1984). The event-related potential as an index of the processing demands of a complex target acquisition task. *Ann. N Y Acad. Sci.* 425, 295–299. doi: 10.1111/j.1749-6632.1984.tb23550.x

Wickens, C. D., Kramer, A. F., Vanasse, L., and Donchin, E. (1983). Performance of concurrent tasks: a psychophysiological analysis of the reciprocity of information-processing resources. *Science* 221, 1080–1082. doi: 10.1126/science.6879207

Wickens, C. D., and Yeh, Y.-Y. (1983). The dissociation between subjective workload and performance: a multiple resource approach. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* 27, 244–248. doi: 10.1177/154193128302700314

Yago, E., Escera, C., Alho, K., Giard, M. H., and Serra-Grabulosa, J. M. (2003). Spatiotemporal dynamics of the auditory novelty-P3 event-related brain potential. *Brain Res. Cogn. Brain Res.* 16, 383–390. doi: 10.1016/s0926-6410(03)00052-1

Young, M. S., Brookhuis, K. A., Wickens, C. D., and Hancock, P. A. (2015). State of science: mental workload in ergonomics. *Ergonomics* 58, 1–17. doi: 10.1080/00140139.2014.956151

Zollner, H. G. H., Pool, D. M., Damveld, H. J., Van Paassen, M. M., and Mulder, M. (2010). "The effects of controlled element break frequency on pilot dynamics during compensatory target-following," in *AIAA Modeling and Simulation Technologies Conference* (Toronto, ON), 790–801.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Scheer, Bülthoff and Chuang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution and reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Human Neuroscience

METHODS
published: 12 November 2015
doi: 10.3389/fnhum.2015.00617

Toward a Wireless Open Source Instrument: Functional Near-infrared Spectroscopy in Mobile Neuroergonomics and BCI Applications

Alexander von Lühmann[1,2]*, *Christian Herff*[3], *Dominic Heger*[3] *and Tanja Schultz*[3]

[1] *Machine Learning Department, Computer Science, Technische Universität Berlin, Berlin, Germany,* [2] *Institute of Biomedical Engineering, Karlsruhe Institute of Technology, Karlsruhe, Germany,* [3] *Cognitive Systems Lab, Karlsruhe Institute of Technology, Karlsruhe, Germany*

Brain-Computer Interfaces (BCIs) and neuroergonomics research have high requirements regarding robustness and mobility. Additionally, fast applicability and customization are desired. Functional Near-Infrared Spectroscopy (fNIRS) is an increasingly established technology with a potential to satisfy these conditions. EEG acquisition technology, currently one of the main modalities used for mobile brain activity assessment, is widely spread and open for access and thus easily customizable. fNIRS technology on the other hand has either to be bought as a predefined commercial solution or developed from scratch using published literature. To help reducing time and effort of future custom designs for research purposes, we present our approach toward an open source multichannel stand-alone fNIRS instrument for mobile NIRS-based neuroimaging, neuroergonomics and BCI/BMI applications. The instrument is low-cost, miniaturized, wireless and modular and openly documented on www.opennirs.org. It provides features such as scalable channel number, configurable regulated light intensities, programmable gain and lock-in amplification. In this paper, the system concept, hardware, software and mechanical implementation of the lightweight stand-alone instrument are presented and the evaluation and verification results of the instrument's hardware and physiological fNIRS functionality are described. Its capability to measure brain activity is demonstrated by qualitative signal assessments and a quantitative mental arithmetic based BCI study with 12 subjects.

Keywords: open source, functional near-infrared spectroscopy (fNIRS), brain computer interface (BCI), modularity, wearable devices, neuroergonomics

OPEN ACCESS

Edited by:
Srikantan S. Nagarajan, University of California, San Francisco, USA

Reviewed by:
Hasan Ayaz, Drexel University, USA
Barak A. Pearlmutter, National University of Ireland Maynooth, Ireland
Laurens Ruben Krol, Technische Universität Berlin, Germany

***Correspondence:**
Alexander von Lühmann
a.vonluehmann@campus.tu-berlin.de

Received: *26 May 2015*
Accepted: *26 October 2015*
Published: *12 November 2015*

Citation:
von Lühmann A, Herff C, Heger D and Schultz T (2015) Toward a Wireless Open Source Instrument: Functional Near-infrared Spectroscopy in Mobile Neuroergonomics and BCI Applications. Front. Hum. Neurosci. 9:617. doi: 10.3389/fnhum.2015.00617

1. INTRODUCTION

Functional Near-Infrared Spectroscopy (fNIRS) is an increasingly established technology pioneered by Jöbsis (1977) that allows non-invasive, comparatively low-cost, compact and hazard-free continuous measurement of cerebral oxygenation levels using near-infrared light.

While first generation instruments were rather bulky and expensive, using Laser Diodes with Photo Multiplier Tubes (PMTs) (Cope and Delpy, 1988; Cope, 1991; Rolfe, 2000;

Schmidt et al., 2000) and later Avalanche Photo Diodes (APDs) (Boas et al., 2001; Coyle et al., 2004, 2007), today's devices often take advantage of Light Emitting Diodes (LED) and Photo Diodes (PDs) (Vaithianathan et al., 2004; Bunce et al., 2006; Chenier and Sawan, 2007; Ayaz et al., 2013; Safaie et al., 2013; Piper et al., 2014) which allow safe, more compact and mobile applications. After the initial development of laboratory and bedside-monitoring devices for monitoring of local oxygenation levels e.g., in newborn infants (Cope and Delpy, 1988; Cope, 1991), in the 2000s many research groups focused on the design of imaging instruments for brain activity mapping from topographic information [functional Near-Infrared Imaging (fNIRI)] (Schmidt et al., 2000; Boas et al., 2001; Vaithianathan et al., 2004). Recently, fNIRS and fNIRI have entered neuroscience as a reliable and trustworthy research tool for research based on investigating groups of subjects (Scholkmann et al., 2014), offering potentially complementary information to fMRI, PET and EEG (e.g., oxygenation information or cytochrome oxidase as marker of metabolic demands; Strangman et al., 2002). But also in adjacent fields such as Brain Computer Interfaces (BCI) and neuroergonomics, defined as the study of the human brain in relation to performance at work, (Parasuraman, 2003, 2011), fNIRS technology opens new possibilities (see e.g., Matthews et al., 2008, for an introduction in hemodynamics for Brain-Computer Interfaces). It is increasingly built for and used in single-trial fNIRS applications for BCIs for control (Naseer and Hong, 2013; Schudlo and Chau, 2015) and rehabilitation (Kanoh et al., 2009; Yanagisawa et al., 2010) and has successfully been used for cognitive workload assessment (Son and Yazici, 2006; Ayaz et al., 2012), brain dynamics monitoring during working memory training and expertise development (Ayaz et al., 2013), hybrid NIRS-EEG based signal processing tasks (Safaie et al., 2013; Putze et al., 2014) and recently also in combination with trans-cranial direct current stimulation (tDCS; McKendrick et al., 2015). Furthermore, fNIRS is found to be a promising multimodal expansion to EEG-based BCI (Pfurtscheller et al., 2010; Biessmann et al., 2011; Fazli et al., 2012). The major limitation of fNIRS is the relatively slow onset of hemodynamic processes. However, especially in the field of passive BCI (Zander and Kothe, 2011), reaction times do not necessarily have to be extremely fast.

An increasing number of approaches using fNIRS in the field of mobile brain imaging (e.g., Piper et al., 2014) and neuroergonomics (Fairclough, 2009) shows the demand for wireless, miniaturized and customizable fNIRS technology. So far, researchers either relied on costly and mainly static commercial devices, or designed their own fNIRS equipment from scratch. Trying to overcome the restrictions of commercial tabletop instruments, the latter was done by groups such as Ayaz et al. (2013), Safaie et al. (2013), Lareau et al. (2011) and Atsumori et al. (2007), using LEDs with Si PDs/APDs for new generation instruments that generally enable a mobile use. To the best of our knowledge however, only very few of even newer generation devices (Safaie et al., 2013) are truly miniaturized, stand-alone, unobtrusive and mobile and can be carried on the body without a backpack while still enabling free movement and data transmission/processing at the same time, as often external static instrumentation such as DAQ-equipment, lock-in amplifiers and power sources are required. Also, in many cases, signal extraction technologies like lock-in amplification seem to be sacrificed for the sake of miniaturization or complexity. Probe and attachment designs proposed in the last years, such as the use of flexible PCBs (Vaithianathan et al., 2004; Bozkurt et al., 2005; Bunce et al., 2006; Son and Yazici, 2006; Rajkumar et al., 2012), eeg-cap like optodes (Kiguchi et al., 2012; Piper et al., 2014) and mechanical mounting structures (Coyle et al., 2007) are usually limited to static applications and/or in case of flexible PCB to a fixation on the forehead due to obstruction by hair. An often reported issue in the field of mobile applications, that seems not to be resolved satisfactorily so far, is the optode attachment to the head for both stable optical contact, sufficient light levels and comfortable wearing.

While the recent trend of new system designs for portability and mobility can also increasingly be observed in commercial devices, researchers will always need custom solutions for innovative approaches. To help reducing the time and effort in these cases, we present the design and a first evaluation of a configurable, miniaturized, modular, fully mobile (wireless) multichannel fNIRS system that is provided open source on www.opennirs.org under a CC BY-NC 4.0 license with a detailed documentation. It is a customizable low-cost research tool that enables both stand-alone use and the combination with custom or external DAQ equipment. Also, the device makes use of a new detailed spring-loaded optode fixation concept to tackle the above mentioned optode attachment issue.

2. MATERIALS AND METHODS

2.1. Instrument Requirements

We identified aspects that are crucial to be fulfilled for a fNIRS device in the context of mobile BCI and neuroergonomics. Besides the criterion for the hardware to be comparatively low cost, these can be assigned to four groups:

- **Usability:** Miniaturization and mobility of the device, unobtrusiveness and robustness of the optode attachment.
- **Signal Quality:** Low inter channel crosstalk, low drifts of light sources and overall system signals, high signal sensitivity/amplifier precision, robustness to background light and high dynamic range.
- **Safety:** Low heat development, harmless light intensities and galvanic isolation to power lines.
- **Configuration/Customization:** Scalability of channel number, modularity, configuration of light intensities and receiver gain, interface to custom hard-/software.

The following subsections will provide detailed information on our approach to fulfill these requirements on a concept, hardware and software level.

2.2. Instrumentation Design

2.2.1. System Concept

The system concept of the modular open instrument is shown in **Figure 1**. It consists of one or more stand alone 4-channel

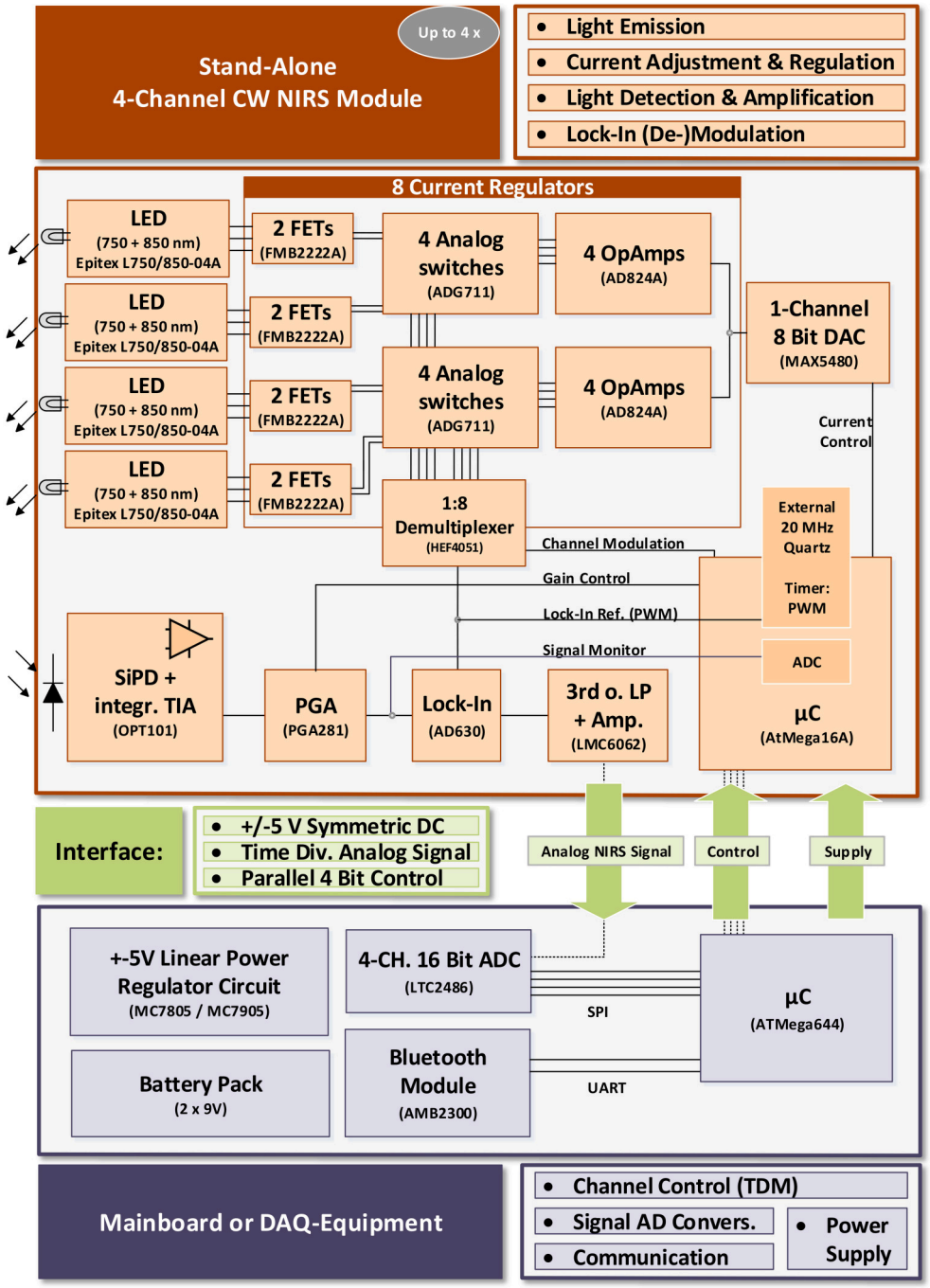


FIGURE 1 | System Concept.

Continuous Wave NIRS modules and a mainboard. Each module is controlled by the mainboard via a simple parallel 4 Bit control interface. The mainboard provides the power supply rails, AD-conversion of the NIRS signals and an UART communication interface and can be replaced by any custom data acquisition (DAQ-) equipment when the control interface and symmetric ±5 V power rail are supplied. This enables full customization of the instrument with respect to physical channel number, power consumption and conversion rate and depth, while spatially distributing the hardware components (and weight), and performing local hardware signal amplification and processing, thus minimizing noise and interferences.

The fNIRS modules were designed considering the current understanding of fNIRS instrumentation technology as reviewed

by Scholkmann et al. (2014) and others (Obrig and Villringer, 2003; Son and Yazici, 2006) with special regard to hardware design and wavelength-selection for SNR maximization/crosstalk minimization and considering potential hazards as identified by Bozkurt and Onaral (2004).

Each module provides four dual wavelength fNIRS channels using 750 and 850 nm multi-wavelength *Epitex L750/850-04A* LEDs. While the LEDs have a broader emission spectrum ($\Delta\lambda = 30/35$ nm) than sharp peaked laser diodes (typically $\Delta\lambda \approx 1$ nm), their incoherent and uncollimated light allows for a higher tissue interrogation intensity and direct contact with the scalp due to less heating and is safer with the human eyes.

The LED current is regulated by adjustable current regulator circuits based on high precision amplifiers (*Analog Devices AD824A*) and field effect transistors (FMB2222A). Channel activation and current modulation for lock-in amplification is performed by analog switches (*Analog Devices ADG711*) that are accessed via an analog 1:8 demultiplexer (*NXP HEF4051*). After tissue interrogation, NIR light is detected by a central Si photo detector with integrated trans-impedance amplifier for output noise minimization (*Texas Instruments OPT101*, 1 MΩ feedback resistor, bandwidth 14 kHz) and is then amplified and lock-in demodulated (using Analog Devices AD630). An 8 Bit *Atmel Corp. AtMega16A* microcontroller's PWM module creates the 3.125 kHz square wave reference for lock-in (de-)modulation using an external 20 MHz crystal for jitter minimization. It also processes incoming control signals from the 4 Bit control interface and operates and configures the on board hardware. For adjustment of the LED currents, an 8 Bit digital-to-analog converter (DAC; *Maxim MAX5480*) is implemented. It supplies the voltage level at the current regulator inputs that is the command variable for the current regulation level. A programmable gain amplifier (*Texas Instruments PGA281*) is implemented for pre lock-in amplification of the detected NIR signal with a variable gain from $G = 0.688$ to 88.

During lock-in demodulation, the signal is filtered by a 3rd-order Butterworth low-pass and is then again amplified ($G = 5.1$) and stabilized by a set of two high precision amplifiers (*Texas Instruments LMC6062*) before leaving the fNIRS module for external AD conversion.

The system is designed for Time-Division Multiplexing (TDM) of the fNIRS channels. This is a trade-off between minimizing inter-channel crosstalk, heating (Bozkurt and Onaral, 2004) and battery consumption on the one hand and sacrificing SNR, which is limited by the width of the applied time windows. For demultiplexing of the locked-in output branches, a variable (sample rate dependent) dwell time is inserted after each onset of a single channel activation before sampling the steady state photo detector signal on the mainboard or with custom DAQ equipment.

Configurable PGA gain ($G = 0.6875$–88) and LED-intensity (256 DAC levels) in combination with a feedback "signal monitor" line allow the signal dependent adaption for maximum amplification in the lock-in demodulation process without reaching the dynamic range limit of one of the components.

Modularity: The above described design of the fNIRS modules allows operation in many configurations—only requiring compatibility with the above mentioned interface consisting of 4 Bit control, power supply and analog output. For an extension of the total channel count, several modules can be used. Changes in set-up and module count only affect the control unit and its routines chosen by the user, which activate the time division multiplexed channels and convert the analog fNIRS signals from the modules:

- As the objective of this work was the design of miniaturized fNIRS modules for mobile applications, a microcontroller (*Atmel AtMega644*) based mainboard was developed for mobile data acquisition and module control. Using a 4 channel 16 Bit analog-to-digital converter (ADC; *Linear Technologies LTC2486*) and a Bluetooth wireless controller (*Amber Wireless AMB2300*), the mainboard acquires the fNIRS signal(s) from up to 4 modules (16 channels), transmits the data to a computer via serial protocol and processes incoming user controls. To scale the number of channels, the user connects the desired number of modules and configures the channel administrator routine on the mainboard's microcontroller (see also Section 2.2.3). The symmetric ±5 V power rail is created from battery DC voltage using a stabilized linear power regulator circuit (based on *ON Semiconductor MC7805 and MC7905* ICs). Running on batteries and using only low voltages also ensures user safety.
- The mainboard is a placeholder for any (custom) peripheral acquisition and control hardware. With DAQ-devices providing digital I/Os and an external power supply, any number of modules (limited by the desired dwell time and sampling rate) can be used and controlled by control- and acquisition routines written and customized by the user.

2.2.2. Selection of Hardware Design Aspects

***Emitter Branch*:** For a high accuracy of the fNIRS instrument, a careful design of the NIR-light emitting circuit is crucial, as fluctuations in the radiation intensity cannot be discriminated from changes in absorption due to changes in chromophore concentrations in the tissue.

To keep the current through the LED semiconductor junctions constant and independent from variations in supply voltage and temperature, and at the same time allow intensity adjustment and current modulation for the lock-in amplification process, a customized current regulator circuit was designed (see **Figure 2**).

Similar to a solution proposed by Chenier and Sawan (2007), an analog switch is used in the OpAmp based regulation circuit for square-wave modulation of the current. However, instead of disrupting the regulation process at the transistor base, analog switches (ADG711) are used at the inputs of the regulator circuits to pull the regulator inputs low when deactivated. fNIRS channel activation and modulation is thus realized by simply feeding through the square wave reference to the corresponding current regulator switch selected by the multiplexer.

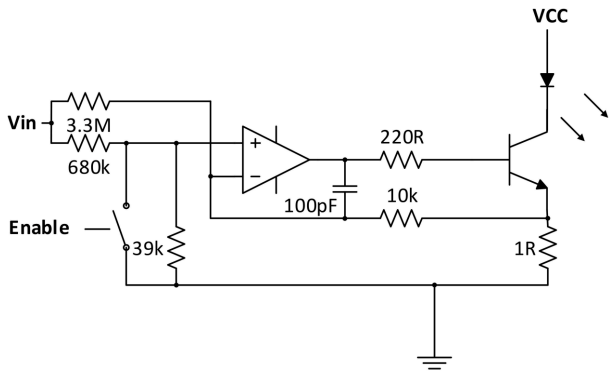


FIGURE 2 | Current regulator/modulator circuit.

As the regulator is modulated in the kHz-range, over- and undershoots influence the ideally square-wave shape of the current. To optimize the shape, a passive negative RC feedback was added and evaluated for best performance.

***Receiver Branch*:** The receiver branch was designed to maximize SNR by minimizing noise influences from shot, thermal and $1/f$ noise, dark currents and stray light from external light sources.

Shot noise is based on the quantum nature of the photons and therefore unavoidable and, for detectors without internal amplification, proportional to the square root of the average incident intensity (Scholkmann et al., 2014). To maximize SNR, the instrument is operated using the maximum NIR-light intensity level for the current regulators that is feasible in the experimental situation. Opaque cell rubber tubes are used to cover the sides of the NIR emitters and detector and the fNIRS module housing is covered with opaque paint to minimize shot noise influences from background radiation.

To reduce thermal noise influences, a Si photo diode with integrated trans-impedance amplifier circuitry (OPT101) was selected for detection. Lock-in extraction of the detected signal further reduces stray light, dark current and 1/f noise influences. Placing the PGA between the detection and lock-in extraction unit enables maximum pre-amplification of the signal while amplifier noise components added in the amplification process are reduced by the subsequent lock-in demodulation. Non-physiological high frequency components of the signal are attenuated by the 3rd order low pass filter of the lock-in demodulation unit.

2.2.3. Interfaces and Software Design

Figure 3 shows the software concept. The fNIRS module software sets up hardware components (PGA, DAC, MUX,...) and is controlled by an interrupt-based architecture that receives its control signals from the 4 Bit parallel interface. Therefore, interface operation and analog signal conversion can be done by the mainboard or any custom or standard DAQ-equipment with 4Bit programmable digital outputs (such as e.g., NI USB600x series). Using the mainboard, a channel administration routine both supervises data acquisition and acts as interface between the fNIRS modules and the PC by processing received user

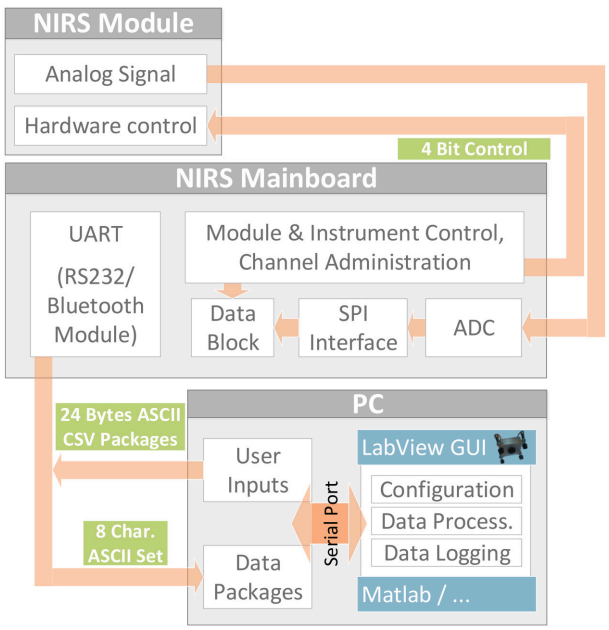


FIGURE 3 | Software and Interface Concept. Stand alone fNIRS module operated via parallel control interface by the mainboard or any custom control and data acquisition device. Function of the 4 Bit interface (3:RST, 2:TRIG, 1:CH1, 0:CH0): Bits CH1:CH0 select one of the four physical NIRS channels. A rising edge on the TRIG line activates the selected channel, always beginning with wavelength 750 nm of the corresponding LED. Each subsequent rising edge toggles the activation between 750 and 850 nm. When the RST line is pulled up, all channels are turned off. The next rising edge on the TRIG line starts the process again, beginning with 750 nm.

commands (configuration, start, stop...), translating them into signals for the 4 Bit fNIRS module interface(s) and sending acquired data packages via the UART interface. On the PC's operating system side, the user can control the instrument and directly read out the data packages in ASCII CSV format via a simple serial port command console or access the serial port with any software such as LabView or Matlab. A LabView graphical user interface was developed for easy configuration and control as well as display and logging of raw and modified Beer-Lambert Law data.

2.2.4. Mechanical and Probe Design

In the fNIRS instrument's mechanical design, the idea of modularity/scalability and robust fixation is continued by providing independent custom 3D printed solutions for the single fNIRS modules and the mainboard:

The Mainboard, Bluetooth module and batteries are worn on the upper arm of a subject in a chained multiple-unit housing (see also **Figure 5**, in the next section).

For the single fNIRS modules, a new mechanical spring-loaded design was approached to optimize signal quality, sensitivity and light penetration depth together with easy and robust, adaptive fixation of the optodes (see **Figure 4**). Based on a spherical approximation of the head with diameter $D = 20$ cm, the central NIR light detector and the four NIR LEDs are placed perpendicular to the scalp with a source-detector distance of $d =$

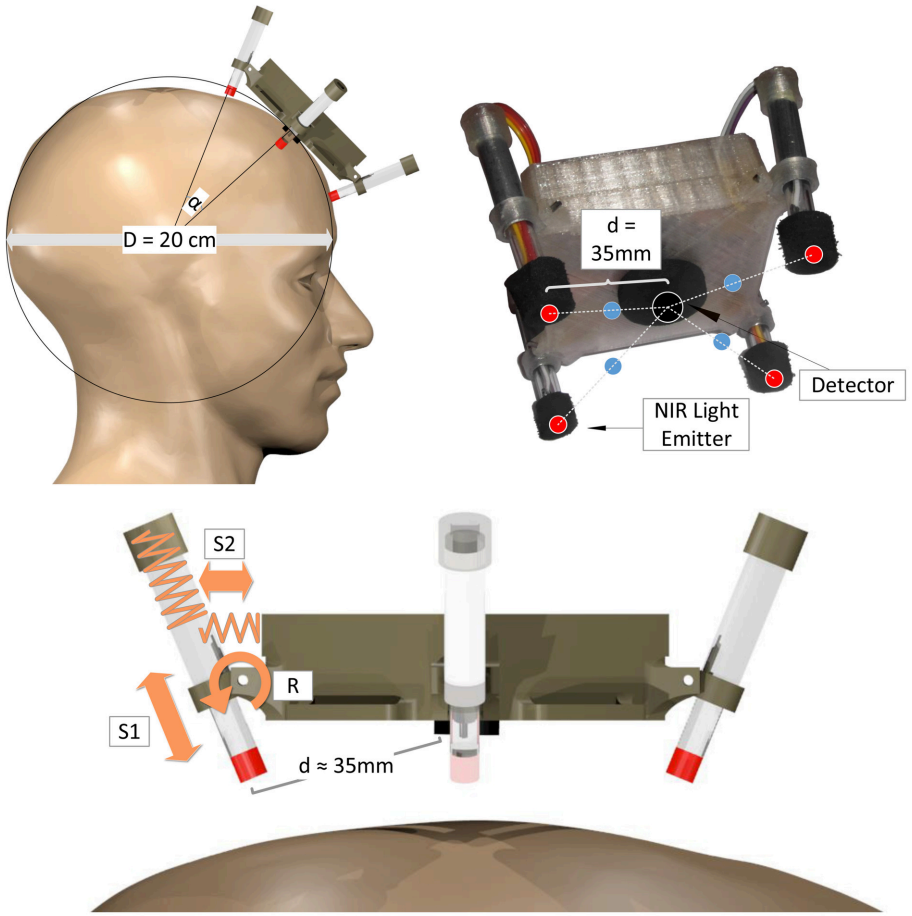


FIGURE 4 | Mechanical spring-loaded concept: Spherical head approximation (top left), geometric channel arrangement (red: NIR LEDs, black: photo detector, blue: measurement points of highest sensitivity; top right), spring-loaded mechanical design illustrated on one LED-holder (bottom). Spring S1 for alignment and buffering, spring S2 and rotatory joint R for perpendicular alignment.

35 mm. To enable perpendicular fixation of the emitters/detector and at the same time allow alignment to the natural unevenness of the head and its deviations from the spherical approximation, the NIR light LEDs are not stiffly connected to the module body housing but integrated in movable spring-loaded LED holders. These holders are based on two nested tubes that are spring-loaded against each other (S1) and against the module housing (S2) and are able to rotate around an axis (R): Spring S1 presses the LED toward the surface of the head, thus enabling alignment and preventing the loss of contact during movements. Spring S2 and the rotary joint R keep the LED perpendicular to the surface while enabling small deviations for comfort and alignment.

To minimize stray light influences and for cushioning purposes, the detector and emitters are encased by an opaque cell rubber tubing. To fixate a single module to the head, a flexible ribbon with hook-and-loop fastener can be used that is sewed to the module housing.

The mechanical concept was designed to allow the modules to be used on the forehead as well as over haired regions of the head: The single spring-loaded optodes are easily accessible due to their modular fixation without a cap or other concealing elements. This enables the user to manually brush aside obstructing hair from under the optodes for better optical contact. Even though we successfully conducted measurements over hairy regions of the head, it has to be pointed out that the usability of the modules on other regions than the forehead has not been proven under controlled conditions so far.

2.3. System Evaluation

2.3.1. Hardware Analysis

To enable a differentiated characterization of the instrument's hardware according to functional units, evaluation and analysis was split into emitter branch (current regulation and modulation), receiver branch (lock-in module), power supply stability and overall drift characteristics:

- *Current regulator/modulator speed and current shape/oscillation characteristics*: To evaluate and optimize the current regulator design characteristics for a stable and

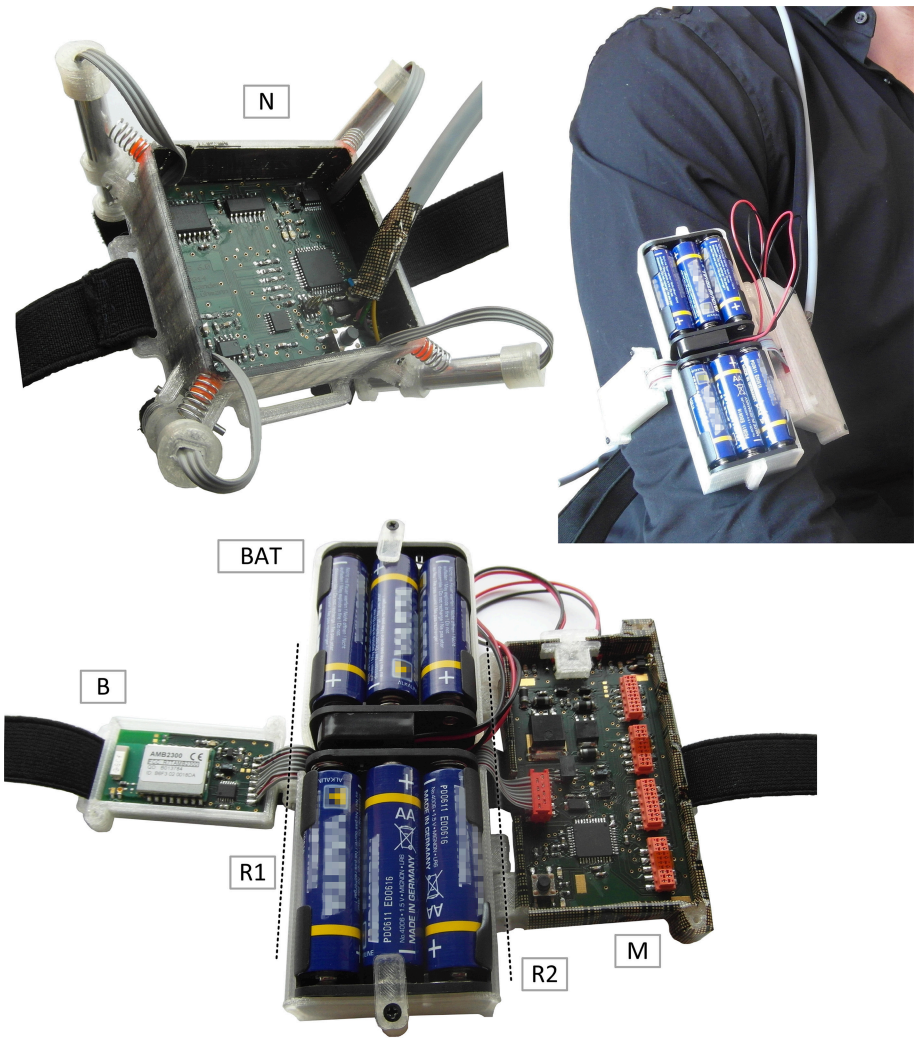


FIGURE 5 | Final System. (N): single 4 channel fNIRS module. Top right/bottom: Chained mainboard module. B, bluetooth module; BAT, batteries; R1/R2, Rotatory joints; M, mainboard.

minimally oscillating but steep square wave shape of the regulated current signal, both LTSpice simulations and measurements were conducted and the regulator design parameters iteratively improved using two high-precision operational amplifiers (Analog Devices AD824A and Linear Technologies LMC6064). To minimize transient oscillation and settling times, a negative feedback decoupling capacitor C was introduced to the regulator design. For the determination of its optimal value, the shape of the regulated square wave current signal was investigated in a range from $C = 0$ pF to $C = 330$ pF at different current levels.

- *Lock-in performance*: The sum of propagation delays that result from each hardware component in the emitter-detector-signal path leads to an overall phase shift between input and reference signal in the analog lock-in amplification process. Such a phase shift results in an attenuation of the signal during demodulation (Meade, 1982, 1983). To minimize this effect, all hardware elements in the signal path were selected with respect to high-speed/low delay times. The remaining overall phase shift $\Delta\Phi = \frac{\Delta t}{T} \cdot 2\pi$ between the reference signal (with period T) and the detected pre-amplified signal was measured before demodulation. Using the established straight forward mathematical model for square wave reference lock-in demodulation, as in Meade (1983), a phase shift dependent attenuation factor

$$A = cos(\Delta\Phi) \tag{1}$$

was used to estimate the resulting attenuation.

For an estimation of the receiver sensitivity using the noise equivalent power (NEP), dark voltage noise levels (no incident light to the photo detector) were measured at the output of the lock-in-module.

- *System drifts*: The following possible sources of system drift were considered: Changes in the 1 Ω LED current regulation resistance due to temperature changes, changes in the total

radiated power of the LEDs due to semiconductor junction temperature and changes despite constant currents and supply voltage variations. Changes in stray light, amplifier and thermal resistor noise are strongly suppressed by the lock-in amplification process. To minimize signal drifts resulting from changes in the 1 Ω current regulator resistance, Panasonic current sensing resistors with a low temperature coefficient of resistance ($TCR = \pm 50 \cdot 10^{-6}/^{\circ}\mathrm{C}$) were chosen.

The overall system drift of a single fNIRS module was specified with 20 min continuous acquisition windows of a single active channel at maximum intensity (100 mA) with the PGA set to $G = 44$ and the module being placed at a fixed position in an opaque closed box.

- *Mainboard power supply stability*: DC supply voltage drifts during 20 min signal acquisition periods and current modulation impacts on the supply voltage were evaluated. As the 100 mA (max.) square wave 3.125 kHz modulation can influence the power supply voltage stability and noise it can degrade the performance of the signal detection and amplification elements. Their output signals during active modulation were acquired while zero optical input to the photo detector was ensured by encasing the active LED with an opaque metal box. For customization, the layout of the fNIRS module allows both separate and common supply of the LED currents and module hardware.

2.3.2. Physiological Verification

Simple qualitative experiments were conducted using a channel at 10–20 point Fp1 to verify significant strength of physiological information in the raw signal and its power spectrum. Amongst others, visibility and strength of pulse artifacts are indicators for the signal quality and have been widely documented in fNIRS literature with the pulse artifact's amplitude being in the order of metabolic variations due to brain activity (Boas et al., 2004; Lareau et al., 2011; Scholkmann et al., 2014). Thus, with the fNIRS module pressed firmly against the head to reduce the sensitivity to scalp signals (decreased blood flow under the optodes), a clearly visible pulse artifact is a first indicator for sufficient signal quality to measure brain activation. The pulse rate was verified with conventional reference pulse measurements.

For verification and quantification of the device's capability to measure metabolic brain activity, a mental arithmetic BCI experiment was conducted with 12 subjects. In this experiment, it is shown that the measured hemodynamic responses can be classified on a single-trial basis, i.e., each trial can be classified as containing mental arithmetic or relaxation, instead of measuring only the difference in the average hemodynamic response.

Mental arithmetic tasks are known to illicit strong hemodynamic reactions in frontal brain areas and have been investigated in a variate of studies with fNIRS (Ang et al., 2010; Herff et al., 2013; Bauernfeind et al., 2014). Here, 30 trials of mental arithmetic data were recorded for each participant. During each 10 s trial, participants were asked to repeatedly subtract a number between 7 and 19 (excluding 10) from a number between 501 and 999). Both numbers were presented on a screen at a distance of roughly 50 cm. After each mental arithmetic trial, participants were asked to relax for 25–30 s. These pause intervals were indicated by a fixation cross on the screen. A longer resting period of variable length was included after 15 trials to allow participants to rest and drink. No data of these extended resting periods were used in our analysis.

The open fNIRS device was placed on the forehead and fixated around the head with the flexible ribbon with hook-and-loop fastener sewed to its housing. It was placed such that both active emitters were placed on the locations Fp1 and Fp2 of the international 10–20-system. The light detector was placed on AFz resulting in an emitter-detector distance of approximately 3.5 cm.

All subjects were informed prior to the experiment and gave written consent.

The signal processing of the recorded data was performed in a straight-forward and simple manner, since we focus on the developed hardware in this paper. More advanced methods have been shown to improve accuracies for classification in neuroimaging (Calhoun et al., 2001; Blankertz et al., 2008; Lemm et al., 2011; Heger et al., 2014). The raw optical densities were transferred to concentration changes of oxygenated and deoxygenated hemoglobin (HbO and HbR, respectively) using the modified Beer-Lambert Law (Sassaroli and Fantini, 2004). HbO and HbR values were then linearly detrended in windows of 300 s. Low frequency noise was attenuated by subtracting a moving average of the mean of 30 s prior and after every sample. Finally the data was low-pass filtered using an elliptic IIR filter with filter order 6 and a cut-off frequency of 0.5 Hz to reduce high-frequency systemic noise like pulse artifacts.

After preprocessing, trials were extracted based on the experiment timings. For the pause blocks, we extracted the last 10 s of the 25–30 s pause intervals, to ensure that hemoglobin levels have returned to baseline. For each mental arithmetic trial, we extracted 10 s of data starting 5 s after stimulus presentation, to ensure that the hemodynamic response has already developed. Labels were assigned to the trials referring to either mental arithmetics or pause data. For each trial, we extracted the slope of a straight line fitted to the HbO and HbR data of each channel as a feature. The line was fitted using linear regression with a least-squares approach. Slope features have been shown to work well in previous studies (Herff et al., 2014).

Evaluation was performed using a 10-fold cross-validation and classification by Linear Discriminant Analysis. In addition to the single trial analysis, the average hemodynamic response is calculated by averaging over all mental arithmetics or all pause trials.

3. RESULTS

The developed open modular multichannel fNIRS system (see **Figure 5**) proved functionality, fast set up and easy application in all testing conditions. Wearing the mainboard module on the upper arm and the fNIRS module on the head using flexible ribbons and hook-and-loop fastener, the user can move freely and is bothered minimally by the instrument while signal quality and robustness to movement showed promising results. It should be noted however, that the physiological signals used for evaluation results in this paper were acquired from sitting

subjects to reduce possible error sources and thus allow a more explicit first performance assessment of the new open source hardware.

The final instrument is characterized by:

- Modularity, customization and stand-alone functionality.
- Optimized adjustable current regulation and modulation with negative decoupling.
- A lock-in-based signal extraction module with programmable amplification.
- A 4-channel spring-loaded mechanical concept for fNIRS probe attachment to improve user comfort and robustness against movement artifacts.
- A microcontroller and Bluetooth based mainboard as interchangeable peripheral control, acquisition and transmission hardware with a Bluetooth range of max. 20 m (optimal open field conditions).

Low cost components were used for the design. The total cost of the instrument's hardware for one 4 channel fNIRS module and one mainboard mainly depends on PCB fabrication costs and is approximately 200 *EUR*/250 *USD*.

A full documentation including detailed descriptions, schematics, and evaluation can be found in the supplementary materials for this article/on the web: www.opennirs.org. In the following, we present the main evaluation results of the steps described in section 2.

3.1. Current Regulation/modulation Circuit

Mostly due to its higher slew rate, the AD824A showed a much faster current regulation and lower transient oscillations than the LMC6064. The experimental results for the minimization of oscillation and settling times with different decoupling capacitor values C (see **Figure 6**) showed higher settling oscillations for low C at low current levels, higher transient oscillation for high C at higher current levels and allowed the identification of the optimal C: $C = 100$ pF showed the best tradeoff between minimal oscillations and maximal edge steepness for all current levels.

3.2. System Drifts

The signal drift of a continuously active channel in Volts per second was calculated using linear least squares regression on the acquired 20 min. raw signals, yielding a negative drift coefficient of $C_D = -1 \cdot 10^{-6}$ V/s (measured fNIRS signals typically—dependent on the device configuration—being in the order of several hundred mV) and a respective long term stability coefficient of $< -0.42\%$ for both wavelengths. It was observed that independent from the fNIRS module, power supply heating on the peripheral hardware can add additional drifts of up to one order of magnitude by effecting the analog-to-digital converter. This points out the importance of careful selection/design of peripheral hardware for the acquisition of the fNIRS module's analog signal.

3.3. Lock-in Amplification, SNR and Dynamic Range

For an approximation of the total effective phase shift between reference and demodulator input signal in the lock-in unit, the delays between both signals were measured at the 50% levels of both respective rising (t_{dr}) and falling edges (t_{df}). It is the sum of times where the logical levels of both signals do not match and was measured as $\Delta t = t_{dr} + t_{df} = 18.5 + 7.2$ µs. To estimate the attenuation caused by non-phase-synchronous demodulation of the signal A, Equation (1) is used with the measured Δt and reference signal cycle duration $T = 320$ µs, and yields $A \approx 0.875$, which does not affect the overall accuracy significantly. Evaluation of the single component phase delays in the signal path revealed that further minimization approaches should first target the PGA ($\Delta t_{PGA} = 7.0 + 4.5$ µs).

For the evaluation of the detector's sensitivity and dynamic range, the mean dark voltage signal μ_d (no incident light on the

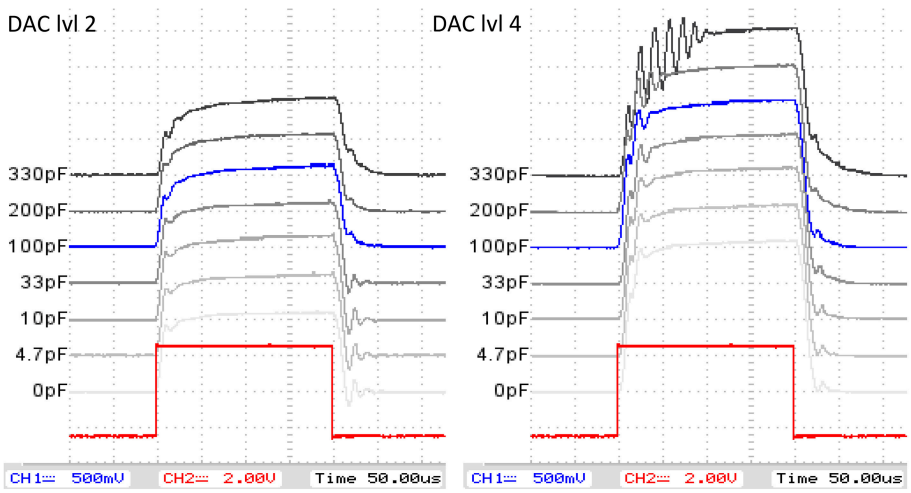


FIGURE 6 | Current regulator/moulator evaluation: 3.125 kHz PWM modulation reference (red) and regulated LED currents with different decoupling capacitor values and DAC levels (blue: selected value for design).

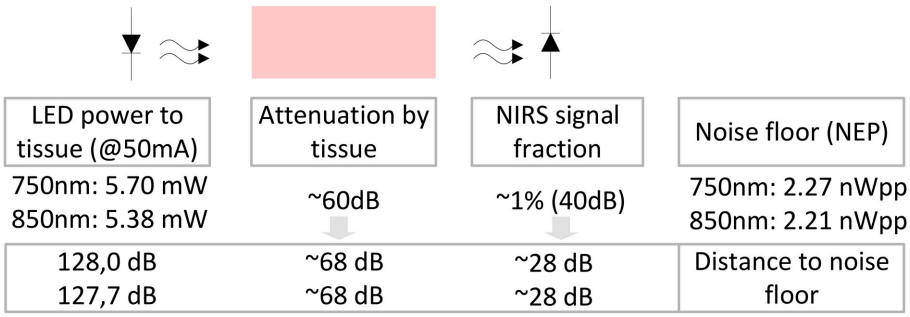


FIGURE 7 | Estimation of signal and noise in the instrument: With the NEPs identified to be 2.27/2.21 nWpp, the distance between these optical powers equivalent to the noise floor of the detection circuit and the measured powers incident to the tissue (5.70 mW/5.38 mW) at medium LED illumination can be determined to be approximately 128 dB/127.7 dB, respectively. The degree of optical loss of the incident light in the tissue is subject dependent. Here, we estimate it to be in the order of at least 40–60 dB. With the actual metabolic fNIRS signal being in the order of 1% of the measured optical signal, the distance of the fNIRS signals to the noise floor is approximately 28 dB.

photodetector) at the output of the lock-in amplifier and post amplification branch was measured to be $\mu_d = 0.101$ Vrms with a standard deviation of $\sigma_d = 3.99$ mVrms at a typical PGA gain of $G = 44$ and fixed lock-in filter gain of $G = 5.1$. Using the mean dark voltage plus standard deviation and the responsivities R_λ of the OPT101 photodiode ($R_{750} = 0.55$ V/μW and $R_{850} = 0.60$ V/μW), the Noise Equivalent Powers of the whole detector branch for a SNR of one

$$NEP_\lambda = \frac{\mu_d + \sigma_d}{R_\lambda \cdot G_{total}} \quad (2)$$

were estimated to be $NEP_{750} = 2.27$ nWpp $= 0.80$ nWrms and $NEP_{850} = 2.21$ nWpp $= 0.78$ nWrms.

The optical powers radiated by the LED at medium intensity ($I_F = 50$ mA) were measured to be 5.70 mW for 750 nm and 5.38 mW for 850 nm. Using these incident powers and the NEPs allows an estimation of the signal to noise distances (for an overview see **Figure 7**): The wavelength dependent ratio of incident light to light not longer detectable as its signal is drowning in noise, yields signal to noise distances of 128 dB_{750} and 127.7 dB_{850}. These distances are largely decreased by the optical loss due to tissue scattering and absorption that is subject dependent and assumed to be in the order of $>$ 60 dB. With the physiological fNIRS signal usually being around 1% of the measured optical signal, the distance between the fNIRS signal components and the noise floor of the detection circuit is further decreased by ≈40 dB and estimated to be in the order of 28 dB.

Saturation of the detection branch occurs, when the upper input voltage limit of the ADC, here 2.5 Vpp, is reached for the lowest PGA gain setting of $G = 0.6875$, which is the case at 1.296/1.188 μWpp incident light (750 nm/850 nm). Using these results, the minimum system dynamic range, expressed as the ratio of signal saturation to the NEPs, is estimated to be in the order of 55.13/54.6 dB. It should be stated, that the configuration of the LED intensities (25–100 mA) on the emitter side can further increase the dynamic range of the instrument. **Table 1** summarizes the performance characteristics.

TABLE 1 | Performance characteristics of the fNIRS Module.

Parameter	Value
OPT101 Dark noise voltage	300 μVrms @0.1–20 kHz
OPT101 Responsivity	$R_{750} \approx 0.55$ V/μW
	$R_{850} \approx 0.60$ V/μW
Noise equivalent power	$NEP_{750} \approx 0.80$ nWrms
	$NEP_{850} \approx 0.78$ nWrms
Power to tissue (at peak wavelength)	
$@I_{LED} = 50$ mA	$P_{750} = 5.70$ mW
	$P_{850} = 5.38$ mW
$@I_{LED} = 100$ mA	$P_{750} = 11.10$ mW
	$P_{850} = 10.30$ mW
Est. signal to noise distance	
$@I_{LED} = 50$ mA	≈ 28 dB
Eff. dynamic range	$>$ 55 dB
Signal drift	$< -1 \cdot 10^{-6}$ V/s/ $< -0.5\%$
Sampling rate	Variable, dwell time dependent

3.4. Mainboard Power Supply

The ±5 V DC supply voltage drift measurements showed a stable supply voltage of +4.959 V and −4.960 V with less than 500 μV total drift in 20-min measurement periods. Evaluation of the maximum impact of the current modulation on the detecting components via the power supply revealed that current modulation flanks can create a ±2 mV high-frequency(kHz) noise around the photo detector output baseline signal that is further amplified by the PGA to strong ±100 mV peaks (at $G = 44$) when supplying the LED current either directly from the battery or from the regulated +5 V rail for the other fNIRS module hardware. However, as the supply variations are synchronous with the signals and as high-frequency noise is effectively suppressed by the 3rd-order lock-in low-pass of

the fNIRS module, influences on the baseline of the lock-in demodulated signal were not observed.

3.5. Physiological Measurements

Qualitative physiological experiments showed very clear signals and proved the basic functionality of the instrument. **Figure 8** shows a representative raw signal (750 nm) during three mental arithmetics trials performed by a subject (a) and the power spectrum computed over the whole session for the same subject (b). The latter shows the typical power law appearance and peaks by systemic artifacts that have widely been reported for fNIRS signals in the literature (e.g., Fekete et al., 2011).

The average hemodynamic response (see **Figure 9A**) over all subjects of the mental arithmetics experiment shows the expected behavior, i.e., an increase in HbO peaking after approximately 10 s during mental arithmetics. During the average pause interval, HbO levels still slowly return to baseline after the preceding activation.

Discrimination between pause and mental arithmetics yielded an average of 65.14% accuracy. Of the 12 recorded participants, 9 yielded accuracies significantly higher than chance level (one-sided t-test, $p < 0.05$). Classification results for all participants can be seen in **Figure 9B**). In a similar study by Herff et al. (2013), mental arithmetics could be discriminated from pause with 71.17% using 8 channels and 67.26% when using only two channels at similar positions as in this study.

4. DISCUSSION AND CONCLUSION

4.1. Key Findings

In the beginning of this paper, we identified system requirements for mobile fNIRS based neuroergonomics/BCI applications. The results indicate, that the presented open source device satisfies the requirements.

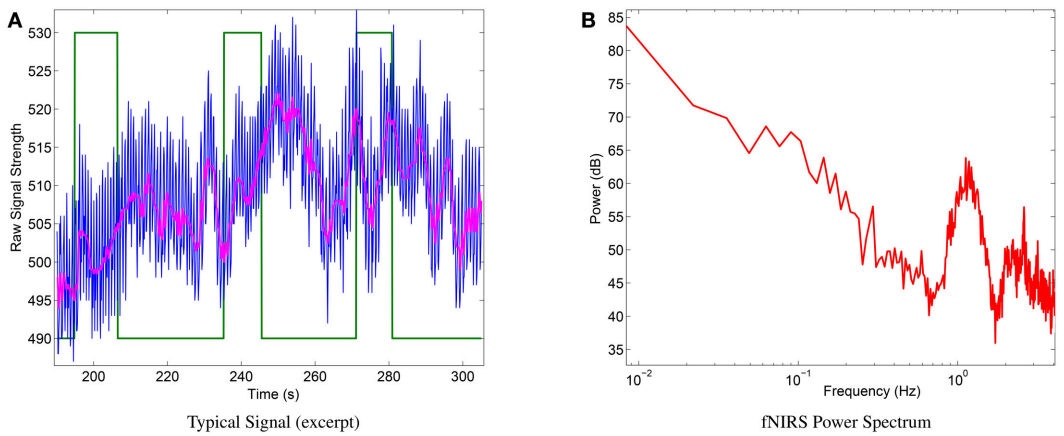


FIGURE 8 | (A) Excerpt of typical raw signal (blue) during mental arithmetics. Green line: binary label (high states: m. arithmetics. low states: relax), magenta line: median filtered signal). **(B)** Typical power spectrum of raw signal of the same subject and complete session, showing expected shape (power spectrum follows power law) and deviations caused by systemic artifacts.

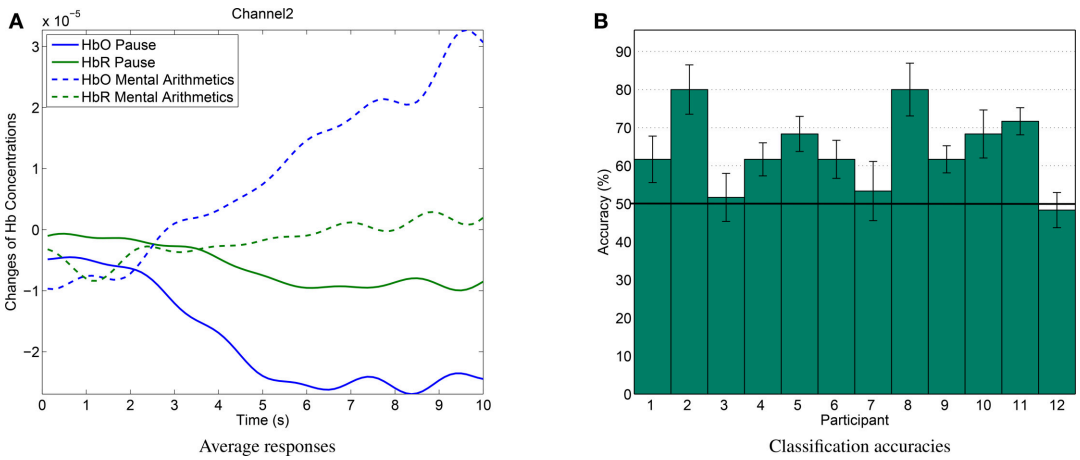


FIGURE 9 | (A) Average hemodynamic response over all subjects during mental arithmetics and pause. **(B)** Classification results for single-trial discrimination between pause and mental arithmetics. Whiskers indicate standard errors. Solid line shows chance level.

In the course of the experiments, both, experimentators and subjects, appraised the usability of the device to be high. Miniaturization of the modules and mobility through Bluetooth based wireless transmission allowed free movement, the use of commercial reference systems usually required longer preparation times for optode fixation and was often uncomfortable and static because of the weight of the optical fiber guides and the lack of cushioning of the optodes. In contrast, the new wearable system could be applied within several seconds and was generally perceived less cumbersome during the experiments.

The hardware evaluation results and physiological verification of the designed miniaturized fNIRS instrument indicated a sufficient signal quality and system performance for brain activity measurements with an approximated signal to noise distance of 28 dB. The lock-in amplifier, detector sensitivity, current modulation precision and drift evaluation of the device showed satisfying results comparable to other documented fNIRS devices. The physiological measurements showed the expected hemodynamic responses, classification accuracies in single-trial analysis exceeded chance level for 9 out of 12 participants and yielded results comparable to those measured with a commercial device in a similar study (Herff et al., 2013) using 2 of 8 channels at similar positions (65.14 vs. 67.26%). The open fNIRS device can thus be used for mobile fNIRS-based BCI and neuroergonomics applications.

Battery supply and wireless communication, low heating due to time multiplexing of the channels and the use of LEDs as light sources assured a safe usage of the device.

The scalable modular concept, configurable light intensities and detector amplification gains and the flexible parallel interface of the fNIRS modules allow easy customization and configuration of the hardware.

However, there are still several elements in the design that can be optimized to further improve instrument performance in the future.

4.2. Limitations and Next Steps

4.2.1. Mainboard/Data Acquisition and Control

An obvious but crucial component for the use of the fNIRS module is the data acquisition unit. When using custom hardware for data acquisition, the design and selection of the analog-to-digital converter (ADC) determine not only quantization depth but also the frequency resolution of the time division multiplexed fNIRS channels, as the ADC sampling rate has to be shared by the up to 4 active channels of one module. The ADC (LTC2486) first used on the mainboard offered 16 Bit conversion depth and exceptional DC accuracy but significantly limited time resolution due to a conversion time of type 80.3 ms. Additional experiments indicated that, using ADCs with significantly higher sample rate but lower resolution, down to 10 Bit quantization depth can suffice for reliable brain activation measurements. Future designs of the mainboard/DAQ hardware should therefore aim to use a better suited (faster) ADC to prevent the sampling frequency bottleneck. Here, the modular concept is advantageous, as the DAQ-unit can be customized and optimized independent from the hardware of the fNIRS modules.

Power supply and current modulation impact evaluation showed, that even though the implemented linear-voltage-regulator-based symmetric supply appeared to be sufficient, several improvements can be suggested for use with the fNIRS module:

To minimize crosstalk between the modulated NIR-LED current and the regulated ± 5 V supply voltage rail for the detection hardware, supplying the LEDs with a separate additional voltage regulator circuit is preferable over the use of a common regulator or direct battery connection in the design. Implementation of additional high-frequency filters and enhanced stabilization are also recommended in future approaches to reduce noise pickup from external sources and further minimize LED current modulation influences on the rest of the system. The use of voltage regulators with higher efficiency can further enhance battery life and decrease heating effects, which also can—dependent on the supplying and acquisition hardware's design and layout—influence system drifts.

4.2.2. fNIRS Module

The phase delay dependent attenuation of approximately 0.875 by the lock-in detector is acceptable as it does not significantly decrease overall system accuracy. However, it can be further minimized: To improve the lock-in performance, an analog adjustment of the PWM reference phase could be implemented for overall phase shift compensation. Alternatively, a potentially superior approach for a next-generation design would be digital lock-in demodulation based on a microcontroller/DSP. This bears several advantages: reduced cost of hardware components, reduced power consumption and an adjustable phase shift correction and thus higher precision.

The four channel set up per module using four LEDs and one photodiode was necessary for this first approach using a single-channel analog lock-in receiver branch for a simple interface in favor of modularity. However, to further reduce energy consumption and increase channel density, future approaches should utilize configurations with more PDs measuring simultaneously. Additionally, although the fNIRS module is already compact and provides stand-alone functionality, further miniaturization is possible. A next step will be the development of entirely stand-alone modules to redundantize peripheral hardware such as the mainboard. Integrating the above mentioned insights and data acquisition, digital lock in, power management and wireless transmission components onto a further miniaturized multichannel fNIRS module could enable even more applications in and out of the lab.

The instrument can be improved and evaluated in several more ways. However, providing this fNIRS device open source, we hope that aspects of this work will be helpful to further simplify and reduce time and effort in future custom fNIRS based mobile BCI and neuroergonomics approaches.

ACKNOWLEDGMENTS

We acknowledge support by Deutsche Forschungsgemeinschaft and Open Access Publishing Fund of Karlsruhe Institute of Technology.

REFERENCES

Ang, K., Guan, C., Lee, K., Lee, J., Nioka, S., and Chance, B. (2010). "A brain-computer interface for mental arithmetic task from single-trial near-infrared spectroscopy brain signals," in *International Conference on Pattern Recognition* (Istanbul), 3764–3767.

Atsumori, H., Kiguchi, M., Obata, A., Sato, H., Katura, T., Utsugi, K., et al. (2007). "Development of a multi-channel, portable optical topography system," in *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE* (Lyon), 3362–3364.

Ayaz, H., Onaral, B., Izzetoglu, K., Shewokis, P. A., McKendrick, R., and Parasuraman, R. (2013). Continuous monitoring of brain dynamics with functional near infrared spectroscopy as a tool for neuroergonomic research: empirical examples and a technological development. *Front. Hum. Neurosci.* 7:871. doi: 10.3389/fnhum.2013.00871

Ayaz, H., Shewokis, P. A., Bunce, S., Izzetoglu, K., Willems, B., and Onaral, B. (2012). Optical brain monitoring for operator training and mental workload assessment. *Neuroimage* 59, 36–47. doi: 10.1016/j.neuroimage.2011.06.023

Bauernfeind, G., Steyrl, D., Brunner, C., and Muller-Putz, G. R. (2014). "Single trial classification of fnirs-based brain-computer interface mental arithmetic data: a comparison between different classifiers," in *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE* (Chicago, IL: IEEE), 2004–2007.

Biessmann, F., Plis, S., Meinecke, F., Eichele, T., and Muller, K. (2011). Analysis of multimodal neuroimaging data. *Biomed. Eng. IEEE Rev.* 4, 26–58. doi: 10.1109/RBME.2011.2170675

Blankertz, B., Tomioka, R., Lemm, S., Kawanabe, M., and Muller, K.-R. (2008). Optimizing spatial filters for robust eeg single-trial analysis. *Signal Process. Mag. IEEE* 25, 41–56. doi: 10.1109/MSP.2008.4408441

Boas, D., Brooks, D., Miller, E., DiMarzio, C., Kilmer, M., Gaudette, R., et al. (2001). Imaging the body with diffuse optical tomography. *Signal Process. Mag. IEEE* 18, 57–75. doi: 10.1109/79.962278

Boas, D. A., Dale, A. M., and Franceschini, M. A. (2004). Diffuse optical imaging of brain activation: approaches to optimizing image sensitivity, resolution, and accuracy. *Neuroimage* 23(Suppl.1), S275–S288. doi: 10.1016/j.neuroimage.2004.07.011

Bozkurt, A., and Onaral, B. (2004). Safety assessment of near infrared light emitting diodes for diffuse optical measurements. *Biomed. Eng. Online* 3:9. doi: 10.1186/1475-925X-3-9

Bozkurt, A., Rosen, A., Rosen, H., and Onaral, B. (2005). A portable near infrared spectroscopy system for bedside monitoring of newborn brain. *Biomed. Eng. Online* 4:29. doi: 10.1186/1475-925X-4-29

Bunce, S., Izzetoglu, M., Izzetoglu, K., Onaral, B., and Pourrezaei, K. (2006). Functional near-infrared spectroscopy. *Eng. Med. Biol. Mag. IEEE* 25, 54–62. doi: 10.1109/MEMB.2006.1657788

Calhoun, V., Adali, T., Pearlson, G., and Pekar, J. (2001). Spatial and temporal independent component analysis of functional mri data containing a pair of task-related waveforms. *Hum. Brain Mapp.* 13, 43–53. doi: 10.1002/hbm.1024

Chenier, F., and Sawan, M. (2007). "A new brain imaging device based on fnirs," in *Biomedical Circuits and Systems Conference, 2007. BIOCAS 2007.* (Montreal, QC: IEEE), 1–4.

Cope, M. (1991). *The Application of Near Infrared Spectroscopy to Non Invasive Monitoring of Cerebral Oxygenation in the Newborn Infant*. Ph.D. thesis, Department of Medical Physics and Bioengineering, University College London.

Cope, M., and Delpy, D. (1988). System for long-term measurement of cerebral blood and tissue oxygenation on newborn infants by near infra-red transillumination. *Med. Biol. Eng. Comput.* 26, 289–294. doi: 10.1007/BF02447083

Coyle, S., Ward, T., and Markham, C. (2007). Brain-computer interface using a simplified functional near-infrared spectroscopy system. *J. Neural Eng.* 4, 219–226. doi: 10.1088/1741-2560/4/3/007

Coyle, S., Ward, T., Markham, C., and McDarby, G. (2004). On the suitability of near-infrared systems for next generation brain computer interfaces. *Physiol. Meas.* 25, 815–822. doi: 10.1088/0967-3334/25/4/003

Fairclough, S. H. (2009). Fundamentals of physiological computing. *Interact. Comput.* 21, 133–145. doi: 10.1016/j.intcom.2008.10.011

Fazli, S., Mehnert, J., Steinbrink, J., Curio, G., Villringer, A., Müller, K.-R., et al. (2012). Enhanced performance by a hybrid nirs-eeg brain computer interface. *Neuroimage* 59, 519–529. doi: 10.1016/j.neuroimage.2011.07.084

Fekete, T., Rubin, D., Carlson, J. M., and Mujica-Parodi, L. R. (2011). The nirs analysis package: noise reduction and statistical inference. *PLoS ONE* 6:e24322. doi: 10.1371/journal.pone.0024322

Heger, D., Herff, C., and Schultz, T. (2014). "Combining feature extraction and classification for fnirs bcis by regularized least squares optimization," in *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE* (IEEE), 2012–2015.

Herff, C., Heger, D., Fortmann, O., Hennrich, J., Putze, F., and Schultz, T. (2014). Mental workload during n-back task-quantified in the prefrontal cortex using fnirs. *Front. Hum. Neurosci.* 7:935. doi: 10.3389/fnhum.2013.00935

Herff, C., Heger, D., Putze, F., Hennrich, J., Fortmann, O., and Schultz, T. (2013). "Classification of mental tasks in the prefrontal cortex using fnirs," in *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE* (Osaka), 2160–2163.

Jöbsis, F. (1977). Noninvasive, infrared monitoring of cerebral and myocardial oxygen sufficiency and circulatory parameters. *Science* 198, 1264–1267. doi: 10.1126/science.929199

Kanoh, S., Murayama, Y.-M., Miyamoto, K.-I., Yoshinobu, T., and Kawashima, R. (2009). "A nirs-based brain-computer interface system during motor imagery: system development and online feedback training," in *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE* (Minneapolis, MN), 594–597.

Kiguchi, M., Atsumori, H., Fukasaku, I., Kumagai, Y., Funane, T., Maki, A., et al. (2012). Note: wearable near-infrared spectroscopy imager for haired region. *Rev. Sci. Instrum.* 83, 056101. doi: 10.1063/1.4704456

Lareau, E., Simard, G., Lesage, F., Nguyen, D., and Sawan, M. (2011). "Near infrared spectrometer combined with multichannel eeg for functional brain imaging," in *Medical Information Communication Technology (ISMICT), 2011 5th International Symposium on* (Montreux), 122–126.

Lemm, S., Blankertz, B., Dickhaus, T., and Müller, K.-R. (2011). Introduction to machine learning for brain imaging. *Neuroimage* 56, 387–399. doi: 10.1016/j.neuroimage.2010.11.004

Matthews, F., Pearlmutter, B. A., Ward, T., Soraghan, C., and Markham, C. (2008). Hemodynamics for brain-computer interfaces. *Signal Process. Mag. IEEE* 25, 87–94. doi: 10.1109/MSP.2008.4408445

McKendrick, R., Parasuraman, R., and Ayaz, H. (2015). Wearable functional near infrared spectroscopy (fnirs) and transcranial direct current stimulation (tdcs): expanding vistas for neurocognitive augmentation. *Front. Syst. Neurosci.* 9:27. doi: 10.3389/fnsys.2015.00027

Meade, M. L. (1982). Advances in lock-in amplifiers. *J. Phys. E* 15, 395. doi: 10.1088/0022-3735/15/4/001

Meade, M. L. (1983). *Lock-in Amplifiers: Principles and Applications*. London, UK: Peter Peregrinus Ltd.

Naseer, N., and Hong, K.-S. (2013). Classification of functional near-infrared spectroscopy signals corresponding to the right-and left-wrist motor imagery for development of a brain–computer interface. *Neurosci. Lett.* 553, 84–89. doi: 10.1016/j.neulet.2013.08.021

Obrig, H., and Villringer, A. (2003). Beyond the visible - imaging the human brain with light. *J. Cereb. Blood Flow Metab.* 12, 1–18. doi: 10.1097/00004647-200301000-00001

Parasuraman, R. (2003). Neuroergonomics: research and practice. *Theor. Issues Ergon. Sci.* 4, 5–20. doi: 10.1080/14639220210199753

Parasuraman, R. (2011). Neuroergonomics: brain, cognition, and performance at work. *Curr. Dir. Psychol. Sci.* 20, 181–186. doi: 10.1177/0963721411409176

Pfurtscheller, G., Bauernfeind, G., Wriessnegger, S. C., and Neuper, C. (2010). Focal frontal (de)oxyhemoglobin responses during simple arithmetic. *Int. J. Psychophysiol.* 76, 186–192. doi: 10.1016/j.ijpsycho.2010.03.013

Piper, S. K., Krueger, A., Koch, S. P., Mehnert, J., Habermehl, C., Steinbrink, J., et al. (2014). A wearable multi-channel fNIRS system for brain imaging in freely moving subjects. *Neuroimage* 85, 64–71. doi: 10.1016/j.neuroimage.2013.06.062

Putze, F., Hesslinger, S., Tse, C.-Y., Huang, Y., Herff, C., Guan, C., et al. (2014). Hybrid fnirs-eeg based classification of auditory and visual perception processes. *Front. Neurosci.* 8:373. doi: 10.3389/fnins.2014.00373

Rajkumar, E. R., Safaie, J., Gupta, R., Pattnaik, D., Abrishamimoghaddam, H., Grebe, R., et al. (2012). "Development of an autonomic portable single-board computer based high resolution nirs device for microcirculation analysis," in *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE* (San Diego, CA), 3235–3238.

Rolfe, P. (2000). *In vivo* near infra-red spectrophotometry. *Annu. Rev. Biomed. Eng.* 2, 315–354. doi: 10.1146/annurev.bioeng.2.1.715

Safaie, J., Grebe, R., Moghaddam, H. A., and Wallois, F. (2013). Toward a fully integrated wireless wearable eeg-nirs bimodal acquisition system. *J. Neural Eng.* 10:056001. doi: 10.1088/1741-2560/10/5/056001

Sassaroli, A., and Fantini, S. (2004). Comment on the modified beer-lambert law for scattering media. *Phys. Med. Biol.* 49, N225–N257. doi: 10.1088/0031-9155/49/14/n07

Schmidt, F., Fry, M., Hillman, E., Hebden, J., and Delpy, D. (2000). A 32-channel time-resolved instrument for medical optical tomography. *Rev. Sci. Instrum.* 71, 256–265. doi: 10.1063/1.1150191

Scholkmann, F., Kleiser, S., Metz, A. J., Zimmermann, R., Pavia, J. M., Wolf, U., et al. (2014). A review on continuous wave functional near-infrared spectroscopy and imaging instrumentation and methodology. *Neuroimage* 85, 6–27. doi: 10.1016/j.neuroimage.2013.05.004

Schudlo, L. C., and Chau, T. (2015). Single-trial classification of near-infrared spectroscopy signals arising from multiple cortical regions. *Behav. Brain Res.* 290, 131–142. doi: 10.1016/j.bbr.2015.04.053

Son, I.-Y., and Yazici, B. (2006). "Near infrared imaging and spectroscopy for brain activity monitoring," in *Advances in Sensing with Security Applications*, ed J. Byrnes (NATO Security Through Science Series-A: Chemistry and Biology, Springer), 341–372. Available on line at: http://www.springer.com/de/book/9781402042843?wt_mc=ThirdParty.SpringerLink.3.EPR653.About_eBook

Strangman, G., Culver, G., Thompson, J., and Boas, D. (2002). A quantitative comparison of simultaneous bold fmri and nirs recordings during functional brain activation. *Neuroimage* 17, 719–731. doi: 10.1006/nimg.2002.1227

Vaithianathan, T., Tullis, I. D. C., Everdell, N., Leung, T., and Gibson, A. (2004). Design of a portable near infrared system for topographic imaging of the brain in babies. *Rev. Sci. Instrum.* 75, 3276–3283. doi: 10.1063/1.1775314

Yanagisawa, K., Asaka, K., Sawai, H., Tsunashima, H., Nagaoka, T., Tsujii, T., et al. (2010). "Brain-computer interface using near-infrared spectroscopy for rehabilitation," in *Control Automation and Systems (ICCAS), 2010 International Conference on* (Gyeonggi-do), 2248–2253.

Zander, T. O., and Kothe, C. (2011). Towards passive brain–computer interfaces: applying brain–computer interface technology to human–machine systems in general. *J. Neural Eng.* 8:025005. doi: 10.1088/1741-2560/8/2/025005

Conflict of Interest Statement: The Review Editor Laurens Ruben Krol declares that, despite being affiliated with the same institution as the Author Alexander Von Lühmann, the review process was handled objectively. The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 von Lühmann, Herff, Heger and Schultz. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Human Neuroscience

OPINION
published: 31 May 2016
doi: 10.3389/fnhum.2016.00250

Why a Comprehensive Understanding of Mental Workload through the Measurement of Neurovascular Coupling Is a Key Issue for Neuroergonomics?

Kevin Mandrick[1], *Zarrin Chua*[1], *Mickaël Causse*[1], *Stéphane Perrey*[2] *and Frédéric Dehais*[1]*

[1] *Département Conception et conduite des véhicules Aéronautiques et Spatiaux, Institut Supérieur de l'Aéronautique et de l'Espace, Toulouse, France,* [2] *EuroMov, University of Montpellier, Montpellier, France*

Keywords: mental workload, mental resources, neurovascular coupling, neuroergonomics, electroencephalography, near-infrared spectroscopy

Raja Parasuraman, the father of *Neuroergonomics* (the crossroads of Ergonomics and Neuroscience, **Figure 1**) has opened the doors to new discoveries and techniques for advancing understanding of human behavior with the underlying brain mechanisms (Parasuraman, 1998). As of his death in 2015, a precise and objective definition of the concept of mental workload (MWL) had still not yet been formulated. In this opinion piece, we posit that MWL is associated through the measurement of neurovascular coupling (NVC); innovative neuroimaging methods is now capable of measuring such a phenomenon; all while highlighting Parasuraman's many contributions to this field.

OPEN ACCESS

Edited by:
Stephen Fairclough, Liverpool John Moores University, UK

Reviewed by:
Ranjana K. Mehta, Texas A&M Health Science Center, USA

***Correspondence:**
Frédéric Dehais
frederic.dehais@isae.fr

Received: *14 December 2015*
Accepted: *13 May 2016*
Published: *31 May 2016*

Citation:
Mandrick K, Chua Z, Causse M, Perrey S and Dehais F (2016) Why a Comprehensive Understanding of Mental Workload through the Measurement of Neurovascular Coupling Is a Key Issue for Neuroergonomics? Front. Hum. Neurosci. 10:250. doi: 10.3389/fnhum.2016.00250

BEYOND THE CONCEPT OF MENTAL WORKLOAD AND TOWARD MENTAL RESOURCES IN NEUROERGONOMICS

MWL measurement is an important issue in the Human Factors field, as seen through its ubiquitous presence in the literature. It is well acknowledged that an accurate assessment of MWL could help to reduce human error while improving human performance. The recently founded field of *Neuroergonomics* may help to reduce the ambiguity surrounding the MWL concept by providing data on its underlying neural processes. Neuroergonomics allows for the study of the human brain structure and function with respect to behavior during physical or cognitive performances in the workplace (Mehta and Parasuraman, 2013). The main goal of this interdisciplinary field is to integrate our understanding of the neural basis of cognition in relation to technologies and settings in complex daily life tasks.

However, Neuroergonomics does not yet provide a consensual and comprehensive explanation of the MWL. Despite being a roughly defined concept, there have been some formal attempts. Generally, MWL reflects how hard one's mind is working (under- over-loaded or occupied) at any given moment or how much mental effort it will cost for brain to meet given task demands (Parasuraman, 2003). Furthermore, Parasuraman and Caggiano (2002) and Kramer and Parasuraman (2007) defined MWL as a set of mental and composite brain states that modulate human performance in different perceptual, cognitive, and/or sensorimotor skills. It is also considered as a construct used to reflect the relation between the demands of the environment (input load), the human characteristics (capacities), and the task performances (output performance). However, the notion of MWL is dissociated from performance as suggested

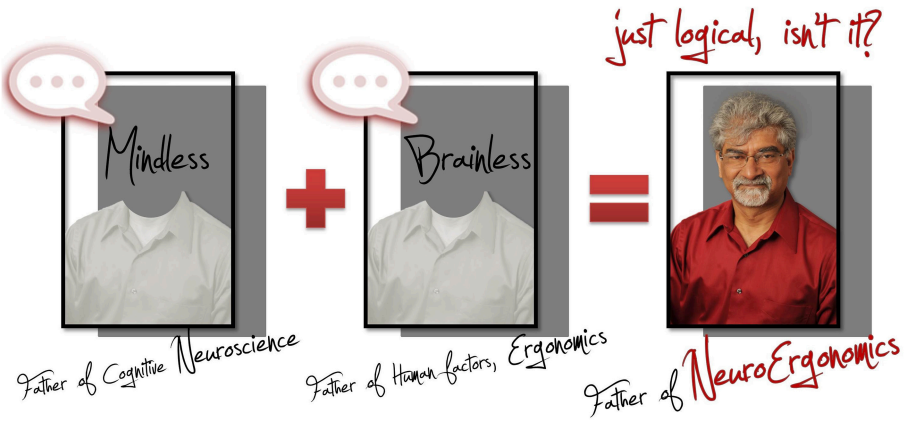


FIGURE 1 | Illustration of Raja Parasuraman as he himself wrote in an article: "One of us (Parasuraman, 1998, 2003) has therefore coined the term "neuroergonomics" to refer to the inclusion of neuroscience in Human Factors/Ergonomics (HF/E). Neuroergonomics can be defined as the study of brain and behavior at work. Traditionally, ergonomics has not paid much attention to neuroscience or to the results of studies of the brain mechanisms underlying human perceptual, cognitive, affective and motor processes. To paraphrase the philosopher Mario Bunge (1980), until recently psychology (and HF/E) has been "brainless," whereas neuroscience has been "mindless." At the same time, neuroscience and its more recent offshoot, cognitive neuroscience, have been only partially concerned with whether their findings bear any relation to human functioning in real (as opposed to laboratory) settings, with the exception of applications to clinical disorders. Neuroergonomics is a response to this twin disregard.

by Ayaz et al. (2012). MWL presupposes that the consumption of true brain resources supports brain activity during work, suggesting a possible link between MWL and the key concept of *mental resources*. These two concepts can be treated by the intensity of the mental costs and be measured by the mental effort of performing tasks to predict operator performance. As stated by Cain (2007) "*As such, [MWL] is an interim measure and one that should provide insight into where increased task demands.*" Therefore, it is not possible to define MWL without also clearly characterizing mental resources.

Though it is generally admitted that mental resources are appreciable, multiple, independent, and limited (Wickens, 2008), most studies remain vague on their exact nature. One perspective is to think of mental resources as neural pathways. However, this oversimplification ignores the fact that mental resources exists in other forms. As a metaphor, an army may have efficient firepower, but without ammunition, a supply corps, and roads, it is useless. Similarly, the army of the brain has mental resources composed of neural pathways, energy supply, and irrigation (communication channel) to fuel mental effort, implemented by the mobilization of neurophysiological cellular processes in the operator's brain.

ENERGY MOBILIZATION OF NEUROVASCULAR COUPLING FOR THE OPERATOR'S BRAIN MACHINERY

The absence of consideration of the neurophysiological mechanisms in Neuroergonomics is certainly due to the difficulty in investigating them. Yet, there are real energy mobilizations that occur within the operator's *brain machinery* across several cellular levels to meet task demand. As previously compared to a super calculator or a computer, the brain machinery supports mental processors that need substantial and constant energy requirements. But the human brain is devoid of intra-cellular capacity for energy storage in oxygen, lactate, and glucose (even if small parts of glycogen exist). Fortunately, the demand for high-metabolic energy of the brain tissue is mainly regulated by complex but adequate energetic substrate delivery via a dense and redundant network of microvessels. Hence, metabolic demands are orchestrated by the blood supply hemodynamic response.

Since the first discoveries by Roy and Sherrington (1890), it has been possible to better understand the close spatiotemporal dynamics between the electrical activity of neuronal cells and the hemodynamic phenomenon that boost the local bloodstream circulation in localized arterioles and capillaries. The intimate neurofunctional relationship that concomitantly links the metabolically active neurons with the increasing oxygenation of the blood flow near of these cells reflects the functional hyperemia and is more widely known as *neurovascular coupling* (NVC). Simply, NVC is a tight temporal association of the neuronal activity with regional cerebral blood flow delivery. Understanding the fundamental cellular mechanisms underlying NVC is necessary to measure a dimension of the local brain machinery expenditure at work. The appraisal of the energetic costs required by NVC implies the assessment of mental resources. For instance, when an operator is engaged in a task, the mobilization of the neural pathways needs a synergistic support of massive astrocyte glial cells to fuel neurons and interneurons with oxygen and nutriments furnished by close capillaries.

NVC is observable due to changes in neuronal-astroglial and microvasculature activities, which occur in several steps. First, the measurable electrical neuronal activity (spiking and postsynaptic potential activity) is accompanied by synaptic neurotransmitter

release (glutamate, GABA) with a neuronal-astroglial regional cerebral metabolic rate of oxygen consumption, mainly for regional cerebral metabolic rate of glucose demand. Second, this activity induces a cascading pathway involving the production and the release of powerful vasodilator metabolites by neurons and astrocytes and drives a chemical signal up to the vascular smooth muscle and pericytes cells along the microvessels which dilate the microvasculature. Third, the microvessels dilatation significantly modulates the regional cerebral blood activity (flow, volume, and oxygenation) which greatly exceeds the neuronal-astroglial oxygen requirements, and results in a measurable overabundance of blood flow, hence, a local hyperoxygenation. Yet, the role of NVC as it contributes to the comprehension of the energy mobilization in response to mental resources is not common knowledge. The cellular measures of energy production, delivery, and utilization are crucial to understanding and interpreting NVC activity. How to clearly establish the role of NVC into the operator's brain machinery? One possible way would be to associate the level of correlates of NVC while interpreting the degree of task demand. It seems thus fairly possible that an accurate measurement of NVC, spatially and temporally and in terms of amplitude, would be a valuable neurophysiological marker for quantifying changes in brain activation. Although this statement is still reductionist (that NVC activity is proportional to operator's brain activity), this approach links the concept of human MWL and mental resources to objective neurophysiological measures for Neuroergonomics purposes.

Recent Neuroergonomics research has progressed in neurocognitive or neuroimaging-sensing instrumentation for determining operator states through the measurement of NVC activity associated with the degree of mental processes (Parasuraman and Wilson, 2008). Tremendous advances have been made toward establishing approaches for portable neuroimaging equipment and brain activation measurements to assess sensitivity to NVC in human operators acting in realistic work environments. This development is especially the case in ambulatory functional neuroimaging methods such as functional near-infrared spectroscopy (fNIRS) and electroencephalography (EEG). To date, the aforementioned non-invasive brain imaging techniques are beginning to be well-established in the Neuroergonomics community. These advantages will be even more beneficial in the future as the coupling between these methods becomes more widespread.

ASSESSING NEUROVASCULAR COUPLING WITH FNIRS-EEG METHODS: AN OBJECTIVE NEUROERGONOMICS APPROACH FOR EVALUATION OF THE OPERATOR'S BRAIN ACTIVITY

Technological advances in opto- and electronic miniaturization have improved the portability and operational flexibility in brain imaging sensors, allowing for greater comprehension of the brain at work in real-world applications (aeronautics, automotive, robotics). fNIRS provides a continuous monitoring of the hemodynamic activity using near-infrared light transmitted between optodes. It infers the changes in the concentrations of oxygenated and deoxygenated hemoglobin in the cortical regions from scattering and absorption properties of light probing beneath the surface of the skull (Perrey, 2008). These two fNIRS signals have their origins in the metabolic response corresponding to a shift of oxygen consumed and the vascular response linked to a modulation of the microvasculature activity (dilatation). This hemodynamic response disrupts the regional cerebral blood flow and volume which exceeds oxygen intake (functional hyperemia) consumed by the recruited neuronal population. fNIRS responses characterize the operator's brain activity related to cerebral blood flow and cerebral tissue oxygenation changes over time (Mandrick et al., 2013; Durantin et al., 2014; Fishburn et al., 2014). Good spatial localization can be derived if a high number of optodes are used in an array, but temporal resolution is coarse by the delayed nature of the hemodynamic response to cortical activity (few seconds).

On the other hand, EEG offers a fine temporal resolution (milliseconds) thus enabling detection of brief neuronal processes, but is limited in its capacity for spatial resolution, at least in real time even though dense array EEG permits source propagation localization. EEG uses scalp electrodes to capture weak electrical current fluctuations generated by inhibitory or excitatory postsynaptic potentials of a pool of neurons firing simultaneously in response to a stimulus. The electrophysiological roots of these signals correspond to the summation of the spontaneously and synchronously recruited neuronal population that contributes to the neuronal activity of the superficial layers of the cortex. EEG waves and event-related potentials signals are particularly strong candidates for objective measures of operator's brain activity at the workplace (Parasuraman and Rizzo, 2006). In general, fNIRS and EEG are complementary as they improve on each other's measurement weaknesses in terms of information content (Fazli et al., 2012). Additionally, there is no noise cross-interference between fNIRS and EEG (light and electrical, respectively; Karanasiou, 2012). Therefore, simultaneous fNIRS-EEG signal acquisition would be suitable for assessing NVC in order to evaluate the operator's brain activity in ecological contexts (Hirshfield et al., 2009; Safaie et al., 2013).

However, it not should limit our understanding of the brain activity to only one perspective; looking at the brain at work with new tools and new eyes we could have new NVC comprehension during ecological context. Readers must note that the multimodality using fNIRS-EEG methods is a very promising approach in the investigation of where, when, and how much NVC exhibits energy mobilization during work. The spatiotemporal evolution of the functional neural connectivity and blood flow regulation through the scalp is permitted due to the recording of temporal electrical activity and spatial hemodynamic activity. Consequently, the evaluation of NVC distribution throughout the head becomes accessible. This measurement makes it possible to dynamically map the brain activity and identify the brain areas with the activated main NVC. Additionally, the assessment of the power of the

electrical signal by EEG coupled with the amplitude of the hemodynamic signal by fNIRS will enable a better depiction of the intensity of the NVC, thus extrapolating the effectiveness of the metabolic effort of performing tasks. This view of the degree of extrapolated metabolic correlates as an indicator of the level of mental resources seems straightforward at first glance. However, the metabolic expenditure that fuels cognitive processes is the prerequisite for any mental resources and the assessment of operator's brain activity. The challenge now is to enhance the reliability of NVC measurement *in situ* with fNIRS-EEG methods.

THE FUTURE FOR NEUROERGONOMICS

It is clear that the extensive work of Parasuraman has left the scientific community in an excellent position to objectively define MWL and subsequently, mental resources, through the measurement of NVC activity. It is our opinion that NVC measurement could be achieved through the use of an efficient fNIRS-EEG coupling. In particular, there needs to be greater characterization of the energy mobilization of NVC with respect to neurophysiological mechanisms (neuronal-astroglial, metabolic and hemodynamic activity) and methods for its assessment in work settings (Parasuraman, 2011). There rests a great deal of work in Neuroergonomics before the development of a standard assessment approach of NVC with innovative neuroimaging technology for the evaluation of the operator's brain activity at work. In other words, there are still opportunities for the technological deployment of coupled hybrid devices (dry-electrodes EEG within a high density headset of fNIRS optodes). From a broader perspective, emerging devices must meet several criteria: discriminate different levels of workload; not interfere with the subject's work and environment; be accepted by the individual; be low cost with high portability; be easy to implement and to evaluate; be reproducible and reliable; and dissociate the mental workload from emotional processes (sensitivity and specificity). Theoretically, a multimodal fNIRS-EEG approach should help to investigate the interactions between different mental states and user behavior while taking into account the physiological processes. Further investigations are warranted to address newer assessments of the neurophysiological events of the operator's brain at work.

AUTHOR CONTRIBUTIONS

Each of the authors has read and concurs with the content in the final manuscript. The first author (KM) wrote the majority of the manuscript. The other authors (ZC, MC, SP, and FD) have extensively reviewed and revised the manuscript from the first draft before giving final approval of the version to be submitted.

FUNDING

This work was funded by the French Research National Agency, the French Defence Procurement Agency (ASTRID), and the AXA Research Fund.

REFERENCES

Ayaz, H., Shewokis, P. A., Bunce, S., Izzetoglu, K., Willems, B., and Onaral, B. (2012). Optical brain monitoring for operator training and mental workload assessment. *Neuroimage* 59, 36–47. doi: 10.1016/j.neuroimage.2011.06.023

Cain, B. (2007). *RTO-TR-HFM-121-Part-II-Human Factors Considerations in the Design, Use, and Evaluation of AMVE-Technology. A Review of the Mental Workload Literature*. North Atlantic Treaty Organisation (NATO), Research and Technology Organisation (RTO), Neuilly-sur-Seine. Available online at: http://ftp.rta.nato.int/public//PubFullText/RTO/TR/RTO-TR-HFM-121-PART-II///TR-HFM-121-Part-II-04.pdf

Durantin, G., Gagnon, J.-F., Tremblay, S., and Dehais, F. (2014). Using near infrared spectroscopy and heart rate variability to detect mental overload. *Behav. Brain Res.* 259, 16–23. doi: 10.1016/j.bbr.2013.10.042

Fazli, S., Mehnert, J., Steinbrink, J., Curio, G., Villringer, A., Müller, K.-R., et al. (2012). Enhanced performance by a hybrid NIRS-EEG brain computer interface. *Neuroimage* 59, 519–529. doi: 10.1016/j.neuroimage.2011.07.084

Fishburn, F. A., Norr, M. E., Medvedev, A. V., and Vaidya, C. J. (2014). Sensitivity of fNIRS to cognitive state and load. *Front. Hum. Neurosci.* 8:76. doi: 10.3389/fnhum.2014.00076

Hirshfield, L. M., Chauncey, K., Gulotta, R., Girouard, A., Solovey, E. T., Jacob, R. J. K., et al. (2009). "Combining electroencephalograph and functional near infrared spectroscopy to explore users' mental workload," in *Foundations of Augmented Cognition. Neuroergonomics and Operational Neuroscience*, eds D. D. Schmorrow, I. V. Estabrooke, and M. Grootjen (Berlin; Heidelberg: Springer), 239–247. doi: 10.1007/978-3-642-02812-0_28

Karanasiou, I. (2012). "Functional brain imaging using non-invasive non-ionizing methods: towards multimodal and multiscale imaging," in *Neuroimaging - Methods*, ed P. Bright (InTech). Available online at: http://www.intechopen.com/books/neuroimaging-methods/functional-brain-imaging-using-non-invasive-non-ionizing-methods-towards-multimodal-and-multiscale-i

Kramer, A. F., and Parasuraman, R. (2007). "Neuroergonomics: application of neuroscience to human factors," in *Handbook of Psychophysiology, 3rd Edn*., eds J. T. Cacioppo, L. G. Tassinary, and G. G. Berntson (Boston, MA: Cambridge University Press), 704–722.

Mandrick, K., Derosiere, G., Dray, G., Coulon, D., Micallef, J.-P., and Perrey, S. (2013). Prefrontal cortex activity during motor tasks with additional mental load requiring attentional demand: a near-infrared spectroscopy study. *Neurosci. Res.* 76, 156–162. doi: 10.1016/j.neures.2013.04.006

Mehta, R. K., and Parasuraman, R. (2013). Neuroergonomics: a review of applications to physical and cognitive work. *Front. Hum. Neurosci.* 7:889. doi: 10.3389/fnhum.2013.00889

Parasuraman, R. (1998). *Neuroergonomics: The Study of Brain and Behavior at Work*. Washington, DC: Cognitive Science Laboratory. Available online at: www.psychology.cua.edu/csl/neuroerg.html

Parasuraman, R. (2003). Neuroergonomics: research and practice. *Theor. Issues Ergon. Sci.* 4, 5–20. doi: 10.1080/14639220210199753

Parasuraman, R. (2011). Neuroergonomics brain, cognition, and performance at work. *Curr. Dir. Psychol. Sci.* 20, 181–186. doi: 10.1177/0963721411409176

Parasuraman, R., and Caggiano, D. (2002). "Mental workload," in *Encyclopedia of the Human Brain*, Vol. 3, *ed* V. S. Ramachandran (San Diego, CA: Academic Press), 17–27. doi: 10.1016/B0-12-227210-2/00206-5

Parasuraman, R., and Rizzo, M. (2006). *Neuroergonomics: The Brain at Work*. New York, NY: Oxford University Press.

Parasuraman, R., and Wilson, G. F. (2008). Putting the brain to work: neuroergonomics past, present, and future. *Hum. Factors* 50, 468–474. doi: 10.1518/001872008X288349

Perrey, S. (2008). Non-invasive NIR spectroscopy of human brain function during exercise. *Methods* 45, 289–299. doi: 10.1016/j.ymeth.2008.04.005

Roy, C. S., and Sherrington, C. S. (1890). On the regulation of the blood supply of the brain. *J. Physiol.* 11, 85–108. doi: 10.1113/jphysiol.1890.sp000321

Safaie, J., Grebe, R., Moghaddam, H. A., and Wallois, F. (2013). Toward a fully integrated wireless wearable EEG-NIRS bimodal acquisition system. *J. Neural Eng.* 10:056001. doi: 10.1088/1741-2560/10/5/056001

Wickens, C. D. (2008). Multiple resources and mental workload. *Hum. Factors* 50, 449–455. doi: 10.1518/001872008X288394

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Mandrick, Chua, Causse, Perrey and Dehais. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 20 April 2016
doi: 10.3389/fnhum.2016.00174

Acute Supramaximal Exercise Increases the Brain Oxygenation in Relation to Cognitive Workload

Cem Seref Bediz[1,2], Adile Oniz[2], Cagdas Guducu[2], Enise Ural Demirci[1], Hilmi Ogut[2], Erkan Gunay[2,3], Caner Cetinkaya[3] and Murat Ozgoren[2]*

[1] Department of Physiology, Faculty of Medicine, Dokuz Eylul University, Izmir, Turkey, [2] Department of Biophysics, Faculty of Medicine, Dokuz Eylul University, Izmir, Turkey, [3] School of Sport Sciences and Technology, Dokuz Eylul University, Izmir, Turkey

Single bout of exercise can improve the performance on cognitive tasks. However, cognitive responses may be controversial due to different type, intensity, and duration of exercise. In addition, the mechanism of the effect of acute exercise on brain is still unclear. This study was aimed to investigate the effects of supramaximal exercise on cognitive tasks by means of brain oxygenation monitoring. The brain oxygenation of Prefrontal cortex (PFC) was measured on 35 healthy male volunteers via functional near infrared spectroscopy (fNIRS) system. Subjects performed 2-Back test before and after the supramaximal exercise wingate anerobic test (WAnT) lasting 30-s on cycle ergometer. The PFC oxygenation change evaluation revealed that PFC oxygenation rise during post-exercise 2-Back task was considerably higher than those in pre-exercise 2-Back task. In order to describe the relationship between oxygenation change and exercise performance, subjects were divided into two groups as high performers (HP) and low performers (LP) according to their peak power values (PP) obtained from the supramaximal test. The oxy-hemoglobin (oxy-Hb) values were compared between pre- and post-exercise conditions within subjects and also between subjects according to peak power. When performers were compared, in the HP group, the oxy-Hb values in post-exercise 2-Back test were significantly higher than those in pre-exercise 2-Back test. HP had significantly higher post-exercise oxy-Hb change (Δ) than those of LP. In addition, PP of the total group were significantly correlated with Δoxy-Hb.The key findings of the present study revealed that acute supramaximal exercise has an impact on the brain oxygenation during a cognitive task. Also, the higher the anerobic PP describes the larger the oxy-Hb response in post-exercise cognitive task. The current study also demonstrated a significant correlation between peak power (exercise load) and post-exercise hemodynamic responses (oxy-, deoxy- and total-Hb). The magnitude of this impact might be related with the physical performance capacities of the individuals. This can become a valuable parameter for future studies on human factor.

Keywords: PFC, human factor, fNIRS, cognition and exercise, N-back test

OPEN ACCESS

Edited by:
Hasan Ayaz,
Drexel University, USA

Reviewed by:
Ranjana K. Mehta,
Texas A&M Health Science Center, USA
Frederic Dehais,
Institut Supérieur de l'Aéronautique et de l'Espace, France

****Correspondence:***
Cem Seref Bediz
cem.bediz@deu.edu.tr

Received: *07 October 2015*
Accepted: *05 April 2016*
Published: *20 April 2016*

Citation:
Bediz CS, Oniz A, Guducu C, Ural Demirci E, Ogut H, Gunay E, Cetinkaya C and Ozgoren M (2016) Acute Supramaximal Exercise Increases the Brain Oxygenation in Relation to Cognitive Workload. Front. Hum. Neurosci. 10:174. doi: 10.3389/fnhum.2016.00174

INTRODUCTION

Individuals usually feel a mental arousal and define an increase in their cognitive abilities after exercise. The relation between exercise and cognitive function has been investigated in several decades. It has been proposed that exercise durations, exercise intensity and the differences in cognitive tasks could cause some contradicting results in previous studies (Ando et al., 2011; Endo et al., 2013; Schmit et al., 2015). Despite the large variations of the reported results, there is meta-analytic evidence that describes significant beneficial effects of acute exercise on cognition in the general population (Lambourne and Tomporowski, 2010; McMorris and Hale, 2012). The positive effects of erobic exercise on various cognitive tasks have been reported mostly. However there are rarely any reports about the effect of anerobic exercise (*i.e., the protocol in the current study*), characterized as short term and highly intensive effort, on cognition. The specific mechanisms by which exercise affects cognitive functions remain largely unclear (Rupp and Perrey, 2008). One of the mechanisms proposed is the relationship between prefrontal cortex (PFC) oxygenation and cognitive function. With the development of imaging methods, it has been possible to show the brain oxygenation during or post exercise conditions. Most of the studies demonstrate an increment in oxygenation of PFC following exercise (Rupp and Perrey, 2008; Jung et al., 2015).

In the literature, effects of exercise on cognitive performance were investigated during exercise and after exercise period. It is largely accepted that the metabolic activity in brain increases during cognitive tasks. The consumption of energy in the neurons depending on the metabolic activity increase leads an increase in the cerebral blood flow to meet the increased demand of oxygen and glucose. For this reason, the increment in brain oxygenation is accepted to be/as the physiological indicator for cognitive workload. Near infrared spectroscopy (NIRS) is a non-invasive measurement based sensitive method that indicates cerebral hemodynamic response (Ozgoren et al., 2012) to cognitive tasks by means of changing levels of oxy- and deoxy-HB (Albinet et al., 2014). Some studies reported a relationship between cerebral oxygenation level and cognitive test scores (Ayaz et al., 2007, 2012; McMorris et al., 2011; McMorris and Hale, 2012). Additionally, Li et al. (2005) and Gateau et al. (2015), describe a higher dorsolateral PFC activation during working memory task by using functional near infrared spectroscopy (fNIRS). Inadequate increase in cerebral oxygenation during cognitive task defined as an indicator of cerebral fatigue (Nybo and Rasmussen, 2007; Mandrick et al., 2013; Mehta and Parasuraman, 2014). Dietrich (2003) reported that, aside from cerebral fatigue, the hypo-frontality may occur due to the challenge of source allocation between the areas responsible for physical and cognitive workload.

We hypothesized that the brain oxygenation during cognitive task after the acute supramaximal exercise is higher than the pre-exercise cognitive task. We also expected a higher behavioral performance after the acute supramaximal exercise. This study aimed to evaluate the hemodynamic and behavioral changes of cognitive processes following an acute (anerobic) supramaximal exercise.

TABLE 1 | Ages, heights, body weights, peak powers and maximal heart rates (Max-HR) of low performers (LP) and high performers (HP) groups were presented as mean ± SD.

	Age (yrs)	Height (m)	Weight (kg)	Peak Power (Watts)	Max-HR (beats/min)
HP (n = 18)	21.0 ± 2.6	165.9 ± 22.5	72.7 ± 9.6	800.7 ± 77.3	165.4 ± 22.8
LP (n = 17)	20.5 ± 2.1	168.9 ± 11.5	71.4 ± 11.1	634.3 ± 89.3	168.9 ± 11.5

MATERIALS AND METHODS

Participants and Experimental Design

Thirty-five male healthy and physically active subjects (1.78 ± 0.07 cm height and 72.1 ± 10.3 kg weight) aged between 18 and 23 years participated to the study (**Table 1**). All subjects were informed about the procedures and signed a written consent. Experiments were conducted in two visits. In first visit, the subjects were informed and became familiar to both exercise and N-Back test protocols. Within 3 days, subjects re-visited the laboratory to perform a short term-supramaximal acute exercise and cognitive tasks (2-Back tests) before and after exercise (**Figure 1**). Brain oxygenation was continuously measured via fNIRS during cognitive tasks and exercise. In order to evaluate the performance dependent results, an average peak power value was calculated for all subjects. Subjects who had higher power values than the average (751 Watt) was considered as high performers (HP), (N = 17 subjects) while who had lower power values than average were considered as low performers (LP), (N = 18 subjects). There were no statistically significant differences in terms of age, weight and height between the HP and LP groups. The Ethics Committee of Dokuz Eylul University approved all procedures and experimental design.

Cognitive Task Procedure

Cognitive performance was evaluated by N-Back test before and after exercise. N-Back test, which mainly evaluates the working memory as well as sustained attention, was administrated to evaluate working memory and inhibitory control of irrelevant information (Jonides et al., 1997; Jaeggi et al., 2010). N-Back test was developed in OpenGL by using C programming language, has been administered to the participants by a laptop. A pseudorandom sequence of 120 letters consisting "K, Q, H, X, M, F, R, and B" were displayed on the center of the screen, one by one. Participants were asked to press the button on a response pad only if a letter on the screen was the same as the letter shown "*n*" steps earlier. In the present study "2-Back" condition was employed. The probability of each letter to be same as the two-step earlier letter was 30%. Each letter was presented on the computer screen for 500 ms and the inter-stimulus interval was set to vary between 1500 and

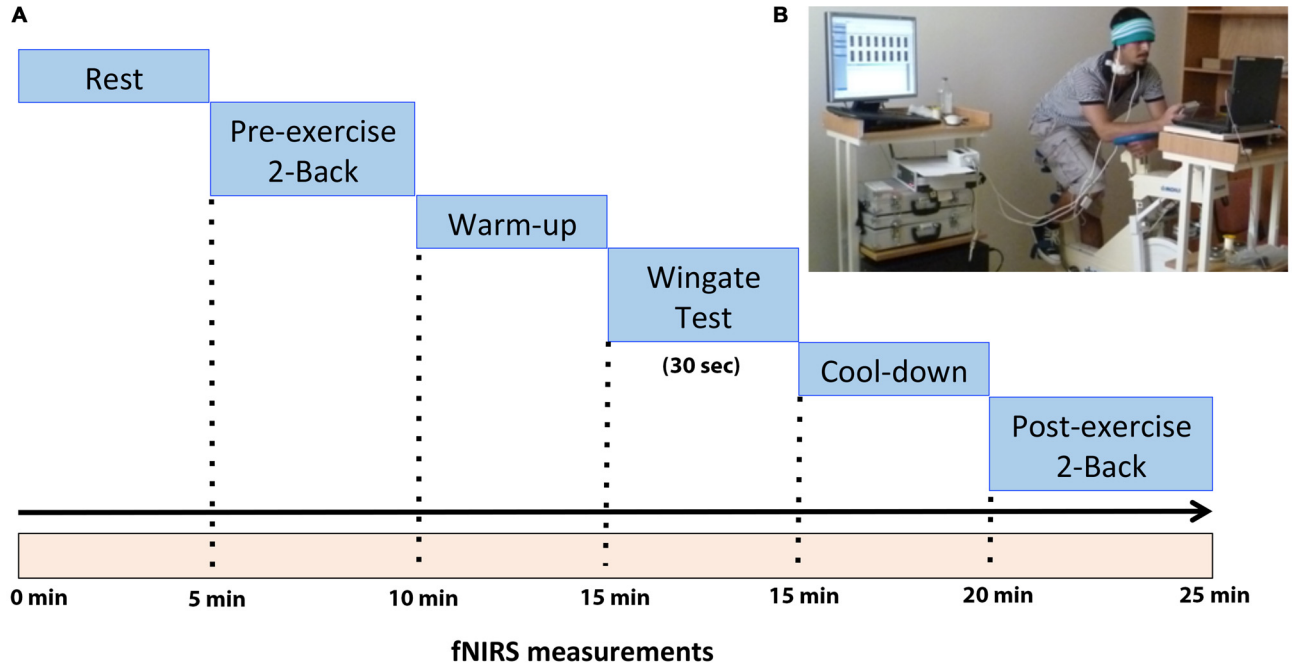


FIGURE 1 | (A) The demonstration of experimental design for the current study. **(B)** Participant seated at cycle ergometer during whole paradigm.

2000 ms. All tests were completed within 5 min. A brief training session, including a practice sequence of 15 letters, was given to all participants before the actual test to familiarize the subjects to the protocols. In order to evaluate working memory; reaction time for all stimuli total reaction time (TRT) and reaction time for the correct answers (CART), number of correct answers (CA), wrong answers (WA), and missed answers (M) have been obtained and measured. Nexus-10 bio-amplifier, which was equipped with appropriate sensors, was used to monitor the appearance of stimuli on the screen and participant responses in order to measure exact reaction times.

Exercise Procedure

Wingate Anerobic test (WAnT) protocol was applied as the supramaximal exercise. WAnT is an exhaustive exercise for assessment of anerobic performance (Dotan and Bar-Or, 1983). WAnT is also accepted as a standardized anerobic exercise model for evaluating physiological responses to supramaximal exercise under controlled conditions (Weinstein et al., 1998). Following a warm up on the bicycle for 5 min at intensity about 50 watts, all subjects performed WAnT on a mechanically braked cycle ergometer with an optical pedal counter (Monark 824E, Sweden). Two unloaded 5-s sprints were performed during warm up. Following the warm up, subjects were instructed to pedal as fast as possible for 30 s against a resistance of 80 g/kg body mass. The subjects were verbally encouraged to maintain high pedaling rate throughout the WAnT. Pedal revolutions were monitored and recorded at 1-s intervals. Power outputs were calculated as described in Weinstein et al. (1998). The highest power output in the first 5 s of the test was used to represent the peak power (PP; PP = Distance × Load/ Time).

fNIRS Recordings

The continuous wave (CW) fNIR system used in the present study (Imager 100, fNIR Devices LLC, MD, USA). System was connected to a flexible sensor pad, Sensor pad contained four light sources with built in peak wavelengths at 730 nm and 850 nm and 10 detectors designed to scan cortical areas underlying the forehead. The forehead area was cleaned with alcohol swap and scrubbing cream (NuPrep, USA). After the cleansing process the sensor was placed to the forehead region and covered by an elastic bandage specifically designed to hold it tightly across the head. Furthermore, a black head bandage was placed on top to eliminate the possible ambient light effects.

This system records two wavelengths and dark current for each of the 16 voxels, totaling 48 measurements for each sampling period (Ayaz et al., 2011, 2013). With a fixed source-detector separation of 2.5 cm, this configuration generates a total of 16 measurement locations (voxels) per wavelength. Data acquisition and visualization were conducted using COBI Studio software (Ayaz and Onaral, 2005). The fNIR device calculates relative changes to baseline values of oxy-hemoglobin (oxy-Hb) and deoxyhemoglobin (deoxy-Hb) molecules by means of a CW spectroscopy system that applies light to tissue at constant amplitude. The mathematical basis of CW-type measurements uses the modified Beer Lambert Law (Cope and Delpy, 1988). In the present study baseline condition was started at the beginning of the each task and lasted 20 s. The summation

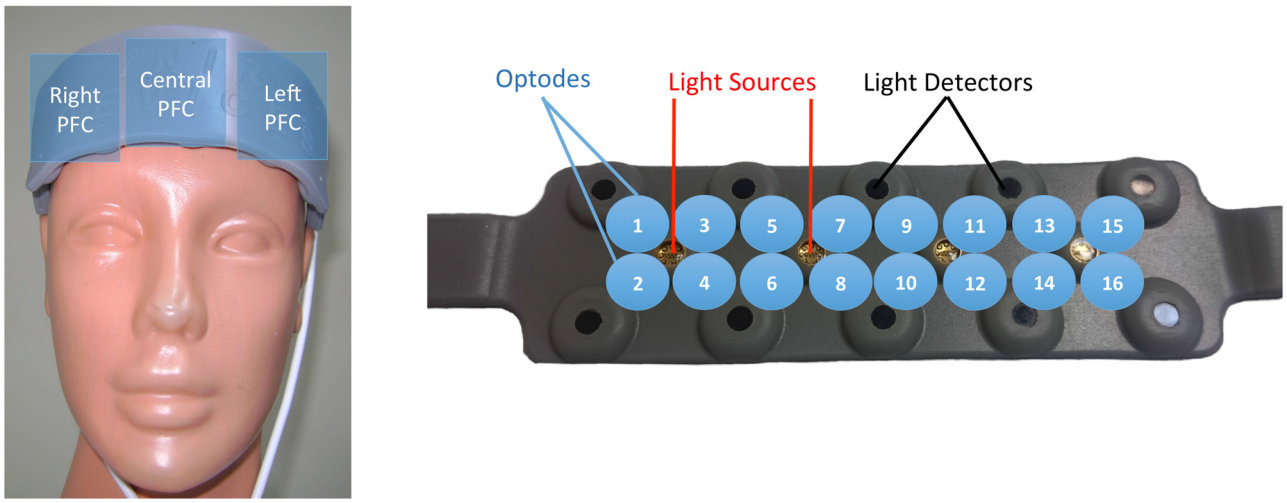


FIGURE 2 | The demonstration of functional near infrared spectroscopy (fNIRS) sensor pad. Brain oxygenation was measured from 16 locations over the forehead area via fNIRS sensor (left side). The sensor pad has 4 light sources, 10 light detectors and 16 optodes (right side). Optodes that are located on the leftmost side of the forehead namely 1, 2, 3 and 4 denotes left prefrontal cortex (PFC), optodes that are centrally located as 7, 8, 9 and 10 denotes central PFC, and optodes that are located on the rightmost side as 13, 14, 15 and 16 denotes right PFC.

of oxy-Hb and deoxy-Hb values was described as total-Hb. Moreover, Δoxy-Hb and Δdeoxy-Hb parameters were used to describe the value differences (oxygenation change; i.e., rise or fall) between average values of two different sessions (i.e. $\Delta OxyHb = OxyHb_{\text{post}} - OxyHb_{\text{pre}}$) for both oxy-Hb and deoxy-Hb.

fNIRS Analysis

The raw intensity measurements at 730, and 850 nm were Butterworth low-pass filtered with MATLAB program (MATLAB and Statistics Toolbox, 2007). Butterworth filter was designed to eliminate possible respiration and heart rate signals and unwanted high frequency noise (Huppert et al., 2009). The artifact removal process has been made according to Ayaz et al. (2012). The PFC oxygenation data retrieved via fNIRS were examined in right, left and central PFC areas and defined as region of interest (ROI; **Figure 2**). Optodes that are located on the leftmost side of the forehead namely 1, 2, 3 and 4 combined to denote left PFC, optodes that are centrally located as 7, 8, 9 and 10 combined to denote central PFC, and optodes that are located on the rightmost side as 13, 14, 15 and 16 combined to denote right PFC. fNIRS signals were calculated and averaged over the whole pre- and post-exercise 2-Back sessions (Endo et al., 2013).

Statistical Analysis

Shapiro-Wilk and Kolmogorov-Smirnov tests were used to control the normality of the data. All of the data were distributed normally. Age, height and weight differences between groups were tested via independent groups *t*-test, and the results were given in **Table 1**. We performed statistical analysis on combined channels as ROI wise.

A two-way group (LP vs. HP) and time (pre vs. post exercise) analysis of variance (ANOVA) was performed on fNIRS data and behavioral data. Moreover, oxy- and deoxy-Hb change were calculated as mentioned before (i.e., ΔOxyHb), and used for the statistical analysis. An independent group *t*-test was performed to test group differences (LP vs. HP) on oxy-, deoxy- and total-Hb changes. Also, the correlations between peak power and oxy-Hb, deoxy-Hb, and total-Hb changes were investigated via Spearman's rank order correlations test.

RESULTS

fNIRS Findings

In the present study, 35 healthy subjects were recruited. In order to evaluate the performance dependent results, pre and post exercise PFC oxygenation levels during 2-Back tests were compared for both HP and LP groups with two-way mixed ANOVA. The oxy-Hb, deoxy-Hb, and total-Hb values of the groups were demonstrated for pre- and post-exercise 2-Back sessions were demonstrated in **Table 2**. Also the demonstration of oxy-Hb and deoxy-Hb levels in central PFC area during pre- and post-exercise 2-Back tests were given for whole group averages in **Figure 3**. All of the figures represent the group averages during related task (pre- and post-exercise 2-back).

For oxygenation related analysis; pre- and post-exercise 2-Back tests oxy-Hb values were selected as within-subject factor and peak power was selected as between-subject factor. According to these analyses, oxygenation in post-exercise 2-Back test was higher than the oxygenation in pre-exercise 2-Back test for all three ROI of PFC [$F_{(1,33)} = 51.82$, $p < 0.001$, $\eta = 0.61$ for left; $F_{(1,33)} = 78.42$, $p < 0.001$, $\eta = 0.70$ for central; $F_{(1,33)} = 54.25$, $p < 0.001$ $\eta = 0.61$ for right]. For right PFC area there was

TABLE 2 | Pre- and post-exercise 2-back oxy-, deoxy, and total-Hb levels (mean ± SD) are given for total, LP and HP group in central prefrontal cortex (PFC) area.

	Pre-exercise 2-Back			Post-exercise 2-Back		
	Oxy-Hb (μMolar)	Deoxy-Hb (μMolar)	Total-Hb (μMolar)	Oxy-Hb (μMolar)	Deoxy-Hb (μMolar)	Total-Hb (μMolar)
Total Group	3.11 (± 3.41)	− 0.99 (± 0.94)	3.01 (± 3.71)	7.32 (± 4.85)***	0.84 (± 2.42)*	8.15 (± 6.56)***
HP	3.30 (± 3.62)	0.11 (± 0.94)	3.42 (± 4.04)	9.18 (± 4.90)***	1.84 (± 2.74)*	11.03 (± 6.88)***
LP	2.92 (± 3.30)	−0.30 (± 0.93)	2.62 (± 3.44)	5.54 (± 3.30)***	−0.11 (± 1.64)	5.43 (± 5.06)**

The significant differences related to comparisons of oxygenation, deoxygenation and total-Hb between pre- and post-exercise 2-Back session are marked with "" (where "*" denotes $p < 0.05$, "**" denotes $p < 0.01$, and "***" denotes $p < 0.001$).*

not any interaction so, the pairwise comparisons were used. According to pairwise comparisons there was not any significant difference between groups, but there were significant differences between pre and post oxygenation levels in right PFC ($p < 0.001$). Because of the significant interaction in left and central PFC areas between within group effects (Pre/Post) and between group effects (HP/LP) in terms of two-way ANOVA results, between group analyses could not be evaluated (McDonald, 2009). Further analyses were made to clarify the statistical significance via within group paired samples *t*-test analysis. Paired samples *t*-test was conducted separately for each group and post-exercise 2-Back test oxy-Hb levels were found significantly higher than the pre-exercise 2-Back test oxy-Hb level for both group [for HP group: in left PFC, $T_{(16)} = -5.99$, $p < 0.001$; in central PFC, $T_{(16)} = -8.12$, $p < 0.001$; for LP group: in left PFC, $T_{(17)} = -3.89$, $p < 0.001$; in central PFC, $T_{(17)} = -4.13$, $p < 0.001$].

Moreover, deoxy-Hb related analyses were conducted. For the analyses; pre- and post-exercise 2-Back tests deoxy-Hb values were selected as within-subject factor and peak power was selected as between-subject factor. According to these analyses, deoxy-Hb values were significantly higher in post-exercise 2-Back test than the deoxy-Hb values in pre-exercise 2-Back test for all three ROI of PFC [$F_{(1,33)} = 9.56$, $p < 0.01$, $\eta = 0.23$ for left PFC; $F_{(1,33)} = 7.03$, $p < 0.05$, $\eta = 0.17$ for central PFC; $F_{(1,33)} = 11.59$, $p < 0.01$ $\eta = 0.26$ for right PFC]. For right and left PFC areas there was not any interaction so, the pairwise comparisons were used. According to pairwise comparisons there was not any significant difference between groups, but there were significant differences between pre and post deoxygenation levels in left and right PFC ($p < 0.01$ for both). Because of the significant interaction in central PFC area between within group effects (Pre/Post) and between group effects (HP/LP) in terms of two-way ANOVA results, between group analyses could not be evaluated (McDonald, 2009). Further analysis was made to clarify the statistical significance via within group paired samples *t*-test analysis. Paired samples *t*-test was conducted separately for each group and post-exercise 2-Back test deoxy-Hb levels were found significantly higher than the pre-exercise 2-Back test deoxy-Hb level for HP group in central PFC area [$T_{(16)} = -2.67$, $p < 0.05$] and there was not any significance for LP group in central PFC area.

Also total-Hb was significantly higher in post-exercise 2-Back test for all three interested ROI of PFC [$F_{(1,33)} = 38.55$, $p < 0.001$, $\eta = 0.54$ for left PFC; $F_{(1,33)} = 47.20$, $p < 0.001$, $\eta = 0.59$ for central PFC; $F_{(1,33)} = 42.15$, $p < 0.001$ $\eta = 0.56$ for right PFC]. For right and left PFC areas there was not any interaction so, the pairwise comparisons were used. According to pairwise comparisons there was not any significant difference between groups, but there were significant differences between pre and post total-Hb levels in left and right PFC ($p < 0.001$ for both). Because of the

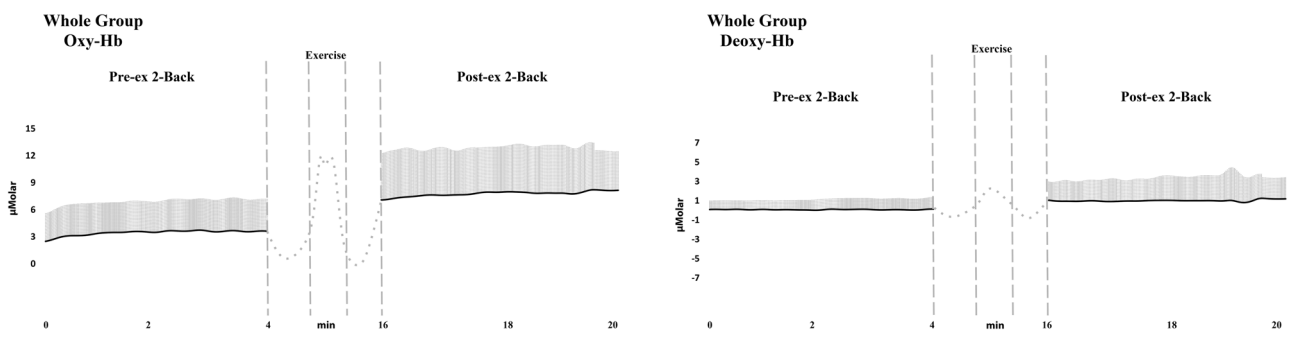


FIGURE 3 | The demonstration of oxy- and deoxy-Hb average levels in central PFC area during pre- and post-exercise 2-Back tests for the whole group. Standard deviations are also marked in line with average as vertical lines (Note that only positive deflection is displayed for the sake of simplicity). Left panel indicated oxy-Hb and right panel indicated deoxy-Hb results. Vertical scale denotes the strength of fNIRS signal in μMolar units, which is normalized to baseline. Horizontal scale denotes the time scale in minutes. Vertical dashed lines denote pre-, during and post- exercise onsets and durations. The dotted lines represent the periods of warm-up and cool-down.

TABLE 3 | Post- and pre-exercise oxy-Hb, deoxy-Hb, and total-Hb changes (Δoxy-Hb, Δdeoxy-Hb, and Δtotal-Hb) are given for LP and HP groups in all PFC area (mean ± SD).

	ΔOxy-Hb (μMolar)			ΔDeoxy-Hb (μMolar)			ΔTotal-Hb (μMolar)		
	Left	Central	Right	Left	Central	Right	Left	Central	Right
HP	5.66 (± 3.89)*	5.88 (± 2.98)**	4.87 (± 3.48)	1.89 (± 1.91)	2.32 (± 2.13)*	1.26 (± 1.18)	7.10 (± 5.59)*	7.61 (± 5.03)**	5.74 (± 4.52)
LP	2.71 (± 2.94)	2.62 (± 2.69)	3.25 (± 3.48)	1.62 (± 1.67)	1.16 (± 0.89)	1.65 (± 1.56)	3.60 (± 4.58)	2.81 (± 3.90)	4.42 (± 4.72)

Significances between groups are marked with "" (where "*" denotes $p < 0.05$, "**" denotes $p < 0.01$).*

significant interaction in central PFC area between within group effects (Pre/Post) and between group effects (HP/LP) in terms of two-way ANOVA results, between group analyses could not be evaluated (McDonald, 2009). Further analyses were made to clarify the statistical significance via within group paired samples *t*-test analysis. Paired samples *t*-test was conducted separately for each group and post-exercise 2-Back test total-Hb levels were found significantly higher than the pre-exercise 2-Back test total-Hb level for HP group [in central PFC, $T_{(16)} = -6.24$, $p < 0.001$], and for LP group [in central PFC, $T_{(17)} = -3.06$, $p < 0.01$].

In addition to these findings, the changes of oxy-Hb, deoxy-Hb, and total-Hb were compared between HP and LP groups via independent samples *t*-test (**Table 3**). While comparing the oxygenation changes (Δoxy-Hb) of the LP and HP group, PFC oxygenation rise in central PFC ($p < 0.01$) and left PFC ($p < 0.05$) areas were found significantly higher in HP group. While comparing the deoxygenation changes (Δdeoxy-Hb) of pre- and post-exercise 2-Back tests, HP group's PFC deoxygenation changes in central PFC area was found significantly higher ($p < 0.05$). While comparing the total-Hb changes (Δtotal-Hb) of pre- and post-exercise 2-Back tests, HP group's PFC total-Hb change in central PFC ($p < 0.01$) and left PFC ($p < 0.05$) areas were found significantly higher. Moreover, the peak power values (PP) of the total group were significantly correlated with Δoxy-Hb, Δdeoxy-Hb, and Δtotal-Hb values for central PFC area ($r = 0.042$ $p < 0.01$; $r = 0.035$ $p < 0.02$; $r = 0.040$ $p < 0.01$, respectively). The demonstration of average oxy-Hb and deoxy-Hb levels in central PFC area during pre- and post-exercise 2-Back Tests for HP and LP group were given in **Figures 4**, **5**.

Behavioral Findings

All subjects have successfully completed the whole procedure. 2-Back test scores (CA, WA, M, CART, TRT) were compared for total group and there were no significant differences between pre- and post-exercise 2-Back test scores. The same comparison was made for HP and LP groups and also there were no significant differences within (pre-exercise vs. post-exercise sessions) and between groups (**Table 4**). Though not statistically significant the reaction times, the correct answer CART were slightly progressed after the exercise.

DISCUSSION

Both physical benefits and cognitive improvements of the exercise have been frequently studied in the field. In the present study, the effects of anerobic exercise on brain oxygenation were investigated in relation to cognitive workload. The results revealed an increment in oxy-, deoxy-, and total-Hb levels on the post-exercise session of cognitive task. The increments in these parameters have been found to be statistically higher than the pre-exercise session of the cognitive task. In the related literature, it has been demonstrated that the oxy-Hb level increases in PFC during both physical exercise and cognitive task. Such rise has been linked to the increase in neural-metabolic activation (González-Alonso et al., 2004; Shibuya et al., 2004; Rupp and Perrey, 2008; Endo et al., 2013). In their study Tam and Zouridakis (2014) describes the oxy-Hb, deoxy-Hb, their summation (oxy-Hb + deoxy-Hb), and their difference (oxy-Hb—deoxy-Hb) measurements corresponding to the changes in oxygen delivery, oxygen extraction, total blood volume delivered, and total oxygenation, respectively. The increasing level of oxy- and deoxy-Hb is accepted as an indicator of an increase in blood flow (Endo et al., 2013). The present study revealed a rise in all oxy-Hb, deoxy-Hb, and total-Hb levels during both pre- and post-exercise cognitive tasks, which might be related to regional increase of cerebral blood flow. In this context, regional blood flow in PFC during post-exercise session of cognitive task could be considered higher than the pre-exercise session of cognitive task. Such increased hemodynamic responses may be an indicator for the additional effort during post-exercise cognitive task (Mandrick et al., 2013).

In the present study, a relationship between brain's hemodynamic responses and physical performance (under anerobic supramaximal exercise conditions) has been revealed. The participants were divided into two groups by means of PP mentioned as high- and low-performers in order to address the performance relationship. Despite the similar hemodynamic response patterns during pre-exercise session of cognitive task, different hemodynamic response patterns have been observed between groups during post-exercise sessions of cognitive task (**Figures 3**, **4**). As the most striking difference the rise of oxy-Hb and total-Hb levels of HP group have been found statistically higher than those of LP group in central and left PFC areas. These findings may demonstrate that the rise of oxygen consumption and demand in PFC is higher in the HP group, as known as higher anerobic capacity, than the LP group. In this context, Drigny et al. (2014), reported that the brain oxygenation could change with training in obese patients. Also, Khan and Hillman (2014) reviewed the connection between training and its effects on brain oxygenation levels and suggested a relationship between erobic fitness and cognitive processes that can be demonstrated

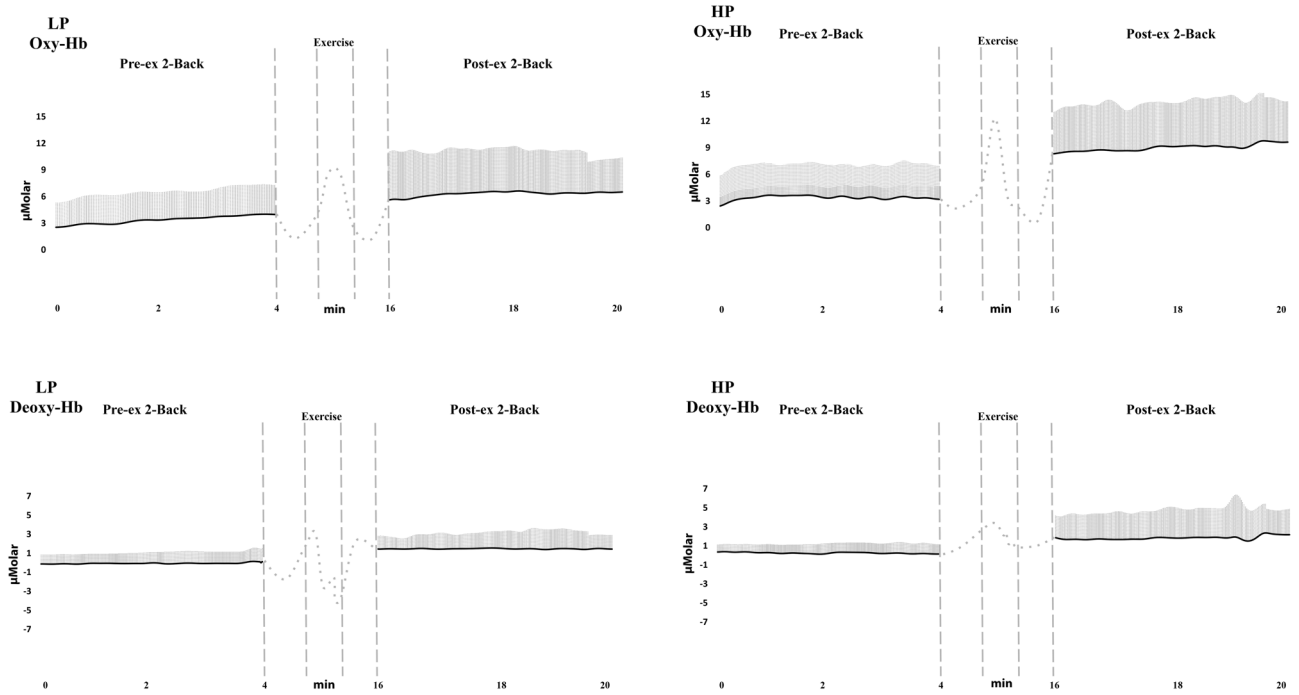


FIGURE 4 | The demonstration of oxy- and deoxy-Hb levels in central PFC area during pre- and post-exercise 2-Back tests for group averages. Standard deviations are also marked in line with average as vertical lines (Note that only positive deflection is displayed for the sake of simplicity). The four panels are divided into performers (low and high performance, left and right consequently) and oxy- and deoxy-Hb (upper and lower consequently). Left panel indicated oxy-Hb and right panel deoxy-Hb results. Vertical scale denotes the strength of fNIRS signal in µMolar units, which is normalized to baseline. Horizontal scale denotes the time scale in minutes Vertical dashed lines denote pre-, during and post- exercise onsets and durations. The dotted lines represent the periods of warm-up and cool-down.

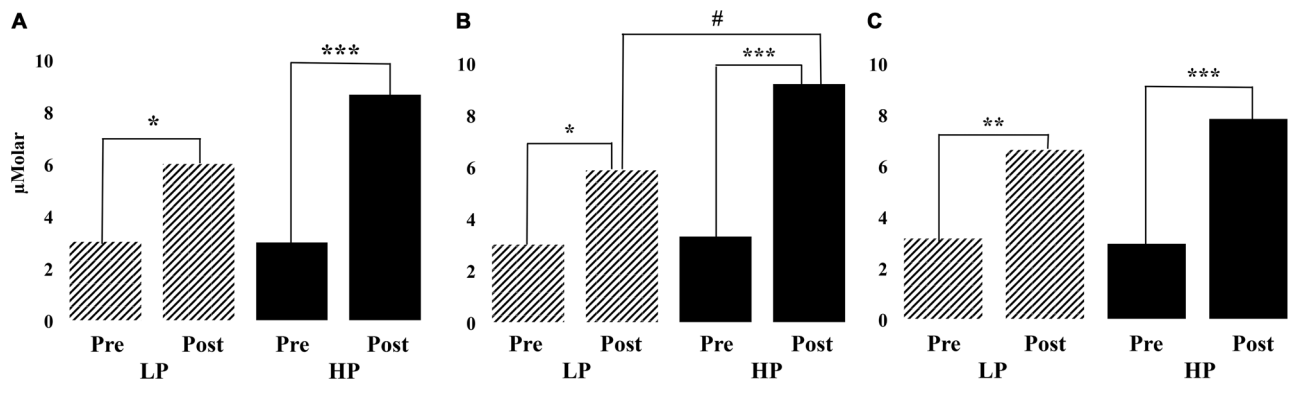


FIGURE 5 | The oxygenation values for pre- and post-exercise 2-Back sessions in left (A), center (B) and right (C) areas of PFC in high and low performance groups. Within group significances are marked with "*" (where "*" denotes $p < 0.05$, "**" denotes $p < 0.01$, and "***" denotes $p < 0.001$); between group significances are marked with "#" (where "#" denotes $p < 0.05$).

via different fMRI and EEG studies. A similar relationship was described between exercise and cognitive functions by Davenport et al. (2012). In the light of results of the present study, level of PFC oxygenation in physically better performing group during post-exercise cognitive task might be higher than the physically lower performing group during post-exercise cognitive tasks.

The current study also demonstrated a significant correlation between peak power (exercise load) and post-exercise change of hemodynamic responses (Δoxy-Hb, Δdeoxy-Hb, and Δtotal-Hb). This correlation also supports the aforementioned assumptions in which the higher performers have higher PFC oxygenation. This can become a valuable parameter for future studies on human factor by means of physical/cognitive

TABLE 4 | The demonstration of correct answers, wrong answers, missed answers; correct answers reaction time, and total reaction time (mean ± SD) for high (HP) and low performers (LP) group during pre- and post-exercise 2-Back sessions.

	Pre exercise 2-Back Session					Post exercise 2-Back Session				
	Correct answers	Wrong answers	Missed answers	Correct answers reaction time (ms)	Total reaction time (ms)	Correct answers	Wrong answers	Missed answers	Correct answers reaction time (ms)	Total reaction time (ms)
Total Group	25.49 (±5.65)	6.37 (±4.56)	10.34 (±5.57)	582.90 (±143.03)	596.60 (±147.39)	25.82 (±6.56)	7.97 (±4.61)	10.17 (±6.56)	576.60 (±186.63)	578.90 (±175.68)
HP	26.0 (±4.23)	6.12 (±4.21)	10.0 (±4.23)	561.76 (±148.61)	575.88 (±154.12)	26.29 (±6.12)	8.12 (±4.61)	9.71 (±6.12)	544.11 (±192.66)	551.17 (±181.71)
LP	25.0 (±6.56)	6.61 (±4.73)	10.67 (±6.45)	602.78 (±130.25)	616.11 (±133.46)	25.39 (±6.75)	7.83 (±4.48)	10.61 (±6.75)	607.22 (±169.70)	605.0 (±160.32)

load. Formerly using a different paradigm, our group has demonstrated the centrofrontal shift of the active brain areas due to cognitive load (Bayazıt et al., 2009). In that study, the cognitive load—such as conflict resolution—displayed a physical drive from the posterior areas towards frontal areas with an increased demand on the frontal regions. Similarly, in the current study, the high peak power task is not solely a motor one but also requires a multitude of cognitive activities. In order to cope with the higher motor (and cognitive) load, the brain has to perform increased levels of concentration, attention, coordination, environment monitoring, and reactive and interactive motor control. The frontal areas are the primary areas for these and similar executive and complex skills. Further clarification is needed in this context to describe the effect of physical fitness on brain oxygenation by means of new study designs and larger size groups.

The effects of acute exercise on cognitive functions have been investigated for decades. However, studies demonstrating the mechanisms of physical exercise on brain oxygenation in relation to cognitive performance via fNIRS are very limited and these mechanism are not well understood (Albinet et al., 2014; Drigny et al., 2014; Dupuy et al., 2015). In previous studies, the participants' performance for decision-making, mental processing speed, selective attention, and reaction time was investigated (Aks, 1998; Arcelin and Brisswalter, 1999; Emery et al., 2001). Moreover, Brisswalter et al. (2002) and Tomporowski (2003) reported that the cognitive performance could be improved by acute exercise. Interestingly, previous studies revealed different findings depending on the type of the exercise, while intense and exhausting exercise causes fatigue (Brisswalter et al., 2002), and light exercise (Varner and Ellis, 1998) causes cognitive performance impairment. Besides, some studies reported a decrease in cognitive functions related with the fatigue caused by the intensity or duration of the exercise (Mehta and Parasuraman, 2014). The physiological basis of this decrement could also be originated from the natural challenge between PFC area related with cognition and motor area related to motion, which causes a hypo-frontality (Dietrich, 2003). The experimental setup of the current study employed a high load but very short (30-sec) anerobic exercise, and such exercise model could cause higher oxygenation responses in PFC similar as erobic exercise models. As mentioned before, the behavioral results displayed only a small but non-significant increase in WA and slight but not significant increase in the CA after the exercise. Our results did not provide a strong support for neither an improvement nor decrement in cognitive scores after acute exercise.

In the related literature, Stroop test was frequently employed as a cognitive task, and generally the improvement of the total test time was accepted as an indicator of improved cognitive functions (Hyodo et al., 2012; Endo et al., 2013). In the present study we employed the 2-Back test as a cognitive task, and evaluated the whole parameters (CA, WA, MA, CART, TRT), but not the total test time. Therefore 2-Back test might not be the right tool as an indicator of the improved cognitive functions. Also the rise of PFC oxygenation could not fully mean an improvement in cognitive function during the

post-exercise session. Another plausible explanation might be that the participants had to work harder "neurally" (indicated by increased oxy-Hb levels) to maintain the cognitive performance during post-exercise session.

In another view, some studies reported an improvement of cognitive functions after the moderate exercise (Potter and Keeling, 2005; Coles and Tomporowski, 2008). However, a recent meta-analysis (McMorris et al., 2011) raised an issue. Those authors examined the effect of acute, moderate intensity exercise on working memory tasks and found that speed and accuracy of processing were differentially affected. In particular, the positive effects of acute exercise seem to be disproportionately influential on executive control processes (i.e., planning, coordination, inhibition, mental flexibility, working memory) relative to tasks of recall and alertness (Chang et al., 2012; McMorris and Hale, 2012). N-Back task used in the current study indeed involves working memory, alertness, planning, and inhibition strategies. Therefore a possible explanation could be that there might be an increase in the cognitive (i.e., executive control process) functions due to prolonged total blood flow capacity increase in the brain metabolic capacity (both oxygen and glucose should be increased) but somehow 2-Back test might not be tough enough to elucidate such improvement. Another explanation could be that the increased metabolic activity in brain might be the mean of more neural effort to maintain the performance, not the indicator of cognitive improvement.

The findings of current literature are not sufficient to explain positive effects of exercise intensity on cognitive functioning (Soga et al., 2015). This study employed a short-term high intensity and all-out exercise model. Following this anerobic exercise, 2-Back test was administered and its results showed a non-significant enhancement in cognitive scores. The different effects of anerobic and erobic exercise on cognitive scores could cause the insignificant differences in test results despite the increase in brain oxygenation. Also the variation in cognitive tests that administered could lead such insignificant results. Finally, the administration time of post-exercise tests could affect the cognitive scores. Soga et al. (2015) reported that there is no clear information in literature about cool-down periods following exercise and administration time of cognitive test. In addition, they (Soga et al., 2015) underlined the importance of assessing both time and heart rate variables post-exercise before administering the cognitive tests for future studies.

In the related literature, it has been suggested that the increment in PFC oxygenation has a positive effect on PFC functions which uses erobic exercise (Endo et al., 2013). Our study has no statistically significant behavioral results to demonstrate such positive effects on PFC functioning. Though the current study used anerobic exercise model, the results (hemodynamic responses) are in line with the studies which consists erobic exercise model (Ando et al., 2011; Endo et al., 2013; Byun et al., 2014). The obtained changes in oxygenation levels are compatible with results of previous studies. But no statistical differences were found in cognitive performance parameters, which were positively affected by erobic exercise. The major dividing line between the erobic and anerobic exercise are the duration and intensity parameters. And these erobic definitions are mostly related to the body and muscle metabolism. As far as we know there is no certain study that has specifically displayed the dividing line between erobic and anerobic capacity in regard to brain only. This might be pointing to the fact that the relationship of the body metabolic capacity and the dynamic patterns might be different than of the brain's.

There are findings demonstrating that cognitive functions, especially following an intense exercise, could be impaired or not augmented. Soga et al. (2015) reported that exercise has different effects on working memory and inhibitory control components. In our study, despite the PFC oxygenation increment, no statistical differences were found between 2-Back test scores. The different effect mechanisms, exercise intensity, and administration time of cognitive tests could be the reason for this insignificancy. In future studies, with an experimental design employing both erobic and anerobic exercises, the hemodynamic and cognitive performance changes can be demonstrated.

Current study may have a considerable limitation. Heart rate may not recover linearly during post exercise (erobic or anerobic), it is likely that cerebral oxygenation, which is regulated by autonomic responses also change drastically from exhaustion to 1 min post recovery to 5 min post recovery. Therefore even though the oxygenation recordings were done post 5 min plus period we may have a bias in the aspect of fluctuations within the 5 min average periods.

CONCLUSION

In summary, acute supramaximal exercise increased the oxy-Hb, deoxy-Hb and total-Hb levels during post-exercise session of 2-Back test. This would indicate that the functions of PFC would increase after acute exercise but the comparison between pre- and post-exercise sessions of 2-Back test scores did not point out any significant improvement. These results could not reveal any cognitive performance augmentation despite the increased PFC oxygenation. As aforementioned in the discussion, such post-exercise oxygenation increase in PFC may not be only derived from cognitive performance. To clarify the basis and effects of oxygenation rise further studies is required with an experimental design, which consist the cognitive tasks in all three phases (pre-, during-, post-exercise).

HP group, who has higher PP, outperformed LP group by means of oxy-, deoxy-, and total-Hb changes. Such difference rise a question if the short-term brain oxygenation responses might be affected by physical performance. The current study could provide a methodological approach for human factor studies that would require combining of behavioral and objective methods (fNIRS, etc).

AUTHOR CONTRIBUTIONS

The study was planned by CSB, AO, and MO, with experimental design by CSB, CG, EUD, EG, CC. Data was collected and analyzed by CG, EUD, HO, CC. Work was drafted by EUD, CC, EG, and HO. Important intellectual support was given by

CSB, AO, MO, EG, and CG throughout the project. All authors prepared the manuscript and approved the final version of the manuscript.

FUNDING

This study was funded partly via Dokuz Eylul University, Department of Scientific Research Projects coded by 2012.KB.SAG.081 and 2014.KB.SAG.012.

ACKNOWLEDGMENTS

Authors would appreciate the research group of Dokuz Eylul University Human Factor Labs. Also authors would like to thank to all volunteers to their attendance to the study and to Ipek Ergonul for editing and commenting of manuscript. The authors would like to thank Dr. Pembe Keskinoglu for her contributions and comments to the statistical analyses.

REFERENCES

Aks, D. J. (1998). Influence of exercise on visual search: implications for mediating cognitive mechanisms. *Percept. Mot. Skills* 87, 771–783. doi: 10.2466/pms.1998.87.3.771

Albinet, C. T., Mandrick, K., Bernard, P. L., Perrey, S., and Blain, H. (2014). Improved cerebral oxygenation response and executive performance as a function of cardiorespiratory fitness in older women: a fNIRS study. *Front. Aging Neurosci.* 6:272. doi: 10.3389/fnagi.2014.00272

Ando, S., Kokubu, M., Yamada, Y., and Kimura, M. (2011). Does cerebral oxygenation affect cognitive function during exercise? *Eur. J. Appl. Physiol.* 111, 1973–1982. doi: 10.1007/s00421-011-1827-1

Arcelin, R., and Brisswalter, J. (1999). Performance stability in simultaneous tasks of pedalling and reaction time. *Percept. Mot. Skills* 88, 1193–1199. doi: 10.2466/pms.88.3.1193-1199

Ayaz, H., Izzetoglu, M., Bunce, S., Heiman-Patterson, T., and Onaral, B. (2007). "Detecting cognitive activity related hemodynamic signal for brain computer interface using functional near infrared spectroscopy," in *Neural Engineering, 2007. CNE '07. 3rd International IEEE/EMBS Conference on* (Kohala Coast, HI: IEEE), 342–345.

Ayaz, H., and Onaral, B. (2005). *Analytical Software and Stimulus-Presentation Platform to Utilize, Visualize and Analyze Near-Infrared Spectroscopy Measures*. Philadelphia, PA: Master's Degree thesis, Drexel University.

Ayaz, H., Onaral, B., Izzetoglu, K., Shewokis, P. A., McKendrick, R., and Parasuraman, R. (2013). Continuous monitoring of brain dynamics with functional near infrared spectroscopy as a tool for neuroergonomic research: empirical examples and a technological development. *Front. Hum. Neurosci.* 7:871. doi: 10.3389/fnhum.2013.00871

Ayaz, H., Shewokis, P. A., Bunce, S., Izzetoglu, K., Willems, B., and Onaral, B. (2012). Optical brain monitoring for operator training and mental workload assessment. *Neuroimage* 59, 36–47. doi: 10.1016/j.neuroimage.2011.06.023

Ayaz, H., Shewokis, P. A., Bunce, S., and Onaral, B. (2011). "An optical brain computer interface for environmental control," in *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE* (Boston, MA: IEEE), 6327–6330.

Bayazıt, O., Öniz, A., Hahn, C., Güntürkün, O., and Özgören, M. (2009). Dichotic listening revisited: trial-by-trial ERP analyses reveal intra- and interhemispheric differences. *Neuropsychologia* 47, 536–545. doi: 10.1016/j.neuropsychologia.2008.10.002

Brisswalter, J., Collardeau, M., and René, A. (2002). Effects of acute physical exercise characteristics on cognitive performance. *Sports Med.* 32, 555–566. doi: 10.2165/00007256-200232090-00002

Byun, K., Hyodo, K., Suwabe, K., Ochi, G., Sakairi, Y., Kato, M., et al. (2014). Positive effect of acute mild exercise on executive function via arousal-related prefrontal activations: an fNIRS study. *Neuroimage* 98, 336–345. doi: 10.1016/j.neuroimage.2014.04.067

Chang, Y. K., Labban, J. D., Gapin, J. I., and Etnier, J. L. (2012). The effects of acute exercise on cognitive performance: a meta-analysis. *Brain Res.* 1453, 87–101. doi: 10.1016/j.brainres.2012.02.068

Coles, K., and Tomporowski, P. D. (2008). Effects of acute exercise on executive processing, short-term and long-term memory. *J. Sports Sci.* 26, 333–344. doi: 10.1080/02640410701591417

Cope, M., and Delpy, D. T. (1988). System for long-term measurement of cerebral blood and tissue oxygenation on newborn infants by near infra-red transillumination. *Med. Biol. Eng. Comput.* 26, 289–294. doi: 10.1007/bf02447083

Davenport, M. H., Hogan, D. B., Eskes, G. A., Longman, R. S., and Poulin, M. J. (2012). Cerebrovascular reserve: the link between fitness and cognitive function? *Exerc. Sport Sci. Rev.* 1, 153–158. doi: 10.1097/JES.0b013e3182553430

Dietrich, A. (2003). Functional neuroanatomy of altered states of consciousness: the transient hypofrontality hypothesis. *Conscious. Cogn.* 12, 231–256. doi: 10.1016/s1053-8100(02)00046-6

Dotan, R., and Bar-Or, O. (1983). Load optimization for the Wingate Anaerobic test. *Eur. J. Appl. Physiol. Occup. Physiol.* 51, 409–417. doi: 10.1007/bf00429077

Drigny, J., Gremeaux, V., Dupuy, O., Gayda, M., Bherer, L., Juneau, M., et al. (2014). Effect of interval training on cognitive functioning and cerebral oxygenation in obese patients: a pilot study. *J. Rehabil. Med.* 46, 1050–1054. doi: 10.2340/16501977-1905

Dupuy, O., Gauthier, C. J., Fraser, S. A., Desjardins-Crèpeau, L., Desjardins, M., Mekary, S., et al. (2015). Higher levels of cardiovascular fitness are associated with better executive function and prefrontal oxygenation in younger and older women. *Front. Hum. Neurosci.* 9:66. doi: 10.3389/fnhum.2015.00066

Emery, C. F., Honn, V. J., Frid, D. J., Lebowitz, K. R., and Diaz, P. T. (2001). Acute effects of exercise on cognition in patients with chronic obstructive pulmonary disease. *Am. J. Respir. Crit. Care Med.* 164, 1624–1627. doi: 10.1164/ajrccm.164.9.2104137

Endo, K., Matsukawa, K., Liang, N., Nakatsuka, C., Tsuchimochi, H., Okamura, H., et al. (2013). Dynamic exercise improves cognitive function in association with increased prefrontal oxygenation. *J. Physiol. Sci.* 63, 287–298. doi: 10.1007/s12576-013-0267-6

Gateau, T., Durantin, G., Lancelot, F., Scannella, S., and Dehais, F. (2015). Real-time state estimation in a flight simulator using fNIRS. *PLoS One* 10:e0121279. doi: 10.1371/journal.pone.0121279

González-Alonso, J., Dalsgaard, M. K., Osada, T., Volianitis, S., Dawson, E. A., Yoshiga, C. C., et al. (2004). Brain and central haemodynamics and oxygenation during maximal exercise in humans: brain circulation and metabolism in exercise. *J. Physiol.* 557, 331–342. doi: 10.1113/jphysiol.2004.060574

Huppert, T. J., Diamond, S. G., Franceschini, M. A., and Boas, D. A. (2009). HomER: a review of time-series analysis methods for near-infrared spectroscopy of the brain. *Appl. Opt.* 48, D280–D298. doi: 10.1364/ao.48.00d280

Hyodo, K., Dan, I., Suwabe, K., Kyutoku, Y., Yamada, Y., Akahori, M., et al. (2012). Acute moderate exercise enhances compensatory brain activation in older adults. *Neurobiol. Aging* 33, 2621–2632. doi: 10.1016/j.neurobiolaging.2011.12.022

Jaeggi, S. M., Buschkuehl, M., Perrig, W. J., and Meier, B. (2010). The concurrent validity of the N-back task as a working memory measure. *Memory* 18, 394–412. doi: 10.1080/09658211003702171

Jonides, J., Schumacher, E. H., Smith, E. E., Lauber, E. J., Awh, E., Minoshima, S., et al. (1997). Verbal working memory load affects regional brain activation as measured by PET. *J. Cogn. Neurosci.* 9, 462–475. doi: 10.1162/jocn.1997.9.4.462

Jung, R., Moser, M., Baucsek, S., Dern, S., and Schneider, S. (2015). Activation patterns of different brain areas during incremental exercise measured by near-infrared spectroscopy. *Exp. Brain Res.* 233, 1175–1180. doi: 10.1007/s00221-015-4201-4

Khan, N. A., and Hillman, C. H. (2014). The relation of childhood physical activity and aerobic fitness to brain function and cognition: a review. *Pediatr. Exerc. Sci.* 26, 138–146. doi: 10.1123/pes.2013-0125

Lambourne, K., and Tomporowski, P. (2010). The effect of exercise-induced arousal on cognitive task performance: a meta-regression analysis. *Brain Res.* 1341, 12–24. doi: 10.1016/j.brainres.2010.03.091

Li, C., Gong, H., Zeng, S., and Luo, Q. (2005). "Verbal working memory load affects prefrontal cortices activation: evidence from a functional NIRS study in humans," in *Proceedings of SPIE - The International Society for Optical Engineering 5696*, San Jose, CA ed. V. V. Tuchin, 33–40.

Mandrick, K., Derosiere, G., Dray, G., Coulon, D., Micallef, J.-P., and Perrey, S. (2013). Prefrontal cortex activity during motor tasks with additional mental load requiring attentional demand: a near-infrared spectroscopy study. *Neurosci. Res.* 76, 156–162. doi: 10.1016/j.neures.2013.04.006

MATLAB and Statistics Toolbox. (2007). Natick, Massachusetts, United States: The MathWorks, Inc.

McDonald, J. H. (2009). *Handbook of Biological Statistics.* Baltimore: Sparky House Publishing.

McMorris, T., and Hale, B. J. (2012). Differential effects of differing intensities of acute exercise on speed and accuracy of cognition: a meta-analytical investigation. *Brain Cogn.* 80, 338–351. doi: 10.1016/j.bandc.2012.09.001

McMorris, T., Sproule, J., Turner, A., and Hale, B. J. (2011). Acute, intermediate intensity exercise and speed and accuracy in working memory tasks: a meta-analytical comparison of effects. *Physiol. Behav.* 102, 421–428. doi: 10.1016/j.physbeh.2010.12.007

Mehta, R. K., and Parasuraman, R. (2014). Effects of mental fatigue on the development of physical fatigue: a neuroergonomic approach. *Hum. Factors* 56, 645–656. doi: 10.1177/0018720813507279

Nybo, L., and Rasmussen, P. (2007). Inadequate cerebral oxygen delivery and central fatigue during strenuous exercise. *Exerc. Sport Sci. Rev.* 35, 110–118. doi: 10.1097/jes.0b013e3180a031ec

Ozgoren, M., Tetik, M., Izzetoglu, K., Oniz, A., and Onaral, B. (2012). "Effect of body position on NIRS based hemodynamic measures from prefrontal cortex," in *Advances in Brain Inspired Cognitive Systems Lecture Notes in Computer Science*, eds H. Zhang, A. Hussain, D. Liu and Z. Wang (Berlin, Heidelberg: Springer), 138–146.

Potter, D., and Keeling, D. (2005). Effects of moderate exercise and circadian rhythms on human memory. *J. Sport Exerc. Psychol.* 27, 117–125.

Rupp, T., and Perrey, S. (2008). Prefrontal cortex oxygenation and neuromuscular responses to exhaustive exercise. *Eur. J. Appl. Physiol.* 102, 153–163. doi: 10.1007/s00421-007-0595-4

Schmit, C., Davranche, K., Easthope, C. S., Colson, S. S., Brisswalter, J., and Radel, R. (2015). Pushing to the limits: the dynamics of cognitive control during exhausting exercise. *Neuropsychologia* 68, 71–81. doi: 10.1016/j.neuropsychologia.2015.01.006

Shibuya, K.-I., Tanaka, J., Kuboyama, N., and Ogaki, T. (2004). Cerebral oxygenation during intermittent supramaximal exercise. *Respir. Physiol. Neurobiol.* 140, 165–172. doi: 10.1016/j.resp.2003.11.004

Soga, K., Shishido, T., and Nagatomi, R. (2015). Executive function during and after acute moderate aerobic exercise in adolescents. *Psychol. Sport Exerc.* 16, 7–17. doi: 10.1016/j.psychsport.2014.08.010

Tam, N. D., and Zouridakis, G. (2014). Temporal decoupling of oxy- and deoxy-hemoglobin hemodynamic responses detected by functional near-infrared spectroscopy (fNIRS). *J. Biomed. Eng. Med. Imaging* 1, 18–28. doi: 10.14738/jbemi.12.146

Tomporowski, P. D. (2003). Effects of acute bouts of exercise on cognition. *Acta Psychol. (Amst)* 112, 297–324. doi: 10.1016/s0001-6918(02)00134-8

Varner, L. J., and Ellis, H. C. (1998). Cognitive activity and physiological arousal: processes that mediate mood-congruent memory. *Mem. Cognit.* 26, 939–950. doi: 10.3758/bf03201174

Weinstein, Y., Bediz, C., Dotan, R., and Falk, B. (1998). Reliability of peak-lactate, heart rate and plasma volume following the Wingate test. *Med. Sci. Sports Exerc.* 30, 1456–1460. doi: 10.1097/00005768-199809000-00017

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Bediz, Oniz, Guducu, Ural Demirci, Ogut, Gunay, Cetinkaya and Ozgoren. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution and reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 16 February 2016
doi: 10.3389/fnhum.2016.00053

Prefrontal Cortex Activation Upon a Demanding Virtual Hand-Controlled Task: A New Frontier for Neuroergonomics

*Marika Carrieri[1], Andrea Petracca[1], Stefania Lancia[1], Sara Basso Moro[1], Sabrina Brigadoi[2], Matteo Spezialetti[1], Marco Ferrari[3], Giuseppe Placidi[1] and Valentina Quaresima[1]**

[1] Department of Life, Health and Environmental Sciences, University of L'Aquila, L'Aquila, Italy, [2] Department of Developmental Psychology, University of Padova, Padova, Italy, [3] Department of Physical and Chemical Sciences, University of L'Aquila, L'Aquila, Italy

Functional near-infrared spectroscopy (fNIRS) is a non-invasive vascular-based functional neuroimaging technology that can assess, simultaneously from multiple cortical areas, concentration changes in oxygenated-deoxygenated hemoglobin at the level of the cortical microcirculation blood vessels. fNIRS, with its high degree of ecological validity and its very limited requirement of physical constraints to subjects, could represent a valid tool for monitoring cortical responses in the research field of neuroergonomics. In virtual reality (VR) real situations can be replicated with greater control than those obtainable in the real world. Therefore, VR is the ideal setting where studies about neuroergonomics applications can be performed. The aim of the present study was to investigate, by a 20-channel fNIRS system, the dorsolateral/ventrolateral prefrontal cortex (DLPFC/VLPFC) in subjects while performing a demanding VR hand-controlled task (HCT). Considering the complexity of the HCT, its execution should require the attentional resources allocation and the integration of different executive functions. The HCT simulates the interaction with a real, remotely-driven, system operating in a critical environment. The hand movements were captured by a high spatial and temporal resolution 3-dimensional (3D) hand-sensing device, the LEAP motion controller, a gesture-based control interface that could be used in VR for tele-operated applications. Fifteen University students were asked to guide, with their right hand/forearm, a virtual ball (VB) over a virtual route (VROU) reproducing a 42 m narrow road including some critical points. The subjects tried to travel as long as possible without making VB fall. The distance traveled by the guided VB was 70.2 ± 37.2 m. The less skilled subjects failed several times in guiding the VB over the VROU. Nevertheless, a bilateral VLPFC activation, in response to the HCT execution, was observed in all the subjects. No correlation was found between the distance traveled by the guided VB and the corresponding cortical activation. These results confirm the suitability of fNIRS technology to objectively

OPEN ACCESS

Edited by:
Hasan Ayaz,
Drexel University, USA

Reviewed by:
Stephane Perrey,
University of Montpellier - EuroMov, France
Murat Perit Cakir,
Middle East Technical University, Turkey

***Correspondence:**
Valentina Quaresima
valentina.quaresima@univaq.it

Received: *04 November 2015*
Accepted: *01 February 2016*
Published: *16 February 2016*

Citation:
Carrieri M, Petracca A, Lancia S, Basso Moro S, Brigadoi S, Spezialetti M, Ferrari M, Placidi G and Quaresima V (2016) Prefrontal Cortex Activation Upon a Demanding Virtual Hand-Controlled Task: A New Frontier for Neuroergonomics. Front. Hum. Neurosci. 10:53. doi: 10.3389/fnhum.2016.00053

evaluate cortical hemodynamic changes occurring in VR environments. Future studies could give a contribution to a better understanding of the cognitive mechanisms underlying human performance either in expert or non-expert operators during the simulation of different demanding/fatiguing activities.

Keywords: functional near-infrared spectroscopy, neuroergonomics, hand-controlled task, LEAP motion controller, virtual reality, remote control, brain activation

INTRODUCTION

The term neuroergonomics was first introduced in 1997 for depicting an interdisciplinary area of research which involves the intersection of two disciplines: neuroscience and ergonomics (Parasuraman and Rizzo, 2007). The studies in this field, previously carried out using either mobile or immobile neuroimaging techniques, have been nicely reviewed by Mehta and Parasuraman (2013). Virtual reality (VR), a computer-based technology that allows the creation of multisensory simulated environments in which users can interact and receive real-time feedbacks on their performance, was claimed by Kearney et al. (2007) to be highly relevant for neuroergonomics. This because VR can replicate, with a greater control than that applicable in the real world, a wide range of conditions that are impractical or impossible to observe in the real situations; then allowing behavioral and neurophysiological observations of the mind and brain at work. Given its peculiarity, VR is also effectively used by human operators to accomplish their work in dangerous environments, thus avoiding any physical risk. For instance, applying gesture-based control interfaces, VR is usually employed for tele-operated applications such as driving robots, rovers and other devices remotely, with the operators at a certain distance from them (Chen et al., 2015; Liu and Zhang, 2015; Wei et al., 2015). The tele-operated systems are very expensive, unique neither replicable nor quickly replaceable, and from their proper use depends the success or failure of long-planned, critical, costly and challenging operations. Taking into account the high degree of responsibility inherent to operators' duties, their considerable physical/cognitive work should be evaluated objectively by neuroimaging techniques in the framework of neuroergonomics (for review, see Gramann et al., 2011, 2014; Mehta and Parasuraman, 2013).

Although the most widely used immobile functional neuroimaging modality has been undoubtedly represented by functional magnetic resonance imaging (fMRI), the development of portable and wearable neuroimaging devices, comprising electroencephalography (EEG) and functional near infrared spectroscopy (fNIRS; for review, see Mehta and Parasuraman, 2013; Gramann et al., 2014; Scholkmann et al., 2014) has considerably facilitated the approach of neuroergonomics. fNIRS, with its high degree of ecological validity and its very limited requirement of physical constraints to subjects, represents a valuable tool for monitoring cortical responses in the research fields of neuroergonomics (for reviews, see Ayaz et al., 2013; Derosière et al., 2013). Furthermore, compared to fMRI, fNIRS is silent, allowing to avoid any bias in the results due to difficulties in focusing on the task because of the high level of noise. Briefly, fNIRS is a non-invasive vascular-based functional neuroimaging technology which assesses, simultaneously from multiple measurement sites, concentration changes in oxygenated-deoxygenated hemoglobin (O_2Hb/HHb, respectively) at the level of the cortical microcirculation blood vessels (Scholkmann et al., 2014). O_2Hb/HHb, indeed, interact differently with near infrared light, so that both physiological indexes can be recovered from the measured signal. This is a further advantage of fNIRS over fMRI, the latter being able to recover only a single physiological index, namely the blood-oxygen level dependent (BOLD) signal. When a specific brain region is activated, cerebral blood flow increases in a temporally and spatially coordinated manner through a complex sequence of coordinated events, tightly linked to changes in neural activity (i.e., neurovascular coupling). The coupling between the neuronal activity and the cerebral blood flow is fundamental to brain function. fNIRS relies exactly on this coupling to reveal the activated cortical region by measuring the associated cortical blood oxygenation changes (i.e., the increase in O_2Hb and the decrease in HHb).

Since 1993, fNIRS has been employed for evaluating the spatiotemporal characteristics of the cortical activation during different motor tasks related to upper and/or lower limb exercise (for a review, see Leff et al., 2011). In most previous fNIRS studies, the activation of the sensory-motor cortex and the PFC has been widely investigated in different tasks of the lower limb like walking (Koenraadt et al., 2014; Mirelman et al., 2014), stepping (Huppert et al., 2013), precision stepping (Koenraadt et al., 2014), etc. Several fNIRS studies have also investigated the PFC activation during hand tasks such as: finger movement (Wriessnegger et al., 2012), passive finger movement (Chang et al., 2014), isometric grasping/grasping (Mandrick et al., 2013), handgrip exercise (Derosière et al., 2014), learning a hand motor skill (Hatakenaka et al., 2007), etc.

Since 2000, fNIRS has been also employed in real-world activities realized in VR environment for evaluating the PFC activation during the simulation of different hand-related demanding/fatiguing activities, like airplane piloting (Ayaz et al., 2012a; Durantin et al., 2014; Gateau et al., 2015), car driving (Tomioka et al., 2009), grasping (Holper et al., 2010, 2012), natural orifice transluminal endoscopic surgery (James et al., 2011), etc. Interestingly, the inferior frontal gyrus (Ayaz et al., 2012a) and the dorsolateral PFC (DLPFC;

Durantin et al., 2014; Gateau et al., 2015) were found activated during airplane piloting tasks. The bilateral ventrolateral PFC (VLPFC) was found activated during natural orifice transluminal endoscopic surgery when the simulation required a more difficult navigation path through an orifice (James et al., 2011).

It is well-known that the PFC, and in particular the VLPFC and DLPFC, are involved in the control of the motor actions. On one hand, it has been demonstrated that the VLPFC is involved in visuo-motor learning tasks (Yamagata et al., 2012; Hoshi, 2013). Moreover the reflexive orienting seems to be controlled by the right VLPFC (Corbetta et al., 2008), whereas the goal relevant information for the action control seems to be maintained and retrieved by the left VLPFC (Badre and Wagner, 2007; Souza et al., 2009). On the other hand, the DLPFC apparently plays a specific role in learning by trial and error (Halsband and Lange, 2006). Furthermore DLPFC, for its involvement in mediating and monitoring of actions, is considered to be the major anatomical correlate of the central executive (Baddeley, 2003; Gateau et al., 2015). Therefore, the VLPFC and the DLPFC are involved in associating visual information with motor responses (Halsband and Lange, 2006; Tanji and Hoshi, 2008).

The PFC plays not only a crucial role in single cognitive or motor tasks, but also in combined sensorimotor-cognitive task (i.e., dual-task; Gentili et al., 2013; Mandrick et al., 2013; Mirelman et al., 2014). Several fNIRS studies have reported that, in comparison to a single task, the attention-demanding dual-tasks (e.g., walking while talking, calculating while stepping, balancing a ball while walking, etc.) induced an increase of the PFC activation due to a greater cognitive load (Holtzer et al., 2011). For instance, Mandrick et al. (2013) investigated how an additional mental load (i.e., arithmetic task) during isometric grasping affects the PFC activation. The performance of the mental task was impaired when the motor task difficulty increased, suggesting that performing a dual-task requires more attentional resources than performing a single task.

In the last few years, the use of VR interfaces, driven by natural hand movements for remote control, is growing-up thanks to the development of innovative optical 3-dimensional (3D) systems for gesture recognition (Erden and Çetin, 2014). The key advantage of gesture recognition technology is that no physical contact is required between the human body and the gesture recognition device, so that the subjects can move freely. One of the most recent optical 3D sensors, based on stereo-vision, is the LEAP Motion Controller® (LEAP). The LEAP is a high-resolution 3D hand-sensing device, which allows the freehand natural interaction crucial for the implementation of real-time, realistic VR systems. This low cost non-bulky device has an extremely accurate reactivity (Bachmann et al., 2014). In addition, the 3D rendering technologies, including state of the art displays and visors specifically designed for VR (Dodgson, 2013; Desai et al., 2014; Nan et al., 2014), have permitted a large development of VR techniques. In particular, their high visual/rendering fidelity and an immersive wide field of view enables the sensation of presence and the feeling being actually inside the virtual scene with the possibility to have multi-sensorial feedbacks.

The aim of the present study was to investigate non-invasively by fNIRS the PFC responses in healthy subjects while performing a complex hand-controlled task (HCT) in a VR environment. This task emulated the interaction with a real, remotely-driven, system operating into a critical environment. The hand movements were captured by the LEAP, a high spatial and temporal resolution 3D hand-sensing device. The subjects were asked to move their right hand/forearm with the purpose of guiding a virtual ball (VB) over a virtual route (VROU). The VROU can be easily and purposely designed to replicate a real track that an operator should travel to carry on a given challenging operation. The HCT-related PFC response was monitored non-invasively by a 20-channel fNIRS system. The HCT involved the control of the hand/forearm movement, and the active interaction with the virtual environment through the hand/forearm motor actions. The attentional resources allocation and the integration of different executive functions (e.g., coordination, planning, decision making, etc.) are needed in performing the HCT. Taking into account the well-known role played by the VLPFC and the DLPFC in motor action preparation and in the allocation of the attentional resources to generate goals from current situations, it was hypothesized that either VLPFC or DLPFC or both of them would be activated bilaterally in subjects while performing the HCT.

MATERIALS AND METHODS

Participants

Fifteen University students (all males, age: 26.6 ± 2.9 years; level of education: 14.4 ± 2.1 years), without neurological or psychiatric illness and normal or corrected-to-normal vision were recruited in the study. In order to prevent any gender differences in emotional responses (Matud, 2004) and in visuo-motor abilities (Wang et al., 2015), only men were enrolled. To exclude left-handed subjects, all participants completed the Edinburgh Handedness Inventory assessing hand dominance. Following a full explanation of the protocol and its non-invasiveness, and prior to the starting of the experimental procedure, a written informed consent was obtained from each participant. All procedures were conducted in accordance with the Declaration of Helsinki and approved by the University Ethics Committee.

Experimental Setup

Hand-Controlled Task (HCT)

A VR HCT was implemented by integrating a LEAP Motion Controller® with a real-time 3D engine. The LEAP provides both a 3D hand model and real-time hand tracking information for enabling subjects to transpose their hand movements within the virtual 3D HCT (**Figure 1**). The LEAP is a small (1.3 cm × 3.2 cm × 8 cm) 3D sensor which uses two internal infra-red (IR) cameras and three IR

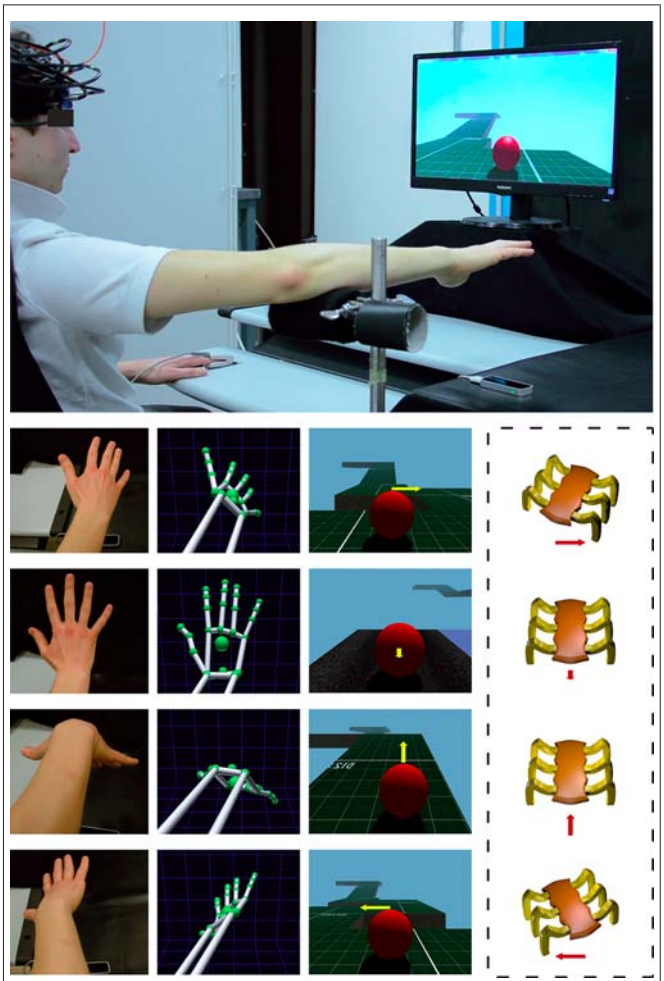


FIGURE 1 | Experimental setting for the execution of a virtual reality (VR) hand-controlled task (HCT). The upper large image shows the subject with the functional near-infrared spectroscopy (fNIRS) probe holder while sitting in front of a PC, and the positioning of the LEAP motion controller below the operator hand. The four columns of small images show (from left to right): the real position of the subject hand/forearm (first column); the corresponding hand/forearm virtual model (second column); the influence of the commands on the virtual ball (VB) during the HCT (third column); and the corresponding effects on a real, remote, spider-like rover (fourth column, in dashed line to indicate that the presence of the rover is just supposed). The yellow arrow indicates the direction of the guided VB and the red arrow indicates the corresponding effect on the rover. The length of the arrows indicates the force amplitude impressed by the operator. Note that, during the HCT, the hand/forearm virtual model (second column) was not shown to the operator and the visualization of the arrows (third column) was disabled.

light emitting diodes to detect objects within a dome of approximately 0.22 m^3 above it. Its spatial and temporal resolution is 1 mm and 15 ms, respectively. The LEAP, connected to a computer via a USB cable, is designed specifically to detect, in real-time, hand and finger motions and gestures, such as pinching fingers, closing hand, tapping, etc. This device, positioned under the palm center of the right hand at a distance of about 25 cm (**Figure 1**), was utilized to: (1) capture the movements of the hand; (2) associate hand movements to a virtual hand model; and (3) translate the movements of the virtual hand model to a set of commands in order to drive a VB within a virtual environment.

In the present study, the adopted virtual environment was aimed at simulating the driving action of a spider-like robot similar to the one developed by the National Aeronautics and Space Administration (NASA). Due to its high stability, equilibrium and ability to change quickly direction, this robot has been proven to be adapted for moving in very rough environments and in environments designed for humans (to go up and down stairs). The movements of the ball, in fact, simulated fairly those ones performed by the considered robot (lateral, ahead, behind, and stop). The aim of this adopted HCT was to move the VB over a VROU of a fixed length (42 m; **Figure 2**), trying to travel as long as possible the distance in a fixed time without falling (2 min). Either in the case of VB falling or in the case of accomplishment of a VROU in advance, subjects were requested to restart the VROU from the beginning. To calculate the whole distance traveled by each subject over the HCT, the completed VROU and the distance traveled until the task end were considered. The VROUs with failure were not considered. The VROU adopted in this study (**Figure 2**) was purposely designed to reproduce a narrow road including some critical points (i.e., stairs, turns, a slippery part, and climbs). The Torque 3D Engine[1], a cross-platform high performance real-time 3D engine, was used both for the editing and the rendering of the whole virtual 3D HCT. A controllable VB and a 3D VROU were created by using a customized version of Marble Motion, a well-known game[2]. In particular, the whole native source code was rewritten and enriched in order to fulfill the requirements of the task: (1) a time driven version of the task in which both the start and the end of the HCT were established by a fixed time interval; (2) the possibility of storing information related to the operator-task interaction process (including the number of times in which the VB was fallen out of the VROU, the position of the VB falls over the VROU, the distance traveled by the VB in the given time). The software allowed also the calculation of some dynamic parameters (followed trajectory, VB speed and acceleration). These last parameters have been considered to calculate the VB speed along the VROU, to get information about resting periods occurred during the HCT, and to evaluate the subject skills while executing the HCT. Moreover, to increase the subject's concentration and motivation during the HCT, the whole technical items, including elements of the VROU (textures and materials), were redefined utilizing some predefined items of the 3D engine editor (Torque 3D Editor[3].

The subject was asked to place his right forearm on a fixed and firm support in order to allow the hand capture (**Figure 1**). This support ensured the maintenance of the correct position of the forearm, and consequently of the hand, during the execution of the HCT. The task started with a stationary

[1]http://www.garagegames.com/products/torque-3d
[2]http://mit.garagegames.com/MarbleMotion-1-0b.zip
[3]http://www.garagegames.com/products/torque-3d/overview/editor

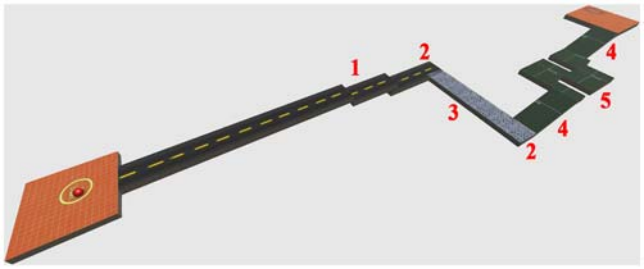


FIGURE 2 | A perspective view of the designed virtual route (VROU) which reproduces a narrow road including some critical points.
1: stairs; 2: turns; 3: slippery part; 4: climbs; and 5: sequential turns.

VB placed at the beginning of the VROU. The subject had to maintain his right hand opened over the LEAP device by keeping his forearm on the support with the center of the palm perpendicular to the center of the device and with all the five fingers extended (**Figure 1**). At the beginning of the HCT, the subject had to guide the VB over the VROU by using four commands (**Figure 1**). The first command (hand flexion) made the VB to proceed forward; the second command (hand extension) made the VB to decrease the speed (up to stop the VB) and to proceed backward; the third and the fourth commands (counter clockwise and clockwise rotations of the wrist) made the VB to move toward left or right, respectively (a combined use of the hand flexion/extension movements or the rotation of the wrist made the VB to stop). These hand movements had a real-time proportional impact on the VB. More specifically, the command chosen by the subject transmitted the direction and the "force" to the VB (e.g., a low hand downward flexion corresponded to a low "force" application to the VB in the forward direction), especially when the VB speed was depending on the inclination degree of the downward flexion and the time during which the hand was maintained at the same position. Thus, the HCT was purposely designed to combine the four main commands and when the subject assumed a pose of his hand halfway between two commands, the system merged both directions and "force" amplitude. In this way, the subject had the feeling to guide the VB without restrictions or constraints.

The effective range of the LEAP tracking system is limited to roughly 60 cm due to the low near infrared light intensity. In this study, the distance between the fNIRS head probe and the LEAP was always greater than 60 cm. Therefore, the LEAP should have not interfered with the fNIRS measurements. Interestingly, several fNIRS studies in combination with other near infrared based tracking systems such as Kinect or eye trackers have been published (Kita et al., 2010; Sukal-Moulton et al., 2014; Urakawa et al., 2015). Although those systems have utilized more powerful light emitters (than the LEAP), and the light emitters have been directly pointed toward the fNIRS head probe, the potential interference with the fNIRS data was not mentioned.

fNIRS Instrumentation and Data Processing

A two-wavelength continuous wave 20-channel fNIRS system (Oxymon Mk III, Artinis Medical Systems, Netherlands) was utilized to map non-invasively the changes in O_2Hb and HHb over the bilateral PFC. The details of this instrumentation have been previously reported (Basso Moro et al., 2013). The O_2Hb/HHb data from the 20 channels were acquired at 10 Hz. The O_2Hb/HHb concentration changes (expressed in $\Delta\mu M$), obtained by using the modified Beer-Lambert law and the age-dependent differential pathlenght factor ($4.99 + 0.067 \times Age^{0.814}$) were displayed in real-time on a PC monitor. Eight optical fiber bundles (length: 3.15 m; diameter: 4.5 mm) were utilized to transport the light to the left and the right PFC (four for each hemisphere), whereas ten optical fiber bundles of the same size (five for each hemisphere) were utilized to collect the light emerging from the PFC. The illuminating and collecting bundles were assembled into a flexible probe holder, consisting of two mirror-like units (9.7 cm × 8.9 cm each) held together by three flexible junctions. In 16 out of the 20 channels the illuminator-detector distance was set at 3.5 cm, while in the remaining four channels the illuminator-detector distance was set at 1 cm (short-separation channels or SS channels). In the 16 channels, the measurement points were defined as the midpoint of the corresponding illuminator-detector pairs. The probe holder was placed over the subject head by a Velcro brand fastener in order to get a stable optical contact with the scalp (**Figure 1**). In particular, the two frontopolar fibers bundles, collecting the light at the bottom of the holder, were centered (according to the International 10–20 system for the EEG electrode placement) on the Fp1 and Fp2 locations for the left and right hemisphere, respectively. The pressure created by the fastener was sufficient to induce a partial transient blockage of the skin circulation during the fNIRS study. The adopted procedure would suggest that a consistent reduction of forehead skin blood flow was occurring as a result of this approach. The Montreal Neurological Institute coordinates of the optodes and the relative 16 measurement points were calculated using a probe placement method. For the details of this procedure see Basso Moro et al. (2013). The measurement points 1, 2, 3, 9, 10, 11 corresponded to the DLPFC, which includes part of the Brodmann's Area (BA) 46; the measurement points 5, 6, 13, 14 corresponded to the frontopolar cortex, which includes part of the BA 10; and measurement points 4, 7, 8, 12, 15, 16 corresponded to the VLPFC, which includes part of the BA 45.

During the data collection procedure, the fNIRS signal quality as well as the absence of movement artifacts were verified on the PC monitor. The subject's heart rate (HR) was monitored by a pulse oximeter (N-600, Nellcor, Puritan Bennett, St. Louis, MO, USA) with the sensor clipped to the index finger of the left hand. The Homer2 NIRS processing package[4] was employed to analyze the data. Raw intensity data in each channel were converted into optical density changes (OD). Channels showing low intensity values were

[4]http://www.nmr.mgh.harvard.edu/PMI/resources/homer2/home.htm

excluded from further analyses. The Wavelet motion correction method was employed to correct motion artifacts. Based on the method developed by Molavi and Dumont (2012), it sets to zero all wavelet detail coefficients exceeding a predefined threshold (iqr = 0.1). The modified Beer-Lambert law was then applied to convert the corrected OD data into concentration changes. A General Linear Model (GLM) approach (hmrDeconvHRF_DriftSS) was utilized to recover the mean hemodynamic response function (HRF) for each subject and channel. The approach, consisting in adding the SS channel signal with the highest correlation with the analyzed standard channel signal in the design matrix, was able to reduce the contribution of the mean arterial blood pressure changes in task-evoked fNIRS signal. The less restrictive set of Gaussian functions with standard deviation (SD) of 3 s and with their means separated by 2 s was chosen as temporal basis functions (ranging between −20 before and 210 s after the starting of the HCT; Gagnon et al., 2011).

Experimental Design

A familiarization/training phase was carried out 3 days before the study. The subjects were informed about the procedures and familiarized with both the experimental setting and the HCT. During this phase, the fNIRS probe holder was placed over the head of the subjects who were trained to stay as firm as possible to avoid movement artifacts in fNIRS measurement during the HCT execution. After evaluating the joint mobility of their hand and wrist, the subjects were requested to pay attention to the presented four commands to be used for guiding properly the VB (hand flexion, hand extension, counter clockwise and clockwise rotations of the wrist). Later, the subjects were asked to place their right hand opened over the LEAP in order to verify the correctness of the estimated hand virtual model. Once this phase was completed, the subjects were asked to guide a VB over a VROU. In order to avoid a potential learning effect, a different VROU was used in this training phase. For each subject, the training phase was considered completed when he demonstrated his ability to guide properly the VB. After 3 days, at the same time of the training phase, the subjects participated in the study. A monetary reward was not given.

This study was carried out in a quiet and dimly lit room. The subject was asked to sit on a comfortable high-backed chair in front of a 17″ PC monitor, and to keep his forearm on the firm support with the right hand opened over the LEAP (see "The Hand-Controlled Task" Section; **Figure 1**). The HCT protocol lasted 6 min. Specifically, the protocol started with a 3 min baseline, during which subjects were asked to relax (observing a white fixation cross presented on a black screen) in order to get stable fNIRS signals. Then, a stationary VB came into view on the PC monitor, and a visual instruction informed the subjects that the 2 min HCT was starting. During the HCT, the subject had to guide the VB over the VROU through the four hand movements (see "The Hand-Controlled Task" Section). When the subject failed in guiding the VB over the VROU or the subject completed the VROU in less than 2 min, the VB was repositioned at the VROU starting point, and the route restarted. At the end of the HCT execution, there was a recovery period (1 min), in which the subject was requested to relax while observing a white fixation cross presented on a black screen. In order to evaluate the potential "state anxiety" provoked by HCT, all the subjects completed the 20-items of the State Trait Anxiety Inventory Form Y-1 (STAI) before and after the protocol.

Data Analysis and Statistics

The integral values of the O_2Hb/HHb ($_{INT}O_2Hb$/HHb) changes of the HRFs were calculated from the beginning (at 0 s) until the end of the HCT (at 120 s), for each measurement point and subject, and were used as metric for the following statistical analysis. The mean values of the HR changes (analyzed as percentage of control) were calculated from the beginning (at 0 s) until the end of the HCT (at 120 s). Both the $_{INT}O_2Hb$/HHb changes and the HR mean values were corrected for the baseline periods, calculated over the last 20 s before the starting of the HCT. The median of the values of the distance traveled by the guided VB was calculated to subdivide the subjects in two groups: best performers (above the median) and worst performers (below the median). The subject of the median value was not included in any group. Student's *t*-test was conducted in order to evaluate the presence of any difference, in terms of the distance traveled by the guided VB between the worst and best performers.

All data were examined for normality and sphericity using Shapiro–Wilk and Mauchly's Sphericity tests, respectively. Each level of the independent variables followed a normal or approximately normal distribution in all the dependent variables (O_2Hb, HHb and HR), permitting the use of parametric statistical analyses. When the sphericity was not assumed, the Greenhouse-Geisser correction was utilized.

In order to investigate the PFC activation in response to HCT, the two-way analysis of variance (ANOVA) was applied to $_{INT}O_2Hb$/HHb changes. The ANOVA included two factors: measurement point (16 levels) and cortical hemodynamic response (CHR; i.e., corrected task period vs. zero; 2 levels). To control for multiple significance tests, the Fisher's least significant difference adjustment was applied. A series of one-way ANOVAs was performed for the HCT in order to evaluate the influence of the CHR (2 levels) on the $_{INT}O_2Hb$/HHb changes. In particular, the one-way series of ANOVAs were performed only for the $_{INT}O_2Hb$/HHb changes related to the measurement points 7, 8, 15 and 16, chosen as descriptive measurement points of the hemodynamic response. For the HCT, the Pearson's correlation coefficient was calculated in order to evaluate the relation between the distance traveled by the guided VB and the $_{INT}O_2Hb$/HHb changes in the 7, 8, 15 and 16 measurement points. A one-way ANOVA was performed for the HCT in order to evaluate the influence of the CHR (2 levels) on the HR changes. The Pearson's correlation coefficient was also calculated for evaluating the relation between the distance traveled by the guided VB and the

HR. Student's *t*-tests were conducted in order to evaluate the presence of any difference in: (1) the anxiety state before and after the protocol; (2) $_{INT}O_2Hb$/HHb changes in the 7, 8, 15 and 16 measurement points between the worst performers and best performers; and (3) the distance traveled by the guided VB of worst and best performers.

All statistical analyses were conducted with SPSS 20.0 (SPSS Inc., Chicago, IL, USA). Data were expressed as mean ± *SD*. The criterion for significance was $p < 0.05$.

RESULTS

The behavioral data analysis revealed the following main results. There was no significant difference ($t = 0.64$, $p = 0.53$) in the anxiety state before (28.8 ± 6.4) and after the protocol (28.9 ± 5.8). The distance traveled by the guided VB in the 15 subjects was: 21, 26, 36, 43, 51, 53, 56, 58, 61, 83, 88, 96, 101, 132, and 148 m, respectively. The median value was 58 m. The mean distance was 70.2 ± 37.2 m. The less skilled subjects failed several times in guiding the VB over the VROU. Therefore, the distance traveled by their guided VB was shorter. Indeed, the distance traveled by the guided VB was significantly different ($t = -4.89$, $p < 0.001$) between the worst performers (40.9 ± 13.7 m) and best performers (101.3 ± 29.7 m). The fNIRS data evidenced a heterogeneous O_2Hb/HHb response over the mapped cortical area in the subjects while performing the HCT (**Figure 3**). In particular, since the beginning of the HCT, a progressive O_2Hb increase and a concomitant progressive HHb decrease were observed in the measurement points 7, 8, 15 and 16, corresponding to the VLPFC, which includes part of the BA 45. About 15 s after the end of the HCT, a gradual return of O_2Hb/HHb to the corresponding baseline values, was observed. This delay is reasonable considering that the cerebral blood flow increase lasts over the period of the HCT. The statistical analysis revealed the following main results. The two-way ANOVA analysis, carried out on the $_{INT}O_2Hb$ changes, revealed a significant main effect of: (1) the measurement point ($F_{(2.67,37.47)} = 9.52$, $p < 0.001$), and (2) the measurement point * CHR interaction ($F_{(2.67,37.47)} = 9.52$, $p < 0.001$). The two-way ANOVA analysis, carried out on the $_{INT}$HHb changes, revealed a significant main effect of the: (1) measurement point ($F_{(2.83,39.67)} = 15.70$, $p < 0.001$); (2) CHR ($F_{(1.00,14.00)} = 13.83$, $p = 0.002$); and (3) measurement point * CHR interaction ($F_{(2.83,39.67)} = 15.70$, $p < 0.001$). The two-way ANOVAs, carried out on the $_{INT}O_2Hb$/HHb changes, revealed the main significant differences between the measurement points 7, 8, 15, 16 and all the others (ps < 0.05). The series of one-way ANOVAs, carried out on the $_{INT}O_2Hb$ changes of the measurement points 7, 8, 15, and 16, revealed a significant activation in all the measurement points ($F_{(1,14)} = 5.32$, $p = 0.037$; $F_{(1,14)} = 23.79$, $p < 0.001$; $F_{(1,4)} = 7.33$, $p = 0.017$; $F_{(1,14)} = 16.89$, $p = 0.001$).The series of one-way ANOVAs carried out on the $_{INT}$HHb changes of the measurement points 7, 8, 15, and 16 revealed a significant cortical activation in all the measurement points ($F_{(1,14)} = 41.61$, $p < 0.001$; $F_{(1,14)} = 58.57$, $p < 0.001$; $F_{(1,14)} = 19.12$, $p = 0.001$; $F_{(1,14)} = 26.23$, $p < 0.001$). In the HCT, no correlation was found between the distance traveled by the guided VB and the corresponding $_{INT}O_2Hb$/HHb changes (ps > 0.05). During the HCT, no differences (ps > 0.05) were found in the $_{INT}O_2Hb$/HHb changes in the 7, 8, 15 and 16 measurement points between the worst performers and best performers. The one-way ANOVA analysis for the HR mean values revealed a significant main effect of the CHR ($F_{(2.88,40.40)} = 8.06$, $p < 0.001$). However, the mean values of the HR changes during the execution of the HCT increased only of about 15% with respect to the mean value of the baseline. No correlation was found between the HR and the distance traveled by the guided VB.

DISCUSSION

In this feasibility study, the bilateral PFC was investigated by a multi-channel fNIRS system while subjects performed a demanding VR HCT, a remotely-driven operation simulated by a high-resolution and low-cost 3D hand-sensing device. The observed involvement of the bilateral VLPFC supports the formulated hypothesis.

The results of the present study have indicated a consistent bilateral VLPFC activation (measured as O_2Hb increase and a concomitant HHb decrease) in response to the execution of the VR HCT (**Figure 3**). It has been reported that VLPFC is involved in associating visual information with motor responses (Tanji and Hoshi, 2008). In fact, the execution of the adopted HCT requires the combination of the contextual visual information in order to coordinate the hand/forearm movements for guiding the VB over the VROU. Although the tested subjects had the same age and level of education (University students), their skills in performing the HCT were different. Indeed, the distance traveled by the VB was heterogeneous: the distance traveled by the VB guided by the best performer was about seven times longer than the distance traveled by the VB guided by the worst performer. These results clearly confirm that the designed VROU (**Figure 2**) was really demanding. The diverse skills of the subjects could not be attributable to emotional factors, because no difference between anxiety state before and after the HCT, and no correlation between the distance traveled by the guided VB and the HR of the subjects were found. However, a bilateral VLPFC activation was observed indiscriminately in all the tested subjects, including the ones who never completed at least one VROU. Therefore, the present fNIRS data did not provide the possibility to discriminate the subjects according to their performance. This could be partly explained by the fact that other cortical areas were not investigated in this study and that subcortical areas and/or cortical-subcortical network are supposed to be responsible of the differences between the best and the worst performers. As well known, PFC is involved in the executive functions (e.g., attention, coordination, planning, decision making, etc.), the same required to perform the HCT regardless of the performer's skills. The combined use of a fNIRS-EEG system has evidenced a greater involvement of the deeper structures (e.g., hippocampus) in the "good performers group" compared with the "bad performers group" while executing a spatial navigation

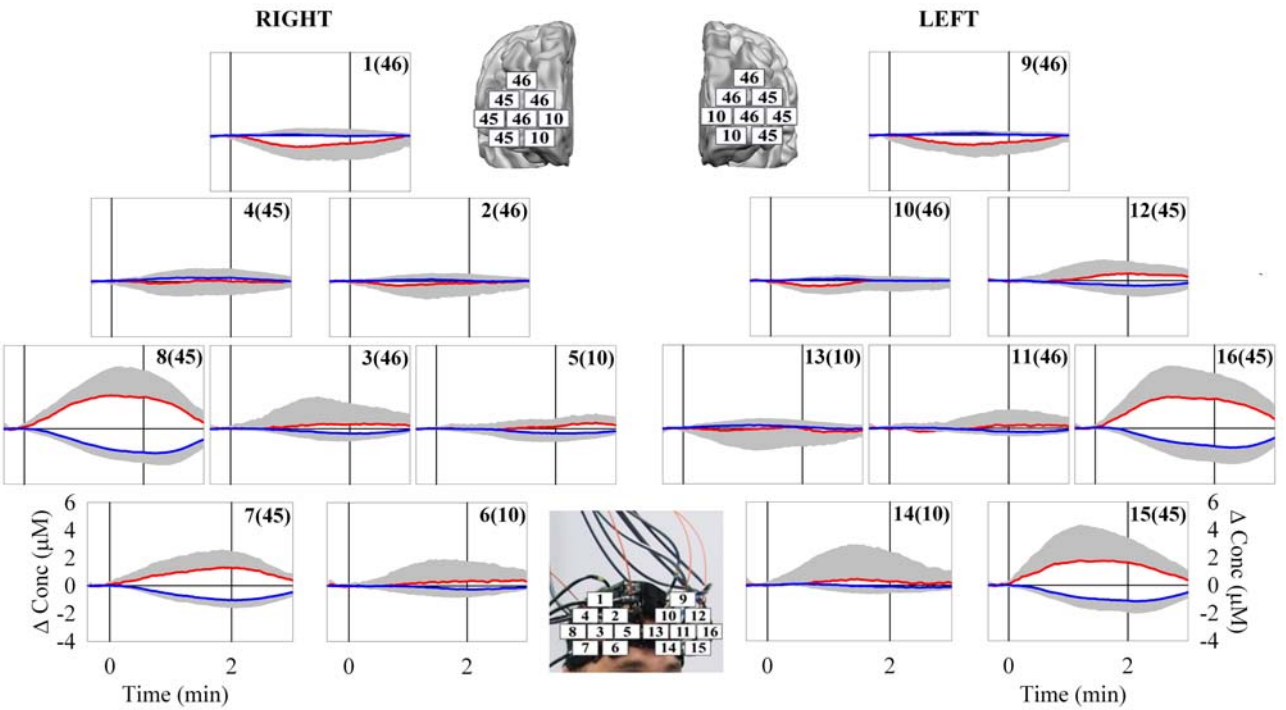


FIGURE 3 | Grand average of O_2Hb (red line) and HHb (blue line) changes observed over the bilateral PFC [16 measurement points of the right (from 1–8) and left (from 9–16) hemisphere] in response to the VR HCT. The corresponding numbers in the brackets refer to the associated Brodmann's Area (BA). The vertical solid lines limit the duration of the task execution. The major cortical activation was observed in the bilateral VLPFC (measurement points 7, 8, 15, and 16). $N = 15$; means ± *SD*.

task (Kober et al., 2013). Moreover, it has been reported that a complex sensorimotor-cognitive task, such as the adopted HCT, would require the involvement of different cortical-subcortical networks including: PFC, spinal cord, brainstem, cerebellum, basal ganglia, and motor cortex (Takakusaki, 2008). This suggests that the degree of the cognitive demand (measured by fNIRS), required for executing the HCT, is not associated with the subjects performance (measured as the distance traveled by the guided VB). This in part confirms the results of other VR fNIRS studies in which a dissociation between the mental work demanded to execute a complex task and the performance output was observed (Ayaz et al., 2012a; Boyer et al., 2015). In the present study, this dissociation could be also explained by the fact that tested subjects had only a short familiarization/training phase. A higher activation of the PFC was usually observed in non-expert subjects while executing a novel VR task compared to expert (Ayaz et al., 2012b). The non-expertise requires the employment of more attentional resources to perform a novel task. On the contrary, the expertise implies an increase of automaticity and does not require the same high level of attention and control (Ayaz et al., 2012b). The VLPFC activation, observed in the present study in non-expert subjects while performing the novel VR HCT, suggests that some of the well-known executive functions (e.g., attention, coordination, planning, decision making, etc.) are required in the learning phase. Then, it could be supposed that the amplitude of the observed VLPFC activation would become lower or even disappear in subjects very familiar with the VR HCT.

Several studies evidenced the advantages of using fNIRS technology for investigating non-invasively cortical responses in subjects while performing different VR tasks; the most representative studies are listed in **Table 1**. The common relevant finding is represented by the activation of different regions of the frontal cortex and the PFC. For example, the involvement of the medial PFC (mPFC) and the frontopolar cortex (Ayaz et al., 2012b) and the inferior frontal gyrus (Harrison et al., 2014) was observed in subjects while executing air traffic control tasks; an activation of the overall frontal cortex was observed in subjects while performing a train piloting task (Kojima et al., 2005). Very recently, the usefulness of fNIRS as a tool to conduct driving research has been nicely reviewed (Liu et al., 2015). For example, an activation of the right PFC (Tomioka et al., 2009) and an activation of the overall PFC (Tsunashima and Yanagisawa, 2009) were found in subjects while executing different simulated car driving tasks. However, in all of the above reported studies, no activation of the VLPFC, and in particular of the BA 45, was found.

The combined use of fNIRS-EEG would be an ideal tool for carrying out studies in the field of neuroergonomics. The pros of lightweight, high-density EEG and fNIRS recording

TABLE 1 | Selected fNIRS studies about the effects of VR tasks on different cortical areas.

VR Tasks	Subjects (patients)	Age (years)	D	Ch	Cortical areas	Main findings	Reference
Air traffic control	24	24–55	D1	16	FP, PFC	Medial FP/PFC activated	Ayaz et al. (2012b)
Air traffic control	12	NA	D1	16	PFC	PFC activated	Harrison et al. (2014)
Airplane piloting	9	36 ± 4	D7	1	FC	FC differently activated by task difficulty	Takeuchi (2000)
Airplane piloting	13	21–28	D1	16	FP, PFC	IFG differently activated by practice	Ayaz et al. (2012a)
Airplane piloting	12	25 ± 5	D1	16	PFC	Correlation DLPFC activation/ performance	Durantin et al. (2014)
Airplane piloting	19	27 ± 6	D1	16	FP, PFC	DLPFC activated bilaterally	Gateau et al. (2015)
Balancing (*swing*)	16	29 ± 5	D2	8	PFC	PFC activated bilaterally	Basso Moro et al. (2014)
Balancing (*tilt board*)	22	26 ± 4	D2	8	PFC	PFC activated bilaterally	Ferrari et al. (2014)
Balancing *(video game)*	9	18–42	D3	32	FC, MC, SC, TC	STG activated	Karim et al. (2012)
Boxing (*video game*)	20	18–40	D4	46	ATC, SMA	ATC and SMA differently activated	Kim et al. (2015)
Car driving	9	NA	D5	42	FC	FC less activated during ACC	Tsunashima and Yanagisawa (2009)
Car driving	14 (12 AD)	NA	D4	52	PFC	PFC less activated in AD	Tomioka et al. (2009)
Dancing	14	22 ± 1	D5	22	(L)PFC, (L)TC	Training-dependent PFC activation	Ono et al. (2015)
Dancing	26	NA	D6	22	(L)PFC, (L)TC	MTG activated	Noah et al. (2015)*
Grasping	23	NA	D8	4	M1, PC PMC, SMA	M1, PMC, SMA, PC activated	Holper et al. (2010)
Grasping	17	26 ± 4	D8	4	PMC, SMA	PMC/SMA differently activated by trials	Holper et al. (2012)
Lathe operation	7	24–26	D6	45	FC, MC	FC and MC activated	Hou and Watanuki (2012)
Line bisection	8	28	D9	20	OC, PC	OC and PC activated	Seraglia et al. (2011)
Line-tracking	2	NA	D8	4	PMC	PMC activated	Brand et al. (2011)
Maze	2	NA	D1	16	PFC	PFC more activated during BLK	Ayaz et al. (2011)
Maze	15 GP 12 BP	24 ± 1 28 ± 1	D4	24	(R)FC, (R)PC	PC activated	Kober et al. (2013)**
Missile defense	30	18–31	D9	2	PFC	PFC activated	Boyer et al. (2015)
Shopping	6 (10 BD)	61 ± 16 23 ± 1	D10	16	PFC	PFC more activated in BD	Okahashi et al. (2014)
Surgery	29	32 ± 6	D4	24	PFC	Lateral PFC more activated in experts	James et al. (2011)
Surgery	20	29 ± 2	D4	24	OC, PC	OC less activated by improved performance	Leff et al. (2015)
Surgery	7	23–26	D4	24	PC	IS activated	Miura et al. (2015)
Train driving	2	NA	D5	44	FC, OC	FC widely activated in manual condition	Kojima et al. (2005)
Walking (*haptic touch on treadmill*)	7 (1 CS)	25 ± 9	D4	44	PFC, PMC, SC, SMA	PFC and SMA widely/locally activated	Sangani et al. (2015)

*ACC, adaptive cruise control; AD, Alzheimer's disease; ATC, anterior temporal cortex; BD, brain damage patients; BLK, blocked order; BP, bad performers; Ch, number of channels; CS, chronic stroke patient; D, Device; D1, Imager 1000 (fNIR Devices, USA); D2, NIRO-200 (Hamamatsu Photonics, Japan); D3, CW6 (TechEn, USA); D4: ETG-4000 (Hitachi, Japan); D5, OMM-3000 (Shimadzu, Japan); D6, FOIRE-3000 (Shimadzu, Japan); D7, NIRO-500 (Hamamatsu Photonics, Japan); D8, Wireless prototype (Zurich University, Switzerland); D9, Imagent (ISS, USA); D10, OEG-16 (Spectratech Inc., Japan); DLPFC, dorsolateral prefrontal cortex; FC, frontal cortex; FP, frontopolar; GP, good performers; IFG, inferior frontal gyrus; IS, intraparietal sulcus; (L), left; M1, primary motor cortex; MC, motor cortex; MTG, middle temporal gyrus; NA, not available; OC, occipital cortex; PC, parietal cortex; PFC, prefrontal cortex; PMC, premotor cortex; (R), right; SC, somatosensory cortex; SMA, supplemental motor area; STG, superior temporal gyrus; TC, temporal cortex; VR, virtual reality; *fNIRS-fMRI study; **combined fNIRS-EEG study.*

to study natural human cognition have been previously reviewed (Gramann et al., 2011, 2014). Simultaneous fNIRS-EEG measurements offer complementary functional information about neuronal activity and hemodynamic changes in order to provide a wider perspective on different aspects of the cortical processes. To the best of our knowledge, Kober et al. (2013) have first compared neuronal responses of good and bad navigators during a VR spatial navigation task by a combined fNIRS-EEG system. The commercial fNIRS systems, utilized either by Kober et al. (2013) or in the present study, are equipped with fiber optic bundles. The disadvantage of using fiber optic bundles is that the fibers are often heavy and with a limited flexibility. Therefore, this kind of fNIRS instrumentation is not the best choice for studies in the neuroergonomics field. Since 2009, different battery operated multi-channel wearable/wireless fNIRS systems have been commercialized (Scholkmann et al., 2014). A 64-measurement point wireless fNIRS system was developed and integrated with simultaneous EEG and electrocardiography (ECG) monitoring in order to record data up to several days (Zhang et al., 2014). These most advanced versions of integrated fNIRS-EEG systems represent a suitable tool for evaluating brain activation in response

to cognitive tasks executed in normal daily activities. To make fNIRS technology more suitable for neuroergonomics studies, in terms of robustness, mobility, user-friendliness and customization, a very recent and successful effort has been made to realize a dedicated fNIRS device (von Lühmann et al., 2015). Further future hardware developments will increase the production of miniaturized wireless integrated fNIRS-EEG devices and their use in neuroergonomics. It is noteworthy to mention that a 16-measurement point wireless fNIRS system has been recently coupled with transcranial direct current stimulation (tDCS) in order to investigate the effects of tDCS on spatial working memory (McKendrick et al., 2015). These authors have suggested the utility of using the combination of simultaneous tDCS and fNIRS techniques for future applications in the field of neuroergonomics: from enhanced/accelerated learning and training of complex human-machine systems to optimization of task load for improved safety and productivity. The neuroergonomics approaches can also provide sensitive and reliable assessment of mental workload in complex tasks and naturalistic work settings (Parasuraman, 2011). The mental workload has been defined as the degree of the effort to be made by the brain to meet the task demands (Young et al., 2015). Recently, Peck et al. (2014) have reviewed the use of fNIRS to measure mental workload in the real world tasks, and the different approaches for automatic detection of the workload.

The strengths and the limitations of the fNIRS technique have been previously discussed in detail (for review, see Scholkmann et al., 2014). The fNIRS equipment is transportable, completely safe and non-invasive. These advantages allow for the investigation of brain activity in natural conditions (e.g., while sitting on a chair) and during daily life activities (e.g., standing and/or walking). Therefore, with respect to other functional neuroimaging methods such as fMRI, fNIRS represents an useful tool for neuroergonomics research (for review, see Ayaz et al., 2013; Derosière et al., 2013), for studies in other fields of neuroscience such as brain-computer interface (for review, see Naseer and Hong, 2015), human-robot interaction (for review, see Canning and Scheutz, 2013), and cognitive states measurements (for review, see Strait and Scheutz, 2014). In addition, very recently the integration of fNIRS with a wearable technology, such as Google Glass, has been demonstrated (Afergan et al., 2015). For an adequate understanding of the current findings, some limitations should be pointed out: (1) this study has been conducted in a small sample of healthy young male adults subjects and the subjective cognitive load was not tested by NASA Task Load Index; (2) the duration of the adopted task and the length of the route were relatively short, hence, the effect of a longer duration of the VR HCT on the PFC hemodynamic response remains unknown; (3) this study did not imply a control session for example including/not including motor task with/without VR; (4) this study did not contemplate repeated trials on separate days in order to verify the reproducibility and the potential learning effect of the HCT; (5) the limited number of measurement points (16) made possible by the utilized fNIRS system did not allow the investigation of the supposed connectivity between the PFC and other cortical areas (e.g., premotor and motor cortices) likely involved in performing HCT; and (6) this study did not take into account the impact of the variability of the skull thickness amongst the 16 measurement points within subject and amongst subjects. This anatomical variability could be examined by acquiring structural T1-weighted MRI scans from each subject.

CONCLUSION

The results of the present study confirm the promising application of fNIRS technology to objectively evaluate cortical hemodynamic changes occurring in VR environments. The ongoing development of fNIRS technology, finalized to deliver more dedicated, sophisticated and wireless devices, together with the most advanced VR solutions, could provide the best combined approach for monitoring operators training and assessing mental work. Future studies could give a contribution to a better understanding of the cognitive mechanisms underlying human performance either in expert or non-expert operators.

AUTHOR CONTRIBUTIONS

Study design and protocols were conceived by MC, AP, SL, SBM, MS, MF, GP and VQ. Data collection was performed by MC, AP, SL, SBM, MS, MF, GP and VQ. Data analysis was performed by MC, AP, SL, SBM, SB, MS, MF, GP and VQ. The manuscript was written by MC, AP, SL, SBM, SB, MS, MF, GP and VQ.

ACKNOWLEDGMENTS

The study has been performed in the framework of the "Interdepartmental Research Centre for Molecular Diagnostics and Advanced Therapies". This work was supported in part by: (1) the 2014 grant from the "Fondazione Cassa di Risparmio della Provincia dell'Aquila", and (2) the "Abruzzo Earthquake Relief Fund" (Toronto, ON; purchase of the Artinis system). The authors wish to thank Dr. Simone Cutini, "Department of Developmental Psychology and Socialization", University of Padua, Italy for helping in the identification of the relative Brodmann's Areas, and Dr. Danilo Avola, "Department of Mathematics and Informatics", University of Udine, Italy for HCT implementation.

REFERENCES

Afergan, D., Hincks, S. W., Shibata, T., and Jacob, R. J. K. (2015). Phylter: a system for modulating notifications in wearables using physiological sensing. *Lect. Notes Comput. Sc.* 9183, 167–177. doi: 10.1007/978-3-319-20816-9_17

Ayaz, H., Cakir, M. P., Izzetoglu, K., Curtin, A., Shewokis, P. A., Bunce, S., et al. (2012a). "Monitoring expertise development during simulated UAV piloting tasks using optical brain imaging," in *Proceedings of the IEEE Aerospace Conference* (Big Sky, MN), 1–11.

Ayaz, H., Shewokis, P. A., Bunce, S., Izzetoglu, K., Willems, B., and Onaral, B. (2012b). Optical brain monitoring for operator training and mental workload assessment. *Neuroimage* 59, 36–47. doi: 10.1016/j.neuroimage.2011.06.023

Ayaz, H., Onaral, B., Izzetoglu, K., Shewokis, P. A., McKendrick, R., and Parasuraman, R. (2013). Continuous monitoring of brain dynamics with functional near infrared spectroscopy as a tool for neuroergonomic research: empirical examples and a technological development. *Front. Hum. Neurosci.* 7:871. doi: 10.3389/fnhum.2013.00871

Ayaz, H., Shewokis, P. A., Curtin, A., Izzetoglu, M., Izzetoglu, K., and Onaral, B. (2011). Using MazeSuite and functional near infrared spectroscopy to study learning in spatial navigation. *J. Vis. Exp.* 56:e3443. doi: 10.3791/3443

Bachmann, D., Weichert, F., and Rinkenauer, G. (2014). Evaluation of the leap motion controller as a new contact-free pointing device. *Sensors (Basel)* 15, 214–233. doi: 10.3390/s150100214

Baddeley, A. (2003). Working memory: looking back and looking forward. *Nat. Rev. Neurosci.* 4, 829–839. doi: 10.1038/nrn1201

Badre, D., and Wagner, A. D. (2007). Left ventrolateral prefrontal cortex and the cognitive control of memory. *Neuropsychologia* 45, 2883–2901. doi: 10.1016/j.neuropsychologia.2007.06.015

Basso Moro, S., Bisconti, S., Muthalib, M., Spezialetti, M., Cutini, S., Ferrari, M., et al. (2014). A semi-immersive virtual reality incremental swing balance task activates prefrontal cortex: a functional near-infrared spectroscopy study. *Neuroimage* 85, 451–460. doi: 10.1016/j.neuroimage.2013.05.031

Basso Moro, S., Cutini, S., Ursini, M. L., Ferrari, M., and Quaresima, V. (2013). Prefrontal cortex activation during story encoding/retrieval: a multi-channel functional near-infrared spectroscopy study. *Front. Hum. Neurosci.* 7:925. doi: 10.3389/fnhum.2013.00925

Boyer, M., Cummings, M. L., Spence, L. B., and Solovey, E. T. (2015). Investigating mental workload changes in a long duration supervisory control task. *Interact. Comput.* 27, 512–520. doi: 10.1093/iwc/iwv012

Brand, J., Geisseler, O., Holper, L., Hepp-Reimond, M. C., Morari, M., Kiper, D., et al. (2011). "The effects of manipulation of visual feedback in virtual reality on cortical activity: a pilot study," in *Proceedings of the IEEE International Conference on Virtual Rehab.*, Zurich, 1–2.

Canning, C., and Scheutz, M. (2013). Functional near-infrared spectroscopy in human-robot interaction. *J. Hum. Robot Interact.* 2, 62–84. doi: 10.5898/jhri.2.3.canning

Chang, P. H., Lee, S. H., Gu, G. M., Lee, S. H., Jin, S. H., Yeo, S. S., et al. (2014). The cortical activation pattern by a rehabilitation robotic hand: a functional NIRS study. *Front. Hum. Neurosci.* 8:49. doi: 10.3389/fnhum.2014.00049

Chen, S., Ma, H., Yang, C., and Fu, M. (2015). "Hand gesture based robot control system using leap motion," in *Intelligent Robotics and Applications*, eds H. Liu, N. Kubota, X. Zhu, R. Dillmann, and D. Zhou (Cham: Springer International Publishing), 581–591.

Corbetta, M., Patel, G., and Shulman, G. L. (2008). The reorienting system of the human brain: from environment to theory of mind. *Neuron* 58, 306–324. doi: 10.1016/j.neuron.2008.04.017

Derosière, G., Alexandre, F., Bourdillon, N., Mandrick, K., Ward, T. E., and Perrey, S. (2014). Similar scaling of contralateral and ipsilateral cortical responses during graded unimanual force generation. *Neuroimage* 85, 471–477. doi: 10.1016/j.neuroimage.2013.02.006

Derosière, G., Mandrick, K., Dray, G., Ward, T. E., and Perrey, S. (2013). NIRS-measured prefrontal cortex activity in neuroergonomics: strengths and weaknesses. *Front. Hum. Neurosci.* 7:583. doi: 10.3389/fnhum.2013.00583

Desai, R. P., Desai, N. P., Ajmera, D. K., and Mehta, K. (2014). A review paper on oculus rift-a virtual reality headset. *Int. J. Eng. Trends Tech.* 13, 175–179. doi: 10.14445/22315381/IJETT-V13P237

Dodgson, N. (2013). Optical devices: 3D without the glasses. *Nature* 495, 316–317. doi: 10.1038/495316a

Durantin, G., Gagnon, J. F., Tremblay, S., and Dehais, F. (2014). Using near infrared spectroscopy and heart rate variability to detect mental overload. *Behav. Brain Res.* 259, 16–23. doi: 10.1016/j.bbr.2013.10.042

Erden, F., and Çetin, A. (2014). Hand gesture based remote control system using infrared sensors and a camera. *IEEE Trans. Consum. Electron.* 60, 675–680. doi: 10.1109/tce.2014.7027342

Ferrari, M., Bisconti, S., Spezialetti, M., Basso Moro, S., Di Palo, C., Placidi, G., et al. (2014). Prefrontal cortex activated bilaterally by a tilt board balance task: a functional near-infrared spectroscopy study in a semi-immersive virtual reality environment. *Brain Topogr.* 27, 353–365. doi: 10.1007/s10548-013-0320-z

Gagnon, L., Perdue, K., Greve, D. N., Goldenholz, D., Kaskhedikar, G., and Boas, D. A. (2011). Improved recovery of the hemodynamic response in diffuse optical imaging using short optode separations and state-space modeling. *Neuroimage* 56, 1362–1371. doi: 10.1016/j.neuroimage.2011.03.001

Gateau, T., Durantin, G., Lancelot, F., Scannella, S., and Dehais, F. (2015). Real-time state estimation in a flight simulator using fNIRS. *PLoS One* 10:e0121279. doi: 10.1371/journal.pone.0121279

Gentili, R. J., Shewokis, P. A., Ayaz, H., and Contreras-Vidal, J. L. (2013). Functional near-infrared spectroscopy-based correlates of prefrontal cortical dynamics during a cognitive-motor executive adaptation task. *Front. Hum. Neurosci.* 7:277. doi: 10.3389/fnhum.2013.00277

Gramann, K., Ferris, D. P., Gwin, J., and Makeig, S. (2014). Imaging natural cognition in action. *Int. J. Psychophysiol.* 91, 22–29. doi: 10.1016/j.ijpsycho.2013.09.003

Gramann, K., Gwin, J. T., Ferris, D. P., Oie, K., Jung, T. P., Lin, C. T., et al. (2011). Cognition in action: imaging brain/body dynamics in mobile humans. *Rev. Neurosci.* 22, 593–608. doi: 10.1515/RNS.2011.047

Halsband, U., and Lange, R. K. (2006). Motor learning in man: a review of functional and clinical studies. *J. Physiol. Paris* 99, 414–424. doi: 10.1016/j.jphysparis.2006.03.007

Harrison, J., Izzetoglu, K., Ayaz, H., Willems, B., Hah, S., Ahlstrom, U., et al. (2014). Cognitive workload and learning assessment during the implementation of a next-generation air traffic control technology using functional near-infrared spectroscopy. *Proc. IEEE Trans. Hum. Mach. Sys.* 44, 429–440. doi: 10.1109/thms.2014.2319822

Hatakenaka, M., Miyai, I., Mihara, M., Sakoda, S., and Kubota, K. (2007). Frontal regions involved in learning of motor skill- A functional NIRS study. *Neuroimage* 34, 109–116. doi: 10.1016/j.neuroimage.2006.08.014

Holper, L., Kobashi, N., Kiper, D., Scholkmann, F., Wolf, M., and Eng, K. (2012). Trial-to-trial variability differentiates motor imagery during observation between low versus high responders: a functional near-infrared spectroscopy study. *Behav. Brain Res.* 229, 29–40. doi: 10.1016/j.bbr.2011.12.038

Holper, L., Muehlemann, T., Scholkmann, F., Eng, K., Kiper, D., and Wolf, M. (2010). Testing the potential of a virtual reality neurorehabilitation system during performance of observation, imagery and imitation of motor actions recorded by wireless functional near-infrared spectroscopy (fNIRS). *J. Neuroeng. Rehabil.* 7:57. doi: 10.1186/1743-0003-7-57

Holtzer, R., Mahoney, J. R., Izzetoglu, M., Izzetoglu, K., Onaral, B., and Verghese, J. (2011). fNIRS study of walking and walking while talking in young and old individuals. *J. Gerontol. A Biol. Sci. Med. Sci.* 66, 879–887. doi: 10.1093/gerona/glr068

Hoshi, E. (2013). Cortico-basal ganglia networks subserving goal-directed behavior mediated by conditional visuo-goal association. *Front. Neural Circuits* 7:158. doi: 10.3389/fncir.2013.00158

Hou, L., and Watanuki, K. (2012). Measurement of brain activity under virtual reality skills training using near-infrared spectroscopy. *J. Adv. Mech. Des. Syst.* 6, 168–178. doi: 10.1299/jamdsm.6.168

Huppert, T., Schmidt, B., Beluk, N., Furman, J., and Sparto, P. (2013). Measurement of brain activation during an upright stepping reaction task using functional near-infrared spectroscopy. *Hum. Brain Mapp.* 34, 2817–2828. doi: 10.1002/hbm.22106

James, D. R. C., Orihuela-Espina, F., Leff, D. R., Sodergren, M. H., Athanasiou, T., Darzi, A. W., et al. (2011). The ergonomics of natural orifice transluminal endoscopic surgery (NOTES) navigation in terms of performance, stress and cognitive behavior. *Surgery* 149, 525–533. doi: 10.1016/j.surg.2010.11.019

Karim, H., Schmidt, B., Dart, D., Beluk, N., and Huppert, T. (2012). Functional near-infrared spectroscopy (fNIRS) of brain function during active balancing using a video game system. *Gait Posture* 35, 367–372. doi: 10.1016/j.gaitpost.2011.10.007

Kearney, J. K., Rizzo, M., and Severson, J. (2007). "Virtual reality and neuroergonomics," in *Neuroergonomics: the Brain at Work*, eds R. Parasuraman and M. Rizzo (New York, NY: Oxford University Press), 253–274.

Kim, S., Biocca, F., Lee, K. M., Lee, S. W., and Chung, J. Y. (2015). "Neural correlates of embodied experiences through avatar self-identification," in *Proceeding of the 21st Annual Meeting of the Organization for Human Brain Mapping*, Hawaii, HI.

Kita, Y., Gunji, A., Sakihara, K., Inagaki, M., Kaga, M., Nakagawa, E., et al. (2010). Scanning strategies do not modulate face identification: eye-tracking

and near-infrared spectroscopy study. *PLoS One* 5:e11050. doi: 10.1371/journal.pone.0011050

Kober, S. E., Wood, G., and Neuper, C. (2013). "Measuring brain activation during spatial navigation in virtual reality: a combined EEG-NIRS study," in *Virtual Environments: Developments, Applications and Challenges*, eds S. Trautman and F. Julien (New York, NY: NOVA Publisher), 1–24.

Koenraadt, K. L., Roelofsen, E. G., Duysens, J., Keijsers, N. L., Kojima, T., and Tsunashima, H. (2014). Cortical control of normal gait and precision stepping: an fNIRS study. *Neuroimage* 85, 415–422. doi: 10.1016/j.neuroimage.2013.04.070

Kojima, T., Tsunashima, H., Shizawa, T., Takada, H., and Sakai, T. (2005). Measurement of train driver's brain activity by functional near-infrared spectroscopy (fNIRS). *Opt. Quant. Electron* 37, 1319–1338. doi: 10.1007/s11082-005-4202-9

Leff, D. R., James, D. R. C., Orihuela-Espina, F., Kwok, K., Sun, L. W., Mylonas, G., et al. (2015). The impact of expert visual guidance on trainee visual search strategy, visual attention and motor skills. *Front. Hum. Neurosci.* 9:526. doi: 10.3389/fnhum.2015.00526

Leff, D. R., Orihuela-Espina, F., Elwell, C. E., Athanasiou, T., Delpy, D. T., Darzi, A. W., et al. (2011). Assessment of the cerebral cortex during motor task behaviours in adults: a systematic review of functional near infrared spectroscopy (fNIRS) studies. *Neuroimage* 54, 2922–2936. doi: 10.1016/j.neuroimage.2010.10.058

Liu, T., Pelowski, M., Pang, C., Zhou, Y., and Cai, J. (2015). Near-infrared spectroscopy as a tool for driving research. *Ergonomics* doi: 10.1080/00140139.2015.1076057 [Epub ahead of print].

Liu, Y., and Zhang, Y. (2015). "Toward welding robot with human knowledge: a remotely-controlled approach," in Vol. 12, *Proceedings of the IEEE Transaction on Automation Science and Engineering* (Gothenburg: IEEE), 769–774.

Mandrick, K., Derosière, G., Dray, G., Coulon, D., Micallef, J. P., and Perrey, S. (2013). Prefrontal cortex activity during motor tasks with additional mental load requiring attentional demand: a near-infrared spectroscopy study. *Neurosci. Res.* 76, 156–162. doi: 10.1016/j.neures.2013.04.006

Matud, M. P. (2004). Gender differences in stress and coping styles. *Pers. Individ. Dif.* 37, 1401–1415. doi: 10.1016/j.paid.2004.01.010

McKendrick, R., Parasuraman, R., and Ayaz, H. (2015). Wearable functional near infrared spectroscopy (fNIRS) and transcranial direct current stimulation (tDCS): expanding vistas for neurocognitive augmentation. *Front. Syst. Neurosci.* 9:27. doi: 10.3389/fnsys.2015.00027

Mehta, R. K., and Parasuraman, R. (2013). Neuroergonomics: a review of applications to physical and cognitive work. *Front. Hum. Neurosci.* 7:889. doi: 10.3389/fnhum.2013.00889

Mirelman, A., Maidan, I., Bernad-Elazari, H., Nieuwhof, F., Reelick, M., Giladi, N., et al. (2014). Increased frontal brain activation during walking while dual tasking: an fNIRS study in healthy young adults. *J. Neuroeng. Rehabil.* 11:85. doi: 10.1186/1743-0003-11-85

Miura, S., Kobayashi, Y., Kawamura, K., Nakashima, Y., and Fujie, M. G. (2015). Brain activation in parietal area during manipulation with a surgical robot simulator. *Int. J. Comput. Assist. Radiol. Surg.* 10, 783–790. doi: 10.1007/s11548-015-1178-1

Molavi, B., and Dumont, G. A. (2012). Wavelet-based motion artifact removal for functional near-infrared spectroscopy. *Physiol. Meas.* 33, 259–270. doi: 10.1088/0967-3334/33/2/259

Nan, X., Zhang, Z., Zhang, N., Guo, F., He, Y., and Guan, L. (2014). vDesign: a CAVE-based virtual design environment using hand interactions. *J. Multimodal User Interfaces* 8, 367–379. doi: 10.1007/s12193-014-0168-x

Naseer, N., and Hong, K. S. (2015). fNIRS-based brain-computer interfaces: a review. *Front. Hum. Neurosci.* 9:3. doi: 10.3389/fnhum.2015.00003

Noah, J. A., Ono, Y., Nomoto, Y., Shimada, S., Tachibana, A., Zhang, X., et al. (2015). fMRI validation of fNIRS measurements during a naturalistic task. *J. Vis. Exp.* 100:e52116. doi: 10.3791/52116

Okahashi, S., Mizumoto, H., Komae, A., Ueno, K., Yokoyama, M., Nagano, A., et al. (2014). An fNIRS-based study on prefrontal cortex activity during a virtual shopping test with different task difficulties in brain-damaged patients. *J. Behav. Brain Sci.* 4, 247–255. doi: 10.4236/jbbs.2014.46026

Ono, Y., Noah, J. A., Zhang, X., Nomoto, Y., Suzuki, T., Shimada, S., et al. (2015). Motor learning and modulation of prefrontal cortex: an fNIRS assessment. *J. Neural Eng.* 12:066004. doi: 10.1088/1741-2560/12/6/066004 [Epub ahead of print].

Parasuraman, R. (2011). Neuroergonomics: brain, cognition and performance at work. *Curr. Dir. Psychol. Sci.* 20, 181–186. doi: 10.1177/0963721411409176

Parasuraman, R., and Rizzo, M. (2007). *Neuroergonomics: The Brain at Work*. New York, NY: Oxford University Press.

Peck, E. M., Afergan, D., Yuksel, B. F., Lalooses, F., and Jacob, R. J. K. (2014). "Using fNIRS to measure mental workload in the real world," in *Advances in Physiological Computing*, eds S. H. Fairclough and K. Gilleade (London: Springer-Verlag), 117–139.

Sangani, S., Lamontagne, A., and Fung, J. (2015). Cortical mechanisms underlying sensorimotor enhancement promoted by walking with haptic inputs in a virtual environment. *Progr. Brain Res.* 218, 313–330. doi: 10.1016/bs.pbr.2014.12.003

Scholkmann, F., Kleiser, S., Metz, A., Zimmermann, R., Mata Pavia, J., Wolf, U., et al. (2014). A review on continuous wave functional near-infrared spectroscopy and imaging instrumentation and methodology. *Neuroimage* 85, 6–27. doi: 10.1016/j.neuroimage.2013.05.004

Seraglia, B., Gamberini, L., Priftis, K., Scatturin, P., Martinelli, M., and Cutini, S. (2011). An exploratory fNIRS study with immersive virtual reality: a new method for technical implementation. *Front. Hum. Neurosci.* 5:176. doi: 10.3389/fnhum.2011.00176

Souza, M. J., Donohue, S. E., and Bunge, S. A. (2009). Controlled retrieval and selection of action-relevant knowledge mediated by partially overlapping regions in left ventrolateral prefrontal cortex. *Neuroimage* 46, 299–307. doi: 10.1016/j.neuroimage.2009.01.046

Strait, M., and Scheutz, M. (2014). What we can and cannot (yet) do with functional near infrared spectroscopy. *Front. Neurosci.* 8:117. doi: 10.3389/fnins.2014.00117

Sukal-Moulton, T., de Campos, A. C., Stanley, C. J., and Damiano, D. L. (2014). Functional near infrared spectroscopy of the sensory and motor brain regions with simultaneous kinematic and EMG monitoring during motor tasks. *J. Vis. Exp.* 94:e52391. doi: 10.3791/52391

Takakusaki, K. (2008). Forebrain control of locomotor behaviors. *Brain Res. Rev.* 57, 192–198. doi: 10.1016/j.brainresrev.2007.06.024

Takeuchi, Y. (2000). Change in blood volume in the brain during a simulated aircraft landing task. *J. Occup. Health* 42, 60–65. doi: 10.1539/joh.42.60

Tanji, J., and Hoshi, E. (2008). Role of the lateral prefrontal cortex in executive behavioral control. *Physiol. Rev.* 88, 37–57. doi: 10.1152/physrev.00014.2007

Tomioka, H., Yamagata, B., Takahashi, T., Yano, M., Isomura, A. J., Kobayashi, H., et al. (2009). Detection of hypofrontality in drivers with Alzheimer's disease by near-infrared spectroscopy. *Neurosci. Lett.* 451, 252–256. doi: 10.1016/j.neulet.2008.12.059

Tsunashima, H., and Yanagisawa, K. (2009). Measurement of brain function of car driver using functional near-infrared spectroscopy (fNIRS). *Comput. Intell. Neurosci.* 2009:164958. doi: 10.1155/2009/164958

Urakawa, S., Takamoto, K., Ishikawa, A., Ono, T., and Nishijo, H. (2015). Selective medial prefrontal cortex responses during live mutual gaze interactions in human infants: an fNIRS study. *Brain Topogr.* 28, 691–701. doi: 10.1007/s10548-014-0414-2

von Lühmann, A., Herff, C., Heger, D., and Schultz, T. (2015). Towards a wireless open source instrument: functional near-infrared spectroscopy in mobile neuroergonomics and BCI applications. *Front. Hum. Neurosci.* 9:617. doi: 10.3389/fnhum.2015.00617

Wang, L., Krasich, K., Bel-Bahar, T., Hughes, L., Mitroff, S. R., and Appelbaum, L. G. (2015). Mapping the structure of perceptual and visual-motor abilities in healthy young adults. *Acta Psychol. (Amst)* 157, 74–84. doi: 10.1016/j.actpsy.2015.02.005

Wei, J. L., Sen, W. L., and Sani, Z. (2015). Leap motion underwater robotic arm control. *J. Teknologi* 74, 153–159. doi: 10.11113/jt.v74.4819

Wriessnegger, S. C., Bauernfeind, G., Schweitzer, K., Kober, S., Neuper, C., and Müller-Putz, G. R. (2012). The interplay of prefrontal and sensorimotor cortices during inhibitory control of learned motor behavior. *Front. Neuroeng.* 5:17. doi: 10.3389/fneng.2012.00017

Yamagata, T., Nakayama, Y., Tanji, J., and Hoshi, E. (2012). Distinct information representation and processing for goal-directed behavior in the dorsolateral and ventrolateral prefrontal cortex and the dorsal premotor cortex. *J. Neurosci.* 32, 12934–12949. doi: 10.1523/JNEUROSCI.2398-12.2012

Young, M. S., Brookhuis, K. A., Wickens, C. D., and Hancock, P. A. (2015). State of science: mental workload in ergonomics. *Ergonomics* 58, 1–17. doi: 10.1080/00140139.2014.956151

Zhang, Q., Ivkovic, V., Hu, G., and Strangman, G. E. (2014). Twenty-four-hour ambulatory recording of cerebral hemodynamics, systemic hemodynamics, electrocardiography and actigraphy during people's daily activities. *J. Biomed. Opt.* 19:047003. doi: 10.1117/1.JBO.19.4.047003

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Carrieri, Petracca, Lancia, Basso Moro, Brigadoi, Spezialetti, Ferrari, Placidi and Quaresima. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution and reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 18 May 2016
doi: 10.3389/fnhum.2016.00216

Into the Wild: Neuroergonomic Differentiation of Hand-Held and Augmented Reality Wearable Displays during Outdoor Navigation with Functional Near Infrared Spectroscopy

Ryan McKendrick[1], Raja Parasuraman[1], Rabia Murtza[1], Alice Formwalt[1], Wendy Baccus[1], Martin Paczynski[1] and Hasan Ayaz[2,3,4]**

[1] Psychology Department, Human Factors and Applied Cognition, George Mason University, Fairfax, VA, USA, [2] School of Biomedical Engineering, Science and Health Systems, Drexel University, Philadelphia, PA, USA, [3] Department of Family and Community Health, University of Pennsylvania, Philadelphia, PA, USA, [4] Division of General Pediatrics, Children's Hospital of Philadelphia, Philadelphia, PA, USA

OPEN ACCESS

Edited by:
Richard A. P. Roche, National University of Ireland, Maynooth, Ireland

Reviewed by:
Sean Commins, National University of Ireland, Maynooth, Ireland
Antonia Hamilton, University of Nottingham, UK

***Correspondence:**
Ryan McKendrick
rmckend2@gmu.edu;
Hasan Ayaz
hasan.ayaz@drexel.edu

Received: *13 December 2015*
Accepted: *26 April 2016*
Published: *18 May 2016*

Citation:
McKendrick R, Parasuraman R, Murtza R, Formwalt A, Baccus W, Paczynski M and Ayaz H (2016) Into the Wild: Neuroergonomic Differentiation of Hand-Held and Augmented Reality Wearable Displays during Outdoor Navigation with Functional Near Infrared Spectroscopy. Front. Hum. Neurosci. 10:216. doi: 10.3389/fnhum.2016.00216

Highly mobile computing devices promise to improve quality of life, productivity, and performance. Increased situation awareness and reduced mental workload are two potential means by which this can be accomplished. However, it is difficult to measure these concepts in the "wild". We employed ultra-portable battery operated and wireless functional near infrared spectroscopy (fNIRS) to non-invasively measure hemodynamic changes in the brain's Prefrontal cortex (PFC). Measurements were taken during navigation of a college campus with either a hand-held display, or an Augmented reality wearable display (ARWD). Hemodynamic measures were also paired with secondary tasks of visual perception and auditory working memory to provide behavioral assessment of situation awareness and mental workload. Navigating with an augmented reality wearable display produced the least workload during the auditory working memory task, and a trend for improved situation awareness in our measures of prefrontal hemodynamics. The hemodynamics associated with errors were also different between the two devices. Errors with an augmented reality wearable display were associated with increased prefrontal activity and the opposite was observed for the hand-held display. This suggests that the cognitive mechanisms underlying errors between the two devices differ. These findings show fNIRS is a valuable tool for assessing new technology in ecologically valid settings and that ARWDs offer benefits with regards to mental workload while navigating, and potentially superior situation awareness with improved display design.

Keywords: fNIRS, situation awareness, mental workload, spatial navigation, working memory, head-mounted display, neuroergonomics

INTRODUCTION

The availability and use of highly mobile computing devices is increasing. Examples include fitness trackers, smartwatches, and smartphones; however, there are also devices such as Google Glass, Occulus Rift and Microsoft Hololens which promise not just mobile computing but the coexistence of real world objects with supplementary computer generated objects (i.e., augmented reality; Azuma et al., 2001). Augmented reality wearable displays (ARWD) are already being put into service by the National Aeronautics and Space Administration (NASA). It is believed that these devices will help astronauts on the international space station improve their training and performance in highly demanding situations (Schierholz et al., 2015). While it is clear that having a hands-free display can improve physical ergonomics, especially when both hands are required for adequate task execution, ARWDs could also enhance cognitive ergonomics through augmentation of mental workload and situation awareness.

Ideal task performance is dependent on optimizing mental workload. Mental workload refers to the limited information processing capacity of the brain that is demanded by a task (Parasuraman et al., 2008). When demands exceed the brains maximum information processing capacity, further increases in mental workload lead to ever increasing decrements in performance (Hancock and Parasuraman, 1992). This can be realized as incorrect responses, missed responses or even the "shedding" of secondary tasks (Wickens et al., 2013). ARWDs have the potential to reduce mental workload by reducing the distance and time between visual fixations. Reducing fixation time and distance could reduce the amount of information needed to be held in working memory. For example, during simulated emergency braking, drivers using Google Glass to send text messages experienced less mental workload relative to drivers using a smartphone (Sawyer et al., 2014). ARWDs have also been used to improve operator comfort and procedure efficiency during cardiac surgery (Opolski et al., 2015). An ARWD allowed cardiologists to view reconstructed tomographic images while performing catheterization, improving landmark visualization and verification of surgical tools.

Situation awareness, the perception of critical information (stage 1), comprehension of its meaning (stage 2), and the projection of this information into the future (stage 3; Endsley, 1995a) is also critical for complex task performance (Wickens et al., 2013). High situation awareness, while not guaranteeing successful performance, increases the probability of successful performance. Like mental workload, situation awareness is dependent on working memory and highly dependent on attention (Endsley, 1995a). In this regard ARWDs have the potential to both enhance and degrade situation awareness. ARWDs may enhance situation awareness by freeing up working memory capacity. Conversely, ARWDS may reduce situation awareness from degradation of divided attention. Divided attention relates to the optimal allocation of attention to different inputs by splitting or rapidly shifting the focus of attention (Parasuraman, 1998). The compellingness of ARWD symbology is more likely to exogenously capture the focus of attention and hold it (Thomas and Wickens, 2001, 2004). This results in increased focused attention to display elements, and reduced or eliminated attention to task relevant information outside of the ARWD display. This phenomena of increased focused attention to a display coinciding with decreased divided attention to an external scene is referred to as cognitive tunneling (Fischer et al., 1980). Cognitive tunneling is often implicated in aviation studies where a failure to perceive and act on an unexpected event reduces performance (Crawford and Neal, 2006).

Measurement of situation awareness and mental workload in ARWDs is problematic. Traditionally situation awareness and workload are assessed with questionnaires administered during artificial pauses (Situation Awareness Global Assessment Technique (SAGAT); Endsley, 1995b), in task probes (Situation Present Assessment Measure (SPAM); Durso and Dattel, 2004), or upon task completion NASA Task Load Index (TLX; Hart and Staveland, 1988). Within dynamic environments such assessments can be intrusive, thereby reducing ecological validity, or underrepresenting time critical signals, such as abrupt changes in workload. Workload can also be objectively assessed via dual-task secondary task decrements. In the dual task paradigm, interference on a cognitive process is anticipated between the primary task and the secondary task. This results in a decrement in performance on the secondary task, due primarily to the mental resource demands of the secondary task exceeding the mental resources that can be allocated. This secondary task performance decrement can be used as an index of the cognitive workload required of the primary task (Gopher, 1993; Wickens, 2008; Wickens et al., 2013). However, dual-task decrements have been criticized with regard to circularity; as performance varies with resource allocation, but resources are only inferred from performance (Navon, 1984).

An objective, non-invasive, motion artifact robust and portable method is needed to measure situation awareness and mental workload in ARWDs. Functional near infrared spectroscopy (fNIRS) provides an attractive method for continuous monitoring of brain dynamics in both seated and mobile participants (Ayaz et al., 2013). fNIRS is safe, highly portable, user-friendly and relatively inexpensive, with rapid application times and near-zero run-time costs (Villringer and Chance, 1997; Ayaz et al., 2012a; Ferrari and Quaresima, 2012). fNIRS uses specific wavelengths of light to provide measures of cerebral oxygenated and deoxygenated hemoglobin that are correlated with the blood-oxygen-level dependent (BOLD) contrast used in functional magnetic resonance imaging (fMRI; Cui et al., 2011; Sato et al., 2013). Importantly fNIRS measurements are objective and non-invasive to the mental task being measured. fNIRS for mobile neural measurement is also relatively robust to motion artifacts and allows wearable sensors to be physically untethered to the acquisition module (Ayaz et al., 2013; McKendrick et al., 2015). Mobile fNIRS allows for a freedom of movement not previously possible in neuroimaging, providing the opportunity to monitor mental workload and situation awareness in dynamic mobile tasks.

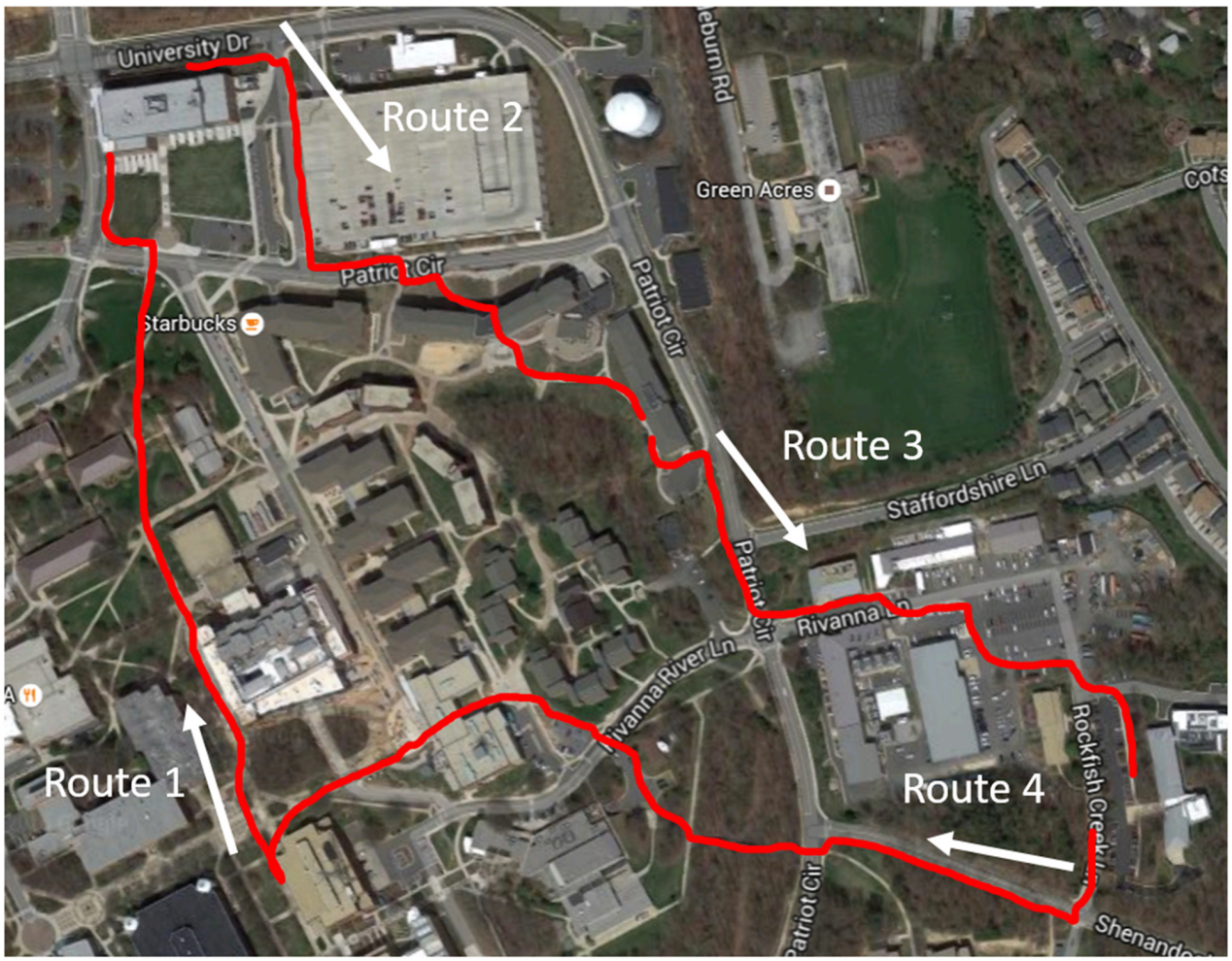


FIGURE 1 | Map depicting the four routes followed by participants. Exact routes depicted in red, white arrows indicate walking direction. Image © 2015 DigitalGlobe.

Hemodynamic indexes of mental workload as used by fNIRS and fMRI assume that activity related metabolic changes in specific functional brain regions are useful indexes of mental workload. Prefrontal cortex (PFC) is commonly monitored due to its functional relationship with working memory (Braver et al., 1997; Cohen et al., 1997), decision making (Ramnani and Owen, 2004; Figner et al., 2010), and executive control (Badre et al., 2005; Badre and Wagner, 2007). A growing body of research has found fNIRS hemodynamic measurements of PFC to be a useful index of mental workload in a number of complex cognitive and real world tasks (Ayaz et al., 2011, 2012b; Abibullaev and An, 2012; Naseer and Keum-Shik, 2013; Bogler et al., 2014; Derosière et al., 2014; Herff et al., 2014; Schudlo and Chau, 2014; Pinti et al., 2015; Solovey et al., 2015). Divided attention has also been associated with activity in PFC (Corbetta et al., 1991; Herath et al., 2001; Loose et al., 2003; Fagioli and Macaluso, 2009; Mizuno et al., 2012). Divided attention is a key component of dual tasking (Pashler, 1994), and superior dual-tasking has been associated with decreased activity/more efficient processing in PFC (Rypma et al., 2002; Grabner et al., 2006; McKendrick et al., 2014). Reduced demands on working memory capacity and superior dual-tasking are factors that influence greater situation awareness (Endsley, 1995a). Therefore, reduced PFC activity may be implicative of greater situation awareness during ARWD use.

TABLE 1 | List of situation awareness queries.

Query: "Did You See"	Correct response
A drinking fountain	No
A large metallic sculpture	Yes
A children's playground	Yes
A red "Do Not Enter" sign	No
An american flag	Yes
A black bike rack	No
A tree wrapped in multi-colored yarn	Yes
Two satellite dishes	Yes
Two figures dancing	No

The present study implemented a neuroergonomics approach (Parasuraman, 2003) to examine the cognitive differences between an ARWD (Google Glass) and a handheld display (Smartphone). We used mobile fNIRS to monitor lateral PFC and complimented it with two separate secondary tasks assessing differences in mental workload and situation awareness during navigation. Superior performance on the secondary tasks is anticipated to reflect reduced mental workload and greater situation awareness respectively. Reduced PFC activity is anticipated to index reduced mental workload and improved situation awareness in the absence of secondary task errors. Specifically, the ARWD was expected to show reduced mental workload and superior situation awareness across both behavioral and hemodynamic indices.

MATERIALS AND METHODS

Participants

Twenty participants (12 female adults) volunteered for the study. All participants were right handed and aged 18–29 years. Each participant was randomly assigned to one of two experimental groups. The two experimental groups each contained 10 participants. If complications were experienced viewing the Google Glass display, these individuals were moved into the other experimental condition (two such complications occurred). All participants reported normal or corrected to normal vision. All participants also reported average or greater cardiovascular health, and had no history of cardiovascular abnormalities. Each participant gave informed consent via a form approved by the George Mason University Institutional Review Board prior to study participation.

Primary Task

Route Following

Participants were given a visual map of a route to walk along. The visual map was generated via Google Maps and presented to the subject via an Apple Iphone 4 s (which the participant held in hand) or Google Glass (affixed to the participant's head). In both devices Google Maps presents a birds-eye-view of the route with a digital arrow indicating the direction to be followed, as well as written turn-by-turn instructions. Google maps also provides auditory turn-by-turn instructions but these were muted in both devices. Four different routes were used (route one = 1500 ft, route two = 1400 ft, route three = 1600 ft, route four = 2000 ft; **Figure 1**) and each participant walked all four routes, total experiment time was between 45 and 60 min. The route following took place on a North American college campus. Portions of the routes were familiar to the participants, however the majority were unfamiliar and selected specifically because these regions are not frequented by university undergraduates. The routes also contained portions that simulated urban and rural environments. Each route was entered into either device by the experimenter. Participants in the hand held device (Smartphone; Apple Iphone 4 s) group were asked to hold the device in their right hand and lift the device near their field of view when confirmation of the correct route was needed (to avoid excessive motion artifacts in the fNIRS signal from tilting the head down). Participants in the ARWD (Google Glass) group were instructed to keep their right index finger on the Google Glass touchpad. This was done to ensure that Google Glass did not enter "sleep mode" during route following and to control for physical load in the right arm across devices. Once the route navigation began participants had no interaction with the devices other than viewing the generated route. Participants were instructed to walk at the pace they felt most comfortable with. This was done to minimize variability in the physical load of the walking task via self-adaptation. If errors were made during route following, participants were tapped on the shoulder and instructed as to the correct direction of the route. Only two such errors occurred throughout the experiment, one in each display group in the same navigation route, the error was related to a poor GPS signal.

Secondary Tasks

Auditory 1-Back

While following the route, participants simultaneously completed 37 blocks of an auditory 1-back lasting 60 s each. The auditory stimuli consisted of tone triplets randomly composed from fundamental frequencies of 493.88, 554.36, 698.45 and 880 Hz presented via Bluetooth in-ear headphones. The tones were created from bandpass filtered white noise and a tone overlay. The triplets were presented randomly in one of three spatial locations; left, right and central (balanced sound distribution). Five triplets were presented for each block. Participants were asked to compare the triplet they had just heard to the triplet they had previously heard. If the two triplets were of the same frequencies presented in the same sequence, then the trial was considered a match. At the end of a block participants were prompted by the experimenter to verbally indicate how many matches they heard. The experimenter recorded the response within the program administering the auditory task and participants were immediately given feedback regarding the accuracy of their response. An fNIRS measurement block began with each 1-back block and ended just prior to the participant being prompted to respond.

Scenery Probe

While route following, participants were also asked 10 questions about their surroundings to assess and help maintain an accurate awareness of the environment. After a prompt from the experimenter to be "situationally aware", participants maintained this search disposition for approximately 30 s after which the experimenter asked them to stop moving and face forward. During this time the experimenter queried whether the participant had seen a particular object in the environment. The participant was previously informed to respond verbally with a response of either "yes" or "no". Queried objects could either have been present in the environment or not present, and there were six instances where the queried object was

present and four where it was not. When the object was present in the environment the participant was stopped and queried 5 s after the object was no longer visible. Participants were given immediate feedback regarding the accuracy of their responses. The query list is presented in **Table 1**. An fNIRS measurement trial began when participants were prompted to be situationally aware and ended when the participant was asked to stop walking just before the scenery probe query.

Procedures

fNIRS Setup

Participants were seated and asked to remove any makeup from their forehead with an alcohol swab and or adjust their hair prior to affixing the wireless and battery operated fNIRS neuroimaging device, Model 1100W (fNIR Devices, LLC[1]). The hardware unit was connected to headband sensor pads via cable and transmitted the data wirelessly to a remote tablet computer. Both the pocket sized control hardware (that contains the battery and antenna) and sensor pads were affixed to the subject making the participant completely mobile during recording. Two separate sensor headband pads were placed approximately 3 cm above the participant's brow and centered approximately with respect to the eye pupil of the corresponding side, laterally symmetric from the midline of the participant's forehead, one pad for left and the other for right hemisphere monitoring. The positioning was intended to capture hemodynamic changes in bilateral dorsolateral PFC. Draw strings attached to the sensor pads were used to prevent the pads from moving once positioned on the participant. A 9 cm wide self-adhesive bandage of length approximately the circumference of the participant's head was folded width-wise and secured around the participant's head across the brow just below the fNIRS sensor pads. Next a sheet of aluminum foil approximately half the circumference of the participant's head and folded width-wise was form fitted over the bandage and fNIRS sensor pads. Care was taken to ensure that the fNIRS sensor pads were fully encapsulated by the aluminum foil sheet. This was done to ensure that while imaging in sunlight infrared light from the sun would not contaminate the fNIRS signal. Once the foil was affixed to the participant two more self-adhesive bandages of length approximately the circumference of the participant's head were used. One bandage folded twice width-wise was wrapped around the participant's head just below the fNIRS sensor pads, over the participant's brow and over the aluminum foil. The second bandage was folded once width-wise and wrapped around the participant's head just above the fNIRS sensor pads and over the foil. These bandages were used to ensure that the foil did not shift during walking, and special care was taken to minimize constrictive pressure over the fNIRS sensor as initial pilot tests showed this to be extremely uncomfortable for the participants after only a few minutes of walking. Once the sensors, foil, and bandages were positioned, the fNIRS device was turned on and the received light signal was adjusted by light source brightness and detector gain for signal quality. Also, an ambient light channel was captured to further assess signal quality. When the signal was deemed adequate, the participant was asked to put the

[1]www.fnirdevices.com.

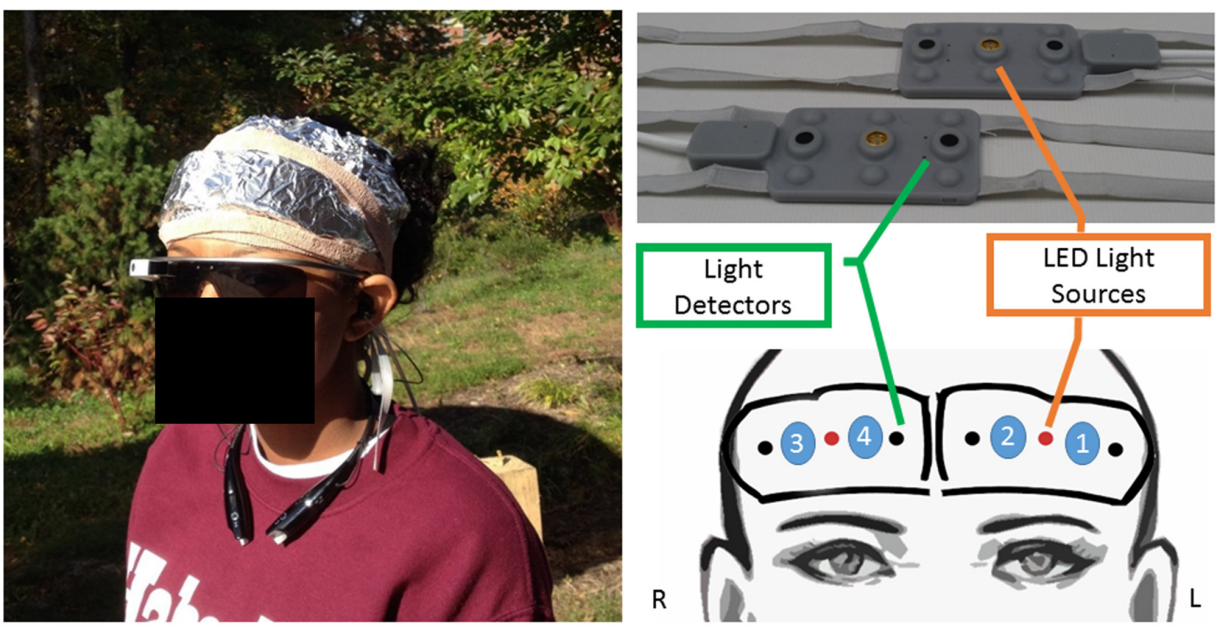


FIGURE 2 | Participant in augmented reality wearable displays (ARWD) group wearing battery operated wireless functional near infrared spectroscopy (fNIRS) sensor over the forehead, Google Glass and Bluetooth headphones (left) wireless fNIRS sensor pads (right, top) and placement sketch (right, bottom) with four optodes identified between light source and detectors.

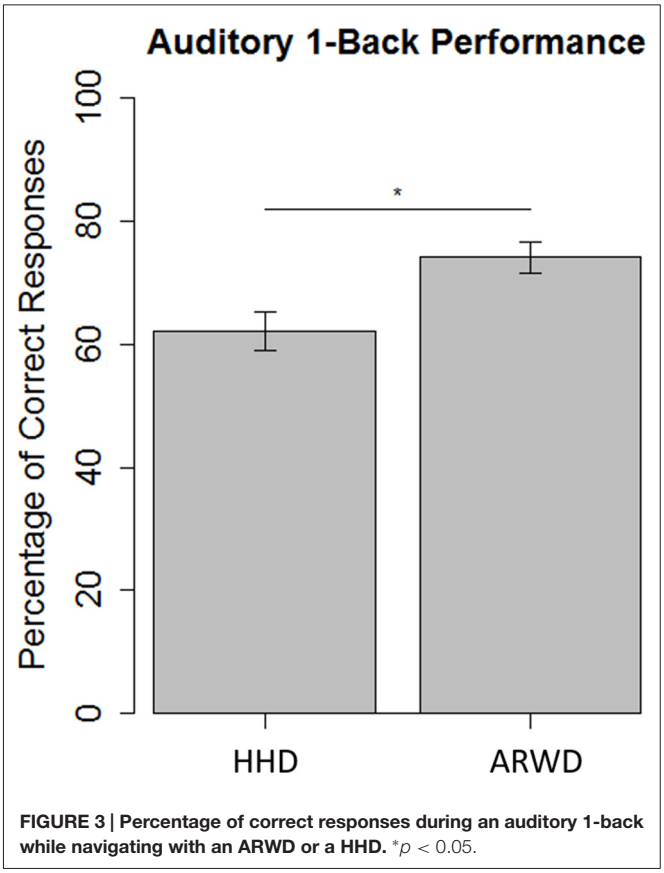


FIGURE 3 | Percentage of correct responses during an auditory 1-back while navigating with an ARWD or a HHD. *$p < 0.05$.

TABLE 2 | Auditory 1-back secondary-task hemodynamics as a function of device and accuracy.

	Left lateral prefrontal cortex			
	HbO		HbR	
	B	*t*-value	*B*	*t*-value
Fixed				
Correct	−0.298**	−2.88	−0.064	−1.76
Incorrect	−0.118	−1.00	−0.008	−0.17
Incorrect-Correct	0.180**	2.60	0.056	1.39
ARWD	−0.252	−1.82	−0.023	−0.48
HHD	−0.164	−1.03	−0.050	−0.87
HHD-ARWD	0.088	0.42	−0.027	−0.36
ARWD: Correct	−0.486**	−3.54	−0.001	−0.02
ARWD: Incorrect	−0.018	−0.12	−0.045	−0.74
ARWD: Incorrect-Correct	0.469***	5.38	−0.044	−0.87
HHD: Correct	−0.110	−0.71	−0.128	−2.29
HHD: Incorrect	−0.218	−1.21	0.029	0.39
HHD: Incorrect-Correct	−0.108	−1.01	0.157*	2.49
HHD: Correct-ARWD: Correct	0.376	1.82	−0.127	−1.73
HHD: Incorrect-ARWD: Incorrect	−0.201	−0.85	0.074	0.77
	Var	*Std. Dev*	*Var*	*Std. Dev*
Random				
Intercept	0.150	0.387		
Block slope	0.002	0.048	0.008	0.090
Residual	1.102	1.050	0.327	0.610

Notes. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

fNIRS transmitter in their pocket. Final setup can be viewed in **Figure 2**.

Experimental Paradigm

Once the fNIRS neuroimaging setup was complete, participants were given the Bluetooth head phones and instructed to place the earbuds in their ears. Prior to this, the earbuds were cleaned with alcohol swabs. If the ear buds did not fit, a new size bud was used to optimize the setup for the participant. Once the headphones were set up, participants were introduced to the auditory 1-back and scenery probe tasks as described in sections "Auditory 1-Back" and "Scenery Probe" respectively. For the auditory task, participants were informed as to the type of stimuli they would hear, and what was considered a correct response, after which participants performed one practice block to ensure they understood the task. If participants were still unclear as to the nature of the task following the practice block, a second practice was given. No participant required more than two practice blocks in order to understand the principal of the auditory task. For the scenery probe task, participants were told they would be prompted to be "situationally aware" at which point they should be acutely aware of their surroundings. They were also informed that after being in this state for a brief period they would be questioned as to whether an object was or was not present in the environment during this time. Participants were informed that both of these tasks would take place while they were route following, but that the auditory 1-back and scenery probe task would never occur simultaneously. Once participants acknowledged they understood the nature of the two secondary tasks, the experimenters and participant relocated outdoors. All testing sessions took place between 7 and 11 am to minimize fatigue from midday heat. The participant was told they would navigate a predetermined route, the route will be displayed via a navigation device (dependent upon their group assignment) and programed into the device by the experimenter. The navigation task is described in detail in section "Route Following". The first secondary task was prepared and the participant was instructed to begin. The secondary task orders were randomized within-subjects across the four routes, at least 15 s of navigation occurred between secondary task blocks. The start and end positions along the routes for each secondary task were preplanned so that each participant would experience the same secondary task at the same place along their navigation routes. Start and end times of the secondary tasks were synchronized with the fNIRS signal via manual entry of timing markers in the data acquisition program at the preplanned start and end positions. Upon completing a route, the participant was instructed to relax and asked whether they were still comfortable and if they wished to continue. No participant indicated they would like to stop participation due to discomfort. The navigation device was then taken by the experimenter, a new route was inputted, and the next route began.

TABLE 3 | Auditory 1-back secondary-task hemodynamics as a function of device and accuracy.

	Left medial prefrontal cortex			
	HbO		HbR	
	B	*t*-value	*B*	*t*-value
Fixed				
Correct	−0.087	−0.77	−0.201***	−3.82
Incorrect	0.223	1.78	−0.029	−0.42
Incorrect-Correct	0.310***	4.17	0.172**	2.73
ARWD	0.063	0.46	−0.092	−1.46
HHD	0.074	0.41	−0.137	−1.68
HHD-ARWD	0.011	0.05	−0.045	−0.44
ARWD: Correct	−0.240	−1.74	−0.113	−1.77
ARWD: Incorrect	0.366*	2.42	−0.071	−0.87
ARWD: Incorrect-Correct	0.607***	6.83	0.042	0.55
HHD: Correct	0.067	0.38	−0.288***	−3.45
HHD: Incorrect	0.080	0.40	0.014	0.13
HHD: Incorrect-Correct	0.013	0.11	0.303**	3.01
HHD: Correct-ARWD: Correct	0.308	1.37	−0.176	−1.67
HHD: Incorrect-ARWD: Incorrect	−0.286	−1.14	0.086	0.64
	Var	*Std. Dev*	*Var*	*Std. Dev*
Random				
Intercept	0.126	0.354		
Block slope	0.006	0.080	0.002	0.043
Residual	0.876	0.936	0.654	0.809

Notes. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

TABLE 4 | Auditory 1-back secondary-task hemodynamics as a function of device and accuracy.

	Right medial prefrontal cortex			
	HbO		HbR	
	B	*t*-value	*B*	*t*-value
Fixed				
Correct	−0.111	−0.80	−0.061	−1.24
Incorrect	0.181	1.19	0.005	0.08
Incorrect-Correct	0.292***	3.52	0.066	1.14
ARWD	0.010	0.50	0.013	0.19
HHD	−0.029	−0.15	−0.069	−1.00
HHD-ARWD	−0.129	−0.46	−0.082	−0.84
ARWD: Correct	−0.195	−0.99	0.016	0.23
ARWD: Incorrect	0.395	1.83	0.010	0.11
ARWD: Incorrect-Correct	0.590***	4.92	−0.006	−0.07
HHD: Correct	−0.026	−0.13	−0.137	−2.00
HHD: Incorrect	−0.032	−0.15	−0.001	−0.01
HHD: Incorrect-Correct	−0.006	−0.05	0.137	1.72
HHD: Correct-ARWD: Correct	0.169	0.61	−0.153	−1.57
HHD: Incorrect-ARWD: Incorrect	−0.427	−1.40	−0.011	−0.08
	Var	*Std. Dev*	*Var*	*Std. Dev*
Random				
Intercept	0.192	0.438		
Block Slope	0.001	0.027	0.003	0.053
Residual	0.975	0.987	0.473	0.688

Notes. $^{***}p < 0.001$.

fNIRS Signal Processing

For each participant, raw light intensity fNIRS data (4 optodes × 2 wavelengths per optode) that were sampled at 4 Hz were low-pass filtered with a finite impulse response, linear phase filter with order 20 and cut-off frequency of 0.1 Hz to attenuate high frequency noise, respiration and cardiac cycle effects (Ayaz et al., 2011). Each participant's data was checked for any potential saturation (when light intensity at the detector was higher than the analog-to-digital converter limit) and motion artifact contamination by means of a coefficient of variation based assessment (Ayaz et al., 2010) and for each optode, a separate channel that recorded ambient light, provided for additional verification. The light intensity changes for 730 and 850 nm wavelengths for each optode for each task block were extracted using time synchronization markers of task onset and end marked during the experiment and hemodynamic changes during each block were calculated separately using the Modified Beer-Lambert Law as described in Ayaz et al. (2012b). Ten seconds (10 s) local baselines were used in the modified Beer-Lambert law to calculate oxygenation for each task condition to look at the relative changes in oxygenated and deoxygenated hemoglobin within each task condition. The local baselines were taken at the beginning of each secondary task, during that time participants were mobile and performing the primary task. The time series for each block was further binned, the hemodynamic response at each optode across the trial was temporally divided into sub-blocks of 10 s each and each sub-block was averaged across time to provide a down-sampled hemodynamic response at each optode for each block. The final output of each optode was mean block deoxygenated hemoglobin (HbR), mean block oxygenated hemoglobin (HbO).

fNIRS Analysis

Generalized and Linear Mixed Effects Models

All forthcoming statistical tests employ either linear mixed effects, or generalized linear mixed effects models implemented in R (R Core Team, 2012) via lme4 (Bates et al., 2014). Denominator degrees of freedom and *p*-values were estimated via Sattherwaite corrections implemented via lmerTest (Kuznetsova et al., 2013). These models offer several advantages as extensions of the general linear model (GLM). Such as, analysis of binomial outcomes, treatment of effects as simultaneously fixed and random, hierarchical modeling, analysis of unbalanced designs, and robustness to missing data (Pinheiro and Bates, 2000; Baayen et al., 2008; Jaeger, 2008; Verbeke and Molenberghs, 2009; Demidenko, 2013).

Fixed and Random Effects Selection

Bayesian information criterions was used to select the fixed and random effects in the final models for each dependent variable. Competing models were constructed by adding potentially meaningful random and fixed effects to a null model. The null model was specified in each case as having no fixed effects and a random effect of participant intercept. All competing

TABLE 5 | Auditory 1-back secondary-task hemodynamics as a function of device and accuracy.

	Right lateral prefrontal cortex			
	HbO		HbR	
	B	*t*-value	*B*	*t*-value
Fixed				
Correct	−0.287***	−5.23	−0.065	−1.98
Incorrect	−0.122	−1.73	−0.064	−1.45
Incorrect-Correct	0.165**	2.76	0.001	0.03
ARWD	−0.156	−1.95	0.058	1.24
HHD	−0.252**	−3.28	−0.186***	−4.18
HHD-ARWD	−0.097	−0.87	−0.245***	−3.77
ARWD: Correct	−0.366***	−4.68	0.066	1.44
ARWD: Incorrect	0.055	0.54	0.050	0.78
ARWD: Incorrect-Correct	0.421***	4.89	−0.016	−0.27
HHD: Correct	−0.207*	−2.69	−0.196***	−4.24
HHD: Incorrect	−0.298**	−3.08	−0.177**	−2.99
HHD: Incorrect-Correct	−0.091	−1.10	0.019	0.32
HHD: Correct-ARWD: Correct	0.159	1.45	−0.262***	−4.01
HHD: Incorrect-ARWD: Incorrect	−0.353*	−2.51	−0.227**	−2.60
	Var	*Std. Dev*	*Var*	*Std. Dev*
Random				
Intercept	0.001	0.001		
Block Slope	0.003	0.054	0.001	0.025
Residual	0.707	0.841	0.352	0.594

Notes. $*p < 0.05$; $**p < 0.01$; $***p < 0.001$.

models were estimated with maximum likelihood to allow for testing of fixed effects. The competing models were tested simultaneously with BIC and the strength of evidence criterion described by Kass and Raftery (1995) was employed. In the procedure deviations of greater than two BIC are viewed as a meaningful difference. The final model was selected based on having the lowest BIC, with no other models of interest having a BIC deviance of less than two. This procedure serves to both minimize over fitting of the models random effects, and to act as an omnibus test of variance for fixed effects and interactions between fixed effects, as passing this procedure ensures that these variables accounted for a meaningful amount of variance in the data.

Multiple Comparisons Corrections

In all forthcoming analyses of fNIRS data multiple comparisons were corrected for across hypotheses and optodes but within secondary tasks and chromophores by adjusting *p*-value criterion with false discovery rate (FDR) corrections. Controlling for FDR can increase statistical power relative to correcting for multiple comparisons via controlling for the familywise error rate (FWER). The Benjamini-Hockberg FDR procedure, employed here for controlling the FDR is adaptive in that the threshold for rejecting the null hypothesis is dependent on the size of the initial *p*-value and the number of hypotheses tested (Benjamini and Hochberg, 1995; Lindquist, 2008). Adjustments were made with alpha set to 0.05 in the Benjamini-Hockberg equation.

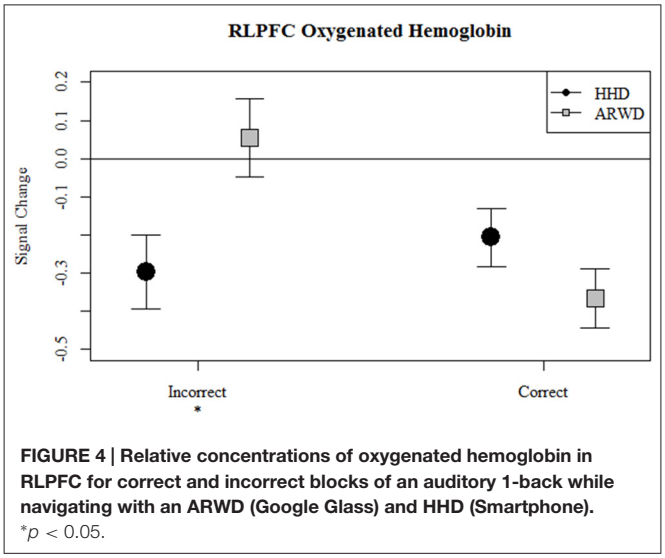


FIGURE 4 | Relative concentrations of oxygenated hemoglobin in RLPFC for correct and incorrect blocks of an auditory 1-back while navigating with an ARWD (Google Glass) and HHD (Smartphone). $*p < 0.05$.

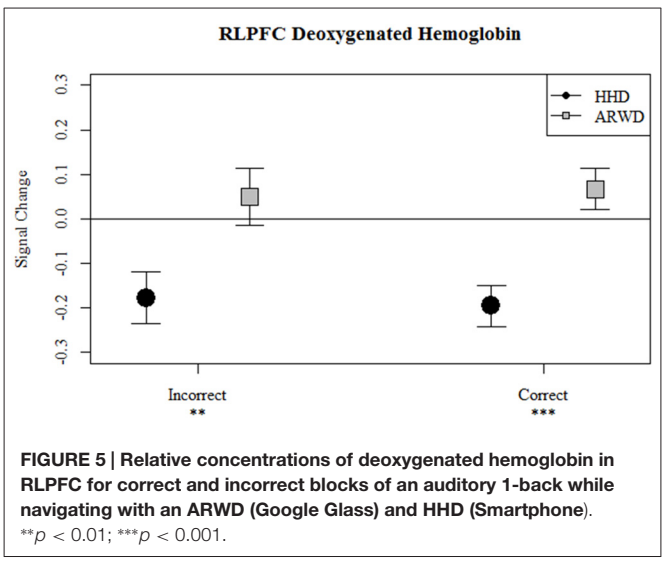


FIGURE 5 | Relative concentrations of deoxygenated hemoglobin in RLPFC for correct and incorrect blocks of an auditory 1-back while navigating with an ARWD (Google Glass) and HHD (Smartphone). $**p < 0.01$; $***p < 0.001$.

RESULTS

Auditory 1-Back

Behavioral

The results of the auditory 1-back were submitted to a generalized linear mixed effects regression. The link function was specified as binomial and parameter estimates were calculated using maximum likelihood. The tested fixed effects included condition (ARWD vs. HHD with smartphone coded as the reference factor), trial and the interaction between the two. The trial component was included to determine if there were any accommodation effects within and between the two devices. Participant intercepts were specified as the random effect. Parameter estimates are reported here as log odds ratios (as they are linear and non-conditional within this analysis). Participants in the HHD group were more likely to correctly than incorrectly report the number of matches heard ($b = 0.528$, $SE = 0.178$, $p < 0.005$). Participants in the ARWD group were more likely

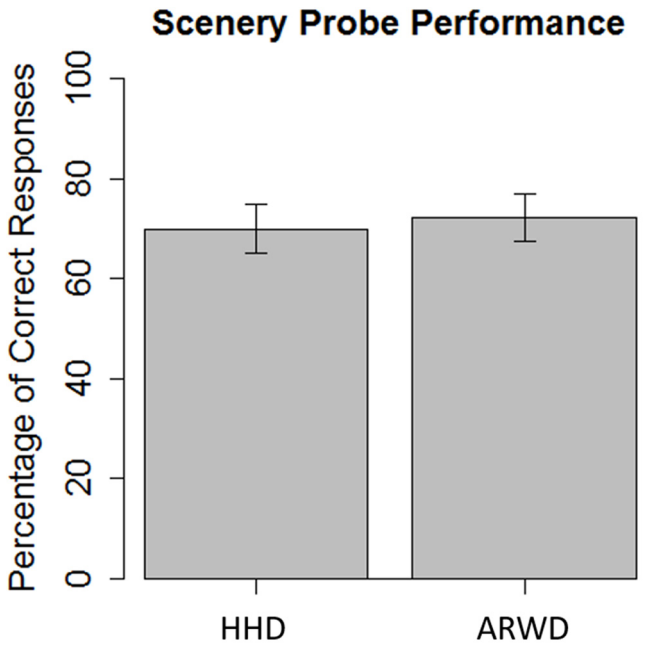


FIGURE 6 | Percentage of correct responses during a scenery probe task while navigating with ARWD or a HHD.

than the HHD group to correctly report the number of matches heard ($b = 0.551$, $SE = 0.257$, $p < 0.05$; **Figure 3**). The effect of trial was non-significant ($b = 0.019$, $SE = 0.013$, $p = 0.137$), and this did not differ between the two device groups ($b = 0.013$, $SE = 0.019$, $p = 0.492$). These results suggest that participants in the ARWD condition experienced lower levels of cognitive load relative to participants in the HHD condition when route following. Furthermore, there is no evidence that this level of load changed throughout the experiment both within and between conditions.

fNIRS

Relative measures of HbO and HbR acquired during the auditory 1-back were submitted to a linear mixed effects regression. Parameter estimates in the model selected from the procedure described in section "Fixed and Random Effects Selection" were calculated with restricted maximum likelihood. Fixed effects were condition (ARWD vs. HHD) and an interaction with performance (correct vs. incorrect). Participant intercepts and block slope were specified as random effects for HbO, and participant block slope was specified for HbR. The results of models for optodes over left lateral, left medial, right medial, and right lateral PFC are reported in **Tables 2–5**.

Left Lateral PFC (LLPFC)

Correct blocks while using an ARWD were associated with a decrease in the hemodynamic response as evidenced by a reduction in oxygenated hemoglobin relative to the null hypothesis. Furthermore, correct blocks while using a HHD were associated with an increase in the hemodynamic response as evidenced by a decrease in deoxygenated hemoglobin relative to incorrect blocks.

Left Medial PFC (LMPFC)

Correct blocks while using a HHD were associated with an increase in the hemodynamic response as evidenced by a reduction in deoxygenated hemoglobin. Furthermore, there is evidence to suggest that incorrect blocks while using an ARWD were associated with an increase in the hemodynamic response. Specifically, relative to the null hypothesis and correct blocks, incorrect blocks were related to an increase in oxygenated hemoglobin.

Right Medial PFC (RMPFC)

Incorrect blocks while using an ARWD were associated with an increase in the hemodynamic response as evidenced by increased oxygenated hemoglobin relative to correct blocks.

Right Lateral PFC (RLPFC)

Correct blocks while using an ARWD were associated with a decrease in the hemodynamic response as evidenced by reduced oxygenated hemoglobin relative to the null hypothesis and incorrect blocks. Furthermore, HHD use during correct and incorrect blocks reduced total hemoglobin as evidenced by the reductions in oxy and deoxygenated hemoglobin. Of particular note for workload comparison between the display conditions is that during correct blocks, HHD deoxygenated hemoglobin was less than ARWD deoxygenated hemoglobin. Finally, during incorrect blocks HHD use was associated with decreases in oxygenated and deoxygenated hemoglobin relative to ARWD use.

Overall, correct auditory memory performance while using ARWD was associated with a reduction in the hemodynamic response in bilateral PFC. Interestingly, incorrect responses were associated with an increase in the hemodynamic response at the more medial measurement sites. Effects of HHD use were mainly observed in left medial PFC, where correct auditory memory performance while using a HHD was associated with an increase in the hemodynamic response. Workload differences as inferred from errors on secondary tasks are most apparent in RLPFC. Where auditory errors during HHD use were associated with reductions in oxygenated (**Figure 4**) and deoxygenated (**Figure 5**) hemoglobin relative to ARWD use.

Situation Awareness

Behavioral

The results of the scenery probe task were submitted to a generalized linear mixed effects regression. The link function was specified as binomial and parameter estimates were calculated using maximum likelihood. The tested fixed effects were condition (ARWD vs. HHD with HHD coded as the reference factor), trial and the interaction between the two. The trial component was included to determine if there were any accommodation effects within and between the two devices. Participant intercepts and uncorrelated trial slopes were specified as the random effects. Parameter estimates are reported here

TABLE 6 | Scenery probe secondary-task hemodynamics as a function of device and accuracy.

	Left lateral prefrontal cortex			
	HbO		HbR	
	B	*t*-value	*B*	*t*-value
Fixed				
Correct	−0.165**	−2.82	−0.051	−1.67
Incorrect	0.008	0.10	−0.096	−2.36
Incorrect-Correct	0.173	2.18	−0.045	−1.19
ARWD	0.015	0.19	0.017	0.44
HHD	−0.172	−1.94	−0.164***	−3.53
HHD-ARWD	−0.187	−1.59	−0.181**	−2.97
ARWD: Correct	−0.238**	−3.18	−0.004	−0.11
ARWD: Incorrect	0.268*	2.45	0.039	0.72
ARWD: Incorrect-Correct	0.506***	4.80	0.043	0.85
HHD: Correct	−0.092	−1.02	−0.098	−2.08
HHD: Incorrect	−0.252	−2.09	−0.230***	−3.78
HHD: Incorrect-Correct	−0.160	−1.35	−0.133	−2.34
HHD: Correct-ARWD: Correct	0.146	1.25	−0.094	−1.54
HHD: Incorrect-ARWD: Incorrect	−0.520**	−3.19	−0.269***	−3.31
	Var	*Std. Dev*	*Var*	*Std. Dev*
Random				
Trial slope	0.009	0.097	0.005	0.072
Residual	0.568	0.753	0.128	0.358

Notes. $^*p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

as log odds ratios. Participants in the HHD group were more likely to correctly respond to the scenery probe ($b = 0.914$, $SE = 0.235$, $p < 0.001$) than not. Participants in the ARWD group showed no significant difference relative to the HHD group in correctly responding to the scenery probe ($b = -0.155$, $SE = 0.335$, $p = 0.644$; **Figure 6**). The effect of trial was non-significant ($b = -0.098$, $SE = 0.091$, $p = 0.282$), and this did not differ between the two device groups ($b = -0.148$, $SE = 0.132$, $p = 0.260$). These results suggest that participants in the both conditions were able to effectively perform the task. However, there is no measureable difference in situation awareness for environmental objects between the two conditions. Furthermore, there is no evidence that situation awareness changed throughout the experiment both within and between conditions.

fNIRS

Relative measures of HbO and HbR acquired during the scenery probe task were submitted to a linear mixed effects regression. Parameter estimates in the model selected from the procedure described in section "Fixed and Random Effects Selection" were calculated with restricted maximum likelihood. Fixed effects were condition (ARWD vs. HHD) and an interaction with performance (correct vs. incorrect). Participant random trial slopes were specified as random effects. The results of models for optodes over left lateral, left medial, right medial, and right lateral PFC are reported in **Tables 6–9**.

Left Lateral PFC

Correct trials while using an ARWD, were associated with a decrease in oxygenated hemoglobin. Incorrect ARWD trials were associated with an increase in oxygenated hemoglobin, and the difference in relative oxygenated hemoglobin between the two outcomes was significant. Incorrect trials while using a HHD were related to reduced deoxygenated hemoglobin. Finally, incorrect trials while using a HHD reduced oxygenated and deoxygenated hemoglobin relative to incorrect trials while using an ARWD. This is either representative of only a reduction in total hemoglobin or a reduction in total hemoglobin and a reduction in brain activity in this region as the decline in oxygenated hemoglobin is greater than that of deoxygenated hemoglobin.

Left Medial PFC

Correct trials while using an ARWD were associated with a decrease in the hemodynamic responses as evidenced by the reduction in oxygenated hemoglobin. Furthermore, incorrect trials while using a HHD were associated with an increase in the hemodynamic response as evidenced by reduced deoxygenated hemoglobin.

Right Medial PFC

No significant differences in hemodynamics were observed in regards to accuracy, or device use.

Right Lateral PFC

Incorrect trials while using a HHD were associated with a decrease in the hemodynamic response as evidenced by reduced oxygenated hemoglobin relative to the null hypothesis and correct trials. Furthermore, incorrect trials while using an ARWD were associated with an increase in the hemodynamic response as evidenced by an increase in oxygenated hemoglobin relative to correct trials, and incorrect trials while using a HHD.

Overall, high situation awareness while using glass was associated with a reduced hemodynamic response in left PFC. Low situation awareness while using glass was related to an increase in the hemodynamic response in bilateral PFC. Conversely, low situation awareness while using a smartphone was associated with a reduced hemodynamic response in bilateral PFC (**Figure 7**).

DISCUSSION

ARWDs are increasing in use and it is important that we understand how such devices affect mental workload and situation awareness. NASA plans to use ARWDs to improve training and performance in highly demanding situations (Schierholz et al., 2015). Objectively measuring mental workload and situation awareness in ARWDs can be difficult due to the immersive and mobile nature of the technology. To circumvent issues of mobility and immersion we used wireless fNIRS to examine hemodynamic differences in mental workload and situation awareness between an ARWD (i.e., Google Glass) and a hand-held display (i.e., a smartphone) during real-world navigation and dual-tasking.

TABLE 7 | Scenery probe secondary-task hemodynamics as a function of device and accuracy.

	Left medial prefrontal cortex			
	HbO		HbR	
	B	*t*-value	*B*	*t*-value
Fixed				
Correct	−0.146	−2.02	−0.080	−1.34
Incorrect	−0.172	−1.58	−0.168	−1.94
Incorrect-Correct	−0.026	−0.24	−0.087	−1.11
ARWD	−0.155	−1.68	−0.038	−0.49
HHD	−0.163	−1.38	−0.210	−2.09
HHD-ARWD	−0.008	−0.05	−0.173	−1.37
ARWD: Correct	−0.218*	−2.60	−0.107	−1.52
ARWD: Incorrect	−0.092	−0.66	0.032	0.29
ARWD: Incorrect-Correct	0.125	0.91	0.139	1.38
HHD: Correct	−0.074	−0.63	−0.054	−0.55
HHD: Incorrect	−0.251	−1.51	−0.367**	−2.73
HHD: Incorrect-Correct	−0.176	−1.07	−0.313*	−2.58
HHD: Correct-ARWD: Correct	0.021	0.24	0.053	0.44
HHD: Incorrect-ARWD: Incorrect	−0.158	−0.73	−0.399	−2.31
	Var	*Std. Dev*	*Var*	*Std. Dev*
Random				
Trial slope	0.006	0.079	0.011	0.105
Residual	0.637	0.798	0.336	0.579

Notes. $*p < 0.05$; $**p < 0.01$.

TABLE 8 | Scenery probe secondary-task hemodynamics as a function of device and accuracy.

	Right medial prefrontal cortex			
	HbO		HbR	
	B	*t*-value	*B*	*t*-value
Fixed				
Correct	−0.134	−1.98	−0.044	−0.80
Incorrect	−0.118	−1.20	−0.002	−0.02
Incorrect-Correct	0.016	0.16	0.042	0.50
ARWD	−0.080	−0.79	0.077	0.91
HHD	−0.172	−1.77	−0.122	−1.54
HHD-ARWD	−0.091	−0.65	−0.199	−1.72
ARWD: Correct	−0.141	−1.50	−0.041	−0.54
ARWD: Incorrect	−0.019	−0.13	0.194	1.53
ARWD: Incorrect-Correct	0.122	0.86	0.236	1.87
HHD: Correct	−0.126	−1.30	−0.047	−0.59
HHD: Incorrect	−0.217	−1.65	−0.198	−1.76
HHD: Incorrect-Correct	−0.091	−0.72	−0.151	−1.35
HHD: Correct-ARWD: Correct	0.015	0.11	−0.006	−0.05
HHD: Incorrect-ARWD: Incorrect	−0.197	−1.00	−0.392	−2.31
	Var	*Std. Dev*	*Var*	*Std. Dev*
Random				
Trial slope	0.012	0.108	0.004	0.067
Residual	0.448	0.669	0.360	0.600

Behavioral differences between the ARWD and HHD while navigating and performing an auditory working memory task suggest differences in experienced workload. While dual-tasking, both tasks were preformed successfully across displays types. However, individuals using an ARWD showed superior working memory recall relative to HHD users. The dual-task method of assessing mental workload (Ogden et al., 1979; O'Donnell and Eggemeier, 1986) dictates that higher performance observed in secondary tasks represents reduced workload during the primary task. The increased working memory performance observed while using an ARWD suggests that relative to hand-held displays ARWDs induce less mental workload while being used for navigation.

Mental workload and the hemodynamic response representative of brain activity are positively related, especially in working memory tasks (Braver et al., 1997; Cohen et al., 1997; Culham et al., 2001; Ayaz et al., 2012b); from our behavioral results, we expected a lower hemodynamic response for ARWD users relative to HHD users. In accordance with our behavioral results ARWD blocks were associated with a reduction of oxygenated hemoglobin representative of a reduction of brain activity in bilateral PFC. Furthermore, HHD trials were associated with a reduction of deoxygenated hemoglobin representative of an increase in brain activity in left medial and right lateral PFC. A direct comparison of the two conditions hemodynamics in RLPFC revealed reduced deoxygenated hemoglobin during HHD use relative to ARWD use during correct auditory working memory performance. This provides further evidence that even when the interference between the auditory working memory task and the navigation task was not *overloading*, neural activity was higher while using an HHD.

With regard to the scenery probe task, we observed no performance differences between ARWDs and hand-held displays, but hemodynamic differences were observed. Both display groups performed the scenery probe and navigation tasks successfully. However, unlike when working memory and navigation co-occurred, dual-task assessment could not differentiate between the two displays in terms of mental workload during the scenery probe task. This was not the case for hemodynamic measurements made with wireless fNIRS. The difference between the display conditions is strongest in left lateral PFC. In this region there was a reduction in oxygenated hemoglobin during ARWD use on correct trials. A decrement was not present in left lateral PFC during hand-held display use. While not as large, a similar trend can be seen between ARWD and hand-held displays in right lateral PFC as well. While inconclusive, considering the non-significant differences in oxygenated hemoglobin on correct trials between ARWD and hand-held displays, the trend is for reduced brain activity during ARWD use. Taking the scenery probe task as a proxy for level 1 and 2 situation awareness, less mental resources were required during landmark perception and comprehension while navigating with an ARWD relative to a hand-held display.

Scenery probe and working memory errors were associated with changes in ARWD hemodynamics. Lower situation awareness during ARWD use was associated with increased

TABLE 9 | Scenery probe secondary-task hemodynamics as a function of device and accuracy.

	Right lateral prefrontal cortex			
	HbO		HbR	
	B	*t*-value	*B*	*t*-value
Fixed				
Correct	−0.045	−1.04	−0.060	−1.56
Incorrect	−0.114	−1.98	−0.100	−1.92
Incorrect-Correct	−0.069	−1.28	−0.039	−0.79
ARWD	0.043	0.75	−0.054	−1.03
HHD	−0.2019**	−3.17	−0.106	−1.86
HHD-ARWD	−0.2453**	−2.85	−0.053	−0.68
ARWD: Correct	−0.055	−0.95	0.017	0.32
ARWD: Incorrect	0.142	1.83	−0.124	−1.75
ARWD: Incorrect-Correct	0.1969**	2.65	−0.141	−2.06
HHD: Correct	−0.034	−0.54	−0.137	−2.41
HHD: Incorrect	−0.3696***	−4.37	−0.076	−0.99
HHD: Incorrect-Correct	−0.3354***	−4.28	0.062	0.86
HHD: Correct-ARWD: Correct	0.021	0.24	−0.154	−1.99
HHD: Incorrect-ARWD: Incorrect	−0.5114***	−4.45	0.048	0.46
	Var	*Std. Dev*	*Var*	*Std. Dev*
Random				
Trial slope	0.008	0.091	0.005	0.073
Residual	0.216	0.464	0.182	0.427

Notes. $^{**}p < 0.01$; $^{***}p < 0.001$.

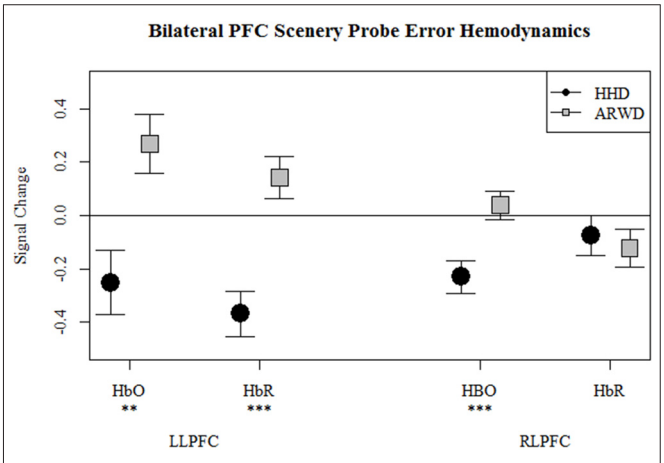


FIGURE 7 | Relative oxygenated and deoxygenated hemoglobin at bilateral optodes for incorrect trials of the scenery probe task while navigating with an ARWD (Google Glass) and HHD (Smartphone). $^{**}p < 0.01$, $^{***}p < 0.001$.

oxygenated hemoglobin in bilateral PFC. Similarly, incorrect working memory trials and ARWD use were associated with increased oxygenated hemoglobin across PFC. Effectively, poor secondary task performance was associated with an increase in PFC activity while navigating with an ARWD. This increase in activity coincides with the increase in workload expected due to dual-task interference. Stimulus driven attention capture is related to increased activity in PFC (Fockert et al., 2004; Serences et al., 2005; Asplund et al., 2010). Furthermore, head-up display symbology is known to negatively affect performance from unnecessary attention capture (Thomas and Wickens, 2001, 2004). The presence of cognitive tunneling during ARWD use can parsimoniously explain the presence of an error, the increase in brain activity and the increase in mental workload observed across both secondary tasks. Also considering that the display symbology was unchanged between the ARWD and HHD conditions, and that the symbology was originally designed for the HHD; the presence of cognitive tunneling was expected.

The association of secondary task errors on HHD hemodynamics was the opposite of that observed during ARWD use. Across both secondary tasks, errors were associated with decreases in brain activity. Working memory errors were associated with an increase in left PFC deoxygenated hemoglobin. Lower situation awareness was associated with a decrease in bilateral PFC oxygenated hemoglobin and RLPFC deoxygenated hemoglobin. It is probable that HHD errors were related to task shedding, the abandonment of one of the two tasks being performed; a common strategy during dual-tasks that overload mental resources (Schneider and Detweiler, 1988; Raby and Wickens, 1994; Hancock and Szalma, 2003; Grier et al., 2008; Schulte and Donath, 2011). Task shedding should produce a reduction in brain activity due to reducing mental workload. Therefore, we would expect a reduced hemodynamic response during correct secondary task trials if the primary task was shed. This effect was not observed. Instead, brain activity decreased during incorrect trials. Continuing with the logic that reduced activity is related to reduced mental workload, reduced activity during incorrect secondary trials suggests that the secondary-tasks may have been shed. This explanation is consistent with the emphasis we placed on the navigation task as well as our observed behavioral and hemodynamic effects.

LIMITATIONS

Due to the nature of the wireless fNIRS, and the miniaturized design of our imaging unit we are limited to four optodes imaging the PFC. Therefore, other cortical regions may have shown significant hemodynamic differences between the two devices that we could not measure. Furthermore, given the current design, we could not account for all factors that might influence difference in mental workload between the two devices. We could only measure differences in mental workload that manifest as dual-task interference from increased working memory load, or increased perceptual load.

CONCLUSION

Taking a neuroergonomic approach combining dual-task interference and wireless fNIRS, we were able to examine differences in mental workload, and situation awareness between a hand-held display (smartphone) and an augmented reality wearable display (Google Glass) while navigating an outdoor environment. ARWDs show few downsides with regards to dual tasking while route following. Relative to a HHD, mental workload while navigating with an ARWD was reduced, both during a working memory and situation awareness secondary

task; performance was also enhanced during the working memory dual-task. Hemodynamic effects induced during errors also suggest ways in which ARWDs can be improved, specifically by reducing unwanted attention capture and cognitive tunneling. Future work should identify other hemodynamic biomarkers induced by cognitive tunneling. From an applied perspective development of tunneling biomarkers could greatly advance display design for navigation, training and other tasks ARWDs are expected to enhance.

DEDICATION

In Memory of Professor Raja Parasuraman, this article is dedicated to Professor Parasuraman for his guidance on past and present neuroergonomic studies as well as the inspiration he provides for studies to come.

AUTHOR CONTRIBUTIONS

RMcK designed the study, collected the data, analyzed the data and wrote the manuscript. RP designed the study. RM designed the study and collected data. AF designed the study and collected data. WB designed the study, analyzed data and edited the manuscript. MP analyzed data and edited the manuscript. HA designed the study, analyzed data and wrote the manuscript.

FUNDING

This research was supported by Air Force Office of Scientific Research (AFOSR) Grant No. FA9550-10-1-0385, and the Center of Excellence in Neuroergonomics, Technology, and Cognition (CENTEC).

REFERENCES

Abibullaev, B., and An, J. (2012). Classification of frontal cortex haemodynamic responses during cognitive tasks using wavelet transforms and machine learning algorithms. *Med. Eng. Phys.* 34, 1394–1410. doi: 10.1016/j.medengphy.2012.01.002

Asplund, C. L., Todd, J. J., Snyder, A. P., and Marois, R. (2010). A central role for the lateral prefrontal cortex in goal-directed and stimulus-driven attention. *Nat. Neurosci.* 13, 507–512. doi: 10.1038/nn.2509

Ayaz, H., Çakır, M. P., Izzetoğlu, K., Curtin, A., Shewokis, P. A., Bunce, S. C., et al. (2012a). Monitoring expertise development during simulated UAV piloting tasks using optical brain imaging. *IEEE Aerospace Conference* 2012, 1–11. doi: 10.1109/aero.2012.6187350

Ayaz, H., Shewokis, P. A., Bunce, S., Izzetoglu, K., Willems, B., and Onaral, B. (2012b). Optical brain monitoring for operator training and mental workload assessment.*Neuroimage* 59, 36–47. doi: 10.1016/j.neuroimage.2011.06.023

Ayaz, H., Izzetoglu, M., Shewokis, P. A., and Onaral, B. (2010). Sliding-window motion artifact rejection for functional near-infrared spectroscopy. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 2010, 6567–6570. doi: 10.1109/iembs.2010.5627113

Ayaz, H., Onaral, B., Izzetoglu, K., Shewokis, P. A., McKendrick, R., and Parasuraman, R. (2013). Continuous monitoring of brain dynamics with functional near infrared spectroscopy as a tool for neuroergonomic research: empirical examples and a technological development. *Front. Hum. Neurosci.* 7:871. doi: 10.3389/fnhum.2013.00871

Ayaz, H., Shewokis, P. A., Curtin, A., Izzetoglu, M., Izzetoglu, K., and Onaral, B. (2011). Using MazeSuite and functional near infrared spectroscopy to study learning in spatial navigation. *J. Vis. Exp.* 56, 10–3791. doi: 10.3791/3443

Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S., and Macintyre, B. (2001). Recent advances in augmented reality. *IEEE Comput. Graphics Appl.* 21, 34–47. doi: 10.1201/b18703-18

Baayen, R. H., Davidson, D. J., and Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *J. Mem. Lang.* 59, 390–412. doi: 10.1016/j.jml.2007.12.005

Badre, D., Poldrack, R. A., Paré-Blagoev, E. J., Insler, R. Z., and Wagner, A. D. (2005). Dissociable controlled retrieval and generalized selection mechanisms in ventrolateral prefrontal cortex. *Neuron* 47, 907–918. doi: 10.1016/j.neuron.2005.07.023

Badre, D., and Wagner, A. D. (2007). Left ventrolateral prefrontal cortex and the cognitive control of memory. *Neuropsychologia* 45, 2883–2901. doi: 10.1016/j.neuropsychologia.2007.06.015

Bates, D., Maechler, M., Bolker, B., and Walker, S. (2014). lme4: Linear mixed-effects models using 'Eigen' and S4. Version, 1.1-10. Retrievable from: https://cran.r-project.org/web/packages/lme4/index.html

Bogler, C., Mehnert, J., Steinbrink, J., and Haynes, J. D. (2014). Decoding vigilance with NIRS. *PLoS One* 9:e101729. doi: 10.1371/journal.pone.0101729

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Series B. Stat. Methodol.* 57, 289–300.

Braver, T. S., Cohen, J. D., Nystrom, L. E., Jonides, J., Smith, E. E., and Noll, D. C. (1997). A parametric study of prefrontal cortex involvement in human working memory. *Neuroimage* 5, 49–62.

Cohen, J. D., Peristein, W. M., Braver, T. S., Nystrom, L. E., Noll, D. C., Jonides, J., et al. (1997). Temporal dynamics of brain activation during a working memory task. *Nature* 386, 604–607. doi: 10.1038/386604a0

Corbetta, M., Miezin, F. M., Dobmeyer, S., Shulman, G. L., and Petersen, S. E. (1991). Selective and divided attention during visual discriminations of shape, color and speed: functional anatomy by positron emission tomography. *J. Neurosci.* 11, 2383–2402. doi: 10.1002/9780470514184.ch10

Crawford, J., and Neal, A. (2006). A review of the perceptual and cognitive issues associated with the use of head-up displays in commercial aviation. *Int. J. Aviat. Psychol.* 16, 1–19. doi: 10.1207/s15327108ijap1601_1

Cui, X., Bray, S., Bryant, D. M., Glover, G. H., and Reiss, A. L. (2011). A quantitative comparison of NIRS and fMRI across multiple cognitive tasks. *Neuroimage* 54, 2808–2821. doi: 10.1016/j.neuroimage.2010.10.069

Culham, J. C., Cavanagh, P., and Kanwisher, N. G. (2001). Attention response functions: characterizing brain areas using fMRI activation during parametric variations of attentional load. *Neuron* 32, 737–745. doi: 10.3410/f.1003266.29857

Demidenko, E. (2013). *Mixed Models: Theory and Applications with R.* New York, NY: John Wiley & Sons.

Derosière, G., Dalhoumi, S., Perrey, S., Dray, G., and Ward, T. (2014). Towards a near infrared spectroscopy-based estimation of operator attentional state. *PLoS One* 9:e92045. doi: 10.1371/journal.pone.0092045

Durso, E. T. and Dattel, A. R. (2004). "SPAM: the real-time assessment of SA" in *A Cognitive Approach to Situation Awareness: Theory and Application*, eds S. Banbury and S. Tremblay (Hampshire: Ashgate Publishing Ltd), 137–154.

Endsley, M. R. (1995a). Toward a theory of situation awareness in dynamic systems. *Hum. Fact.* 37, 32–64. doi: 10.1518/001872095779049543

Endsley, M. R.. (1995b). Measurement of situation awareness in dynamic systems. *Hum. Fact. J. Hum. Fact. Ergon. Soc.* 37, 65–84. doi: 10.10.1518/001872095779049499

Fagioli, S., and Macaluso, E. (2009). Attending to multiple visual streams: interactions between location-based and category-based attentional selection. *J. Cogn. Neurosci.* 21, 1628–1641. doi: 10.1162/jocn.2009.21116

Ferrari, M. and Quaresima, V. (2012). A brief review on the history of human functional near-infrared spectroscopy (fNIRS) development and fields

of application. *Neuroimage* 63, 921–935. doi: 10.1016/j.neuroimage.2012.03.049

Fischer, E., Haines, R. F., and Price, T. A. (1980). *Cognitive Issues in Head-Up Displays. (NASA Tech. Paper No. 1711)*. Moffett Field, CA: NASA Ames, Research Center.

Figner, B., Knoch, D., Johnson, E. J., Krosch, A. R., Lisanby, S. H., Fehr, E., et al. (2010). Lateral prefrontal cortex and self-control in intertemporal choice. *Nat. Neurosci.* 13, 538–539. doi: 10.1038/nn.2516

Fockert, J. D., Rees, G., Frith, C. D., and Lavie, N. (2004). Neural correlates of attentional capture in visual search. *J. Cogn. Neurosci.* 16, 751–759. doi: 10.1162/089892904970762

Gopher, D. (1993). Mental workload. *Hand. Hum. Percept. Per.* 13, 197–198). doi: 10.1007/978-1-4757-0884-4_1

Grabner, R. H., Neubauer, A. C., and Stern, E. (2006). Superior performance and neural efficiency: the impact of intelligence and expertise. *Brain Res. Bull.* 69, 422–439. doi: 10.1016/j.brainresbull.2006.02.009

Grier, R., Wickens, C., Kaber, D., Strayer, D., Boehm-Davis, D., Trafton, J. G., et al. (2008). The red-line of workload: theory, research and design. *Proc. Hum. Fact. Ergon. Soc. Annu. Meet.* 52, 1204–1208. doi: 10.1037/e578252012-011

Hancock, P. A., and Parasuraman, R. (1992). Human factors and safety in the design of Intelligent Vehicle-Highway Systems (IVHS). *J. Safe. Res.* 23, 181–198.

Hancock, P. A., and Szalma, J. L. (2003). Operator stress and display design. *Ergon. Des.* 11, 13–18. doi: 10.1177/106480460301100205

Hart, S. G., and Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. *Adv. Psychol.* 52, 139–183. doi: 10.1016/s0166-4115(08)62386-9

Herath, P., Klingberg, T., Young, J., Amunts, K., and Roland, P. (2001). Neural correlates of dual task interference can be dissociated from those of divided attention: an fMRI study. *Cereb. Cortex* 11, 796–805. doi: 10.1093/cercor/11.9.796

Herff, C., Heger, D., Fortmann, O., Hennrich, J., Putze, F., and Schultz, T. (2014). Mental workload during n-back task-quantified in the prefrontal cortex using fNIRS. *Front. Hum. Neurosci.* 7, 935–940. doi: 10.3389/fnhum.2013.00935

Jaeger, T. F. (2008). Categorical data analysis: away from ANOVAs (transformation or not) and towards logit mixed models. *J. Mem. Lang.* 59, 434–446. doi: 10.1016/j.jml.2007.11.007

Kass, R. E., and Raftery, A. E. (1995). Bayes factors and model uncertainty. *J. Amer. Statist. Assoc.* 90, 773–795.

Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2013). lmerTest: tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package). lmerTest Version 1.2-0. Retrievable from: http://www2.uaem.mx/r-mirror/web/packages/lmerTest/index.html

Lindquist, M. A. (2008). The statistical analysis of fMRI data. *Statist. Sci.* 23, 439–464. doi: 10.1007/978-1-60327-919-2_6

Loose, R., Kaufmann, C., Auer, D. P., and Lange, K. W. (2003). Human prefrontal and sensory cortical activity during divided attention tasks. *Hum. brain mapp.* 18, 249–259. doi: 10.1002/hbm.10082

McKendrick, R., Ayaz, H., Olmstead, R., and Parasuraman, R. (2014). Enhancing dual-task performance with verbal and spatial working memory training: continuous monitoring of cerebral hemodynamics with NIRS. *Neuroimage* 85, 1014–1026. doi: 10.1016/j.neuroimage.2013.05.103

McKendrick, R., Parasuraman, R., and Ayaz, H. (2015). Wearable functional near infrared spectroscopy (fNIRS) and transcranial direct current stimulation (tDCS): expanding vistas for neurocognitive augmentation. *Front. Syst. Neurosci.* 9:27. doi: 10.3389/fnsys.2015.00027

Mizuno, K., Tanaka, M., Tanabe, H. C., Sadato, N., and Watanabe, Y. (2012). The neural substrates associated with attentional resources and difficulty of concurrent processing of the two verbal tasks. *Neuropsychologia* 50, 1998–2009. doi: 10.1016/j.neuropsychologia.2012.04.025

Naseer, N., and Keum-Shik, H. (2013). 'Functional near-infrared spectroscopy based discrimination of mental counting and no-control state for development of a brain-computer interface. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 2013, 1780–1783.

Navon, D. (1984). Resources: a theoretical soup stone. *Psychol. Rev.* 91, 216–234.

O'Donnell, R. D., Eggemeier, F. T. (1986). "Workload assessment methodology," in *Handbook of Perception and Human Performance (Vol. II Cognitive Processes and Performance*, eds K. R. Boff, L. Kaufman, and J. P. Thomas (New York, NY: John Wiley & Sons), 42–41–41–49.

Ogden, G. D., Levine, J. M., and Eisner, E. J. (1979). Measurement of workload by secondary tasks. *Hum. Factors* 21, 529–548. doi: 10.1177/001872087902100502

Opolski, M. P., Debski, A., Borucki, B. A., Szpak, M., Staruch, A. D., Kepka, C., et al. (2015). First-in-man computed tomography-guided percutaneous revascularization of coronary chronic total occlusion using a wearable computer: proof of concept. *Can. J. Cardiol.* doi: 10.1016/j.cjca.2015.08.009 [Epub ahead of print].

Parasuraman, R. (1998). "The attentive brain: issues and prospects," in *The Attentive Brain*, ed. R. Parasuraman. (Cambridge, MA: MIT Press), 3–15.

Parasuraman, R. (2003). Neuroergonomics: research and practice. *TIES* 4, 5–20. doi: 10.1080/14639220210199753

Parasuraman, R., Sheridan, T. B., and Wickens, C. (2008). Situation awareness, mental workload and trust in automation: viable, empirically supported cognitive engineering constructs. *J. Cogn. Eng. Decis. Making* 2, 140–160. doi: 10.1518/155534308x284417

Pashler, H. (1994). Dual-task interference in simple tasks: data and theory. *Psychological bulletin* 116, 220–244. doi: 10.1037/0033-2909.116.2.220

Pinheiro, J. C., and Bates, D. M. (2000). *Mixed-Effects Models in S and SPLUS*. New York, NY: Springer.

Pinti, P., Aichelburg, C., Lind, F., Power, S., Swingler, E., Merla, A., et al. (2015). Using fiberless, wearable fNIRS to monitor brain activity in real-world cognitive tasks. *J. Vis. Exp.* 106, e53336–e53336. doi: 10.3791/53336

Raby, M., and Wickens, C. D. (1994). Strategic workload management and decision biases in aviation. *Int. J. Aviat. Psychol.* 4, 211–240. doi: 10.1207/s15327108ijap0403_2

Ramnani, N., and Owen, A. M. (2004). Anterior prefrontal cortex: insights into function from anatomy and neuroimaging. *Nat. Rev. Neurosci.* 5, 184–194. doi: 10.1038/nrn1343

R Core Team. (2012). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rypma, B., Berger, J. S., and D'Esposito, M. (2002). The influence of working-memory demand subject performance on prefrontal cortical activity. *J. Cogn. Neurosci.* 14, 721–731. doi: 10.1162/08989290260138627

Sato, H., Yahata, N., Funane, T., Takizawa, R., Katura, T., and Atsumori, H. (2013). A NIRS-fMRI investigation of prefrontal cortex activity during a working memory task. *Neuroimage* 83, 158–173. doi: 10.1016/j.neuroimage.2013.06.043

Sawyer, B. D., Finomore, V. S., Calvo, A. A., and Hancock, P. A. (2014). Google glass: a driver distraction cause or cure? *Hum. Factors* 56, 1307–1321. doi: 10.1177/0018720814555723

Schierholz, S., Huot, D., and McGregor, V. (2015). *NASA, Microsoft Collaborate to Bring Science Fiction to Science Fact*. Retrieved from: https://nasa.gov.

Schneider, W., and Detweiler, M. (1988). The role of practice in dual-task performance: toward workload modeling a connectionist/control architecture. *Hum. Factors* 30, 539–566. doi: 10.1037/e512682013-165

Schudlo, L. C., and Chau, T. (2014). Dynamic topographical pattern classification of multichannel prefrontal NIRS signals: II. Online differentiation of mental arithmetic and rest. *J. Neural Eng.* 11:016003. doi: 10.1088/1741-2560/11/1/016003

Schulte, A., and Donath, D. (2011). "Measuring self-adaptive UAV operators' load-shedding strategies under high workload," in *Engineering Psychology and Cognitive Ergonomics*, ed. D. Harris (Germany: Springer, Berlin, Heidelberg), 342–351.

Serences, J. T., Shomstein, S., Leber, A. B., Golay, X., Egeth, H. E., and Yantis, S. (2005). Coordination of voluntary and stimulus-driven attentional control in human cortex. *Psychol. Sci.* 16, 114–122. doi: 10.1111/j.0956-7976.2005.00791.x

Solovey, T. E., Afergan, D., Peck, E. M., Hincks, S. W., and Jacob, R. J. (2015). Designing implicit interfaces for physiological computing: guidelines and lessons learned using fNIRS. *ACM Trans. Comput.-Hum. Interact.* 21, 1–27. doi: 10.1145/2687926

Thomas, L. C., and Wickens, C. D. (2001). Visual displays and cognitive tunneling: frames of reference effects on spatial judgments and change detection. *Proc. Hum. Fact. Ergon. Soc. Annu. Meet.* 45, 336–340. doi: 10.1037/e446242006-001

Thomas, L. C., and Wickens, C. D. (2004). Eye-tracking and individual differences in off-normal event detection when flying with a synthetic vision system display. *Proc. Hum. Fact. Ergon. Soc. Annu. Meet.* 48, 223–227. doi: 10.1177/154193120404800148

Verbeke, G., and Molenberghs, G. (2009). *Linear Mixed Models for Longitudinal Data.* Germany: Springer Science + Business Media

Villringer, A., and Chance, B. (1997). Non-invasive optical spectroscopy and imaging of human brain function. *Trends. Neurosci.* 20, 435–442. doi: 10.1016/s0166-2236(97)01132-6

Wickens, C. D. (2008). Multiple resources and mental workload. *Hum. Factors* 50, 449–455. doi: 10.1518/001872008x288394

Wickens, C. D., Hollands, J. G., Banbury, S., and Parasuraman, R. (2013). *Engineering Psychology and Human Performance.* Upper Saddle River, NJ: Prentice Hall.

Conflict of Interest Statement: fNIR Devices, LLC manufactures the optical brain imaging instrument and licensed IP and know-how from Drexel University. HA was involved in the technology development and thus offered a minor share in the new startup firm fNIR Devices, LLC. The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The Review Editor SC and handling Editor declared their shared affiliation, and the handling Editor states that the process nevertheless met the standards of a fair and objective review.

Copyright © 2016 McKendrick, Parasuraman, Murtza, Formwalt, Baccus, Paczynski and Ayaz. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution and reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 19 January 2016
doi: 10.3389/fnhum.2015.00707

Processing Functional Near Infrared Spectroscopy Signal with a Kalman Filter to Assess Working Memory during Simulated Flight

Gautier Durantin[1,2,3]*, *Sébastien Scannella*[1], *Thibault Gateau*[1], *Arnaud Delorme*[2,3] and *Frédéric Dehais*[1]

[1] *Département Conception et Conduite des Véhicules Aéronautiques et Spatiaux, Institut Supérieur de l'Aéronautique et de l'Espace (ISAE-Supaéro), Toulouse, France,* [2] *Centre de Recherche Cerveau et Cognition, Université Toulouse III - Paul Sabatier, Toulouse, France,* [3] *Centre National de la Recherche Scientifique, Centre de Recherche Cerveau et Cognition, Toulouse, France*

Working memory (WM) is a key executive function for operating aircraft, especially when pilots have to recall series of air traffic control instructions. There is a need to implement tools to monitor WM as its limitation may jeopardize flight safety. An innovative way to address this issue is to adopt a Neuroergonomics approach that merges knowledge and methods from Human Factors, System Engineering, and Neuroscience. A challenge of great importance for Neuroergonomics is to implement efficient brain imaging techniques to measure the brain at work and to design Brain Computer Interfaces (BCI). We used functional near infrared spectroscopy as it has been already successfully tested to measure WM capacity in complex environment with air traffic controllers (ATC), pilots, or unmanned vehicle operators. However, the extraction of relevant features from the raw signal in ecological environment is still a critical issue due to the complexity of implementing real-time signal processing techniques without a priori knowledge. We proposed to implement the Kalman filtering approach, a signal processing technique that is efficient when the dynamics of the signal can be modeled. We based our approach on the Boynton model of hemodynamic response. We conducted a first experiment with nine participants involving a basic WM task to estimate the noise covariances of the Kalman filter. We then conducted a more ecological experiment in our flight simulator with 18 pilots who interacted with ATC instructions (two levels of difficulty). The data was processed with the same Kalman filter settings implemented in the first experiment. This filter was benchmarked with a classical pass-band IIR filter and a Moving Average Convergence Divergence (MACD) filter. Statistical analysis revealed that the Kalman filter was the most efficient to separate the two levels of load, by increasing the observed effect size in prefrontal areas involved in WM. In addition, the use of a Kalman filter increased the performance of the classification of WM levels based on brain signal. The results suggest that Kalman filter is a suitable approach for real-time improvement of near infrared spectroscopy signal in ecological situations and the development of BCI.

Keywords: fNIRS, Kalman filtering, Neuroergonomics, working memory, SVM

OPEN ACCESS

Edited by:
Klaus Gramann,
Berlin Institute of Technology, Germany

Reviewed by:
Daniele Marinazzo,
University of Ghent, Belgium
Tao Liu,
Sun Yat-Sen University, China

***Correspondence:**
Gautier Durantin
gautier.durantin@isae.fr

Received: *25 September 2015*
Accepted: *17 December 2015*
Published: *19 January 2016*

Citation:
Durantin G, Scannella S, Gateau T, Delorme A and Dehais F (2016) Processing Functional Near Infrared Spectroscopy Signal with a Kalman Filter to Assess Working Memory during Simulated Flight. Front. Hum. Neurosci. 9:707. doi: 10.3389/fnhum.2015.00707

1. INTRODUCTION

The development of passive Brain Computer Interfaces (BCI) is a key topic of research in Neuroergonomics. In contrast with active ones, Passive BCI (Cutrell and Tan, 2008) allows the use of unintentionally produced brain activity to derive various cognitive states (Blankertz et al., 2010) such as excessive mental workload. Such states inference provides an interesting insight as they aim at dynamically adapting the nature of the human-system interactions to overcome cognitive limitations (Zander and Kothe, 2011; Brouwer et al., 2013). In the field of BCI design to enhance user performance, there is a growing interest for functional near infrared spectroscopy (fNIRS) based BCI (Coyle et al., 2004; Derosière et al., 2014; Strait et al., 2014). This brain imaging device uses near infrared light absorption properties to estimate local variations of cortical hemodynamics. It uses a modified Beer-Lambert law to link light transmittance through brain tissues to variations in local concentrations in oxygenated hemoglobin (HbO_2) and deoxygenated hemoglobin (HHb) (Villringer and Obrig, 2002). fNIRS has a good spatial resolution (around 1 cm^2) and interesting signal-to-noise ratio. Moreover, this technique has the advantage to be easy and fast to set over the participant's head with a short calibration process (Naseer and Hong, 2015). However, the processing of fNIRS signal faces a lack of methodological consensus and thus still represents a great challenge (Bashashati et al., 2007). The extraction of the relevant activity from brain signals requires complex techniques (van Erp et al., 2012), and most efficient ones often rely on long calibration times [e.g., in subspace artifact removal techniques (von Bünau et al., 2009), adaptive filtering (Zheng et al., 2002)]. The complexity of these methods limits their applicability for Neuroergonomics purpose, as the signal has to be useable in real-time.

Most BCI designs rely on classical linear bandpass filtering techniques such as Infinite Impulse Response (IIR) (Naseer and Hong, 2015), although current research focuses on the investigation of alternative signal processing techniques, such as the Moving Average Convergence Divergence (MACD) filter (Durantin et al., 2014b; Gateau et al., 2015). On this basis, the improvement of signal quality in real-world conditions as suggested in the Neuroergonomics approach, makes Kalman filtering an ideal candidate. This signal processing and estimation technique relies both on the measurements performed on a system and on a modeling of its dynamics to improve signal quality (Kalman, 1960). The use of a Kalman filter including a physiological model of brain function to improve signal usability has been previously applied to EEG (Georgiadis et al., 2005; Callan et al., 2015) or fMRI (Diamond et al., 2005). However, concerning fNIRS, this technique has been limited to the estimation of model parameters (Abdelnour and Huppert, 2009) or the correction of motion artifacts (Izzetoglu et al., 2010), therefore not requiring the use of a physiological model of hemodynamic response to stimulation.

One of the greatest challenges regarding Kalman filter design is the tuning of its parameters, i.e., to evaluate the level of measurement noise (R) affecting the signal and the state noise (Q) in the model (Diamond et al., 2005). The value of the ratio Q/R greatly influences the behavior of the Kalman filter. Indeed, a Kalman filter with a low value of Q/R will put confidence in the dynamical model, whereas a Kalman filter with a high value of Q/R will put confidence in the measurements. In practice, the value of this ratio often has to be chosen empirically (Abdelnour and Huppert, 2009; Callan et al., 2015), as there exists no efficient way to evaluate it. Consequently, the dynamics of the Kalman filter may not be adapted to the data needed to be improved. The challenge of this study was to design a Kalman filter suitable for fNIRS that includes a physiological model of hemodynamic response (Boynton et al., 1996). By applying this filter to fNIRS data collected during both controlled and ecological experiments, we also aimed at testing the improvements such a filter could bring to fNIRS signal toward the implementation of a passive BCI. To that end, we first designed a Kalman filter relying on a model of the hemodynamic response (Boynton et al., 1996) to improve signal quality. We then conducted a first experiment with a prefrontal fNIRS, involving a digit sequence memorization task used to measure Working Memory (WM) storage and update capacity. Provided that the development of a signal improvement technique usable in realistic operational settings was the objective of this study, this basic task was chosen as WM is a key executive function to operate complex systems (Causse et al., 2011). Data collected during the first experiment were used to select the value of the filter parameter Q/R using an optimization procedure. Finally, the improvement of the signal by the optimal Kalman filter was evaluated with formal classification during an ecological experiment which involved pilots performing a realistic WM task (i.e., recalling air traffic instructions) in a flight simulator.

2. KALMAN FILTER DESIGN

The functional model used to design the Kalman filter for fNIRS signal was inspired by the Hemodynamic Response Function (HRF) proposed by Boynton et al. (1996). This function is simple enough to be represented by a low order state-space model. This model assumes a third order impulsional response to stimulation, and has the following transfer function :

$$HRF(p) = \frac{\tau^3 e^{-\delta p}}{(p+\tau)^3} \quad (1)$$

As shown in Equation (1), the response shape depends on two parameters : δ represents the pure delay between stimulation and the start of HbO_2 increase ; τ influences the time-to-peak delay. Typical values that were chosen here were extracted from Boynton et al. (1996), and are $\delta = 2\,s$ and $\tau = 1.5\,s$. This choice leads to a time-to-peak delay from pulse stimulation of around $5\,s$ (Handwerker et al., 2004). Then, the Kalman filter principle requires the addition to the model of a state noise w (defined as the amount of noise affecting the model, i.e., the amount of errors in it) and of a measurement noise v (defined as the amount of noise affecting the measures). As shown on **Figure 1**, we chose to represent the state noise as a perturbation affecting the stimulus (i.e., the input of the model). This choice led us to consider that

state noise represents a stimulus perception (or internalization) bias.

The perception bias perturbing the stimulus is noted b. In the nominal model, Kalman filter assumptions impose that $b = w$, where w is the state noise following a gaussian centered distribution. This model, in addition to the choice of a Q/R value (where Q is the variance of the state noise, and R is the variance of the measurement noise), allowed us to design a *Kalman filter* for fNIRS signal improvement. The inputs of the Kalman filter were the stimuli onsets and the fNIRS raw signal.

One of the main limitations of this approach is the fact that the stimulus perception bias has to be centered (i.e., $b = 0$ on average), which can be erroneous when the subject sustainably disengages from the task and doesn't pay attention to the stimuli. To take this element into account, we built a second model, which is an augmentation of the nominal model, and in which $\dot{b} = w$. Thus, as the first derivative of b follows the gaussian centered distribution, it is still possible to design a Kalman filter, without assuming that b is null on average. This augmented model, along with the value of Q/R, allowed the computation of the *augmented Kalman filter* for fNIRS signal processing. For both filters, the value of the ratio Q/R was fixed according to an optimization process (see next section).

3. FIRST STEP : SETTING THE FILTER PARAMETERS

3.1. Material and Methods

Nine healthy participants from the Institut Superieur de l'Aeronautique et de l'Espace (ISAE ; Mean age = 21.6; $SD = 1.5$; eight males, eight right handed) participated in the experiment. The volunteers performed a computer-based digit sequence memorization task, while fNIRS measurements of the prefrontal cortex were recorded. Data were recorded using a Biopac® fNIR100 device, composed of 16 optodes placed on the forehead (see **Figure 2**). Each optode of the device records hemodynamics at a frequency of 2 Hz in term of oxygenated hemoglobin (HbO_2) and deoxygenated hemoglobin (HHb) level variations in comparison to a baseline.

Each trial of the experiment consisted in the memorization of a sequence of 5, 7, or 9 randomly chosen digits. The size of the sequence defined a level of difficulty. **Figure 3** summarizes the time sequence of a trial. During each trial, the subjects were asked to look at a fixation cross at the center of the screen. The digit sequence was presented through the loudspeakers of the computer using prerecorded audio tracks, at a rate of one digit per second. After the presentation of the last digit, the fixation cross was replaced by three crosses, indicating that the subjects had 8 *s* to type the memorized sequence on the keyboard. Between two consecutive trials, the subjects looked passively at the fixation cross at the center of the screen for 6 to 9 *s* (the inter-trial interval was chosen randomly to avoid task periodicity). The experiment consisted of 27 trials (nine trials for each of the three levels of difficulty), presented in a randomized order.

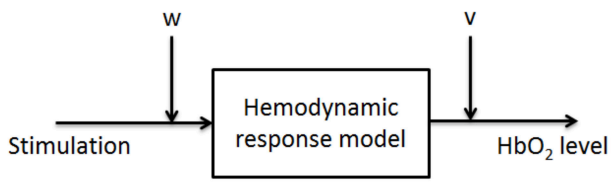


FIGURE 1 | Proposed approach for Kalman filtering of fNIRS. The state noise is considered as a perturbation of the input of the system.

3.2. Data Processing

Data were processed using Matlab®. Two different types of Kalman filters were applied to the data, the *nominal Kalman filter* (in which we assumed that the stimulus perception bias is null on average) and the *augmented Kalman filter* (without this assumption). The inputs for both filters were the stimuli onsets and the raw fNIRS data. For each filter, the value of the Q/R ratio chosen for the Kalman filter tuning ranged from 10^{-5} to 10^5, in order to look for the optimal results. Simultaneously, we also applied the MACD filter (Durantin et al., 2014b) to raw data in order to compare Kalman results with classical filtering.

For each trial, we computed the HbO_2 peak response (noted ΔHbO_2), i.e., the difference between the maximum value of HbO_2 in the 30 s following the trial onset and the value of HbO_2 at onset time. We similarly computed the HHb peak response (noted ΔHHb).

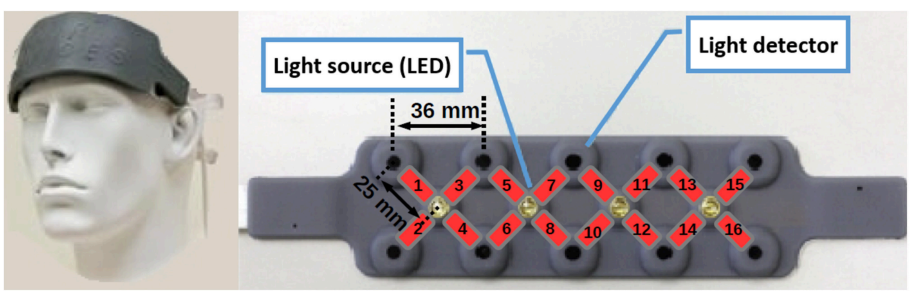


FIGURE 2 | fNIRS device optodes location. The device is composed of four light sources and 10 light detectors. The association of one light source and one light detector composes the optodes. The disposition of the sources and detectors leads to 16 optodes over the prefrontal cortex.

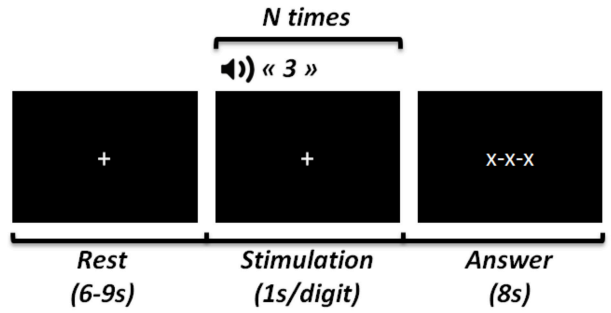


FIGURE 3 | Time course of one trial of the experiment. The experiment consisted of a total of 27 trials (three for each level of difficulty), presented in a random order. *N* = digit sequence size (5, 7, or 9).

Preliminary, the potential good values for Q/R (i.e., those leading to improvement in the signal) were isolated by computing an *Effect Size Index* (ESI, illustrated on **Figure 4**. For each filter (MACD, nominal Kalman or augmented Kalman with a given Q/R value) and each digit sequence size N, we computed the mean (μ_N) and standard deviation (σ_N) of the level of ΔHbO_2 or ΔHHb measured at each optode. We used these values to compute a confidence interval corresponding to one standard deviation as $[\mu_N - \sigma_N; \mu_N + \sigma_N]$. The *ESI* was defined as the gap between the confidence intervals of each condition (negative if the confidence intervals are overlapping), i.e.,

$$ESI = ((\mu_7 - \sigma_7) - (\mu_5 + \sigma_5)) + ((\mu_9 - \sigma_9) - (\mu_7 + \sigma_7))$$

We then proceeded to visual inspection to find the best values for Q/R ratio, by finding the parameters leading to higher ESI values. Each set of data was finally tested using a Two-way analysis of variance ANOVA, with two factors (16 optodes, three levels of difficulty), performed using STATISTICA® software. The strength of the statistical effect of the difficulty level, evaluated using the partial η^2, was used to compare the results of the different filters.

3.3. Results

As shown on **Figure 5**, the optimal results were obtained for HbO_2 at optode 2 recording mainly from the left inferior frontal gyrus, when using the nominal Kalman filter with $Q/R = 3.98$ or the augmented Kalman filter with $Q/R = 0.50$. **Table 1** summarizes the effect sizes obtained for each of the signal processing techniques tested (those effects showed an increase in the level of ΔHbO_2 with growing sequence sizes).

The frequency and phase responses (Bode diagram) of the nominal Kalman filter ($Q/R = 3.98$) and of the augmented Kalman filter ($Q/R = 0.50$) are given on **Figure 6**. As the two filters exhibit similar Bode diagrams (and therefore similar filtering properties), we retained only the augmented Kalman filter for testing on new data.

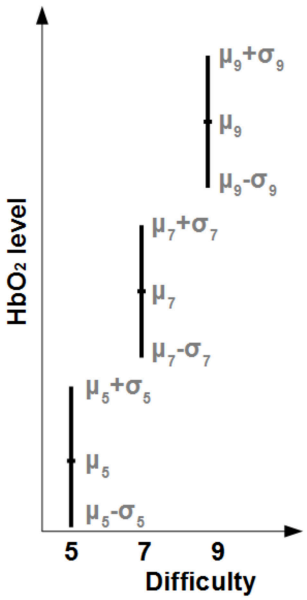


FIGURE 4 | Illustration of the computation of the Effect Size Index (ESI) for HbO_2. We computed the mean (μ) and standard deviation (σ) of the level of HbO_2 across subjects for each difficulty. For a difficulty *N*, a confidence interval corresponding to one standard deviation was computed as $[\mu_N - \sigma_N; \mu_N + \sigma_N]$. The corresponding ESI was computed as the sum of the gaps between the confidence intervals (negative if the confidence intervals are overlapping).

TABLE 1 | Effect sizes obtained for the effect of difficulty over all the subjects for the level of ΔHbO_2 measured at optode 2, depending on the type of filter used for signal processing.

Filter type	Q/R ratio	partial η^2
MACD		0.21
Nominal Kalman	3.98	0.32
Augmented Kalman	0.50	0.34

4. SECOND STEP : TESTING THE APPLICABILITY OF THE FILTER IN ECOLOGICAL CONDITIONS

4.1. Material and Methods

Data used for testing the Kalman filter were extracted from a second experiment involving a digit sequence memorization in a realistic flight simulator (see **Figure 7** for an illustration of the setup). The experiment was similar to Gateau et al. (2015), and included 18 healthy subjects (Mean age = 27.1; *SD* = 6.4; six women). Pilots heard prerecorded Air Traffic Controller (ATC) messages and were asked to dial the corresponding flight parameters in the Flight Control Unit (FCU) using the four knobs (i.e., speed, heading, altitude, and vertical speed knobs) of the FCU. The ATC messages were delivered at 78 dB SPL trough a Sennheiser® headset. We defined two levels of difficulty depending on the complexity of the message:

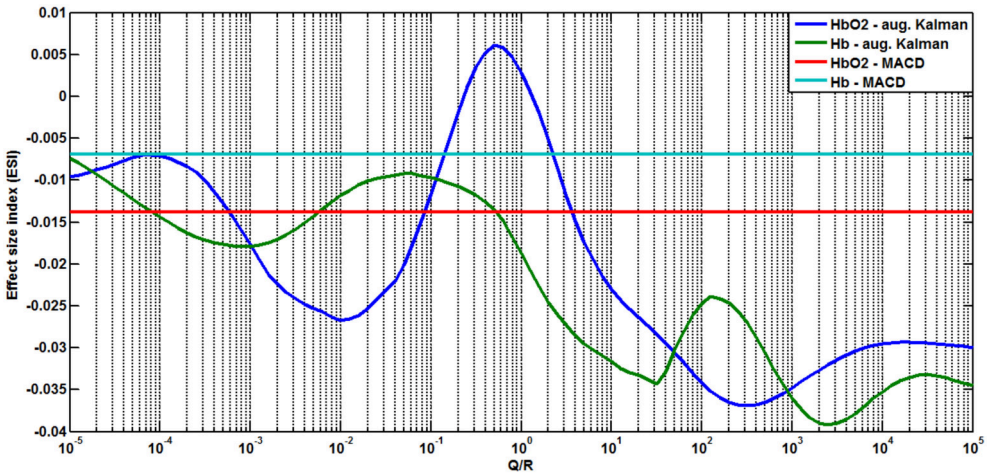


FIGURE 5 | Estimated Effect Size Index (ESI) for the augmented Kalman filter (for *HHb* –in green– and HbO_2 –in dark blue–) in function of the value of Q/R, compared to the ESI of the MACD filter (in red for HbO_2, in light blue for *HHb*). The index is estimated using the data from optode 2.

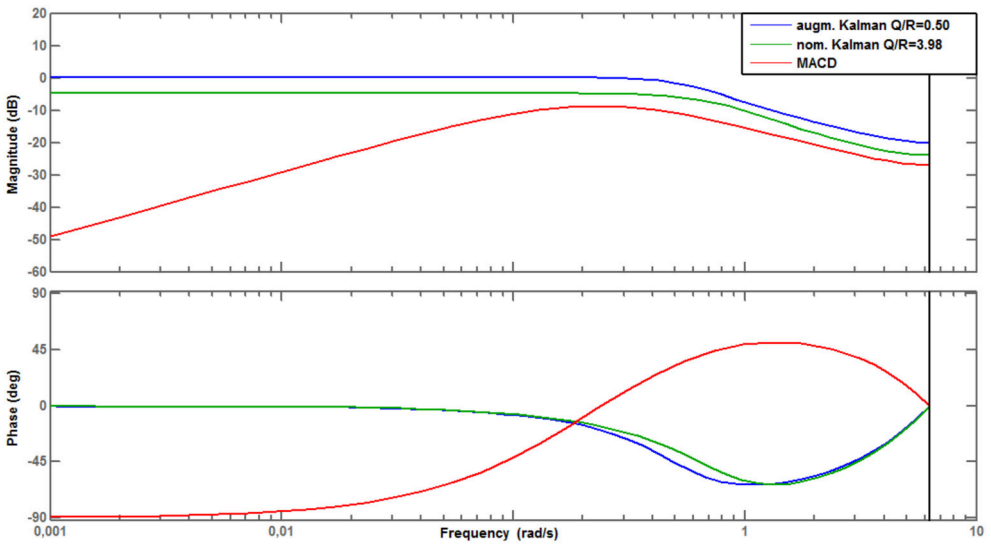


FIGURE 6 | Bode diagrams (frequency response, phase response) of the MACD filter, the nominal, and augmented Kalman filters.

- Low Load trial difficulty: only one major value per trial was used to set each flight parameter (e.g., 15 for "speed 150, heading 150, altitude 1500, vertical speed +1500").
- High Load trial difficulty: each flight parameter value was different from the previous one and composed of different digits to maximize the complexity (e.g., "speed 164, heading 235, altitude 8700, vertical speed −1600").

The task consisted in 20 repetitions of each difficulty for a total of 40 trials. Each ATC message started with the airplane call sign (i.e., "Supaero 32"), followed by the sequence of flight parameters. It ended with the message "over." The subjects were instructed to set the parameters strictly only after they heard the "over" message. A practice session was conducted for each subject before the actual experiments to allow them to become familiar with the experiment and the interface. After each message, the pilots had 18 seconds to enter the flight parameters. Trials were separated by 11 to 13 s of rest. During the experiment, hemodynamics of the prefrontal cortex were recorded using the same device than in the first experiment.

4.2. Data Processing and Classification

The raw HbO_2 data measured at each optode were filtered using three types of filter. First, we used the MACD filter and the augmented Kalman filter retained from the optimization phase ($Q/R = 0.50$). We also used a classical IIR Butterworth bandpass filter ($0.02\ \text{Hz} < f < 0.1\ \text{Hz}$), in order to compare the results to classical filtering. The statistical effect sizes of the level of ΔHbO_2 (computed in the same way than in the first experiment)

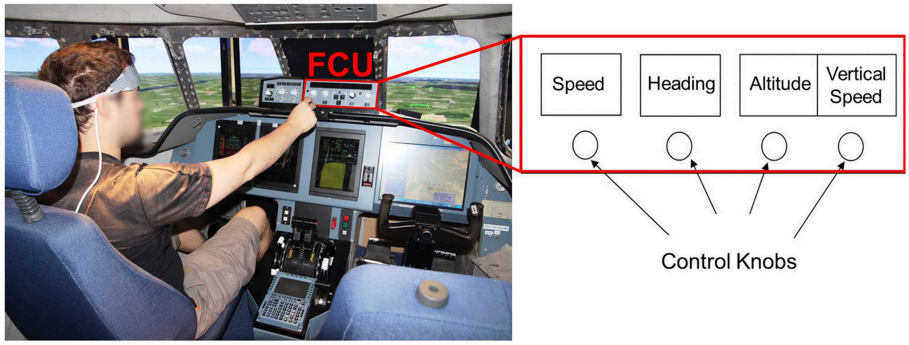


FIGURE 7 | Pilot's interaction with the auto flight system. The participants controlled the aircraft simulator of ISAE from the left seat. The red rectangle corresponds to the FCU (Flight Control Unit) dedicated to set the autopilot using the four control knobs, accordingly to ATC (Air Traffic Control) clearances (speed, heading, altitude, and vertical speed selection). Adapted from Gateau et al. (2015).

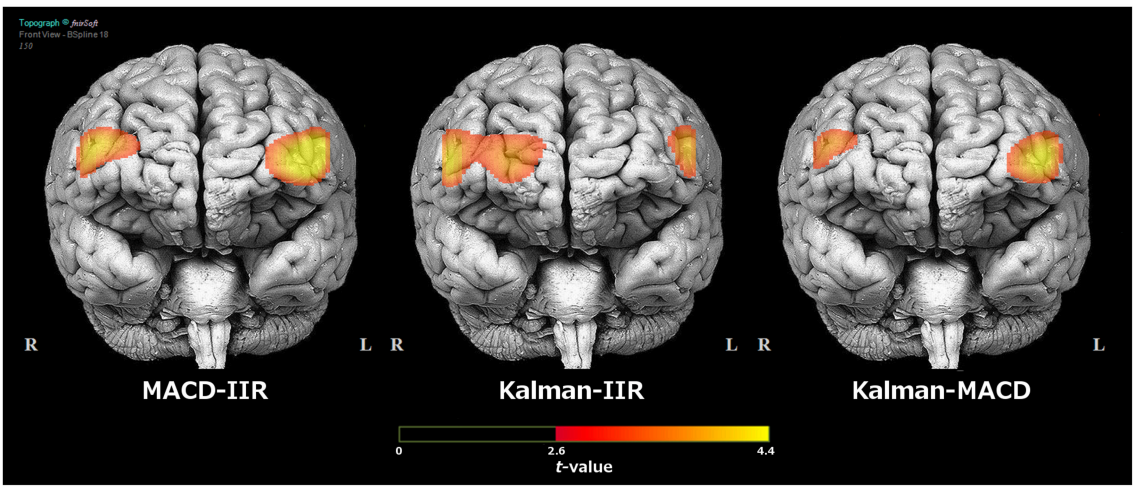


FIGURE 8 | Comparison of the *t*-maps for the contrast *High load - Low load* on the level of HbO_2 over the prefrontal cortex obtained in function of the type of filter used for signal processing (classical IIR, MACD, or Kalman). The topographical view was extracted from fNIRSoft® and the threshold was fixed at the statistical significance level with $\alpha = 0.01$, to account for multiple comparisons.

were evaluated using repeated measures ANOVA performed with STATISTICA®. The performance of the different filters were compared in terms of partial η^2. In addition, we computed the statistical *t*-maps representing the differences in the contrasts between high and low load conditions in terms of level of HbO_2 for each type of filters. This computation was done using Matlab and plotted using the topograph tool from fNIRSoft®.

The improvement of the signal depending on the type of filter used for processing was also evaluated by performing formal classification on the data. This analysis was performed using the Statistics and Machine Learning toolbox from Matlab. The ΔHbO_2 values extracted from each optode were used to train and test a Linear Support Vector Machine (SVM) classifier through a 10-fold cross validation process : for each subject, data from all trials were randomly divided in 10. The difficulty (high or low load) of the trials of each 10% of data was predicted by a SVM classifier that was previously trained on the 90% remaining data. The predicted labels were then examined to evaluate the *Accuracy* (probability of good classification), *Sensitivity* (probability of good classification for high load trials), and *Specificity* (probability of good classification for low load trials) of the classifier.

4.3. Results

The partial η^2 obtained for each type of filters are given in **Table 2** (for each optode and across all the optodes). The results show that the use of MACD elicits a better statistical effect size than the classical IIR filter. Similarly, the use of Kalman filter yields better results than both MACD and IIR filters. This result is true not only when filtering data from optode 2, but present notably at all optodes located in the bilateral dorsolateral areas of the prefrontal cortex (optodes 1, 2, 3, 4 and 13, 14 ,15, 16).

The effect of trial difficulty on the level of HbO_2 measured over the prefrontal cortex is shown on **Figure 8**. On this figure, we observe that both the MACD and Kalman filter over classical IIR

TABLE 2 | Effect sizes (partial η^2) obtained for the effect of difficulty over all the subjects for the level of HbO_2 measured for each optodes (plus main effect size over all optodes), depending on the type of filter used for signal processing.

Optode number	1	2	3	4	5	6	7	8	
IIR	0.34	0.36	**0.48**	0.35	0.15	0.32	0.32	0.00	
MACD	**0.51**	**0.44**	**0.50**	**0.53**	0.24	0.26	0.15	0.01	
Augmented Kalman ($Q/R = 0.50$)	**0.64**	**0.55**	**0.57**	**0.55**	0.38	0.36	0.39	0.15	
Optode number	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**All**
IIR	0.29	0.26	0.32	0.11	**0.41**	0.34	0.33	0.30	0.36
MACD	0.15	0.01	0.31	0.12	**0.50**	0.25	0.39	0.38	0.33
Augmented Kalman ($Q/R = 0.50$)	**0.42**	0.35	**0.53**	0.39	**0.55**	**0.55**	**0.52**	**0.49**	**0.52**

The effect sizes corresponding to a significant effect ($p < 0.05$) after correction for multiple comparisons are reported in bold font.

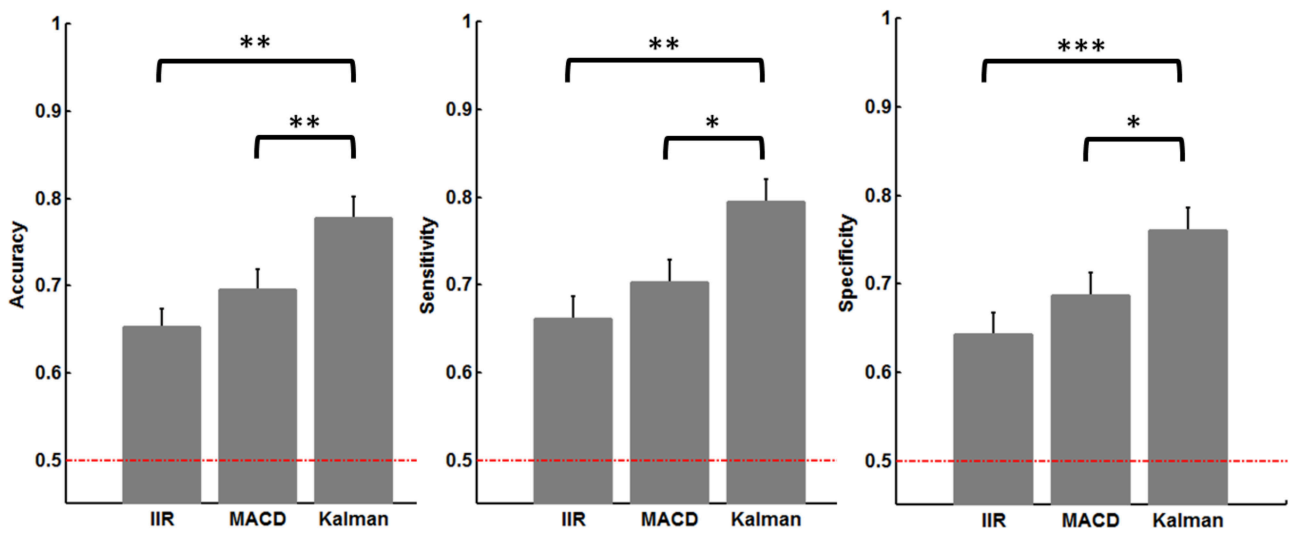


FIGURE 9 | Accuracy, sensitivity, and specificity obtained from the 10-fold cross-validation procedure for each type of filter. The dashed red line represents chance level (50%). The error bars represent the standard error of the mean across subjects, and statistical significance after correction for multiple comparisons is indicated by stars (*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$).

filter improve the discriminability between the two conditions in the lateral areas of the prefrontal cortex.

Ultimately, the cross-validation procedure performed on the data to classify low-load vs. high-load trials are presented in **Figure 9** in terms of accuracy, sensitivity, and specificity. The classification results were all significantly better than chance, although Kalman filter led to statistically better results than IIR and MACD filters. Using Kalman filtering, the classification accuracy reached 77.8%, with a sensitivity of 79.4% and a specificity of 76%.

5. DISCUSSION

The objective of the study was to design a Kalman filter to improve fNIRS signal for Neuroergonomics applications. In particular, the main challenge concerned the tuning of the parameters Q and R (Diamond et al., 2005), representing the state noise and measurement noise variances. Based on a simple model of the hemodynamic response to neuronal stimulation (Boynton et al., 1996), we designed a Kalman filter model taking into account both the measurement noise and the stimulus perception bias that can occur in periods of disengagement or when the level of attention varies. During an optimization process, we showed that it was possible to find values for the parameters which leads to better statistical results ($Q/R = 0.50$) with an augmented model. Interestingly, the relatively low value of the Q/R ratio in the second model suggests that the Kalman filter put more confidence in the dynamical model of hemodynamics response than in fNIRS data. The higher optimal value obtained for this ratio when using the first choice of model ($Q/R = 3.98$) suggests this model was less consistent with the actual hemodynamics characteristics.

We applied the optimal results found in the first experiment on new data from an ecological experiment in a flight simulator, and showed that the optimal Kalman filter tuning could be applied generically. This filter led to higher effect sizes when

looking at the effect of task difficulty in both tasks, compared to classical filters (see **Figure 8**). It is argued that the use of a dynamical physiological model by the Kalman filter implies less variability across trials and subjects, therefore explaining the greater stability of the results obtained with this filter. These results suggest that this filter would be suitable to improve the discriminability between the two conditions toward the implementation of a BCI to assist the operator, and would support the use of Kalman filtering to improve fNIRS signal (Izzetoglu et al., 2010). In particular, the Kalman filter helped us perform better during the SVM-based classification procedure between low-load and high-load trials, which confirms its contribution to the improvement of the signal. In addition, the experiment also confirmed that the MACD filter brings good results compared to classical IIR filtering, as it was previously demonstrated (Durantin et al., 2014b; Gateau et al., 2015). Although the discriminability obtained with this filter is not as good as the one obtained with the Kalman filter, it presents the advantage of not requiring any information on the stimulus onsets.

Interestingly, the optimal results for the first experiment were found at optode 2 recording mainly from the left inferior frontal gyrus. More generally, when applying the optimal Kalman filter in the second experiment, the WM solicitation elicited an activation of bilateral areas in the inferior and middle frontal gyri, part of the dorsolateral prefrontal cortex (see **Figure 8** and **Table 2**). This result is in agreement with previous fNIRS studies that have found these regions are sensitive to WM solicitation (Ayaz et al., 2012; Durantin et al., 2014a). Therefore, the improvement of the fNIRS signal collected in this region suggests that this filter could be applied to any experiment recruiting the same functional areas. In particular, the optimization process carried in this study would avoid the need of a calibration phase or of a convergence phase (in case of adaptive filtering) to improve signal quality. However, further investigation is still needed to assess whether this filter could be used with the same model and tuning in experiments recruiting different brain areas. Similarly, further investigation is also needed to assess the usability of these filters in ecological conditions that would differ from a simulated flight (e.g., with higher levels of light variations or motion artifacts).

Nevertheless, some modifications of the model could lead to better usability and performance of the Kalman filter. For instance, the use of a stimulus onset detection technique such as the detection technique based on the MACD filter (Durantin et al., 2014b; Gateau et al., 2015) could replace the stimulus onsets input of the Kalman filter, therefore reducing the complexity of the filter. In addition, it would be interesting to compare the results of the current Kalman model relying on a simple modeling of the hemodynamic response to more complex physiological models (e.g., Buxton et al., 2004). Finally, using an adaptive Q/R gain or realizing an optimization process for each subject instead of using a generic filter could also yield better results, although it would add complexity and a calibration phase to the procedure.

Altogether, the promising results of the study stand in favor of the use of Kalman filtering as a signal improvement technique for fNIRS signals with applications in Neuroergonomics. In particular, the improved signal would be available in real-time and without a calibration phase, and would allow better classification of WM levels in ecological settings.

AUTHOR CONTRIBUTIONS

All the authors contributed to the experiment design, results discussion, and paper redaction. Data collection was made by GD, TG, and SS. Signal processing tools (Kalman filter, MACD), were developed by GD.

ACKNOWLEDGMENTS

This research was supported by the Midi-Pyrénées region and Pôle de Recherche et d'Enseignement Supérieur (Neurocockpit project), and the AXA research fund (Neuroergonomics for flight safety). The work was approved by the Inserm Committee of Ethics Evaluation (Comité d'Évaluation Éthique de l'Inserm - CEEI/IRB00003888).

REFERENCES

Abdelnour, A. F., and Huppert, T. (2009). Real-time imaging of human brain function by near-infrared spectroscopy using an adaptive general linear model. *Neuroimage* 46, 133–143. doi: 10.1016/j.neuroimage.2009.01.033

Ayaz, H., Shewokis, P. A., Bunce, S., Izzetoglu, K., Willems, B., and Onaral, B. (2012). Optical brain monitoring for operator training and mental workload assessment. *Neuroimage* 59, 36–47. doi: 10.1016/j.neuroimage.2011.06.023

Bashashati, A., Fatourechi, M., Ward, R. K., and Birch, G. E. (2007). A survey of signal processing algorithms in brain–computer interfaces based on electrical brain signals. *J. Neural Eng.* 4, R32. doi: 10.1088/1741-2560/4/2/R03

Blankertz, B., Tangermann, M., Vidaurre, C., Fazli, S., Sannelli, C., Haufe, S., et al. (2010). The berlin brain–computer interface: non-medical uses of bci technology. *Front. Neurosci.* 4:198. doi: 10.3389/fnins.2010.00198

Boynton, G. M., Engel, S. A., Glover, G. H., and Heeger, D. J. (1996). Linear systems analysis of functional magnetic resonance imaging in human v1. *J. Neurosci.* 16, 4207–4221.

Brouwer, A.-M., van Erp, J., Heylen, D., Jensen, O., and Poel, M. (2013). "Effortless passive BCIs for healthy users," in *Universal Access in Human-Computer Interaction. Design Methods, Tools, and Interaction Techniques for eInclusion*, eds C. Stephanidis and M. Antona (Berlin: Springer), 615–622.

Buxton, R. B., Uludağ, K., Dubowitz, D. J., and Liu, T. T. (2004). Modeling the hemodynamic response to brain activation. *Neuroimage* 23, S220–S233. doi: 10.1016/j.neuroimage.2004.07.013

Callan, D. E., Durantin, G., and Terzibas, C. (2015). Classification of single-trial auditory events using dry-wireless eeg during real and motion simulated flight. *Front. Syst. Neurosci.* 9:11. doi: 10.3389/fnsys.2015.00011

Causse, M., Dehais, F., and Pastor, J. (2011). Executive functions and pilot characteristics predict flight simulator performance in general aviation pilots. *Int. J. Aviat. Psychol.* 21, 217–234. doi: 10.1080/10508414.2011.582441

Coyle, S., Ward, T., Markham, C., and McDarby, G. (2004). On the suitability of near-infrared (nir) systems for next-generation brain–computer interfaces. *Physiol. Meas.* 25, 815. doi: 10.1088/0967-3334/25/4/003

Cutrell, E., and Tan, D. (2008). "Bci for passive input in hci," in *Proceedings of CHI*, Vol. 8 (Florence: Citeseer), 1–3.

Derosière, G., Dalhoumi, S., Perrey, S., Dray, G., and Ward, T. (2014). Towards a near infrared spectroscopy-based estimation of operator attentional state. *PLoS ONE* 9:e92045. doi: 10.1371/journal.pone.0092045

Diamond, S. G., Huppert, T. J., Kolehmainen, V., Franceschini, M. A., Kaipio, J. P., Arridge, S. R., et al. (2005). "Physiological system identification with the Kalman filter in diffuse optical tomography," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2005*, eds J. S. Duncan and G. Gerig (Charlestown, MA: Springer), 649–656.

Durantin, G., Gagnon, J.-F., Tremblay, S., and Dehais, F. (2014a). Using near infrared spectroscopy and heart rate variability to detect mental overload. *Behav. Brain Res.* 259, 16–23. doi: 10.1016/j.bbr.2013.10.042

Durantin, G., Scannella, S., Gateau, T., Delorme, A., and Dehais, F. (2014b). "Moving average convergence divergence filter preprocessing for real-time event-related peak activity onset detection: application to fnirs signals," in *Engineering in Medicine and Biology Society, 2014. EMBS 2014. 36th Annual International Conference of the IEEE* (Chicago, IL: IEEE).

Gateau, T., Durantin, G., Lancelot, F., Scannella, S., and Dehais, F. (2015). Real-time state estimation in a flight simulator using fnirs. *PLoS ONE* 10:e0121279. doi: 10.1371/journal.pone.0121279

Georgiadis, S. D., Ranta-aho, P. O., Tarvainen, M. P., and Karjalainen, P. A. (2005). Single-trial dynamical estimation of event-related potentials: a Kalman filter-based approach. *Biomed. Eng. IEEE Trans.* 52, 1397–1406. doi: 10.1109/TBME.2005.851506

Handwerker, D. A., Ollinger, J. M., and D'Esposito, M. (2004). Variation of bold hemodynamic responses across subjects and brain regions and their effects on statistical analyses. *Neuroimage* 21, 1639–1651. doi: 10.1016/j.neuroimage.2003.11.029

Izzetoglu, M., Chitrapu, P., Bunce, S., and Onaral, B. (2010). Motion artifact cancellation in nir spectroscopy using discrete Kalman filtering. *Biomed. Eng. Online* 9:16. doi: 10.1186/1475-925X-9-16

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *J. Fluids Eng.* 82, 35–45. doi: 10.1115/1.3662552

Naseer, N., and Hong, K.-S. (2015). fnirs-based brain-computer interfaces: a review. *Front. Hum. Neurosci.* 9:3. doi: 10.3389/fnhum.2015.00003

Strait, M., and Scheutz, M. (2014). What we can and cannot do with near infrared spectroscopy. *Front. Neurosci.* 8:117. doi: 10.3389/fnins.2014.00117

van Erp, J. B. F., Lotte, F., and Tangermann, M. (2012). Brain-computer interfaces: beyond medical applications. *Computer* 45, 26–34. doi: 10.1109/MC.2012.107

Villringer, A., and Obrig, H. (2002). "Near-infrared spectroscopy and imaging," in *Brain Mapping; The Methods*, eds A. W. Toga and J. C. Mazziotta (San Diego, CA: Academic Press), 141–158. doi: 10.1016/B978-012693019-1/50008-3

von Bünau, P., Meinecke, F. C., Király, F. C., and Müller, K.-R. (2009). Finding stationary subspaces in multivariate time series. *Phys. Rev. Lett.* 103:214101. doi: 10.1103/PhysRevLett.103.214101

Zander, T. O., and Kothe, C. (2011). Towards passive brain–computer interfaces: applying brain–computer interface technology to human–machine systems in general. *J. Neural Eng.* 8:025005. doi: 10.1088/1741-2560/8/2/025005

Zheng, Y., Martindale, J., Johnston, D., Jones, M., Berwick, J., and Mayhew, J. (2002). A model of the hemodynamic response and oxygen delivery to brain. *Neuroimage* 16, 617–637. doi: 10.1006/nimg.2002.1078

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Durantin, Scannella, Gateau, Delorme and Dehais. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Human Neuroscience

GENERAL COMMENTARY
published: 01 March 2016
doi: 10.3389/fnhum.2016.00070

Commentary: Cumulative effects of anodal and priming cathodal tDCS on pegboard test performance and motor cortical excitability

Pierre Besson[1], *Stephane Perrey*[1], *Wei-Peng Teo*[2] *and Makii Muthalib*[1]*

[1] EuroMov, University of Montpellier, Montpellier, France, [2] School of Exercise and Nutrition Sciences, Deakin University, Melbourne, VIC, Australia

Keywords: tDCS, motor performance, metaplasticity, neuroergonomics, priming effects

A commentary on

Cumulative effects of anodal and priming cathodal tDCS on pegboard test performance and motor cortical excitability
by Christova, M., Rafolt, D., and Gallasch, E. (2015). Behav. Brain Res. 287, 27–33. doi: 10.1016/j.bbr.2015.03.028

Consistent with a neuroergonomics approach, task performance can be facilitated by non-invasive neuromodulation techniques, such as anodal transcranial direct current stimulation-atDCS (Clark and Parasuraman, 2014; McKendrick et al., 2015). However, robust stimulation parameters and protocols need to be developed for applying atDCS to enhance motor performance in clinical and healthy populations. For instance, protocols using Online atDCS, where the motor task is performed during the stimulation, has greater facilitative effects on motor performance/learning than if the motor task is performed after the stimulation (i.e., Offline atDCS; Stagg and Nitsche, 2011). These greater facilitative effects of Online atDCS on motor performance/learning are likely due to enhanced synaptic efficacy in the simultaneously engaged neural network through a "gating" mechanism (Ziemann and Siebner, 2008). Overall, the interaction of the timing of tDCS application and motor task are crucial parameters to optimize atDCS effects on enhancing motor performance/learning.

The recent study of Christova et al. (2015) aimed to optimize Online atDCS effects on enhancing motor performance/learning by applying a novel cathodal tDCS (ctDCS) priming protocol that harnessed homeostatic metaplastic mechanisms. In the design of the study, healthy subjects were randomly distributed into three priming tDCS groups ($n = 12$) and were required to perform with their non-dominant left hand a grooved pegboard test (GPT) over four training blocks and a retest 2 weeks later. Three priming tDCS conditions were investigated on the right primary motor cortex (M1): (1) Sham: Sham ctDCS (15 min) 10 min before Sham Online atDCS (20 min); (2) Online atDCS: Sham ctDCS (15 min) 10 min before Online atDCS (1mA, 20 min); (3) ctDCS priming: ctDCS (1mA, 15 min) 10 min before Online atDCS (1mA, 20 min). Transcranial magnetic stimulation (TMS) parameters (motor evoked potential-MEP, intracortical facilitation-ICF, and short interval intracortical inhibition-SICI) were assessed before and up to 60 min after the tDCS conditions. The results indicated that although both Online atDCS conditions improved GPT performance (i.e., faster completion time) over Sham after the four training blocks, only the priming ctDCS/Online atDCS condition further enhanced GPT performance 2 weeks later. These latter findings were explained in relation to homeostatic metaplastic mechanisms based on the Bienenstock-Cooper-Munro (BCM) theory that postulates a "sliding threshold" for bidirectional

OPEN ACCESS

Edited by:
Stephen Fairclough,
Liverpool John Moores University, UK

Reviewed by:
Ute Kreplin,
Massey University, New Zealand

***Correspondence:**
Makii Muthalib
makii.muthalib@umontpellier.fr;
makii.muthalib@gmail.com

Received: *22 November 2015*
Accepted: *12 February 2016*
Published: *01 March 2016*

Citation:
Besson P, Perrey S, Teo W-P and Muthalib M (2016) Commentary: Cumulative effects of anodal and priming cathodal tDCS on pegboard test performance and motor cortical excitability. Front. Hum. Neurosci. 10:70. doi: 10.3389/fnhum.2016.00070

synaptic plasticity (Karabanov et al., 2015). Accordingly, priming with ctDCS, which reduced cortical excitability (reduced MEP amplitude and ICF) and increased cortical inhibition (increased SICI) after the ctDCS session, would have reduced post-synaptic activity in the activated neural network. Based on the BCM model, this ctDCS-induced reduction in post-synaptic activity would be expected to reduce the modification threshold for long term potentiation (LTP)-like plasticity during subsequent Online atDCS, and thus further enhanced GPT performance 2 weeks later. The prolonged increase in ICF and reduced SICI for at least 60 min afterwards provides some evidence for this homeostatic metaplastic effect enhancing offline learning of the GPT. However, the authors acknowledged that a limitation of the study design was that a priming ctDCS followed by Sham Online atDCS condition was not tested, which could have confirmed that the results of the priming ctDCS/Online atDCS condition were primarily due to homeostatic metaplastic mechanisms. Nevertheless, Christova et al.'s (2015) novel methodology and findings can be used to optimize tDCS priming protocols to modulate neuroplasticity and enhance motor performance/learning. The following sections will provide a commentary on ways to optimize the timing and polarity of tDCS applications, which could have significant implications for the original paper's conclusion.

An important tDCS parameter that requires further investigation is the influence of the time delay between priming and test tDCS application on homeostatic metaplasticity and its effects on motor performance/learning (Karabanov et al., 2015). A few studies have investigated the effects of altering the delay between repeated tDCS applications of the same polarity on cortical excitability (Fricke et al., 2011; Monte-Silva et al., 2013; Bastani and Jaberzadeh, 2014) and motor performance/learning (Bastani and Jaberzadeh, 2014). However, no clear evidence of the optimal delay time period could be ascertained from their respective priming tDCS protocols. Christova et al. (2015) considered a 10 min delay between ctDCS and Online atDCS to be sufficient to allow homeostatic metaplastic mechanisms to take hold. But it is still not known if a shorter or longer time delay between priming ctDCS and Online atDCS would differentially modulate homeostatic metaplasticity and motor performance/learning. We (Muthalib et al., 2016) have previously postulated a non-homeostatic approach of priming with atDCS immediately before Online atDCS to further facilitate the neuroplastic effects of Online atDCS. We reason that since sub-threshold neuronal membrane depolarization induced by atDCS has an intensity- and time-dependent effect to strengthen synaptic efficacy (Nitsche and Paulus, 2001), performing atDCS (2 mA, 10 min) immediately before Online atDCS would boost the already strengthened synaptic connections through a further "gating" mechanism induced with the concurrent motor task. We have recently shown that this priming atDCS/Online atDCS protocol on the left M1 can reduce bilateral M1 activation to perform a unilateral simple finger sequence task at the same tapping rate (Muthalib et al., 2016). These results could be explained by a non-homeostatic mechanism following the "gating" theory, such that the reduced motor task related bilateral M1 activation during the atDCS suggests a greater efficiency of neuronal transmission (i.e., less synaptic input for the same neuronal output) in the activated neuronal network. Whether this priming atDCS/Online atDCS protocol would enhance motor performance/learning greater than a priming ctDCS/Online atDCS or Online atDCS protocol still requires to be investigated.

Since the design of the Christova et al. (2015) study corresponded to a learning paradigm, it is difficult to differentiate the tDCS effect from the learning effect on improving online GPT performance during the four training blocks. In order to specifically test the tDCS effect, and minimize the effects of learning, on performance would have been to include a familiarization session to allow the GPT task to become "well learned" and performance stabilize at near maximal levels in all individuals prior to starting the tDCS interventions (Hummel et al., 2010). For highly skilled individuals (e.g., elite athletes, expert operators), it is extremely difficult to improve maximal performance levels since learning has reached relative "ceiling" levels. However, this "ceiling" performance can conceivably be modulated directly using neuromodulation protocols. For example, an excitatory TMS protocol to the dominant left M1, which lead to increased M1 excitability, was able to increase dominant right hand maximal finger tapping rate and reduce the decline of the movement rate over 10 s (Teo et al., 2012). In contrast, an inhibitory TMS protocol to the dominant left M1, which decreased M1 excitability, was shown to decrease maximal finger tapping rate of the dominant right hand (Jäncke et al., 2004). We therefore, consider that applying tDCS to the dominant left M1/right hand and utilizing a "well learned" stable motor task, such as a simple finger sequence task performed at maximum rate (Avanzino et al., 2008), may provide a sensitive means to investigate the tDCS effects on task performance.

In conclusion, priming tDCS protocols are promising ways to optimize tDCS facilitatory effects on motor performance/learning, which has relevance from a neuroergonomic standpoint. Thus, future studies are necessary to determine the optimal polarity and timing of tDCS applications to modulate neuroplasticity and enhance performance in clinical, sports, and real-world settings.

AUTHOR CONTRIBUTIONS

All authors listed, have made substantial, direct, and intellectual contribution to the work, and approved it for publication.

ACKNOWLEDGMENTS

We would like to thank Prof. John Rothwell for his helpful comments on the manuscript. MM was supported by a Labex NUMEV Fellowship (Digital and Hardware Solutions, Environmental and Organic Life Modeling, ANR-10-LABX-20). WT is supported by an Alfred Deakin Postdoctoral Fellowship.

REFERENCES

Avanzino, L., Bove, M., Trompetto, C., Tacchino, A., Ogliastro, C., and Abbruzzese, G. (2008). 1-Hz repetitive TMS over ipsilateral motor cortex influences the performance of sequential finger movements of different complexity. *Eur. J. Neurosci.* 27, 1285–1291. doi: 10.1111/j.1460-9568.2008.06086.x

Bastani, A., and Jaberzadeh, S. (2014). Within-session repeated a-tDCS: the effects of repetition rate and inter-stimulus interval on corticospinal excitability and motor performance. *Clin. Neurophysiol.* 125, 1809–1818. doi: 10.1016/j.clinph.2014.01.010

Christova, M., Rafolt, D., and Gallasch, E. (2015). Cumulative effects of anodal and priming cathodal tDCS on pegboard test performance and motor cortical excitability. *Behav. Brain Res.* 287, 27–33. doi: 10.1016/j.bbr.2015.03.028

Clark, V. P., and Parasuraman, R. (2014). Neuroenhancement: enhancing brain and mind in health and in disease. *Neuroimage* 85, 889–894. doi: 10.1016/j.neuroimage.2013.08.071

Fricke, K., Seeber, A. A., Thirugnanasambandam, N., Paulus, W., Nitsche, M. A., and Rothwell, J. C. (2011). Time course of the induction of homeostatic plasticity generated by repeated transcranial direct current stimulation of the human motor cortex. *J. Neurophysiol.* 105, 1141–1149. doi: 10.1152/jn.00608.2009

Hummel, F. C., Heise, K., Celnik, P., Floel, A., Gerloff, C., and Cohen, L. G. (2010). Facilitating skilled right hand motor function in older subjects by anodal polarization over the left primary motor cortex. *Neurobiol. Aging* 31, 2160–2168. doi: 10.1016/j.neurobiolaging.2008.12.008

Jäncke, L., Steinmetz, H., Benilow, S., and Ziemann, U. (2004). Slowing fastest finger movements of the dominant hand with low-frequency rTMS of the hand area of the primary motor cortex. *Exp. Brain Res.* 155, 196–203. doi: 10.1007/s00221-003-1719-7

Karabanov, A., Ziemann, U., Hamada, M., George, M. S., Quartarone, A., Classen, J., et al. (2015). Consensus paper: probing homeostatic plasticity of human cortex with non-invasive transcranial brain stimulation. *Brain Stimul.* 8, 442–454. doi: 10.1016/j.brs.2015.01.404

McKendrick, R., Parasuraman, R., and Ayaz, H. (2015). Wearable functional near infrared spectroscopy (fNIRS) and transcranial Direct Current Stimulation (tDCS): expanding vistas for neurocognitive augmentation. *Front. Syst. Neurosci.* 9:27. doi: 10.3389/fnsys.2015.00027

Monte-Silva, K., Kuo, M.-F., Hessenthaler, S., Fresnoza, S., Liebetanz, D., Paulus, W., et al. (2013). Induction of late LTP-like plasticity in the human motor cortex by repeated non-invasive brain stimulation. *Brain Stimul.* 6, 424–432. doi: 10.1016/j.brs.2012.04.011

Muthalib, M., Besson, P., Rothwell, J., Ward, T., and Perrey, S. (2016). Effects of anodal high-definition transcranial direct current stimulation on bilateral sensorimotor cortex activation during sequential finger movements: an fNIRS Study. *Adv. Exp. Med. Biol.* 876, 351–359. doi: 10.1007/978-1-4939-3023-4_44

Nitsche, M. A., and Paulus, W. (2001). Sustained excitability elevations induced by transcranial DC motor cortex stimulation in humans. *Neurology* 57, 1899–1901. doi: 10.1212/WNL.57.10.1899

Stagg, C. J., and Nitsche, M. A. (2011). Physiological basis of transcranial direct current stimulation. *Neuroscientist* 17, 37–53. doi: 10.1177/1073858410386614

Teo, W. P., Rodrigues, J. P., Mastaglia, F. L., and Thickbroom, G. W. (2012). Breakdown in central motor control can be attenuated by motor practice and neuro-modulation of the primary motor cortex. *Neuroscience* 220, 11–18. doi: 10.1016/j.neuroscience.2012.06.048

Ziemann, U., and Siebner, H. R. (2008). Modifying motor learning through gating and homeostatic metaplasticity. *Brain Stimul.* 1, 60–66. doi: 10.1016/j.brs.2007.08.003

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Besson, Perrey, Teo and Muthalib. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 07 March 2016
doi: 10.3389/fnhum.2016.00072

Simultaneous tDCS-fMRI Identifies Resting State Networks Correlated with Visual Search Enhancement

Daniel E. Callan[1,2]*, *Brian Falcone*[3], *Atsushi Wada*[1,2] and *Raja Parasuraman*[3]

[1] Center for Information and Neural Networks (CiNet), National Institute of Information and Communications Technology (NICT), Osaka University, Osaka, Japan, [2] Multisensory Cognition and Computation Laboratory, Universal Communication Research Institute, National Institute of Information and Communications Technology, Kyoto, Japan, [3] Center of Excellence in Neuroergonomics, Technology, and Cognition (CENTEC), George Mason University, Fairfax, VA, USA

This study uses simultaneous transcranial direct current stimulation (tDCS) and functional MRI (fMRI) to investigate tDCS modulation of resting state activity and connectivity that underlies enhancement in behavioral performance. The experiment consisted of three sessions within the fMRI scanner in which participants conducted a visual search task: Session 1: Pre-training (no performance feedback), Session 2: Training (performance feedback given), Session 3: Post-training (no performance feedback). Resting state activity was recorded during the last 5 min of each session. During the 2nd session one group of participants underwent 1 mA tDCS stimulation and another underwent sham stimulation over the right posterior parietal cortex. Resting state spontaneous activity, as measured by fractional amplitude of low frequency fluctuations (fALFF), for session 2 showed significant differences between the tDCS stim and sham groups in the precuneus. Resting state functional connectivity from the precuneus to the substantia nigra, a subcortical dopaminergic region, was found to correlate with future improvement in visual search task performance for the stim over the sham group during active stimulation in session 2. The after-effect of stimulation on resting state functional connectivity was measured following a post-training experimental session (session 3). The left cerebellum Lobule VIIa Crus I showed performance related enhancement in resting state functional connectivity for the tDCS stim over the sham group. The ability to determine the relationship that the relative strength of resting state functional connectivity for an individual undergoing tDCS has on future enhancement in behavioral performance has wide ranging implications for neuroergonomic as well as therapeutic, and rehabilitative applications.

Keywords: fMRI, tDCS, resting state, functional connectivity, visual search, neuroergonomics

OPEN ACCESS

Edited by:
Hasan Ayaz,
Drexel University, USA

Reviewed by:
Christophe Phillips,
University of Liége, Belgium
Filippo Brighina,
University of Palermo, Italy
Vincent Clark,
University of New Mexico, USA

***Correspondence:**
Daniel E. Callan
dcallan@nict.go.jp

Received: *07 November 2015*
Accepted: *12 February 2016*
Published: *07 March 2016*

Citation:
Callan DE, Falcone B, Wada A and Parasuraman R (2016) Simultaneous tDCS-fMRI Identifies Resting State Networks Correlated with Visual Search Enhancement. Front. Hum. Neurosci. 10:72. doi: 10.3389/fnhum.2016.00072

INTRODUCTION

In recent years there has been an explosion of research investigating a method by which to augment human cognition by passing a low amplitude direct current (typically in the range of 0.5–2 mA) through a person's head and enhancing human performance and abilities (Coffman et al., 2014). This technique is called transcranial direct current stimulation (tDCS). tDCS has been shown to enhance such abilities as attention and performance on vigilance, threat detection, and

visual search tasks (Falcone et al., 2012; Parasuraman and Galster, 2013; Nelson et al., 2014); to enhance learning and performance on perceptual and cognitive tasks (Clark et al., 2012; Parasuraman and McKinley, 2014); and to improve motor and cognitive function in patients with brain damage, neuropsychiatric, and neurological diseases (Flöel, 2014; Kuo et al., 2014; O'Shea et al., 2014). The underlying neurological processes that allow for these enhancements in ability are under ongoing investigation. It has been shown that anodal DC stimulation decreases neural firing thresholds, and that glutamatergic modulation of long-term potentiation/depression may be involved with the enduring effects of tDCS (Liebetanz et al., 2002; Nitsche et al., 2003; Bikson et al., 2004; Coffman et al., 2014; Hunter et al., 2015). While one may expect these effects to be localized on the cortex near the stimulating electrode, functional MRI (fMRI) studies have also shown modulation in activity in distal brain regions suggesting possible network effects induced by tDCS (Clemens et al., 2014; Ellison et al., 2014; Weber et al., 2014).

It is our goal in this study to use simultaneous tDCS and fMRI to investigate the relationship between modulation in resting state activity as well as resting state functional connectivity of the brain correlated with improved performance as a result of stimulation. Studies have shown that resting state activity and connectivity in the brain can predict various characteristics such as attention (Kelley et al., 2008), learning (Baldassarre et al., 2012), memory (Hampson et al., 2006), language processing (Koyama et al., 2011), personality (Adelstein et al., 2011), and IQ (van der Heuvel et al., 2009; for review, see Stevens and Spreng, 2014). Previous studies using tDCS and fMRI have revealed, that as a result of stimulation, resting state networks can show wide spread changes in activity and connectivity in cortical and subcortical brain regions (Saiote et al., 2013; Clemens et al., 2014).

In our study, we investigate both resting state activity and performance related resting state connectivity in response to tDCS. A visual search task was employed before (pre-training), during (training), and after (post-training) tDCS stimulation to determine its facilitative effects on performance. Resting state fMRI was recorded toward the end of each session after completing the visual search task. We placed the stimulating electrode over the posterior parietal cortex as it has been found in previous tDCS studies to modulate visual search performance (Bolognini et al., 2010; Ellison et al., 2014). We used the fractional amplitude of low frequency fluctuations (fALFF) in the BOLD signal, which has been found to be associated with spontaneous neural activity (Biswal et al., 1995; Zou et al., 2008; Song et al., 2011), as a measure of resting state activity. By comparing fALFF across tDCS stimulation and sham groups we intend to show brain regions in which the spontaneous neural activity is being modulated. Unlike most previous neuroimaging studies, we applied tDCS and fMRI concurrently in order to observe the active effects of tDCS on resting state activity rather than just the after-effects that exist following the cessation of tDCS. Brain regions determined to show tDCS induced activity are then used as seed regions for a functional connectivity analysis (Song et al., 2011). It is hypothesized that resting state functional connectivity related to improvement in behavioral performance on the visual search task will be found to exist for these seed regions for the tDCS group to a greater extent than for the sham group.

Our study addresses many of the future directions concerning the investigation of tDCS on resting state activity and connectivity proposed by Clemens et al. (2014). Specifically, we applied tDCS and fMRI concurrently to investigate the immediate active effects on resting state activity and connectivity. The after-effects of tDCS on resting state functional connectivity were also investigated following a post-training session. In addition, as proposed by Clemens et al. (2014), our study includes the use of sham stimulation. By comparing between tDCS stim and a sham group (unlike other studies that look at tDCS stim vs. pre-stim), our study is able to investigate behaviorally related enhancement in resting state functional connectivity that differs between the two groups that can be attributed to modulation by tDCS rather than changes in resting state connectivity that normally occur as a result of task training.

MATERIALS AND METHODS

Participants

There were 28 participants that took part in this study. All of the participants (14 males, 14 females) were Japanese right-handed adults ranging from 18 to 25 years (mean = 20.7) of age from Osaka University. The participants were pseudo-randomly assigned to the tDCS stim and sham groups such that there were seven females and seven males in each group. All participants were screened for exclusion if there was a history of head injury, history of mental, neurological, alcohol or drug abuse disorders, or using medication that affects central nervous system function. The participants gave written and informed consent to take part in this experiment. The experimental procedures were approved by the National Institute of Information and Communications Technology (NICT) Human Subject Review Committee were carried out in accordance with the principles expressed in the WMA Declaration of Helsinki. Originally there were 18 tDCS stim and 17 sham participants. From the tDCS stim group two participants were excluded because of pressure pain caused by the tight fit of the headphones within the head coil and two participants were excluded because task performance was below chance on session three even after completing the training session. From the sham group one participants was excluded because of pressure pain caused by the tight fit of the headphones within the head coil and two participants were excluded because task performance was below chance on session three even after completing the training session.

Procedure

The experiment consisted of three sessions within the fMRI scanner. During the first part of scanning the participants conducted a visual search task. During the last 4.5 min of fMRI scanning, for each session, resting state activity was acquired. In this study, we will focus only on the resting state fMRI data from these sessions.

The visual search task was based on a search and rescue mission that required participants to locate a red pickup truck located in the search area amongst buildings and other similar looking distractor vehicles. In each trial there were five non-moving vehicles distributed throughout the search area, one of which could be the red truck. There were a total of 60 trials in each session and the target red truck was randomly present on half of the trials. The task was designed so that as the unmanned aerial vehicle UAV loitered in a circle around the search area, all vehicles would remain in constant view despite a continually changing view angle. Each trial lasted 10 s where the participants searched the area looking for the target and were required to make a button press indicating whether the search area contained a target or not.

The three experimental sessions consisted of the following: Session 1: Pre-training session that did not provide performance feedback. Session 2: Training session in which tDCS stimulation or sham stimulation was delivered. In the training session immediate reinforcement error-feedback ("ding" sound correct, "buzz" sound incorrect) after each response. Additionally, for target present trials only, a transparent white sphere would appear over the target at the end of the 10 s trial identifying the target location. This type of auditory reinforcement feedback will allow subjects to know immediately whether the features they were attending to are incorrect, in the case of a false alarm or a miss, or correct, in the case of a correct rejection or a hit. This information together with the visual feedback of the position of the target at the end of the trial when it was present will allow for learning of the relevant features and improve performance. Session 3: Post-training session with no feedback. The total time of the visual search task for session 1 and 3 was approximately 15 min and session 2 was approximately 16 min. After each experimental session resting state activity was recorded for 4.5 min. For session 2 tDCS stimulation was given concurrently with fMRI scanning. The task for the participants during collection of the resting state data was to visually fixate on a white cross mark presented in the center of the display against a black background. Participants were instructed to fixate on the cross on the screen, and to relax without falling asleep.

Transcranial Direct Current Stimulation

TDCS was delivered during the training session (session 2) using the MRI compatible NeuroConn DC-Stimulator MR. Two rectangular-shaped (5.3 × 7.2 cm) MRI compatible conductive rubber electrodes were placed on the participant before entering the MRI scanner (see **Figure 1** for picture of placement of electrode on the head of a participant and a rendered MRI showing the tDCS electrode on the head). The anodal electrode was placed over the right posterior parietal cortex. It was placed over where the P4 electrode is located according to the 10–20 International EEG System. The electrode was held in place by the conductive paste (Ten20 conductive paste gel, Waver and Company) as well as a padded headband. The cathodal electrode was placed over the contralateral left side trapezius muscle on the back of the shoulder.

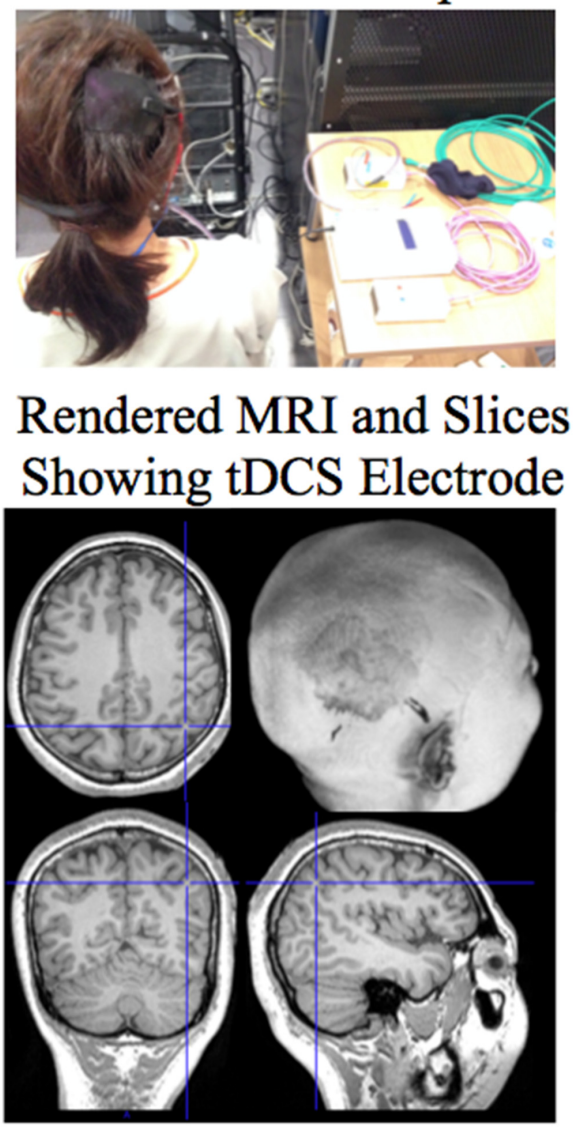


FIGURE 1 | Top: Picture showing the placement of the anodal tDCS electrode on the right posterior parietal cortex of the participant. **Bottom**: The placement of the tDCS electrode can be seen in the rendered MRI of the participant. Sections are shown through the brain at the site of the electrode. For the MRI sections the right side of the image is the right side of the brain.

Participants in the stim group received 1 mA current for a total of 30 min (1 mA was the highest level possible within the MRI scanner with our version of the NeuroConn DC-Stimulator MR). Stimulation was started 5 min before the task in order to ensure that the full modulatory effect of tDCS was active during task performance. Current was ramped up over the initial 10 s and ramped down the last 10 s of

stimulation. The participants in the sham group also received 1 mA current but only for 30 s and then the unit was turned off. This procedure helps to conceal from the participant which group (stim or sham) they belong to as both groups feel the onset of the stimulation. In addition group membership of the participant was not known by the experimenter giving the instructions.

fMRI Data Collection and Analysis

fMRI scanning of resting state activity was acquired for 4.5 min at the end of each session (TR = 2 s; 30 interleaved slices covering the brain and cerebellum, 3 mm × 3 mm × 4 mm voxels; Siemens 3T Trio Scanner; 32 Channel head coil). Preprocessing of fMRI data was conducted using SPM8 (Wellcome Department of Cognitive Neurology, UCL) and included realignment and unwarping, normalization to the template EPI image (2 mm × 2 mm × 2 mm), and smoothing (8 mm × 8 mm × 8 mm). EPI template based normalization was used because the source image upon which the normalization parameters are determined is in the same space as the scans to be normalized. This has advantages in that it avoids additional steps of coregistration to/or from EPI space (resulting in image distortion) that are required when using an anatomical T1 or T2 image to determine the normalization parameters. Because whole brain EPI was acquired the difficulties associated with mapping partial EPI volumes to the template image are avoided.

The REST (Song et al., 2011) Toolkit was used to conduct the resting state spontaneous activity (fALFF) and the functional connectivity analyses. The realignment parameters were used as covariates of non interest and regressed out of the preprocessed EPI data to extract potential confounds related to head movement while scanning. The linear trend was then removed from the data. The parameters for the fALFF analysis included a low frequency fluctuation band of 0.01–0.08 Hz (Biswal et al., 1995) compared to the entire frequency range (0–0.25 Hz). The fALFF results were normalized by dividing by the mean fALFF values within the whole brain mask to be used for second level random effects analyses. The functional connectivity analysis was carried out over the preprocessed covariates removed detrended and filtered (0.01–0.08 Hz) data. Three separate functional connectivity analyses were conducted. The seed regions of interest (ROI) were determined from the results of the fALFF analysis (see "Results" Section). The ROI included the precuneus (MNI 6, −46, 60). A Spherical region with a radius of 8 mm at the given coordinate was used as the seed for the resting state functional connectivity analyses conducted separately for each session. The Pearson linear correlation was used to determine the functional connectivity between the mean of the voxels within the seed ROI and the rest of the voxels in the brain according to the defaults in the REST toolbox (Song et al., 2011). The Fisher's z transform was used to normalize the correlation coefficients to be used for second level random effects analyses. SPM8 was used to conduct the random effects analyses. Correction for multiple comparisons ($p < 0.05$) across the entire brain was carried out using Monte-Carlo simulation of the brain volume to define a voxel contiguity threshold at an uncorrected significance level of $p < 0.005$ (Slotnick et al., 2003; Ellison et al., 2014). Using 10000 Monte-Carlo simulations a cluster extent greater than 154 voxels thresholded at $p < 0.005$ uncorrected, is necessary to correct for multiple comparisons across the whole brain at a threshold $p < 0.05$. Activated brain regions were identified using the SPM Anatomy Toolbox v1.8 (Eichkoff et al., 2005) as well as Talairach Client after transforming from the MNI to the Talairach coordinate system using mni2tal function in Matlab. The substantia nigra, red nucleus, and subthalamic nuclei were identified using the regions specified in Keuken and Forstmann (2015).

RESULTS

Behavioral Results

The behavioral results in terms of percent correct on the visual search task for the tDCS stim and sham groups are as follows: there was a significant enhancement in performance post- relative to pre-training (ANOVA $F_{(2,52)} = 12.47$, $p < 0.05$). The enhancement was statistically significant ($t = 4.05$; $p < 0.05$) for the stim group (pre-training mean = 64.26%; SE = 2.64; post-training mean = 72.06%; SE = 2.36) and was statistically significant ($t = 3.15$; $p < 0.05$) for the sham group (pre-training mean = 64.98%; SE = 2.47; post-training mean = 73.02%; SE = 1.99). There was no significant difference (assessed at $p < 0.05$) between stim and sham groups for either pre- ($T = -0.21$) or post-training ($T = -0.33$) sessions. The interaction between stim and sham groups and pre- and post-training session was not significant ($F_{(2,52)} = 0.3$; $p > 0.05$). Additionally there was no significant difference ($T = -0.83$, $p > 0.05$) between stim (mean = 64.86%; SE = 2.7) and sham (mean = 67.94%; SE = 2.74) groups for the training session 2.

Brain Imaging Results

Resting State Activity: fALFF Analysis

The results of the fALFF analysis are given in **Figure 2** and **Table 1**. Using a random effects between groups *t*-test, statistically significant ($p < 0.05$ corrected for multiple comparisons, see "Materials and Methods" Section) differences in fALFF between the stim and sham groups for session 2 were found to be located in three clusters of activity: Cluster 1 is located around the right superior parietal spreading into the left parietal cortex as well as the neighboring regions of the precuneus, post central gyrus, pre-central gyrus, and supplementary motor area (It should be noted that since the analysis is corrected for multiple comparisons at the cluster level we cannot definitively know which of these regions making up the cluster are activated, only that some of them are); Cluster 2 is located in the right inferior parietal lobule including the temporal parietal junction; Cluster 3 is located in the premotor cortex BA6 (see **Figure 2**, top and **Table 1**, top). Brain regions statistically significant ($p < 0.05$ corrected) for the stim–sham contrast masked by the interaction (random effects ANOVA)

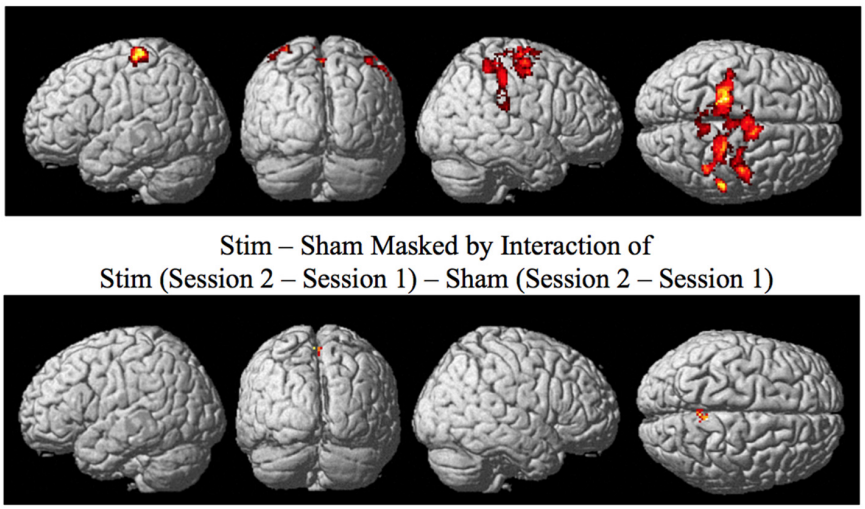


FIGURE 2 | Results of the SPM random effects analysis rendered on the surface of the brain. Top: Differential resting state activity as measured by fractional amplitude of low frequency fluctuations (fALFF) for the stim–sham groups for session 2 corrected for multiple comparisons at the cluster level (p <0.05) using Monte-Carlo simulation (corrected cluster extent threshold greater than 154 contiguous voxels over uncorrected significance threshold of $p < 0.005$). Bottom: Activity in the session 2 stim–sham analysis above that is additionally masked by the interaction of Stim (Session 2 – Session 1) – Sham (Session 2 – Session 1) corrected ($p < 0.05$) for multiple comparisons at the cluster level.

of Stim (Session 2 – Session 1) – Sham (Session 2 – Session 1) with a corrected threshold of $p < 0.05$, consisted of the right precuneus (see **Figure 2**, bottom and **Table 1**, bottom). Masking by the interaction controls against differences in fALFF that may exist between the tDCS stim and sham groups prior to training. In addition the masking also allows for a more focal site that is likely modulated by the tDCS, which can be used as a seed for the resting state functional connectivity analyses. No statistically significant differences in fALFF were found for session 2 for the sham greater than stim contrast when correcting for multiple comparisons.

Resting State Connectivity: Functional Connectivity Analysis

Resting state functional connectivity analyses for each of the three sessions, using post relative to pre behavioral performance

TABLE 1 | Differential resting state activity between stim and sham groups determined by fractional amplitude of low frequency fluctuations.

Brain region	MNI coordinates *x, y, z*	*T* (*df* = 26)	*Z*-score	Cluster size
fALFF (Stim–Sham) session 2				
Precuneus BA7	6, −46, 60	3.76	3.33	1902
SPL BA40	40, −44, 54	3.99	3.76	
Postcentral Gyrus BA3	24, −32, 60	4.48	3.82	
Postcentral Gyrus BA1, 6	−22, −32, 54	4.49	4.49	
SMA BA6	6, 2, 66	3.89	3.42	
IPL BA40	58, −30, 44	5.17	4.25	192
TPJ BA40	50, −30, 22	3.88	3.42	
Precentral Gyrus BA4, 6	46, −14, 62	4.29	3.70	291
Premotor Cortex BA6	26, −16, 74	3.60	3.21	
fALFF (Stim–Sham) session 2 masked by interaction				
Precuneus BA7	6, −46, 60	3.76	3.33	

Top: Brain regions showing significant differential activity for the stim–sham comparison for session 2 corrected for multiple comparisons at the cluster level ($p < 0.05$) using Monte-Carlo simulation (corrected cluster extent threshold greater than 154 contiguous voxels over uncorrected significance threshold of $p < 0.005$). Bottom: Regions that are also significant when masking the results by the interaction of Stim (Session 2 – Session 1) – Sham (Session 2 – Session 1) corrected for multiple comparisons at the cluster level ($p < 0.05$). BA, Brodmann area; IPL, Inferior Parietal Lobule; SPL, Superior Parietal Lobule; SMA, Supplementary Motor Area; TPJ, Temporal Parietal Junction. Negative "x" MNI coordinates denote left hemisphere and positive "x" values denote right hemisphere activity.

as a covariate of interest, were conducted using the precuneus (significant for the fALFF analyses; see **Figure 1**, bottom and **Table 1**, bottom) as a seed. In order to determine differences between the tDCS stim and sham groups in resting state functional connectivity related to enhanced behavioral performance, a random-effects between groups *t*-test using post-minus pre-training behavioral performance as a covariate of interest was conducted. Enhancement in behavioral performance was defined as the percent correct score for session 3 minus session 1 for each participant. The resting state functional connectivity score for each participant was the resultant Fisher's z transformed normalized correlation coefficient of the connectivity analysis for each voxel in the brain.

The results of the behavioral enhancement related functional connectivity analysis for sessions 1, 2, and 3 using the right precuneus as the seed region to the voxels in the entire brain are the following:

For session 1 behavioral enhancement related resting state functional connectivity was not observed for the stim–sham contrast, the stim alone contrast, or the sham alone contrast using a cluster level corrected threshold of $p < 0.05$.

For session 2 a cluster encompassing the substantia nigra, red nucleus, and subthalamic nuclei was found to show statistically significant differences in behaviorally related resting state functional connectivity when correcting for multiple comparisons at the cluster level ($p < 0.05$) for the stim–sham contrast (see **Figure 3** and **Table 2**). For the stim alone contrast three clusters showed statistically significant behavioral enhancement related resting state functional connectivity ($p < 0.05$ corrected). These clusters include the following: (1) the substantia nigra, red nucleus, and subthalamic nuclei; (2) the thalamus; and (3) the cerebellar lobule VIIIa Vermis (see **Figure 4** and **Table 2**, bottom). The sham alone contrast for session 2 did not show any statistically significant behaviorally related resting state functional connectivity using a corrected threshold of $p < 0.05$.

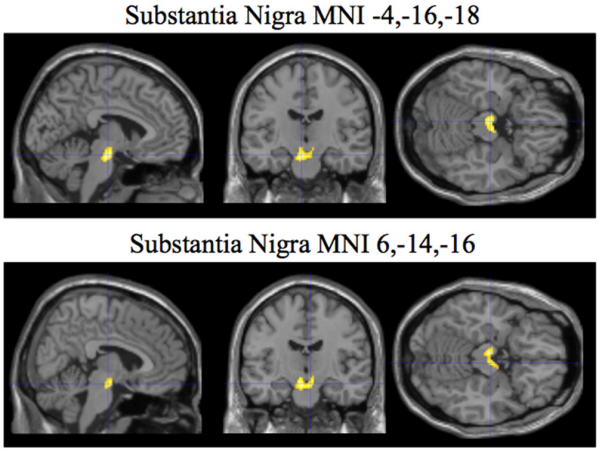


FIGURE 3 | Session 2 results of the SPM random effects between groups *t*-test for stim relative to the sham group for resting state functional connectivity with the precuneus using post-pre behavioral performance as a covariate of interest. Statistically significant ($p < 0.05$ corrected) differences in behaviorally related resting state functional connectivity are rendered on sections of a template T1 MRI scan at MNI coordinates for the peaks in the various significant clusters. Negative "*x*" MNI coordinates denote left hemisphere and positive "*x*" values denote right hemisphere activity. For the MRI sections the right side of the image is the right side of the brain.

For session 3 two clusters (one in the cerebellum lobule VIIa Crus I and the other in the insula) showed statistically significant ($p < 0.05$ corrected) differences in behaviorally related resting state functional connectivity for the stim–sham contrast (see **Figure 5** and **Table 3**). For the stim alone contrast four clusters showed statistically significant ($p < 0.05$ corrected)

TABLE 2 | Behavioral enhancement related resting state functional connectivity with the precuneus session 2.

Brain region	MNI coordinates *x, y, z*	*T* (*df* = 24)	*Z*-score	Stim *r*	Sham *r*	Cluster size
Stim-Sham						
Substantia Nigra	−4, −16, −18	5.20	4.22	0.86**	−0.52	200
	6, −14, −16	4.26	3.64	0.81**	−0.43	
Stim alone						
Substantia Nigra	−6, −14, −16	6.05	4.02	0.87**		387
	6, −14, −16	4.74	3.49	0.81**		
Thalamus	6, −14, 8	5.65	3.87	0.85**		437
	−4, −12, 4	4.3	3.28	0.79**		
	−16, −14, 12	4.37	3.32	0.78**		
Cerebellum	8, −64, −34	4.91	3.57	0.82**		329
Lobule VIIIa Vermis	8, −72, −32	4.74	3.49	0.81**		
	16, −70, −36	4.70	3.47	0.81**		

*Brain regions showing significant differential resting state connectivity for session 2 corrected for multiple comparisons at the cluster level ($p < 0.05$) using Monte-Carlo simulation (corrected cluster extent threshold greater than 154 contiguous voxels over uncorrected significance threshold of $p < 0.005$). BA, Brodmann area. Negative "x" MNI coordinates denote left hemisphere and positive "x" values denote right hemisphere activity. Correlation coefficient r: **denotes ($p < 0.005$).*

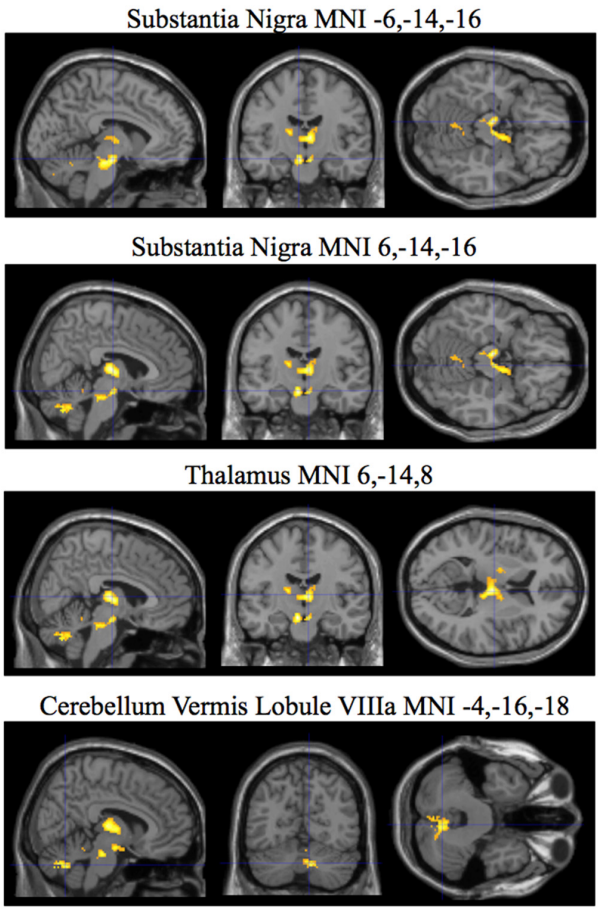


FIGURE 4 | Session 2 results of the SPM random effects *t*-test for stim alone contrast for resting state functional connectivity with the precuneus using post-pre behavioral performance as a covariate of interest. Statistically significant ($p < 0.05$ corrected) behaviorally related resting state functional connectivity is rendered on sections of a template T1 MRI scan at MNI coordinates for the peaks in the various significant clusters. Negative "*x*" MNI coordinates denote left hemisphere and positive "*x*" values denote right hemisphere activity. For the MRI sections the right side of the image is the right side of the brain.

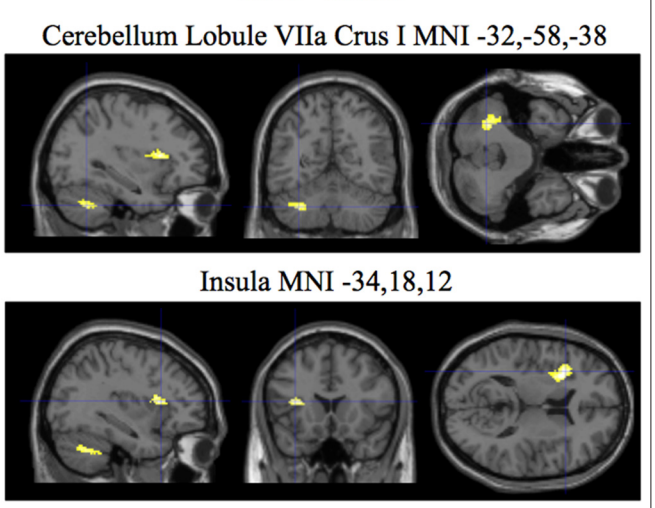


FIGURE 5 | Session 3 results of the SPM random effects between groups *t*-test for stim relative to the sham group for resting state functional connectivity with the precuneus using post-pre behavioral performance as a covariate of interest. Statistically significant ($p < 0.05$ corrected) differences in behaviorally related resting state functional connectivity are rendered on sections of a template T1 MRI scan at MNI coordinates for the peaks in the various significant clusters. Negative "*x*" MNI coordinates denote left hemisphere and positive "*x*" values denote right hemisphere activity. For the MRI sections the right side of the image is the right side of the brain.

behaviorally related resting state functional connectivity. These clusters include the following: (1) the cerebellum lobule VIIa Crus I; (2) the cerebellar lobule VI vermis and hemisphere; (3) the thalamus; and (4) the inferior parietal cortex (see **Figure 6** and **Table 3**, bottom). The sham alone contrast for session 3 did not show any statistically significant behaviorally related resting state functional connectivity using a corrected threshold of $p < 0.05$.

The magnitude and the direction of the correlation between behavioral enhancement and resting state functional connectivity are given in **Tables 2**, **3** (correlation coefficient *r*). For the tDCS stim group an increase in resting state functional connectivity is statistically significantly correlated (see **Tables 2**, **3**) with the increase in behavioral performance. This relation is not found for the sham group.

DISCUSSION

Resting state functional connectivity during tDCS is correlated with future improvement in performance. Our study shows that tDCS affects low amplitude fluctuations in spontaneous brain activity in the precuneus region around the anodal stimulating electrode (see **Figure 2**, bottom and **Table 1**, bottom). Performance enhancement related differences (between tDCS stim and sham groups that is also present in the stim alone analysis) in resting state functional connectivity were found from the precuneus to a cluster encompassing the substantia nigra, red nucleus, and the subthalamic nuclei during concurrent tDCS stimulation for session 2 (see **Figures 3**, **4** and **Table 2**). An after-effect of tDCS stimulation on resting state functional connectivity was measured following a post-training session on the visual search task that occurred approximately 20 min after the session of tDCS stimulation. Performance enhancement related differences (between tDCS stim and sham groups that is also present in the stim alone analysis) in resting state functional connectivity were found from the precuneus to a

TABLE 3 | Behavioral enhancement related resting state functional connectivity with the precuneus session 3.

Brain region	MNI coordinates *x, y, z*	*T* (*df* = 24)	*Z*-Score	Stim *r*	Sham *r*	Cluster Size
Stim–Sham						
Cerebellum Lobule VIIa Crus I	−32, −58, −38	3.54	3.15	0.82**	−0.22	174
Insula BA13	−34, 18, 12	3.77	3.31	0.67*	−0.48	186
Stim alone						
Cerebellum Lobule VIIa Crus I	−32, −62, −36	5.52	3.82	0.85**		337
Cerebellum	−10, −76, −26	3.75	2.99	0.74**		195
Lobule VI Vermis	12, −70, −20	3.71	2.97	0.73**		
and Hemisphere	−4, −72, −22	3.64	2.93	0.72**		
Thalamus	14, −22, 18	4.81	3.52	0.81**		194
	18, −32, 8	4.81	3.52	0.81**		
IPC	42, −42, 30	4.46	3.36	0.79**		198

*Brain regions showing significant differential resting state connectivity for session 3 corrected for multiple comparisons at the cluster level ($p < 0.05$) using Monte-Carlo simulation (corrected cluster extent threshold greater than 154 contiguous voxels over uncorrected significance threshold of $p < 0.005$). BA, Brodmann area; IPC, Inferior Parietal Cortex. Negative "x" MNI coordinates denote left hemisphere and positive "x" values denote right hemisphere activity. Correlation coefficient r: *denotes ($p < 0.05$) and **denotes ($p < 0.005$).*

cluster encompassing the right cerebellum lobule VIIa Crus I for session 3 (see **Figures 5**, **6** and **Table 3**).

The mechanisms behind tDCS-induced enhanced cognition have been associated with that of activity-dependent plasticity. The precuneus region revealed by the fALFF analysis to be specifically modulated by anodal tDCS is most likely the result of increased spontaneous neuronal firing due to excitability changes brought on by tDCS. Spontaneous fluctuations in BOLD signal related to cognitive abilities are known to be present at rest (Biswal et al., 1995; Stevens and Spreng, 2014). Furthermore, studies have shown that resting state activity is modulated by tDCS (Saiote et al., 2013; Clemens et al., 2014).

The results of the fALFF analysis revealed that the precuneus showed significantly greater spontaneous resting state activity for the tDCS stim over the sham group that was not attributed to preexisting differences between the groups in resting state activity present prior to tDCS stimulation (see **Figure 2**, bottom and **Table 1**, bottom). The precuneus has been found to be involved with processes related to the visual search task employed in this experiment. These include attentive tracking of moving targets (Culham et al., 1998), attention orientation (Le et al., 1998; Simon et al., 2002), attention shift between object features (Nagahama et al., 1999), and mental rotation (Suchan et al., 2002). It has been put forward that the precuneus is involved with internally guided attention and manipulation of mental images related to visuospatial processing (Cavanna and Trimble, 2006).

Using the precuneus as a seed, we were able to reveal performance related differences in resting state functional connectivity associated with tDCS stimulation. The resting state functional connectivity analysis assumes that, in the absence of ongoing task related activity, two regions that display spontaneous fluctuations in BOLD signal that are highly temporally synchronized are likely within the same functional network. Using visual search performance post-training (session 3) relative to pre-training (session 1) as a covariate of interest in this analysis allowed us to identify regions that were associated with improved performance as temporal synchrony (functional connectivity) increases with our seed ROI (precuneus). Our results revealed visual search performance enhancement related differences in resting state functional connectivity between the precuneus and a cluster encompassing primarily the substantia nigra for the stim over the sham group (that was also present for the stim group alone contrast; **Figures 3**, **4** and **Table 2**). Interestingly, consistent with the task presented in our study, previous research has implicated the substantia nigra with aspects of visuospatial processing (Matsumoto and Takada, 2013). The study by Matsumoto and Takada (2013), using single cell recordings in monkeys, showed that neurons in the substantia nigra were active when the task required visual search and working memory. Consistent with the findings in these studies, the task in our experiment required the participant to maintain the features of the target truck and distractors in working memory to accomplish the visual search task. Also relevant to our study, the substantia nigra, is part of the dopaminergic system (Björklund and Dunnett, 2007). The dopaminergic system is thought to be intimately involved with value dependent learning (Montague et al., 1996; Schultz, 1998; Doya, 2002; Callan and Schweighofer, 2008). The performance related enhancement in resting state functional connectivity for the stim over the sham group between the precuneus and the substantia nigra is consistent with the hypothesis that tDCS may in part be modulating value dependent learning systems involved with the visual search task (see **Figures 3**, **4** and **Table 2**). While we cannot rule out that the effect of tDCS stimulation alone is responsible for our observed performance related resting state connectivity, given the function of the brain regions involved, it is perhaps more likely that the effect of tDCS stimulation interacting with the visual search task is responsible for the changes in performance related resting state functional connectivity that we observe in our study.

In addition to investigating the active effects of tDCS stimulation on performance related enhancement in resting state functional connectivity we also investigated the after-effect of tDCS stimulation following a post-training session on the visual search task that occurred approximately 20 min after the session of tDCS stimulation. The session 3 results revealed visual search

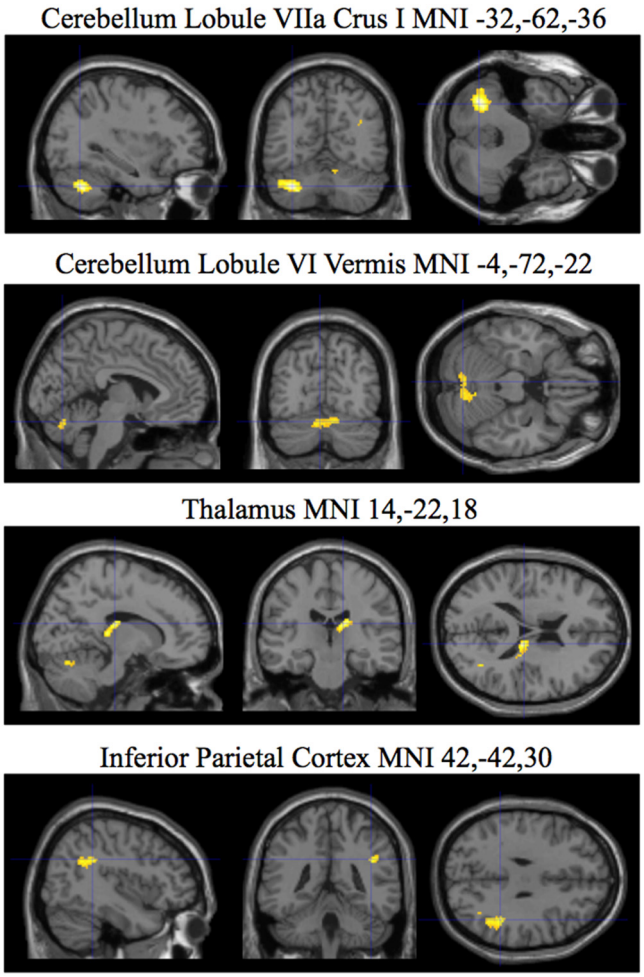


FIGURE 6 | Session 2 results of the SPM random effects *t*-test for the stim alone contrast for resting state functional connectivity with the precuneus using post-pre behavioral performance as a covariate of interest. Statistically significant ($p < 0.05$ corrected) behaviorally related resting state functional connectivity is rendered on sections of a template T1 MRI scan at MNI coordinates for the peaks in the various significant clusters. Negative "*x*" MNI coordinates denote left hemisphere and positive "*x*" values denote right hemisphere activity. For the MRI sections the right side of the image is the right side of the brain.

performance enhancement related differences in resting state functional connectivity between the precuneus and a cluster in the cerebellum lobule VIIa Crus I for the stim over the sham group (that was also present for the stim group alone contrast; **Figures 5**, **6** and **Table 3**). From studies on individuals with localized brain damage, human functional imaging studies, and animal studies the cerebellum is known to be involved with visuospatial processing (for review, see Molinari and Leggio, 2007). Related to our experimental task, the same region of the cerebellum as is present in our study has been found in an fMRI study to be involved with preparatory processes involved with visual search (Bourke et al., 2013). Also relevant is the known presence of anatomical connections between the precuneus and multiple circuits in the cerebellum through the basis pontis (Cavanna and Trimble, 2006).

One interesting finding of the analyses concerning the after-effects of tDCS stimulation is that the locus of differential performance related resting state functional connectivity (between tDCS stim and sham groups) is different from that of active tDCS stimulation. It is unclear why the focus of performance related resting state functional connectivity to the precuneus switches from the substantia nigra during active stimulation to the cerebellum as an after-effect. It is possible that the differences reflect distinctive stages of learning and correspondent modification of resting state networks.

Also of interest was the lack of any significant performance related resting state functional connectivity with the precuneus for the sham group while several regions were found to be significant for the tDCS stim group for active and after-effect analyses (see **Figures 4**, **6** and **Tables 2**, **3**). One possibility is that multiple degenerate networks (Edelman, 1987) are utilized for processing the visual search task for the sham group whereas for the tDCS stim group, as a result of stimulation, specific networks are selectively used. This is potentially why the degree of correlation between the strength of these networks and behavioral improvement in performance is relatively high (see **Tables 2**, **3**) for the tDCS stim group.

In our study as well as in others (e.g., Polanéa et al., 2011) the seed ROIs for the resting state functional connectivity analyses were selected based on previous, although different, analyses of the same data. The advantage of using the results of the fALFF analysis as seed ROIs is that we ensure that we are actually utilizing regions that are showing potential modulation as a result of the tDCS for the functional connectivity analysis instead of arbitrarily selecting a region underneath the stimulation electrode. We do not believe that this will unduly bias the results of the functional connectivity analyses for stim over the sham group comparison because the fALFF (fluctuations in low frequency activity in single voxels) and functional connectivity (correlation in time course between voxels) analyses are quite different. Additionally, we employed the use of improvement in behavioral performance post-relative to pre-training as a covariate of interest in the functional connectivity analyses. There is no *a priori* reason to believe that future improvement in behavioral performance should be predicted by differences in fALFF or functional connectivity unless of course these changes are induced as a result of tDCS.

As with all brain imaging studies, there are many potential limitations and confounds that need to be addressed. One potential limitation of this study is that the changes we see in fALFF and resting state functional connectivity may not be a result of changes in spontaneous neural activity, but rather may be a result of changes in cerebral perfusion or noise induced by tDCS (this is only a potential problem for session two in which concurrent tDCS and fMRI was applied). Previous studies using concurrent tDCS and fMRI have suggested that tDCS induced

distortions on fMRI SNR are minimal (Antal et al., 2011; Zheng et al., 2011). It is unlikely that these tDCS distortion effects would be specific to the low frequency range (0.01–0.08 Hz) used in the fALFF analysis. Since the fALFF compares this low frequency range to the entire range (0–0.25 Hz) it would cancel out any effects induced by tDCS (artifacts on fMRI, etc ...) that are not frequency specific. In terms of the resting state functional connectivity analyses we employed there is no reason why changes in cerebral perfusion or noise induced by tDCS would correlate with post-relative to pre-training behavioral performance. It is much more likely that the behaviorally related resting state functional connectivity we observed is a result of changes in spontaneous neural activity resulting from tDCS stimulation.

An additional confound that we did not test was whether the radio frequency and gradients associated with fMRI EPI scanning influences the tDCS current. The NeuroConn DC-Stimulator MR that we used for this experiment includes an RF filter module with MRI compatible cables and electrodes. These components help to prevent effects of RF on tDCS current. In addition we avoided cable loop formation in the setup that may result in gradient induced currents. Furthermore, the fMRI compatible cables used for tDCS had a high resistance (5 kΩ), which should also decrease the induced current avoiding potential effects of MRI on tDCS. Although we do not believe it to be the case, since we did not measure the current during stimulation, we cannot rule out the possibility that the tDCS current could have been reduced or modulated such that its enhancement effects on behavior were diminished.

One of the biggest limitations of our study was the lack of a robust difference in behavioral performance after training for the tDCS stim over the sham group. There are several reasons why group differences in improved performance between the stim and sham groups may not have been observed in our study when they have been found in many previous studies (Coffman et al., 2014). One reason may be that the task was too difficult when first starting such that many participants were at near chance levels. The effects of training were so great in this situation that the modulatory benefits of tDCS were washed out in the behavioral data. Another reason for a lack of a behavioral performance enhancement difference between tDCS stim and sham groups may be that training was too short for the modulatory effects of tDCS to be revealed in behavioral performance. Relatedly, another limitation in our study was the restriction on the level of tDCS stimulation to be a maximum of 1 mA with our version of the NeuroConn DC-Stimulator MR. Many tDCS studies commonly utilize 2 mA to get robust enhancement in behavioral performance (Coffman et al., 2014). One reason why we did not see an overall difference between tDCS stim and sham groups may be because the level of stimulation was too low to induce robust enhancement in the short training time of approximately 16 min. Given that we do see strong performance related differences in the resting state functional connectivity data it may be the case that we are observing the early stages of tDCS modulated learning. The involvement of dopaminergic brain regions involved with value dependent learning as well as to working memory and visual attention are certainly consistent with this hypothesis. In the future it would be interesting to investigate whether longer training on this same task as well as higher stimulating levels (2 mA) results in enhanced behavioral performance for tDCS stim over sham groups.

CONCLUSION

Participants' who showed greater resting state functional connectivity between parietal regions and dopaminergic subcortical brain regions (substantia nigra, hippocampus, and amygdala) showed greater improvement in visual search task performance. Essentially, future improvement in performance showed a significant linear relationship with resting state functional connectivity (thought to be) induced by tDCS. These results suggest that it may be possible to employ multivariate pattern analysis machine learning decoding techniques to predict future performance, given a certain pattern of resting state functional connectivity. Future experiments need to be conducted to determine if changes in resting state functional activity and connectivity induced by tDCS can be used to predict long-term changes in task performance. One could even use neurofeedback techniques (Fukuda et al., 2015) in conjunction with tDCS to induce greater changes in specific functional networks. These techniques could be utilized to optimize performance benefits resulting from tDCS. Our results have wide ranging implications regarding effective utilization of tDCS for neuroergonomic as well as therapeutic, and rehabilitative applications.

DEDICATION

The present manuscript wishes to be the authors' personal honor to the memory of Raja Parasuraman, as a tribute to his scientific competence, and the contribution that he gave to our research projects in these years of proficient collaboration. Raja was a kind and generous man who was a pioneer in the field of neuroergonomics. He passionately sought out to facilitate and inspire researchers in the field instilling a thirst for knowledge. RP will always be present as a driving force inspiring us all to succeed.

AUTHOR CONTRIBUTIONS

DEC, BF and AW conducted experiment. DEC, BF, AW and RP designed experiment, analyzed the data and wrote the manuscript.

FUNDING

This work was supported by the Center for Information and Neural Networks, National Institute of Information and Communications Technology. Additional support to BF was given by the National Science Foundation sponsored East Asia and Pacific Summer Institutes fellowship award 1414852, funded in collaboration with Japan Society for the Promotion of Science. Additional support to RP was given by the Air Force Office of Sponsored Research grant FA9550-10-1-0385.

ACKNOWLEDGMENTS

We would like to thank Chie Kawakami for her extensive assistance in conducting the experiment and Akiko Callan for her assistance in translating instructions from English to Japanese. We would also like to thank the MRI technicians at CiNet for their assistance in running the experiments.

REFERENCES

Adelstein, J., Shehzad, Z., Mennes, M., Deyoung, C., Zou, X., Kelly, C., et al. (2011). Personality is reflected in the brain's intrinsic functional architecture. *PLoS One* 6:e27633. doi: 10.1371/journal.pone.0027633

Antal, A., Polania, R., Schmidt-Samoa, C., Dechent, P., and Paulus, W. (2011). Transcranial direct current stimulation over the primary motor cortex during fMRI. *Neuroimage* 55, 590–596. doi: 10.1016/j.neuroimage.2010.11.085

Baldassarre, A., Lewis, C., Committeri, G., Snyder, A., Romani, G., and Corbetta, M. (2012). Individual variability in functional connectivity predicts performance of perceptual task. *Proc. Natl. Acad. Sci. U S A* 109, 3201–3521. doi: 10.1073/pnas.1113148109

Bikson, M., Inoue, M., Akiyama, H., Deans, J., Fox, J., Miyakawa, H., et al. (2004). Effects of uniform extracellular DC electric fields on excitability in rat hippocampal slices *in vitro*. *J. Physiol.* 557, 175–190. doi: 10.1113/jphysiol.2003.055772

Biswal, B., Yetkin, F., Haughton, V., and Hyde, J. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planer MRI. *Magn. Reson. Med.* 34, 537–541. doi: 10.1002/mrm.1910340409

Björklund, A., and Dunnett, S. B. (2007). Dopamine neuron systems in the brain: an update. *Trends Neurosci.* 30, 194–202. doi: 10.1016/j.tins.2007.03.006

Bolognini, N., Fregni, F., Casati, C., Olgiati, E., and Vallar, G. (2010). Brain polarization of parietal cortex augments training-induced improvement of visual exploratory and attentional skills. *Brain Res.* 1349, 76–89. doi: 10.1016/j.brainres.2010.06.053

Bourke, P., Brown, S., Ngan, E., and Liotti, M. (2013). Functional brain organization of preparatory attentional control in visual search. *Brain Res.* 1530, 32–43. doi: 10.1016/j.brainres.2013.07.032

Callan, D., and Schweighofer, N. (2008). Positive and negative modulation of word learning by reward anticipation. *Hum. Brain Mapp.* 29, 237–249. doi: 10.1002/hbm.20383

Cavanna, A., and Trimble, M. (2006). The precuneus: a review of its function anatomy and behavioural correlates. *Brain* 129, 564–583. doi: 10.1093/brain/awl004

Clark, V., Coffman, B., Mayer, A., Weisend, M., Lane, T., Calhoun, V., et al. (2012). TDCS guided using fMRI significantly accelerates learning to identify concealed objects. *Neuroimage* 59, 117–128. doi: 10.1016/j.neuroimage.2010.11.036

Clemens, B., Jung, S., Mingoia, G., Weyer, D., Domahs, F., and Willmes, K. (2014). Influence of anodal transcranial direct current stimulation (tDCS) over the right angular gyrus on brain activity during rest. *PLoS One* 9:e95984. doi: 10.1371/journal.pone.0095984

Coffman, B., Clark, V., and Parasuraman, R. (2014). Battery powered thought: enhancement of attention, learning and memory in healthy adults using transcranial direct current stimulation. *Neuroimage* 85, 895–908. doi: 10.1016/j.neuroimage.2013.07.083

Culham, J., Brandt, S. A., Cavanagh, P., Kanwisher, N., Dale, A., and Tootell, R. (1998). Cortical fMRI activation produced by attentive tracking of moving targets. *J. Neurophysiol.* 80, 2657–2670.

Doya, K. (2002). Metalearning and neuromodulation. *Neural Netw.* 15, 495–506. doi: 10.1016/s0893-6080(02)00044-8

Edelman, G. (1987). *Neural Darwinism: The Theory of Neuronal Group Selection.* New York, NY: Basic Books.

Eichkoff, S., Stephan, K., Mohlberg, H., Grefkes, C., Fink, G., Amunts, K., et al. (2005). A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *Neuroimage* 25, 1325–1335. doi: 10.1016/j.neuroimage.2004.12.034

Ellison, A., Ball, K. L., Moseley, P., Dowsett, J., Smith, D. T., Weis, S., et al. (2014). Functional interaction between right parietal and bilateral frontal cortices during visual search tasks revealed using functional magnetic imaging and transcranial direct current stimulation. *PLoS One* 9:e93767. doi: 10.1371/journal.pone.0093767

Falcone, B., Coffman, B. A., Clark, V. P., and Parasuraman, R. (2012). Transcranial direct current stimulation augments perceptual sensitivity and 24-hour retention in a complex threat detection task. *PLoS One* 7:e34993. doi: 10.1371/journal.pone.0034993

Flöel, A. (2014). tDCS-enhanced motor and cognitive function in neurological diseases. *Neuroimage* 85, 934–947. doi: 10.1016/j.neuroimage.2013.05.098

Fukuda, M., Yamashita, A., Kawato, M., and Imamizu, H. (2015). Functional MRI neurofeedback training on connectivity between two regions induces long-lasting changes in intrinsic functional network. *Front. Hum. Neurosci.* 9:160. doi: 10.3389/fnhum.2015.00160

Hampson, M., Driesen, N., Skudlarshi, P., Gore, J., and Constable, R. (2006). Brain connectivity related to working memory performance. *J. Neurosci.* 26, 13338–13343. doi: 10.1523/jneurosci.3408-06.2006

Hunter, M., Coffman, B., Gasparovic, C., Calhoun, V., Trumbo, M., and Clark, V. (2015). Baseline effects of transcranial direct current stimulation on glutamatergic neurotransmission and large-scale network connectivity. *Brain Res.* 1594, 92–107. doi: 10.1016/j.brainres.2014.09.066

Kelley, T., Serences, J., Giesbrecht, B., and Yantis, S. (2008). Cortical mechanisms for shifting and holding visuospatial attention. *Cereb. Cortex* 18, 114–125. doi: 10.1093/cercor/bhm036

Keuken, C., and Forstmann, B. (2015). A probabilistic atlas of the basal ganglia using 7T MRI. *Data Brief* 4, 577–582. doi: 10.1016/j.dib.2015.07.028

Koyama, M., Di Martino, A., Zuo, X., Kelly, C., Mennes, M., Jutagir, D., et al. (2011). Resting-state functional connectivity indexes reading competence in children and adults. *J. Neuroscience* 31, 8617–8624. doi: 10.1523/JNEUROSCI.4865-10.2011

Kuo, M., Paulus, W., and Nitsche, M. (2014). Therapeutic effects of non-invasive brain stimulation with direct currents (tDCS) in neuropsychiatric diseases. *Neuroimage* 85, 948–960. doi: 10.1016/j.neuroimage.2013.05.117

Le, T., Pardo, J., and Hu, X. (1998). 4-T fMRI study of nonspatial shifting of selective attention: cerebellar and parietal contributions. *J. Neurophysiol.* 79, 1535–1548.

Liebetanz, D., Nitsche, M. A., Tergau, F., and Paulus, W. (2002). Pharmacological approach to the mechanisms of transcranial DC–stimulation-induced after-effects of human motor cortex excitability. *Brain* 125, 2238–2247. doi: 10.1093/brain/awf238

Matsumoto, M., and Takada, M. (2013). Distinct representations of cognitive and motivational signals in midbrain dopamine neurons. *Neuron* 79, 1011–1024. doi: 10.1016/j.neuron.2013.07.002

Molinari, M., and Leggio, M. (2007). Cerebellar information processing and visuospatial functions. *Cerebellum* 6, 214–220. doi: 10.1080/14734220701230870

Montague, P., Dayan, P., and Sejnowski, T. (1996). A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *J. Neurosci.* 16, 1936–1947.

Nagahama, Y., Okada, T., Katsumi, Y., Hayashi, T., Yamauchi, H., Sawamoto, N., et al. (1999). Transient neural activity in the medial superior frontal gyrus and precuneus time locked with attention shift between object features. *Neuroimage* 10, 193–199. doi: 10.1006/nimg.1999.0451

Nelson, J. T., McKinley, R. A., Golob, E. J., Warm, J. S., and Parasuraman, R. (2014). Enhancing vigilance in operators with prefrontal cortex transcranial direct current stimulation (tDCS). *Neuroimage* 85, 909–917. doi: 10.1016/j.neuroimage.2012.11.061

Nitsche, M., Fricke, K., Henschke, U., Schlitterlau, A., Liebetanz, D., Lang, N., et al. (2003). Pharmacological modulation of cortical excitability shifts induced by transcranial direct current stimulation in humans. *J. Physiol.* 553, 293–301. doi: 10.1113/jphysiol.2003.049916

O'Shea, J., Boudrias, M., Stagg, C., Bachtiar, V., Kischka, U., Blicher, J., et al. (2014). Predicting behavioural response to TDCS in chronic motor stroke. *Neuroimage* 85, 924–933. doi: 10.1016/j.neuroimage.2013.05.096

Parasuraman, R., and Galster, S. (2013). Sensing, assessing and augmenting threat detection: behavioral, neuroimaging and brain stimulation evidence for the

critical role of attention. *Front. Hum. Neurosci.* 7:273. doi: 10.3389/fnhum.2013.00273

Parasuraman, R., and McKinley, R. (2014). Using noninvasive brain stimulation to accelerate learning and enhance human performance. *Hum. Factors* 56, 816–824. doi: 10.1177/0018720814538815

Polanéa, R., Paulus, W., Antal, A., and Nitsche, M. (2011). Introducting graph theory to track for neuroplastic alterations in the resting human brain: a transcrianial direct current stimulation study. *Neuroimage* 54, 2287–2296. doi: 10.1016/j.neuroimage.2010.09.085

Saiote, C., Turi, Z., Paulus, W., and Antal, A. (2013). Combining functional magnetic resonance imaging with transcranial electrical stimulation. *Front. Hum. Neurosci.* 7:435. doi: 10.3389/fnhum.2013.00435

Schultz, W. (1998). Predictive reward signal of dopamine neurons. *J. Neurophysiol.* 80, 1–27.

Simon, O., Mangin, J., Cohen, L., Le Bihan, D., and Dehaene, S. (2002). Topographical layout of hand, eye, calculation and language-related areas in the human parietal lobe. *Neuron* 33, 475–487. doi: 10.1016/s0896-6273(02)00575-5

Slotnick, S. D., Moo, L. R., Segal, J. B., and Hart, J., Jr. (2003). Distinct prefrontal cortex activity associated with item memory and source memory for visual shapes. *Brain Res. Cogn. Brain Res.* 17, 75–82. doi: 10.1016/s0926-6410(03)00082-x

Song, X., Dong, Z., Long, X., Li, S., Zuo, X., Zhu, C. Z., et al. (2011). REST: a toolkit for resting-state functional magnetic resonance imaging data processing. *PLoS One* 6:e25031. doi: 10.1371/journal.pone.0025031

Stevens, D., and Spreng, N. (2014). Resting-state functional connectivity MRI reveals active processes central to cognition. *Wiley Interdiscip. Rev. Cogn. Sci.* 5, 233–245. doi: 10.1002/wcs.1275

Suchan, B., Yaguez, L., Wunderlich, G., Canavan, A., Herzog, H., Tellmann, L., et al. (2002). Hemispheric dissociation of visuo-spatial processing and visual rotation. *Behav. Brain Res.* 136, 533–544. doi: 10.1016/s0166-4328(02)00204-8

van der Heuvel, M., Mandl, R., Kahn, R., and Hulshoff Poll, H. (2009). Functionally linked resting-state networks reflect the underlying structural connectivity architecture of the human brain. *Hum. Brain Mapp.* 30, 3127–3141. doi: 10.1002/hbm.20737

Weber, M. J., Messing, S. B., Rao, H., Detre, J. A., and Thompson-Schill, S. L. (2014). Prefrontal transcranial direct current stimulation alters activation and connectivity in cortical and subcortical reward systems: a tDCS-fMRI study. *Hum. Brain Mapp.* 35, 3673–3686. doi: 10.1002/hbm.22429

Zheng, X., Alsop, D., and Schlaug, G. (2011). Effects of transcranial direct current stimulation (tDCS) on human regional cerebral blood flow. *Neuroimage* 58, 26–33. doi: 10.1016/j.neuroimage.2011.06.018

Zou, Q., Zhu, C., Yang, Y., Zuo, X., Long, X., Cao, Q. J., et al. (2008). An improved approach to detection of amplitude of low-frequency fluctuation (ALFF) for resting-state fMRI: frational ALFF. *J. Neurosci. Methods* 172, 137–141. doi: 10.1016/j.jneumeth.2008.04.012

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Callan, Falcone, Wada and Parasuraman. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution and reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 09 February 2016
doi: 10.3389/fnhum.2016.00034

Transcranial Direct Current Stimulation Modulates Neuronal Activity and Learning in Pilot Training

Jaehoon Choe[1], *Brian A. Coffman*[1,2,3], *Dylan T. Bergstedt*[1,4], *Matthias D. Ziegler*[1,5] *and Matthew E. Phillips*[1]*

[1] HRL Laboratories LLC, Malibu, CA, USA, [2] Department of Psychiatry, The University of Pittsburgh, Pittsburgh, PA, USA, [3] Psychology Clinical Neuroscience Center, The University of New Mexico, Albuquerque, NM, USA, [4] Department of Sports Medicine, Pepperdine University, Malibu, CA, USA, [5] Advanced Technologies Laboratories, Lockheed Martin, Arlington, VA, USA

Skill acquisition requires distributed learning both within (online) and across (offline) days to consolidate experiences into newly learned abilities. In particular, piloting an aircraft requires skills developed from extensive training and practice. Here, we tested the hypothesis that transcranial direct current stimulation (tDCS) can modulate neuronal function to improve skill learning and performance during flight simulator training of aircraft landing procedures. Thirty-two right-handed participants consented to participate in four consecutive daily sessions of flight simulation training and received sham or anodal high-definition-tDCS to the right dorsolateral prefrontal cortex (DLPFC) or left motor cortex (M1) in a randomized, double-blind experiment. Continuous electroencephalography (EEG) and functional near infrared spectroscopy (fNIRS) were collected during flight simulation, n-back working memory, and resting-state assessments. tDCS of the right DLPFC increased midline-frontal theta-band activity in flight and n-back working memory training, confirming tDCS-related modulation of brain processes involved in executive function. This modulation corresponded to a significantly different online and offline learning rates for working memory accuracy and decreased inter-subject behavioral variability in flight and n-back tasks in the DLPFC stimulation group. Additionally, tDCS of left M1 increased parietal alpha power during flight tasks and tDCS to the right DLPFC increased midline frontal theta-band power during n-back and flight tasks. These results demonstrate a modulation of group variance in skill acquisition through an increasing in learned skill consistency in cognitive and real-world tasks with tDCS. Further, tDCS performance improvements corresponded to changes in electrophysiological and blood-oxygenation activity of the DLPFC and motor cortices, providing a stronger link between modulated neuronal function and behavior.

Keywords: tDCS, EEG, fNIRS, DLPFC, M1, flight simulation, skill learning

OPEN ACCESS

Edited by:
Hasan Ayaz,
Drexel University, USA

Reviewed by:
Makii Muthalib,
University of Montpellier, France
Mickael Causse,
Institut Supérieur de l'Aéronautique et de l'Espace, France

***Correspondence:**
Matthew E. Phillips
mephillips@hrl.com

Received: *06 November 2015*
Accepted: *19 January 2016*
Published: *09 February 2016*

Citation:
Choe J, Coffman BA, Bergstedt DT, Ziegler MD and Phillips ME (2016) Transcranial Direct Current Stimulation Modulates Neuronal Activity and Learning in Pilot Training. Front. Hum. Neurosci. 10:34. doi: 10.3389/fnhum.2016.00034

INTRODUCTION

There has recently been a rapid increase in the number of published studies in the field of neuromodulation due to the availability of non-invasive stimulation technologies such as transcranial direct current stimulation (tDCS). New tools for training enhancement are emerging which target specific, basic cognitive functions, with the goal of increasing performance in high-level, real-world tasks, such as pilot training. For example, Clark et al. (2012) demonstrated enhanced concealed image detection training with tDCS. Others have observed enhanced skill learning with tDCS in spatial and verbal working memories (Martin et al., 2014; Richmond et al., 2014), language acquisition (Flöel et al., 2008) and motor skills development (Banissy and Muggleton, 2013; Reis et al., 2015; Rumpf et al., 2015). For a review of tDCS enhancements (see Coffman et al., 2014).

Computerized cognitive training methods have been only moderately successful in enhancing performance (Ball et al., 2002). However, computerized procedural training (flight simulation) has been an important part of airplane pilot training since the mid 1970's. Commercial and military pilot training programs now utilize flight simulation extensively for training basic flight and combat skills (Bell and Waag, 1998; Rosenkopf and Tushman, 1998). Research on the effectiveness of flight simulator training has historically been limited by the high cost of full flight simulators, and occurs in the context of ongoing pilot training programs, rather than unbiased third-party research programs (Hays et al., 1992; Rosenkopf and Tushman, 1998). The field has recently overcome this limitation by the commercialization of relatively low-cost flight simulator devices available for purchase and use in standard research environments. These personal computer-based flight simulators are also used in various contexts for flight training (Koonce and Bramble, 1998), lending ecological validity to simulator studies.

Piloting an airplane is a demanding task requiring skillful execution of learned procedures. This has been observed as a correlation between flight simulator performance and measures of reasoning and working memory in general aviation pilots (Causse et al., 2011), and a concurrent decline in working memory and flight errors (Dismukes, 2008; Engle, 2010). Furthermore, neurophysiological markers of both short-term (e.g., fatigue) and long-term (e.g., expertise) cognitive functions correlate with behavioral performance (Ayaz et al., 2013; Borghini et al., 2014). Pilot skill development requires a synthesis of multiple cognitive faculties, many of which are enhanced by tDCS and include: dexterity (Boggio et al., 2006), mental arithmetic (Hauser et al., 2013), cognitive flexibility (Chrysikou et al., 2013), visuo-spatial reasoning (Heimrath et al., 2012), and working memory (Gill and Hamilton, 2014)—an important predictor of flight situation awareness in novices (Sohn and Doane, 2004).

Working memory is linked primarily with brain activity in the dorsolateral prefrontal cortex (DLPFC) (Courtney et al., 1996; Braver et al., 1997; Curtis and D'Esposito, 2003), an area often targeted by non-invasive brain stimulation in cognitive research. Most researchers agree that tDCS of DLPFC has substantial effects on working memory (for a review see Coffman et al., 2014); however, Horvath et al. (2015) recently reported disconfirming evidence for this hypothesis in a meta-analysis of selected studies investigating the cognitive effects of tDCS. In this meta-analysis, tDCS did not have a significant effect on any cognitive measure. However, their approach may be confounded by calculation of effect sizes based only on post-stimulation scores, rather than accounting for pre-stimulation differences between groups. Chhatbar and Feng (2015) illustrated this issue in their response paper, where they show substantial effects of tDCS when calculating effect sizes from pre-post difference scores rather than post-stimulation scores alone.

The focality of stimulation is also a critical component of tDCS-driven behavioral changes, and this aspect of experimental design is difficult to capture in meta-study. Large pad-type electrodes used in previous studies have comparatively poor focality and target current intensity as compared to the multiple electrode montage approach (Dmochowski et al., 2011). Finite elements modeling work with MRI-derived brain models performed by various groups demonstrate optimization of currents to the brain that improve focality and intensity to areas of interest by 80 and 98%, respectively (Bikson et al., 2009; Datta et al., 2011, 2012; Dmochowski et al., 2011; Faria et al., 2011; Edwards et al., 2013). The importance of this modeling work is underscored by clinical investigations that show differences in targeting and stimulation intensity results in marked differences in behavioral output and stimulation efficacy (Valle et al., 2009; Moliadze et al., 2010; Mendonca et al., 2011). Finally, Santarnecchi et al. (2015) have suggested that the impact of tDCS on target brain structures is dependent on not only the placement of electrodes and current density, but also the current state of activity in those brain areas. This crucial point is often overlooked in tDCS research, and investigators should carefully consider the cognitive task performed during stimulation to maximize the desired effect.

Despite recent controversy over the effects of tDCS on working memory, tDCS applied to specific brain regions has been reported to improve behavioral performance in a diverse array of cognitive categories: attention (Coffman et al., 2014), reaction time (Teo et al., 2011), object recognition (Clark et al., 2012), memory (Manenti et al., 2013), creativity (Chrysikou et al., 2013), and motor skill acquisition (Nitsche et al., 2003). In addition to acute improvement of various performance measures, some laboratories have also observed persistence of cognitive enhancement even after the electrical current is removed (Snowball et al., 2013; Lefebvre et al., 2014). These results indicate that, in some cases, stimulation need only be applied initially or periodically to achieve continual performance gains. Although the modulation of procedural learning through enhancement of working memory has remained an open question in the field, non-invasive brain stimulation methods are potential vehicles to enhance learning and performance and nootropic benefits for commercial and military applications (Clark et al., 2012; Phillips and Ziegler, 2014).

The application of neuroimaging techniques, such as functional near-infrared spectroscopy (fNIRS) and electroencephalography (EEG), allow the precise measurement of spatial and dynamic functional brain activity. The development

of these non-invasive, low overhead and high-resolution tools have given investigators the ability to observe the activity of the human brain *in vivo* with an unprecedented degree of control (Been et al., 2007; McKendrick et al., 2015).

EEG results confirm tDCS-related modulation of brain processes involved in working memory, as evidenced by increased midline frontal theta-band oscillatory brain activity (MFT) during a working memory task (Miller et al., 2015). MFT is most commonly measured during maintenance of information in working memory, and reflects theta coupling between the DLPFC and anterior cingulate cortex (Sauseng et al., 2004). MFT is positively correlated with attentional demands during mental calculation (Ishii et al., 2014) and working memory load (Jensen and Tesche, 2002), and theta-band synchrony between frontal and parietal areas is directly related to individual working memory capacity (Palva et al., 2010). Further evidence supporting the functional relationship comes from studies temporarily disrupting the DLPFC with transcranial magnetic stimulation—leading to performance decrements in working memory tasks (Grafman et al., 1994; Pascual-Leone and Hallett, 1994). Other frequency bands have also been implicated in working memory and attentional control. For example, tonic increases (and phasic decreases) in parietal alpha-band power reflects greater perceptual involvement for tasks requiring attention to the environment (Klimesch, 1999), suggesting a role of alpha in perception. Furthermore, Sauseng et al. (2009) showed that alpha band activity over sensorimotor areas indicates greater excitability in that region, as measured with transcranial magnetic stimulation. Therefore, stimulation of either M1 or DLPFC could increase tonic alpha band activity in this study compared to sham by enhancing sensorimotor excitability and/or perceptual involvement.

Other imaging studies, employing fNIRS have found significant correlations between cognitive performance and blood oxygenation in the DLPFC (Yanagisawa et al., 2010; McKendrick et al., 2014). fNIRS is an non-invasive imaging technique that measures the relative concentrations of oxygenated (Hboxy) and deoxygenated (Hbdeoxy) hemoglobin to infer neuronal activity. fNIRS relies on differences in the near infrared absorption spectra of oxygenated and deoxygenated hemoglobin along with a neuro-vascular hemodynamic response function to relate relative chances in localized cerebral blood flow to neuronal activity (Villringer et al., 1993).

Hbdeoxy and total hemoglobin concentrations (Hbtot) are linked to levels of cognitive workload in the anterior prefrontal cortex (PFC) (Ayaz et al., 2012). For example, using a Scarborough adaptation of the Tower of London task, Ruocco et al. (2014) found that difficult problems were associated with greater Hboxy concentrations in the DLPFC relative to a baseline condition. The study also found that participants who scored higher in deliberation, or careful thinking, before acting, showed greater activation in this same region, regardless of task difficulty. The magnitude of Hbtot and Hbdeoxy concentration changes in specific brain regions has been used as a proxy for mental workload and expertise. Hbtot levels increase in the PFC during difficult trials in the N-back task, suggesting greater recruitment of neural resources (Herff et al., 2013). In addition, during a complex flight task, Hbtot levels decrease in the PFC over a 9-day learning period with progression from beginner to intermediate and finally advanced levels of performance (Ayaz et al., 2012). Furthermore, blood oxygenation level-dependent (BOLD) responses, which correlate with Hboxy, Hbdeoxy and Hbtot concentrations (Cui et al., 2011), decrease with improvements in response time, suggesting more efficient activation of PFC (Holland et al., 2011). Decreases in hemoglobin concentrations exist in the motor system (Hbdeoxy—Wolf et al., 2007), and in prefrontal cortex where they were correlated with reward value (Hboxy and Hbtot—DiStasio and Francis, 2013).

Although reported effects of primary motor cortex (M1) stimulation on skill acquisition and procedural learning have been promising, these methods have primarily been investigated in standard psychological and motor tasks including the serial reaction time task (Nitsche et al., 2003), the tower of London task (Dockery et al., 2009); and sequential visual isometric pinch task (Reis et al., 2009). Increasing evidence for the application of tDCS to enhance real-world skills has been reported for vehicle control (Beeli et al., 2008; Sakai et al., 2014), golf (Zhu et al., 2015), threat detection in image analysis (Falcone et al., 2012), air traffic control (Nelson et al., 2014). tDCS has also decreased resumption lag after interruption (Blumberg et al., 2014), and maintained vigilance (McIntire et al., 2014) in real-world tasks.

Critical for the acquisition of these real-work skills are both online and offline learning. Online learning is the change in behavioral performance across trials within an experimental session and is analogous to encoding (Reis et al., 2009). Offline learning is the change in performance, between sequential experimental sessions, from the last trial of the n-1th session to the first trial of the n^{th} session, and is analogous to consolidation (Robertson et al., 2004). The modulation of online and offline learning rate for practical, real-world skill acquisition with tDCS of M1 or DLFPC stimulation have remained unexplored.

Here, we investigated changes in skill acquisition and learning rates with tDCS applied to either DLPFC or M1 during custom pilot training exercises developed and administered with a commercially available flight simulator (X-Plane). These results were recently reported in a poster presentation at the Society for Neuroscience Meeting (Choe et al., 2015). We measured task-evoked changes in functional activity using fNIRS and EEG as subjects learned to complete flight simulator and N-back training exercises at increasing levels of expertise across four daily consecutive sessions. We hypothesized that stimulation of DLPFC over the course of flight simulation and N-back training would alter group variability in skill learning, MFT power, and Hboxy and Hbtot concentrations in the DLPFC. Furthermore, we hypothesized that tDCS of M1 will alter tonic alpha-band power over parietal cortex.

MATERIALS AND METHODS

Participants

Thirty-two right-hand dominant, healthy adult HRL Laboratories employees (31 males) participated in this study.

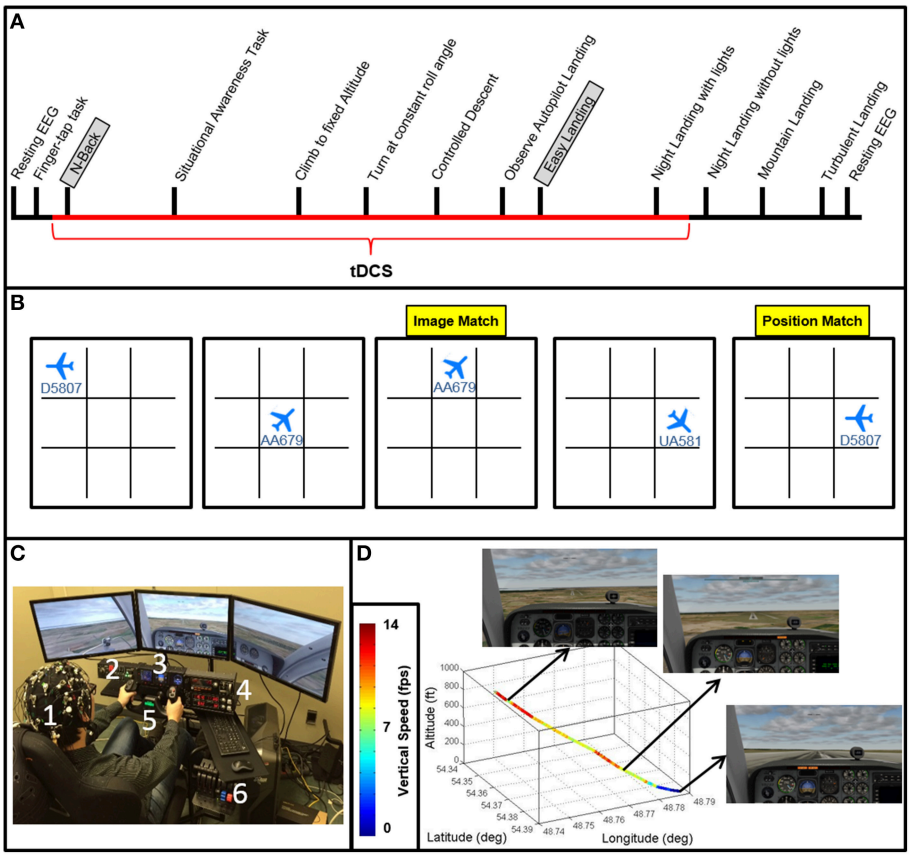


FIGURE 1 | Experimental design. (A) Experiment timeline depicting the relative timing of each task (see **Table 1** for descriptions of each task). The N-Back and Easy Landing tasks are highlighted, and the duration of tDCS is depicted in red. **(B)** An example of 6 trials of the N-Back task is shown. 1-back orientation and location match trials are highlighted in yellow. **(C)** The flight simulator, neuroimaging (EEG and FNIRS) and tDCS setup is shown with on a subject (1). Flight simulator equipment includes three-panel display, a radio panel (2), an instrument panel (3) with (from left to right) compass, altimeter, airspeed indicator, vertical speed indicator, and turn/slip indicator, a multi-panel (4) with (from left to right) autopilot settings, auto throttle switch, flaps switch, and elevator trim wheel, yoke (5), and throttle quadrant system (6). **(D)** Autopilot flight path for the Easy Landing task is shown in 3 dimensions, color-coded by vertical speed. Screenshots for initial descent, approach, and landing are also shown.

Their ages ranged from 21 to 64 (mean ± STD = 38 ± 13). Participants were randomly assigned to one of four groups: DLFPC stim (n = 7, age = 35 ± 11), DLPFC sham (n = 7, age = 42 ± 13), M1 stim (n = 10, age = 41 ± 16), or M1 sham (n = 8, age = 31 ± 5). HRL Laboratories employees are a vulnerable class of subjects for this study. In order to manage the risk of any undue influence, coercion, or confidentiality breach we only allowed individuals who are not directly supervised by the investigators of this study to volunteer, and only performed experiments during normal business hours (9 a.m.–5 p.m.) to mitigate any possibility for recourse or reward for participation in performance evaluation or job advancement. To maintain confidentiality, each subject was assigned a unique number, known only to the investigators of the study and subject identities were not shared. This design is in line with the recommendations of Meyers (1979) on student and employees as a vulnerable population of subjects and complies with DHHS: protected human subject 45 CFR 46; FDA: informed consent 21 CFR 50. Inclusion criteria were: (1) normal or corrected-to-normal vision, (2) no prior history of epileptic seizures or known neurological disorders, and (3) no females who are pregnant or are likely to become pregnant during the course of the study. All participants provided written informed consent to participate in the experiment. JC, MDZ, and MEP are listed as inventors in patent applications on brain stimulation methods.

Materials

Flight Simulator

Flight simulation tasks were designed and administered with the XForce Dream Simulator package (X-Force PC) and the X-plane 10 Flight Simulator software (Laminar Research). A depiction of the XForce Dream Simulator package can be seen in **Figure 1C**, and included a yoke, a radio panel, an instrument panel with compass, attitude indicator, altimeter, airspeed indicator, vertical speed indicator, and turn/slip indicator, a multi-panel with autopilot settings, auto throttle switch, flaps switch, and elevator trim wheel, and a throttle quadrant system. This flight

simulator included an adjustable seat for maximum comfort for the subject. Three monitors were placed at an optimal distance from the subject to avoid any eyestrain. Custom scenarios were designed using the simulator software development kit following a model of flight training (Williams, 2012, see **Table 1**).

Neuroimaging

We recorded continuous EEG and fNIRS data during flight simulation training, N-back, finger tapping, situational awareness, and resting-state assessments. Horizontal and vertical electro-oculogram (EOG) was also recorded. EEG was collected using a 32-channel acti32Champ system, with electrodes placed in a custom, 10-10 based arrangement to accommodate tDCS electrodes (StarStim Neuroelectrics) and fNIRS illuminators/receivers (NIRSport NIRX) within custom headcaps (BrainVision). EEG caps were selected for each subject based on individual head size and aligned to Cz. Conductive gel (Signagel) was applied onto each EEG electrode and ultrasound gel (Aquasonic clear) was applied to each fNIRS source and detector. fNIRS was recorded with dual-wavelength continuous-wave (CW) near infrared (NIR) diffuse tomographic measurements at 760 and 850 nm. A total of 20 fNIRS channels (source-detector pairs) were recorded over the left M1 (10 channels) and right DLPFC (10 channels, see **Figure 2**). The distance between source-detector pairs was <3.5 cm (see **Figure 2**). EEG data were collected at 500 Hz, and fNIRS data were collected at 8 Hz. Locations of EEG electrodes and fNIRS channels can be seen in **Figure 2**.

tDCS

Sham or actual tDCS was applied with the Starstim system (Neuroelectrics) following the finger tapping task (see **Figure 1A**). The total current applied was 2 mA, with scalp current density of 0.04 A/m^2 for active tDCS (for 60 min), or 0.1 mA (0.002 A/m^2) for sham tDCS (for 1 min). Currents were applied with a 1 min ramp-in at initiation and a 1 min ramp-out at termination. Sham stimulation was used as a control condition to induce the physical sensation associated with tDCS (e.g., tingling) without directly stimulating the brain areas located below the electrodes (Coffman et al., 2012b). Silver/silver chloride electrodes were each 3.14 cm^2 in size (total anode area = 6.28 cm^2; total cathode area = 9.42 cm^2). During stimulation the impedance value was limited to 20 kΩ for operation of the device; actual impedance values typically were below 10 kΩ and impedances were observed to be stable throughout the duration of the experiment. tDCS channel impedances were continually monitored at 1 HZ. To achieve maximum focality for the targeted brain regions of interest, electrode placements were derived using HD Targets (Soterix Medical) with stimulation targets in the left M1 (right posterior field orientation model) and right DLPFC (left anterior field orientation model) and possible electrode locations were defined using standard 10-10 electrode locations (see **Figure 2**). HD Targets uses a MRI-derived finite element brain model that provides predictions for current flow and alignment for multiple interacting electrodes; this model was used to calculate maximal focality and intensity for regions of interest. For M1 stimulation, this resulted in current values of: CP1 = 1244 μA, CP3 = 745 μA, FP1 = −417 μA, F8 = −448 μA, and F9 = −1124 μA. For DLPFC stimulation, current values were F6 = 1511 μA, FC6 = 482 μA, AF8 = −271 μA, AF4 = −283 μA, and FP2 = −1439 μA (see **Figure 2**). The predicted field intensities at the target locations were 0.56 V/m (DLPFC) and 0.45 V/m (M1). Groups are denoted as: DLPFC stim, DLFPC sham, M1 stim, and M1 sham.

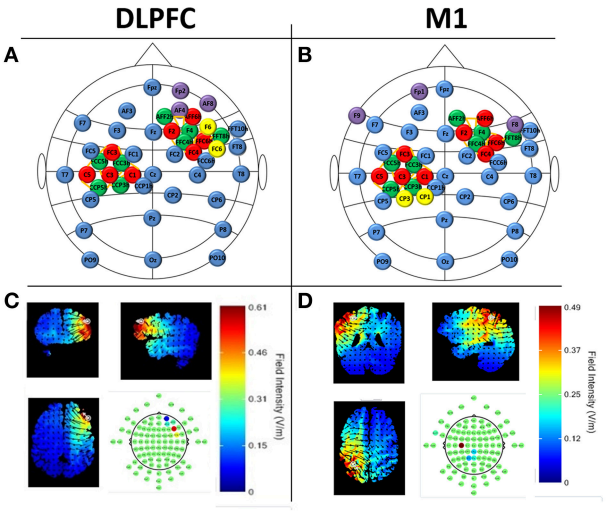


FIGURE 2 | Neuroimaging and tDCS experimental setup for DLPFC (A,C) and M1 stimulation (B,D). (A,B) EEG locations are denoted in blue and follow the 10–20 locations where possible. fNIRS sources (red) and detectors (green) are shown over the left-M1 and right DLPFC with channels depicted as orange lines (M1 channels: FC3-FCC5h, FC3-FCC3h, C5-FCC5h, C3-FCC5h, C3-FCC3h, C1-FCC3h, C5-CCP5h, C3-CCP5h, C3-CCP3h, C1-CCP3h; DLPFC channels: AFF6h-AFF2h, AFF6h-F4, F2-AFF2h, F2-F2, F2-FFC4h, FFC6h-F4, FFC6h-FFT8h, FFC6h-FFC4h, FC4-FFC4h, FC4-F4) tDCS electrodes are denoted in purple (cathodes) and yellow (anodes) and follow the current values specified in Section Neuroimaging CandDBal prefrontal cortex (DLPFC [e confidence bound was >4x the size of the positive confidence bound]). Predicted electric field intensities from the maximum focality montages from the Male 1 model in the Soterix HD Targets software (Soterix Medical).

Procedures

All participants performed flight simulation training, N-back, finger tapping, situational awareness, and resting-state assessments once per day for four consecutive daily sessions (see **Figure 1**). Resting-state brain activity was collected for 1 min both before and after the experiment. During resting scans, subjects observed autopilot flight (level flight at 5000 ft. altitude) and were instructed to keep their eyes open and observe the visual scene while keeping their hands in their laps. Following the pre-experiment resting-state assessment, motor reference scans were taken during a simple motor sequence task in which subjects were instructed to touch each fingertip with the thumb of the right hand in sequence/cycle, continuously for 30 s (**Figure 1A**, finger-tapping task). We analyzed neuroimaging data recorded during the finger-tapping task as a confirmatory

TABLE 1 | Experimental task schedule performed on each of the four consecutive experimental sessions.

Task	# of trials	Task instruction/Description	Average duration (minutes)
Survey	1	Setup and Explanation (day 1) Learn the basic controls via explanation	1
Rest	1	Eyes open, hand resting on yoke Observe straight and level autopilot flight	1
Finger tap	1	Using right hand, touch thumb to each finger in sequence	0.5
N-back	6	Adaptive threshold (at 80% accuracy) N-back on 3 × 3 grid of visual orientation of aircraft, flight number and spatial location (3 × 3 grid)	10
Situational awareness	4	Memorize gauge cluster images with 15 s distractor task and 30 s. for recall	10
Climb to fixed altitude	1	Adjust altitude from 5000 to 6000 ft. and back to 5000 ft. Maintain vertical speed at <1000 ft./min with level roll	5
Turn at constant roll angle	1	Change bearing/azimuth from 90° to 180° and back to 90° with a maximum roll angle of 20°	5
Descend @ constant Rate	1	Adjust altitude from 3000 to 1000 ft. at a 800 ft./min rate of descent	5
Autopilot landing (observe)	1	Observe the autopilot approach on the "Easy Landing" runway in perfect weather and high visibility	2
Easy landing	5	Control all aspects of landing in perfect weather and high visibility	10
Nighttime landing	2	Control all aspects of landing in perfect weather at night, with runway lights (low visibility)	4
No-lights landing	2	Control all aspects in approach with zero visibility	4
Hard landing (Mountains)	1	Difficult visual-only approach over terrain where the runway is initially obstructed from view over a mountain range	6
Hard landing (Turbulence)	1	Control all aspects of landing with turbulence value is set to level 1 out of 10 (instead of 0), with high visibility	2

measure (see Supplementary Figures S4, S5), where sensorimotor network activity was expected to be evident in EEG as increased power in the beta band, and reduced power in the alpha band, compared to baseline, and in fNIRS as an increase in deoxygenated hemoglobin beneath M1 sensors.

Participants then performed the N-back task followed by a series of basic flight training exercises including a situational awareness task, climbing to fix altitude, turning at a constant roll angle, and a controlled descent. Follow these fight control tasks participants performed a series of landing task including the "easy landing" task, nighttime landing, nighttime landing without runway lights, a landing in mountainous terrain, and a landing in turbulent weather (**Figure 1A**). Results for the situational awareness, free flight, climb to fixed altitude, heading change at constant roll angle, descent at constant vertical speed, nighttime landing, no-lights landing, mountain, and turbulence landing task are the subject of subsequent manuscripts.

N-Back

The Brain Workshop N-back task was implemented in this study (Paul Hoskinson, V.4.8.8 http://brainworkshop.sourceforge.net/). Participants monitored position and image for N-back matches without audio feedback. Custom N-back images were used, showing airplanes in eight different orientations with 1 of 3 possible flight numbers (24 total image possibilities, see **Figure 1B**). Subjects completed six blocks of 20 trials each day. Every subject began the N-Back task at the 1-back level, and was instructed at the beginning of each block to focus on a central fixation point. Subjects were free to move their eyes during the task. Upon reaching an upper threshold of accuracy within a given block (>80%), the task difficulty was increased (N + 1) using an adaptive threshold paradigm (Jaeggi et al., 2008). Upon reaching a lower threshold of accuracy (<20%), the task difficulty was decreased (N − 1). Each time this occurred, the changes were explained to the subjects between blocks. At the completion of each block, subjects were allowed to review the rules and ask clarifying questions about the tasks.

Autopilot Landing Observation

Subjects viewed a replay video of the autopilot executing an "optimal" landing from ∼800 ft. altitude onto a runway. Initial aircraft position was aligned with the runway and aircraft was already maintaining proper vertical speed for ideal glide slope. This scenario presents a wide, long, flat runway with no visual obstructions and no landscape features that interfere with landing the aircraft. Subjects were instructed not to manipulate controls or control the simulation in any way, but were told to pay close attention to the flight parameters through the instrumentation, as well as the visual field displayed by the simulator as the aircraft proceeded with landing. Particular emphasis was placed on two key parameters: azimuth (20°) for runway alignment, and vertical speed (∼700–800 ft./min) for appropriate glide slope. Attention was also drawn to the final control input to landing (pitch up at ground contact), and subjects were instructed to minimize landing force (G-force) as a top priority. Once the autopilot landing was viewed in its entirety, subjects were given the opportunity to ask questions about the landings. Most subjects asked very few, if any questions, typically on the first trial day.

Easy Landing Task

Subjects were instructed to complete the landing task as shown by the autopilot under daylight conditions and 100% visibility. Subjects attempted landing under these conditions a total of 5 times per day. As the subject attempted replication of the autopilot landing, the experimenter made observations in three categories: (1) Vertical speed maintenance; (2) Runway alignment; and (3) Final approach dynamics (pitch angle at touchdown). Any large deviations from the autopilot in any of these modalities were noted, then provided as feedback to the subject after the plane had touched down and the simulator paused. When given, feedback was ∼1–2 min in length and conducted in an informal manner. The time duration of feedback also shortened throughout training as the subject made fewer errors. Following feedback, the subject was offered opportunity to ask any questions regarding landing technique, then the scenario was restarted. If subjects passed beyond the terminal end of the runway, the attempt was ended and the landing listed as "missed landing." This counted against the number of subject attempts (i.e., attempts were not repeated due to missed landing). Feedback methods and handling of missed landings was identical for all landing task.

Data Analysis

EEG

EEG data were preprocessed using EEGLAB (Delorme and Makeig, 2004) by applying a 0.5 Hz high-pass filter (Butterworth, 12 dB/oct) and removing bad channels (max = 19%). Adaptive Mixture Independent Components Analysis (AMICA) (Delorme et al., 2012) was then used to detect and remove artifacts associated with eye blinks, vertical and horizontal electrooculogram, electrocardiogram, and tDCS-related voltage fluctuation. Following artifact rejection using AMICA, data were back-reconstructed and channels removed prior to AMICA decomposition were interpolated back into the data by spherical interpolation. Blocks corresponding to N-back, resting-state, and Easy Landing tasks were then segmented from the data.

Frequency decomposition was performed using FieldTrip (Oostenveld et al., 2011) by first segmenting data for each task into sequential 1-s epochs. Data were then windowed using a hanning taper, and frequency content of each trial was assessed at 1 Hz increments from 4 to 7 Hz (theta-band) or 8–12 Hz (alpha-band) using Fast Fourier Transform (multitaper method). After frequency decomposition, epochs with average theta or alpha power greater than two standard deviations from the mean were rejected, and remaining epochs were averaged for each participant, training day, and task. Data missing due to equipment issues (i.e., amplifier battery failure: $N = 4$, stimulus trigger errors: $N = 1$, or excessive noise/artifact during recording which could not be removed with AMICA: $N = 7$) were replaced with the mean for that participant group and training day prior to statistical analysis. We verified sensorimotor network activity during the finger-tapping task on the first day of flight simulator training (prior to tDCS) within baseline-subtracted beta and alpha band power maps, calculated across all subjects (Supplementary Figures S4, S5).

Participants receiving tDCS were compared with sham tDCS participants at each of the 4 days of training using independent-samples *t*-tests, which separately tested differences in alpha-band and theta-band activity at each sensor. Additionally, day 1 was compared to day 4 within each tDCS group and sensor using paired *t*-tests to assess training-related effects on alpha-band and theta-band activity. Statistical tests were corrected for multiple comparisons using cluster-based permutation tests (500 repetitions, data point $\alpha = 0.05$, cluster-level $\alpha = 0.05$, minimum spatial extent = 2 channels). Results from these comparisons are reported separately for each cluster of significant differences between groups/conditions. We calculated mean alpha/theta band power within clusters for use in examining relationships between task-related EEG and fNIRS/behavioral data.

We also examined correlations between behavioral measures, fNIRS beta values, and mean theta/alpha power across clusters identified during cluster-based permutation tests comparing days 1 to 4. fNIRS beta values were unavailable for 7 subjects (4 active M1 subjects and 3 sham M1 subjects) because time stamps could not be parsed from the fNIRS data files; therefore, the number of participants used in this analysis were: M1 stim = 6, M1 sham = 5, DLPFC sham = 7, and DLPFC stim = 7. These correlations were examined only within the stimulation groups where significant clusters were identified. To investigate relationships between midline frontal theta-band activity (Midline frontal theta-band activity was calculated as the mean theta power across electrodes Fz and FC1, the electrodes nearest to medial prefrontal cortex) and behavioral measures in the easy landing and N-Back task, Pearson correlation statistics were examined. We compared midline frontal theta-band activity in the easy landing task with autopilot displacement, g-force at landing, vertical speed at landing, roll at landing, pitch at landing, or online/offline learning rates for number of control inputs, autopilot displacement, vertical speed deviance from autopilot, or vertical speed variance. In the N-Back task, we compared midline frontal theta-band activity with average N level achieved and online/offline learning rates. Correlations were examined separately for DLPFC and M1 groups, stim and sham groups, and days of training. Because of the large number of correlation statistics examined, we used a conservative alpha of 0.001 to determine statistical significance. We additionally report statistics with a relaxed alpha of 0.05; however, these effects will be considered trends in this analysis.

In addition to cluster-based permutation tests across all channels, 3-way split-plot ANOVA was used to compare midline frontal theta-band activity between tDCS conditions (stim and sham), days of training (day 1, 2, 3, and 4), and training block (Block 1, 2, 3, 4, and 5) for the N-back and easy landing tasks. Huynh-Feldt epsilon was used to correct degrees of freedom for assumptions of sphericity, and Fishers Least Significant Difference corrections of alpha were used for simple-effects/pairwise comparisons (Maxwell and Delaney, 2004).

fNIRS

fNIRS data was processed within the nirsLAB analysis package (NIRx Medical Technologies, Glen Head, NY; Xu et al., 2014). The Gratzer Spectrum was used to measure the absorbance

TABLE 2 | Average ± standard deviation of day 1, day 4, and day 4–day 1 Hboxy, Hbdeoxy, and Hbtot concentrations across subjects and channels for M1 and the DLPFC.

			Easy landing			N-Back		
			Day 1	Day 4	Day 4–Day 1	Day 1	Day 4	Day 4–Day 1
DLPFC Stim	M1	Oxy	−0.00060 ± 0.0018	−0.0019 ± 0.0013	**−0.0013 (8)**	0.00059 ± 0.0010	−0.00041 ± 0.00076	−0.00099
		Deoxy	−0.00037 ± 0.00051	0.00012 ± 0.0011	**0.0005 (4)**	−0.00017 ± 0.00012	0.00021 ± 0.00074	0.00039
		Total	−0.00098 ± 0.0017	−0.0017 ± 0.0012	**−0.00077 (8)**	0.00041 ± 0.0010	−0.00019 ± 0.00043	−0.00061
	DLPFC	Oxy	0.0030 ± 0.0022	0.00055 ± 0.0014	**−0.0024 (8)**	0.0021 ± 0.0014	0.00048 ± 0.00091	−0.0016
		Deoxy	−0.00024 ± 0.00058	−0.00034 ± 0.00041	−0.000095	−0.00018 ± 0.00019	−0.00014 ± 0.00021	0.000041
		Total	0.0027 ± 0.0013	0.00021 ± 0.0014	−0.0025	0.0019 ± 0.0014	0.00034 ± 0.00088	0.0016
DLPFC Sham	M1	Oxy	−0.0018 ± 0.0020	−0.0018 ± 0.0019	**−0.0000082 (10)**	−0.000094 ± 0.0012	−0.000033 ± 0.00098	**0.000062 (10)**
		Deoxy	−0.00022 ± 0.00063	0.000065 ± 0.0011	**0.00028 (4)**	−0.000099 ± 0.00053	−0.00011 ± 0.00022	**−0.00001 (4)**
		Total	−0.0020 ± 0.0019	−0.0017 ± 0.0012	**0.00028 (10)**	−0.00019 ± 0.00077	−0.00014 ± 0.00086	**0.000052 (10)**
	DLPFC	Oxy	−0.000033 ± 0.0018	−0.00046 ± 0.0014	**−0.00043 (1)**	0.00067 ± 0.0012	0.00035 ± 0.0012	**−0.00032 (4)**
		Deoxy	−0.00056 ± 0.00043	−0.00053 ± 0.00021	0.000035	−0.00034 ± 0.00020	−0.00036 ± 0.00013	**−0.000017 (8)**
		Total	−0.00059 ± 0.0019	−0.00099 ± 0.0013	**−0.0004 (4)**	0.00032 ± 0.0012	−0.000015 ± 0.0012	**−0.00034 (4)**
M1 Stim	M1	Oxy	0.00024 ± 0.0031	−0.000084 ± 0.0015	**−0.00032 (4)**	0.0013 ± 0.0017	0.00045 ± 0.00088	−0.00084
		Deoxy	−0.00019 ± 0.0014	−0.00049 ± 0.00028	**−0.0003 (5)**	−0.00048 ± 0.00045	−0.00027 ± 0.00031	0.00021
		Total	0.000043 ± 0.0039	−0.00057 ± 0.0014	−0.00062	0.00080 ± 0.0016	0.00017 ± 0.00085	−0.00063
	DLPFC	Oxy	0.00050 ± 0.0026	−0.00060 ± 0.00055	−0.0011	0.00015 ± 0.00051	−0.00031 ± 0.00036	**−0.00046**
		Deoxy	−0.000045 ± 0.00092	−0.00017 ± 0.00020	−0.00012	−0.00027 ± 0.00030	−0.00012 ± 0.000088	0.00015
		Total	0.00046 ± 0.0035	−0.000766 ± 0.00056	−0.0012	−0.00012 ± 0.00078	−0.00043 ± 0.00035	**−0.00031**
M1 Sham	M1	Oxy	−0.00097 ± 0.0010	−0.00080 ± 0.00085	**0.00017 (10)**	0.00022 ± 0.0012	0.00047 ± 0.0016	0.00025
		Deoxy	−0.00017 ± 0.00020	−0.00013 ± 0.00021	**0.000032 (8)**	−0.00015 ± 0.00016	−0.00018 ± 0.00020	−0.000031
		Total	−0.0011 ± 0.00091	−0.00093 ± 0.00089	**0.0002**	0.000074 ± 0.0012	0.00029 ± 0.0017	0.00022
	DLPFC	Oxy	−0.00041 ± 0.00087	−0.00054 ± 0.0019	−0.00013	0.00011 ± 0.00030	−0.00030 ± 0.00034	−0.00041
		Deoxy	−0.000019 ± 0.00038	0.00084 ± 0.0013	0.00086	−0.000087 ± 0.00019	0.00026 ± 0.00071	0.00035
		Total	−0.00043 ± 0.0011	0.00030 ± 0.0032	0.00073	0.000022 ± 0.00029	−0.000043 ± 0.00041	−0.000065

All values are presented are in mM concentration units. Bold numbers indicate a significant difference between days 4 and 1 as determined by a SPM (see Materials and Methods Section fNIRS).

spectra of Hbdeoxy and Hboxy, with average wavelengths of 760 and 850 nm, respectively. The corresponding molar extinction coefficients ε are $\varepsilon_{\text{Hboxy}}$ [1097.0 781.0] cm-1/M and $\varepsilon_{\text{Hbdeoxy}}$ = [645.5 1669.0] cm-1/M, (nirsLAB, NIRx Medical Technologies). The differential path lengths were 5.98 for Hboxy and 7.15 for Hbdeoxy (Essenpreis et al., 1993). In the Beer-Lambert law calculation, the distance between source-detector pair was =< 3.5 cm, and the exact distances were computed within NIRSLab according to the corresponding distances on the headcap.

Hbdeoxy, Hboxy and Hbtot concentration time series were band-pass filter from 0.01 to 0.2 Hz (finite impulse response with least-squares error minimization), to remove slow drifts in the signal and respiratory and cardiac rhythms. Inter-trail data was removed from the time series, and the average baseline concentration values were subtracted from the task-evoked concentration measurements.

The average concentration value of Hbtot, Hboxy, and Hbdeoxy were computed separately for each channel, subject, task, and day. Concentration values were averages within days, across all 20 trails of each of the 6 blocks in the N-back, and all 5 trials of the easy landing task. Individual channel concentration values were then averaged across channels within regions (M1 and DLPFC) and across subjects within each group. Day1 group-averaged concentration values were then subtracted from Day 4 concentrations to compute the change in concentrations across the duration of the experiment.

Statistical significance of group-averaged concentrations changes from days 1 to 4 was determined using Statistical Parametric Mapping (SPM version 8). SPM was performed based on a general linear model of the canonical hemodynamic response function, with a discrete cosine transformation used for temporal filtering. A t-statistic-thresholded, baseline-subtracted Beta image was generated for each subject for baseline-subtracted, task-evoked Hbtot, Hboxy, and Hbdeoxy concentrations for days 1 and 4 (corrected for multiple comparisons across channels using the Bonferroni correction: # channels = 20, $p < 0.0025$). Paired t-statistic maps (subtracting the day 1 from day 4 betas) were generated from baseline-subtracted, trial/block-averaged (within day $n = 5$

Easy landing, $n = 6$ N-back) task betas obtained from individual subjects. If a t-statistic exceeded the corrected p-value threshold of 0.0025 the days 4–1 concentration values were determined to be significant (**Table 2**—denoted by bolded values).

Channel-wise statistical analysis was performed on all channels for measurements of Hboxy, Hbdeoxy, and Hbtot days 4–1 concentrations in easy landing and N-back for all subjects. Significance was determined if the trial-wise average exceeded 3.5 standard deviations from the null hypothesis of no concentration change (Bonferroni corrected, two-sided, Fischer's test $p < 0.00035$).

Behavioral Performance

N-Back

Raw percent accuracy values for each subject and for each block were scaled according to the information content required for each back condition. A 100% score on a 1-back trial requires both an image match: 9 possible plane orientations, 4 possible flight numbers, and a position match: 9 possible

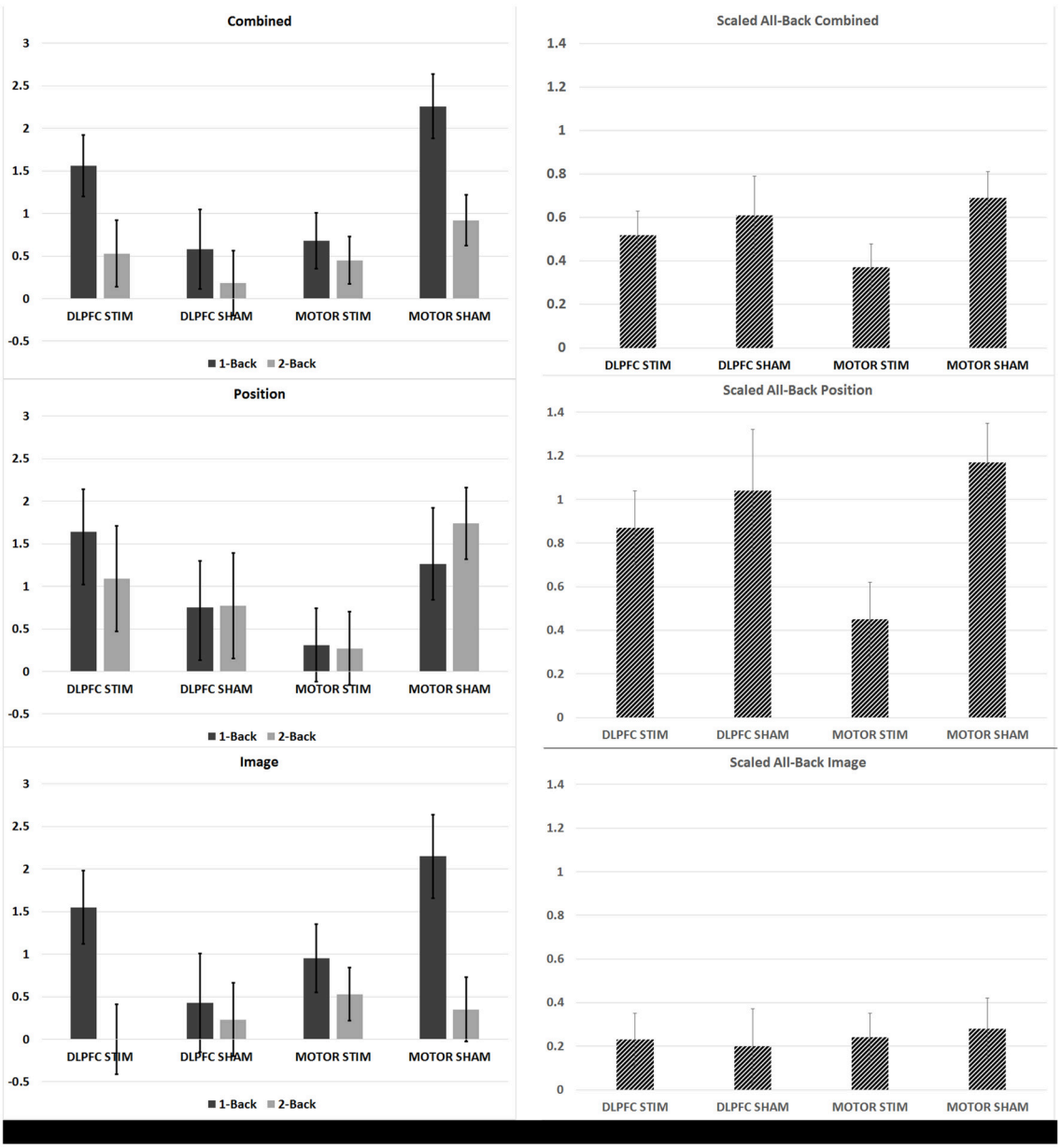


FIGURE 3 | N-back learning rates across experimental groups. The average group-learning rate is shown for each group in 1 and 2 back trials (left) and for all back trials (right column scaled by information content see Section N-Back). Learning rates computed by combining across position and image match trials **(top row)**, for position trails **(middle row)** and image trials **(bottom row)** are shown.

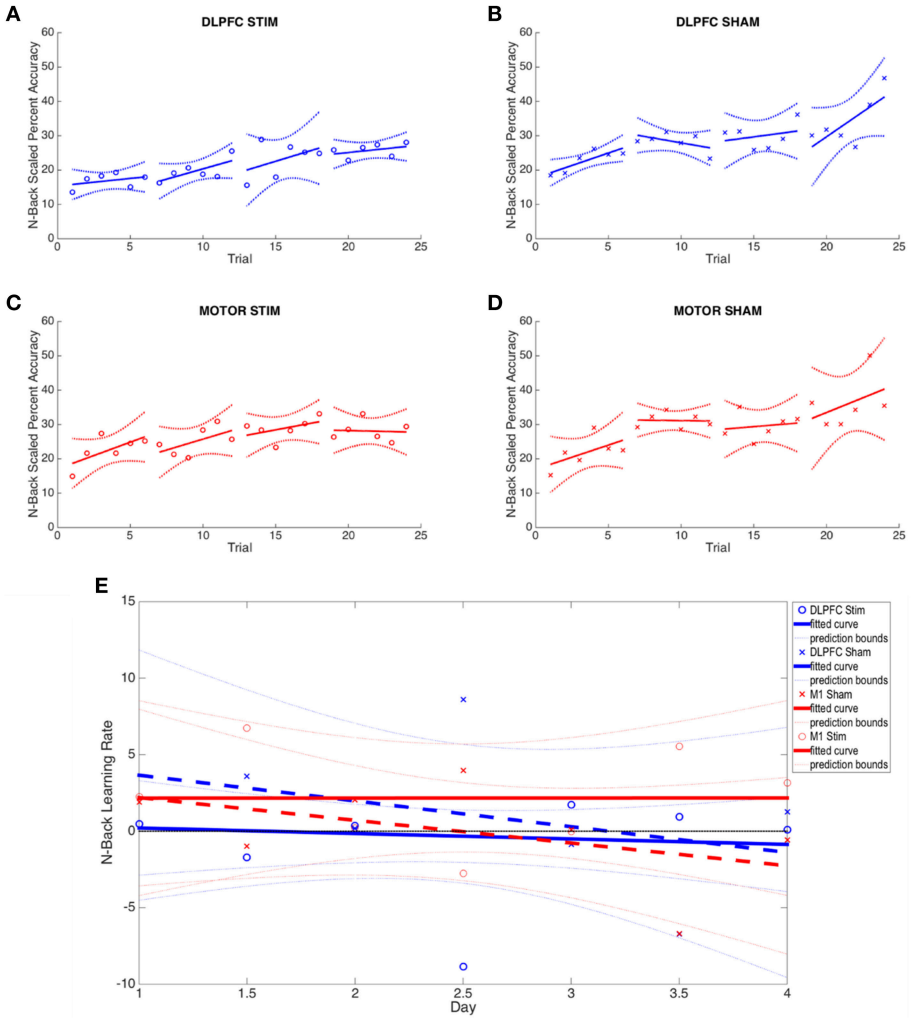


FIGURE 4 | N-back results across all four experimental groups. (A–D) The average percent correct is plotted across all 4 days (6 trials per day) for each group. Percent correct values are scaled based on the required information for a 100% correct response (see Section N-Back). **(E)** Online and offline N-back learning rates are plotted for each experimental group across the duration of the experiment. Whole numbers on the x-axis represent the average online learning rate (slope of scaled percent correct linear regression for each subject across 6 blocks within a day and ½ numbers of the x-axis represent offline learning rates (slope of the percent correct on the last trials of the N-1 day to the first trial of the Nth day).

spatial location, a 100% score on 2-back doubles the required information kept in working memory, and a 100% score on a 3-back trial triples this value. The normalization weights used for the 1–, 2–, and 3–back raw accuracy values were therefore 0.33, 0.66, and 1.0. Alternative normalization schemes (e.g., bit-wise maximum information and log-scaling) did not generate substantial differences in the outcome metrics. Learning rates were determined by the slope (±standard deviation) of a linear regression over block-wise group-averaged scaled percent accuracy: (1) across all 4 days (overall learning rate), (2) within each day independently (online learning rate), and (3) between the accuracy of the first trial of the nth day and the last trial of the n-1th day (offline learning rate) (Reis et al., 2009). Meta-learning rate was determined from the slope of the linear regression over the combined online/offline learning rate time series (the rate of change in the learning rates over time). The average number of trials for each group to reach the 2 and 3 back levels in the adaptive N-back task and the average streak (number of consecutive trials) at 2 and 3 back were calculated for each group. Learning rates were compared using one-sample against zero or paired, two-tailed t-tests (both $\alpha = 0.05$) were noted.

Easy landing

G-force assessment. Flight parameters were sampled from the simulator at 10 Hz, including altitude (above ground level), longitude, and latitude. The derivative of the vertical speed of the aircraft at runway touchdown determined the landing impact g-force (acceleration divided by 9.8 m/s^2). Smaller g-force landings reflected improved skill with the landing task as subjects were asked to minimize this value to the best of their ability for each trial. The impact g-force is a "one-shot" assessment

of landing skill at the most difficult and critical phase of the landing task, while ignoring other factors of landing performance (e.g., approach, glide slope, alignment, aircraft attitude). Online, offline, and meta learning rates are negative indicating a reduction in the applied G-force at landing (Supplementary Table S2, **Figure 4**).

Flight path deviation. Latitude, longitude, and altitude were transformed into Cartesian (X-Y-Z) coordinates and the Euclidean distance between coordinates of subject and autopilot were computed over the flight path $\Sigma sqrt(x^2+y^2+z^2)$ using a moving window average to resample and align the flight paths. The Euclidean distances for each sample were then summed in order to provide the total deviation from the autopilot flight path. Unlike G-force, this metric takes into account the entire approach, including all flight maneuvers leading up to the final descent and touchdown. This measure, however, does not take into account proficiency with aircraft controls or avionics; it merely assesses the ability of the subject to adhere to the reference flight path. Subjects were instructed to replicate the flight path of the autopilot landing observation. With this metric, a better landing would have lower deviation values (Supplementary Table S2, **Figure 5**).

Vertical speed deviation. The vertical speed of the subject throughout the landing trial was subtracted from the vertical speed of the autopilot landing at each time step and summed as in the flight path deviation. The vertical speed profile of the aircraft is stereotypic for an excellent landing and this parameter is visible on the aircraft's instruments. Subjects could therefore be reasonably expected to match the vertical speed of their aircraft with that of the example shown during autopilot

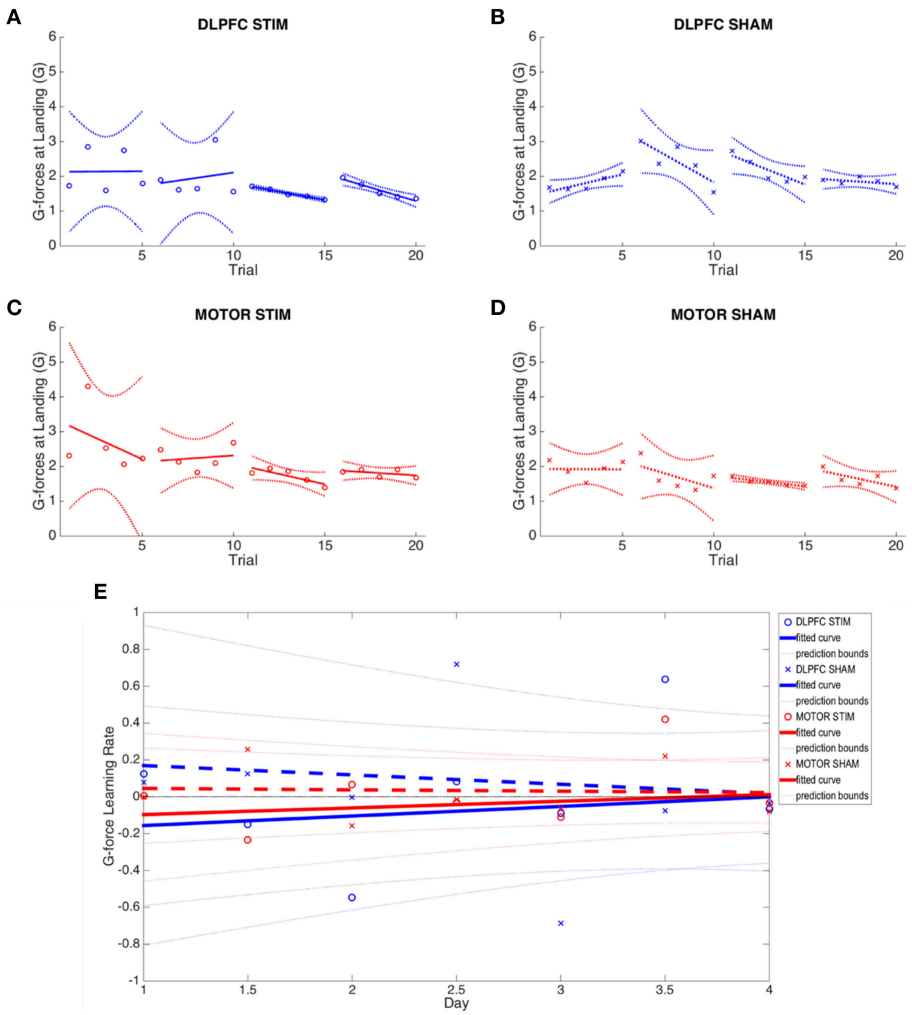


FIGURE 5 | G-force at moment of landing results across all four experimental groups. (A–D) Average g-force at moment of landing across all 4 days is plotted for each group. Note reduction in between-subject variance in days 3 and 4 of the DLPFC stim group. **(E)** Online and offline g-force learning rates are plotted for each experimental group across the duration of the experiment. Whole numbers on the x-axis represent the average online learning rate (slope of scaled percent correct linear regression for each subject across 6 blocks within a day) and ½ numbers on the x-axis represent offline learning rate (slope of the percent correct on the last trials of the N-1 day to the first trial of the Nth day). Smaller G-force indicates improved performance.

observation (vertical speed maintained at 600 ft./min for majority of approach, see **Figure 1D**). Replication of the autopilot-derived demonstration flight should result in lower overall vertical speed deviation values as performance improves (Supplementary Table S2, **Figure 5**).

Vertical speed variance. The amount of vertical speed variation throughout the landing approach was summed across trials to represent the degree to which a subject could maintain a steady, continuous descent. This measure does not penalize the subject for deviating from the ideal flight path, it merely assesses the degree to which the subject can maintain a smooth descent with little variation. This removes the goal-directed aspect of flight parameter maintenance while focusing on the motor aspect of flight parameter maintenance. As a means of comparison, the autopilot flight data only changes vertical speed in the final 5 s before landing, which minimizes this variance in the autopilot. In the ideal scenario, vertical speed stays constant, with only slight changes necessary for the final phase of landing; therefore, smaller variances indicate superior flight performance (Supplementary Table S2, **Figure 6**).

Control input measure. The number of control inputs was computed over landing trials by identifying the number of sign-changes in the vertical speed parameter throughout the landing period. This metric identifies to what extent subjects could maintain a consistent vertical speed profile (negative vertical speed indicates descent, positive ascent). Since maintaining vertical speed with minimal control input adjustment does not require specific planning of actions or prediction of flight path, it was hypothesized to be a primarily motor-processing focused measure. The number of sign-changes in the vertical speed variable was summed between start and end of the landing. As a means of comparison, the autopilot had 1 major control input at the nose flair ∼1 s. before touchdown.

Outlier rejection

For each metric, trial-wise data were examined for outliers across subjects across all groups. If any trial exceeded three standard deviations from the mean, it was determined an outlier and removed from analysis. Outlier rejection was performed on a trial-wise basis for all computed metrics.

Group variance analysis

For each metric, the variance in the average online learning rate was computed as the change in the group's average accuracy treating each subject's performance in a trial as a repeated sample within days. For the n-back task, the metric used was the scaled percent accuracy across 6 trials per day. For the easy landing task, flight metrics, performed over the course of 5 landing trials per day. This measure is the variance in the online learning rate linear regression (Reis et al., 2009). Significant differences in learning rate variance was assessed with a two-sample *F*-test for equal variances. The null hypothesis that two independent samples of two subject pools come from a single normal distribution with the same variance was tested against the alternative that they come from two normal distributions with different variances. F-stat criticality was computed by generating a F cumulative distribution function appropriate to the variance ratio and degrees of freedom of sample pools. The resulting critical values are asymmetric and can be used at either tail. We were then able to determine the distance between computed F and $F = 1$ (null hypothesis). Across days, Bartlett's test was performed to test the hypothesis of equal population variance across groups. This test was performed on the average subject metrics (across trials within days) to preserve sample independence. Reported *p*-values represent the probability of observing the given result by chance if the null hypothesis were true (Snedecor and Cochran, 1989).

RESULTS

Finger-Tapping Task

As expected, the finger-tapping task induced beta band oscillatory activity and increased the concentration of Hbdeoxy over sensorimotor cortex, and reduced alpha band activity over frontal and parietal cortex compared to baseline (Supplementary Figures S4, S5). Power in the beta band was greatest over left sensorimotor cortex, contralateral to the hand used during the finger-tapping task.

N-Back Task

Behavioral Results

DLPFC stimulation

The DLPFC stim group showed significant overall learning in five separate learning rate measures, compared to two significant learning rates observed for the DLPFC sham simulation group (one-sample, two-tailed *t*-test, see **Figures 3**, **4**, and Supplementary Table S1). Significant overall learning was observed for the DLPFC stim group collectively across all trial types (combining image and position match trials—denoted as "combined trials," and for position match trails) aggregating across all 1/2/3 back trials (scaled according to the methods in Section N-Back). Significant overall learning was also observed for the DLPFC stim group in 1-Back combined, position, and image trials. Significant overall learning was observed for the DLPFC sham group for combined and position trials, across all backs (see **Figure 3**). Meta-learning regressions did not show statistically significant changes in learning rates between stim and sham groups.

Neither the initial nor the final behavioral performance were significantly different between DLPFC stimulation and sham groups (Supplementary Table S1). The average trial duration to reach 2-/3-back was not significantly different between stimulation and sham groups. In addition, the average number of trials to reach 2-/3- back and the average 2-/3- back streak durations were not statistically different between groups. Significant differences in online, offline, and combined learning rates were not observed between stimulation and sham groups (see Supplementary Table S1, **Figure 4E**).

The variance in the DLPFC stim group's learning rate was significantly less than the variance of the DLPFC sham group. Examined across days, the DLPFC stim group had significantly reduced variance compared to the DLPFC sham group on

Day 3 of experimentation [$Chi_{(1)} = 5.77$, $p < 0.02$, see **Figures 4A,B**]. Examined at the trial-level, the reduced variance reached statistical significance in >33% of individual N-back trials comparing DLPFC stim with DLPFC sham, and no trials showed greater variance in the DLPFC sham group [Day 1 Trial 2: $F_{(6, 6)} = 0.21$; Day 1 Trial 4: $F_{(6, 6)} = 0.18$; Day 1 Trial 6: $F_{(6, 6)} = 0.23$; Day 2 Trial 5: $F_{(6, 6)} = 0.095$; Day 3 Trial 1: $F_{(5, 5)} = 0.069$; Day 3 Trial 2: $F_{(5, 5)} = 0.17$; Day 3 Trial 6: $F_{(5, 5)} = 0.15$; Day 4 Trial 6: $F_{(5, 4)} = 0.021$; $p < 0.05$]. These results support the hypothesis that tDCS of the right DLPFC would reduce the variability in individual learning rates in a cognitive task.

M1 stimulation

The M1 stim group showed significant overall learning in five separate learning rate measures, compared to six significant overall learning rates observed for the M1 sham stimulation group. Significant overall learning was observed for the M1 stim group for combined, image and position trails aggregating across all 1/2/3 back trials. Significant overall learning was also observed for M1 stim in 1-Back combined and image trails. Significant overall learning was observed for the M1 sham group for combined trails, aggregating across all backs and for 1 and 2-Back trials as well as for position trails (all and 2-Backs) and 2-Back image trials (see **Figure 3**). Meta-learning regressions did not show statistically significant changes in learning rates between stim and sham groups.

As with the DLPFC groups, initial and final behavioral performance between stimulation and sham groups was not significantly different (Supplementary Table S1). The average duration to reach 2-/3-back was not significantly different between stimulation and sham groups. In addition, the average number of trials to reach 2-/3- back and the average 2-/3- back streak durations were not statistically different between groups. Significant differences in online, offline, and combined learning rates were not observed between stimulation and sham groups (see Supplementary Table S1 and **Figures 3**, **4E**).

Unlike the results observed for DLPFC stimulation, M1 stimulation resulted in minimal differences in learning rate variance between stimulation and sham groups. Only 1 trial showed reduced M1 stim variance compared to M1 sham variation, while there were 3/24 trials that indicated smaller M1 sham variance when compared with the values from the M1 stim group. Examined across days, no trials shows significant differences in variance under Bartlett's Test.

FNIRS Results

DLPFC stimulation

Hboxy. Exclusion criterion for individual FNIRS channels were greater than 0.001 mM fluctuations between the maximum and minimum measured concentrations during baseline. We did not observe any concentration fluctuations above this cutoff threshold for any of the 20 FNIRS channels (10 above the DLPFC and 10 above the M1 cortex) across all 4 days of recording for the 25 subjects analyses (7 DLPFC stim, 7 DLPFC sham, 6 M1 stim, 5 M1 sham). Subjects were not included ($n = 3$ in M1 stim, and $N = 3$ in M1 sham) if event time stamps could not be identified robustly within the fNIRS data files.

Average Hboxy concentrations across subjects and channels significantly increased between day 1 and day 4 in M1 channels, and significantly decreased in DLPFC channels for the DLPFC sham group and (see **Table 2**, **Figure 7**). Individual channel analysis showed no significant change in Hboxy concentrations from days 1 to 4.

Hbdeoxy. Average Hbdeoxy concentrations across subjects and channels significantly decreased between day 1 and day 4 in M1 and DLPFC channels for the DLPFC sham group (see **Table 2**).

Hbtot. Like Hboxy, average Hbtot concentrations across subjects and channels significantly increased between day 1 and day 4 in M1 channels for the DLPFC sham group, and significantly decreased in DLPFC channels for the DLPFC sham group (see **Table 2**). Individual channel analysis showed no significant change in Hbtot concentrations from days 1 to 4.

M1 stimulation

Hboxy. The average Hboxy concentration across subjects and channels within the DLPFC channels significantly decreased between days 1 and 4 in the M1 stim group (see **Table 2**). Individual channel analysis shows no significant change in Hboxy concentrations from day 1 to 4.

Hbtot. The average Hbtot concentration across subjects and channels within the DLPFC channels significantly decreased between day 1 and 4 in the M1 stim group (see **Table 2**). Individual channel analysis shows no significant change in Hbtot concentrations from days 1 to 4.

EEG Results

Theta (4–7 Hz)

DLPFC stimulation. In each day, significant differences in theta-band power were found between DLPFC stim and sham tDCS groups in frontal/central electrodes (**Table 3**). In days 1–3, right frontotemporal theta power was higher in DLPFC stim participants, compared to sham. Statistical differences were distributed over midline frontal electrodes in day 4. Comparison of days 1 and 4 revealed a significant increase in midline frontal theta-band power in stim, but not sham participants (see **Figure 8A** and **Table 3**). Split-plot ANOVA comparing MFT in the N-back task revealed a trend-level main effect of tDCS group, with DLPFC stim participants showing a greater effect than DLPFC sham participants [$F_{(1, 12)} = 4.65$, $p = 0.052$]. There was no main effect of training or interaction between tDCS group and day of training ($p > 0.1$).

M1 stimulation. For M1 stimulation, broadly-distributed differences in theta-band power were seen between stim and sham participants during N-Back performance on days 1 and 3, which were mostly left-lateralized, and were strongest near the site of stimulation (**Table 3**). Importantly, no differences between days 1 and 4 were seen for M1 stim or sham participants in the N-Back (see **Table 3** and **Figure 8A**). Split-plot ANOVA

TABLE 3 | Cluster statistics for comparisons of alpha- and theta-band power during the N-back task.

	Electrodes	t^*	p^{**}
THETA (4–7 Hz)			
DLPFC stimulation *Actual vs. Sham* Day 1 Cluster 1	E5, FT8, T8, FC2, CP2, CP6	2.89	0.020
Day 2 Cluster 1	FT8, FC2, E11, CP2, CP6, PO10	3.15	0.006
Day 3 Cluster 1	FT8, E11, T8, CP6	3.11	0.020
Day 4 Cluster 1	AF3, F3, Fz	3.21	0.038
Day 4 vs. Day 1 Actual Cluster 1	FC1, Cz, E19	3.23	0.020
M1 stimulation *Actual vs. Sham* Day 1 Cluster 1	FC5, FC1, T7, Cz, C4, T8, E19, CP2, P7, Pz, P8, PO10	2.82	0.004
Day 3 Cluster 1	E5, T8, Cz, CP5, CP2, CP6, P7, Pz	2.27	0.020
Cluster 2	E5, FC2, C4, CP6	2.35	0.028
ALPHA (8–12 Hz)			
DLPFC stimulation *Actual vs. Sham* Day 2 Cluster 1	E5, FT8, E11, T8	2.85	0.008
Cluster 2	AF3, F7, F3	2.69	0.042
Day 4 Cluster 1	Pz, P8, Oz	2.74	0.048
Day 4 vs. Day 1 Actual Cluster 1	FC5, T7, CP5, P7	−3.12	0.032
M1 stimulation *Actual vs. Sham* Day 1 Cluster 1	F3, Fz, FC5, FC1, FC2, Cz	2.88	0.014
Day 3 Cluster 1	FC5, T7, CP5, P7	2.66	0.034

*Reported t-values are the average t-statistic across all electrodes in a given cluster.
**Reported p-values are corrected for multiple comparisons using cluster-based permutation tests.

comparing MFT revealed a significant main effect of day of training [$F_{(3, 48)} = 3.23$, $p = 0.048$]; however no significant or trend-level pairwise comparisons were found, There was no main effect of tDCS group or interaction between group and day of training.

Alpha (8–12 Hz)

DLPFC stimulation. Alpha-band power differences between DLPFC stim and sham groups were found in days 2 and 4 (**Table 3**). In day 2, frontal alpha power was greater in DLPFC stim than sham participants. In day 4, differences existed in parietal and occipital electrode sites, with DLPFC stim greater than sham. Differences between day 1 and 4 were found only for the DLPFC stim group, characterized by reduced alpha power at left temporoparietal sites (see **Table 3** and **Figure 8B**).

M1 stimulation. Greater alpha-band power was found for M1 stim compared to sham participants in days 1 and 3 (**Table 3**). These differences were distributed over frontal, central, and parietal electrode sites, mostly near the site of stimulation. No differences in alpha power were found in the comparison of days 1 and 4, for either M1 stim or M1 sham (see **Table 3** and **Figure 8B**).

EEG/fNIRS/behavioral correlations

We did not find any significant correlations between MFT or alpha power in the N-Back task and behavioral measures (i.e., average N level achieved and online/offline learning rates, p's $>$ 0.05). No correlations were identified between fNIRS beta values and either MFT or alpha power in the N-Back task for any group (p's > 0.1).

Easy Landing Task

G-Force

DLPFC stimulation

tDCS to DLPFC reduced the variability (standard deviation) of the third and fourth day online learning rates compared to sham (DLPFC stim: day 3 = 0.355, day 4 = 0.583; DLPFC sham: day 3 = 0.846, day 4 = 0.637, see **Figure 5**). First trial comparisons of variance between DLPFC stim and DLPFC sham groups in day 3 showed statistically significant changes in between-subject variance [$F_{(4, 5)} = 0.046$, $p < 0.02$]. There were no significant differences in variance for first trials of day 1 and day 2 ($p >$ 0.1). Trial 1 of day 4 also did not show significant changes in variance ($p > 0.1$). Examined across days, Bartlett's comparisons of variance between DLPFC stim and DLPFC sham groups in day 3 showed statistically significant changes in between-subject variance [$\text{Chi}_{(1)} = 7.33$, $p < 0.01$]. There were no significant differences in variance for days 1, 2, or 4 ($p > 0.1$). These results support the hypothesis that tDCS of the right DLPFC would reduce the variability in individual learning rates in the easy landing task.

Learning rates were determined by computing the rates at which performance improved (i.e., reduction of G Force over time, see **Figures 5A–D**). Meta-learning regressions did not show statistically significant changes in learning rates between stim and sham groups. The DLPFC stim group exhibited positive meta-learning rates (DLPFC stim = 0.052 ± 0.090), where the DLPFC sham group, by contrast, showed overall negative meta-learning (DLPFC sham = −0.051 ± 0.106), but this between-group difference did not reach statistical significance due to large within –group variance ($p > 0.1$; **Figure 5E**).

There were also no statistically significant differences in offline learning rates (p's > 0.1), but DLPFC stim showed a relatively strong offline learning between day 1 and 2 (−0.149 ± 0.52) compared to sham (0.124 ± 0.990).

The number of missed landings (did not land before or during the runway) was not different across groups over days (DLPFC stim: day 1: 2.9%, day 2: 2.9%, day 3: 0%, day 4: 0%; DLPFC sham:

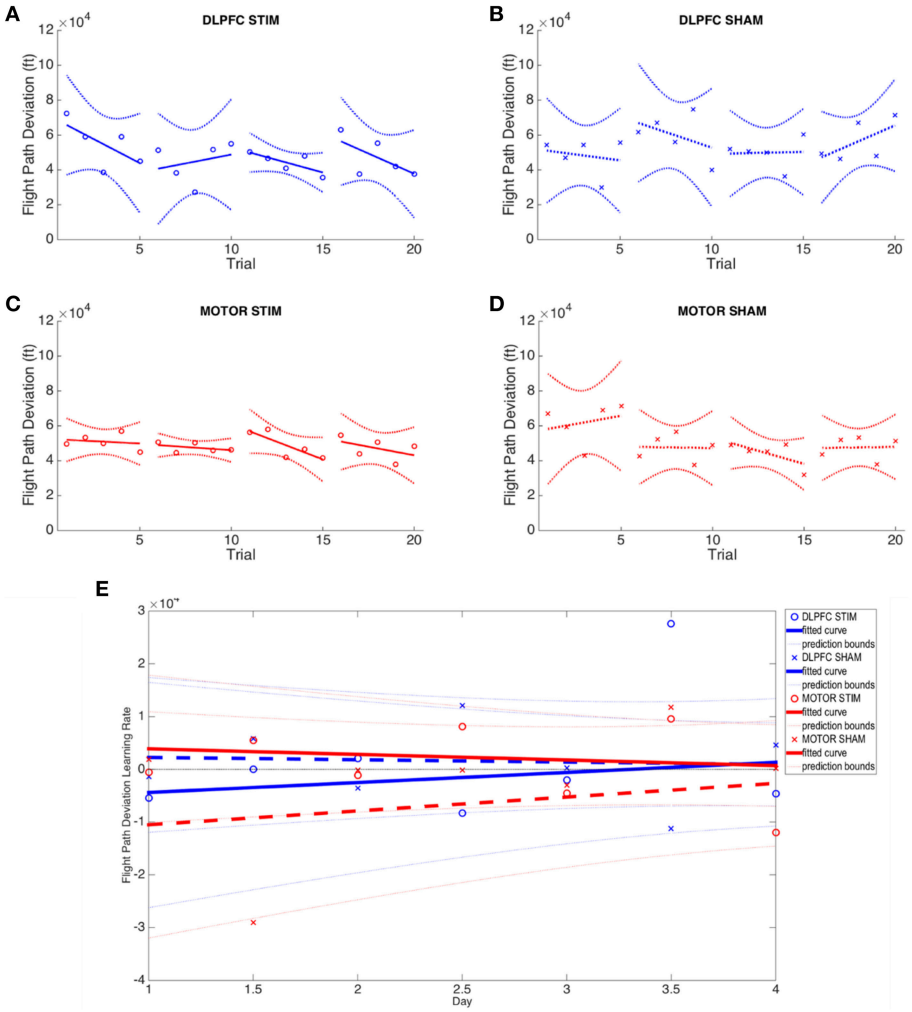


FIGURE 6 | Flight path deviation results across all four experimental groups. (A–D) Average flight path deviation of subjects from ideal autopilot-guided glide slope is shown for each group. **(E)** Online and offline flight path deviation learning rates are plotted for each experimental group across the duration of the experiment. Whole numbers on the x-axis represent the average online learning rate (slope of scaled percent correct linear regression for each subject across 6 blocks within a day) and ½ numbers on the x-axis represent offline learning rate (slope of the percent correct on the last trials of the N-1 day to the first trial of the Nth day). Reduced flight path deviation indicates improved performance.

day 1: 2.9%, day 2: 14.3%, day 3: 5.7%, day 4: 0%). Missed landings typically occurred on the first trial of the day.

M1 stimulation

tDCS to M1 resulted in no significant changes in inter-subject variance when compared to the M1 sham group across days (Bartlett's Test, p's > 0.1). M1 sham appeared to have unusually low variance during day 2 (**Figure 3D**), and this was determined to be a statistically significant reduction of variance when compared to the M1 stim group for three of the five trials of Day 2 [$F_{(8,\ 7)} = 6.32$, $F_{(8,\ 7)} = 3.86$, $F_{(8,\ 7)} = 5.07$; $p < 0.05$]. However, this reduction in variance only applied to Day 2 and in single trials only in Day 1 and 3. All trials on Day 4 had no significant change between M1 stim and sham variances. There were no statistically significant changes in learning rates between stim and sham groups (Supplementary Table S2). As with DLPFC stim group, there were also no statistically significant differences in offline learning rates (p's > 0.1), but M1 stim showed a relatively strong offline learning between day 1 and 2 (-0.235 ± 0.676) compared to sham (0.258 ± 1.330).

Initial starting (day 1 average), group averaged (across all days and the final (day 4 average) G-forces were not significantly different between experimental and control groups (paired t-test $p > 0.1$), which are similar to the results found with the N-back task. Though performance improved in both sham and stimulation cohorts (reduced overall G-force), the ultimate performance of each subject group was similar. It is probable that the landing task was effectively learned over the course of four training days, and subjects reached a G-force performance ceiling.

The number of missed landings were not different across groups over days (M1 stim: day 1: 10%, day 2: 10%, day 3: 2.0%,

day 4: 0%; M1 sham: day 1: 0%, day 2: 0%, day 3: 0%, day 4: 2.5%).

3D Autopilot Displacement

No group (DLPFC stim/sham nor M1 stim/sham) exhibited statistically significant learning (i.e., reduction of flight path displacement over time), as inter-subject variability was very high for this metric (**Figures 6A–D**). All groups showed inconsistent positive and negative learning slopes, and though the M1 stim group had negative online learning slopes for all 4 days, none of these reached statistical significance (p's>0.1). When comparing group variances for 3D autopilot displacement over experiment days, the M1 stim group showed greater variance compared with the M1 sham group during Day 1 [$\text{Chi}_{(1)} = 8.46$, $p < 0.01$]. The variance for the M1 stim group, however, was significantly lower than that of DLPFC stim group for 3 trials across 3 days of training [Day 1, Trial 1: $F_{(5,\ 9)} = 5.1032$; Day 2, Trial 4: $F_{(6,\ 9)} = 6.03$; Day 3 Trial 2: $F_{(6,\ 7)} = 3.99$; $p < 0.05$]. Interestingly, this was also true for M1 sham vs. DLPFC stim [Day 1, Trial 1: $F_{(5,\ 7)} = 13.67$; Day 2 Trial 4: $F_{(6,\ 7)} = 4.74$; $p < 0.001$]. No other variance comparisons yielded statistically significant results (Supplementary Table S2, **Figure 6**). While the combined learning intercept for the M1 sham group was negative (-13140 ± 13300, $p < 0.05$) this resulted from an isolated day 1/2 offline learning rate with a large standard deviation (-29011 ± 41147).

Vertical Speed Variance

The M1 stim group exhibited the lowest average values of vertical speed variance on the final day of training (4.955 ± 0.433; Supplementary Figure S2C). This is similar to the M1 sham (Supplementary Figure S2D) value of 5.051 ± 0.502 and an improvement over DLPFC groups (5.684 ± 0.690 and 5.647 ± 0.718, for stim and sham groups, respectively, but this does not reach significance under 2-way ANOVA ($p > 0.1$; Supplementary Figures S2A,B). This appears to be derived from the relatively higher online/offline learning rates in both M1 groups as compared with the DLPFC groups, though the overall rates were not statistically significant. ANCOVA, covarying the learning rates of vertical speed variance with group identity, performed on this data shows that the slopes appear identical ($p > 0.9$ vs. null hypothesis) but the initial performance (intercept) approaches significance ($p < 0.07$).

Online, offline and meta-learning rates were largely flat, and training effects were not observes within any group (p's < 0.1 level (Supplementary Table S2, Supplementary Figure S2). Because vertical speed variation is primarily a motor-centric task, it may be subject to a different learning curve that was not specifically measured during this study.

Autopilot Vertical Speed Deviation

M1 sham showed significant overall offline learning, with smaller deviations of vertical speed on Day 4 as compared with Day 1 (-12.21 ± 2.51, $p < 0.05$, Supplementary Figure S3D). DLPFC sham (Supplementary Figure S3B) also had a negative slope indicating reduced deviation from ideal vertical speeds, but this was not statistically significant (-69.71 ± 952.145, $p > 0.05$). However, both of these offline learning effects were washed out when combined into overall learning rates across the 4 days (M1 sham: -5.98 ± 20.26; DLPFC sham: -17.09 ± 25.80). Overall performance did not significantly change over the course of the 4 days, and initial/final performance were not significantly different across groups (p's > 0.1; Supplementary Figure S3E).

Unlike tests of G-force and flight path deviation, F-tests do not show any significant difference for inter-subject variance during 1st trial comparisons across all groups (p's > 0.1).

Number of Control Inputs

Variance between subjects for both DLPFC groups appeared larger than that of both M1 groups (Supplementary Figure S1). However, no meta, online, offline, or combined learning rates reach significance, and no significant changes were observed between groups (see Supplementary Table S2 and Supplementary Figure S1).

fNIRS Results

DLPFC Stimulation

Hboxy

Average Hboxy concentrations across subjects and channels significantly decreased between days 1 and 4 in M1 channels for the DLPFC stim and DLPFC sham groups, and decreased between days 1 and 4 in DLPFC channels for the DLPFC stim group (see **Table 2**, **Figure 7**).

Furthermore, individual channel analysis revealed that only two subjects (S7 in the DLPFC stim group, and S7 in the DLPFC sham group) showered a significant change in Hboxy concentrations from days 1 to 4 (−0.01 mM decrease at DLPFC channel: source AFF6h to detector F4 in DLPFC stim subject 7, and 0.02 mM increase at DLPFC channel: source FC4 to detector FFC4h in DLPFC sham subject 7, see **Figure 2** for channel locations). Within the DLPFC stim group, of all 70 channels measured across subjects in the DLPFC region (10 channels per subject, 7 subjects per group) 65 showed a decrease in Hboxy concentration from days 1 to 4 (compared to 40/70, 33/60, 325/50 for DLPFC sham, M1 stim, and M1 sham respectively).

Hbdeoxy

The average Hbdeoxy concentration across subjects and channels within M1 significantly increased between day 1 and day 4 in the DLPFC stim and DLPFC sham groups (see **Table 2**). Individual channel analysis shows no significant change in Hbdeoxy concentrations from days 1 to 4.

Hbtot

Average Hbtot concentrations across subjects and channels significantly decreased between days 1 and 4 in M1 channels for DLPFC stim and in the DLPFC channels for the DLPFC sham group, and increased in M1 channels for the DLPFC sham group (see **Table 2**). Individual channel analysis revealed that only one subject (S7 in the DLPFC stim group) showed a significant change in Hbtot concentrations from days 1 to 4 (−0.01 mM decrease at DLPFC channel: source AFF6h to detector F4, see **Figure 2**). Within the DLPFC stim group 64/70 channels in the DLPFC region showed a decrease in Hbtot concentration from

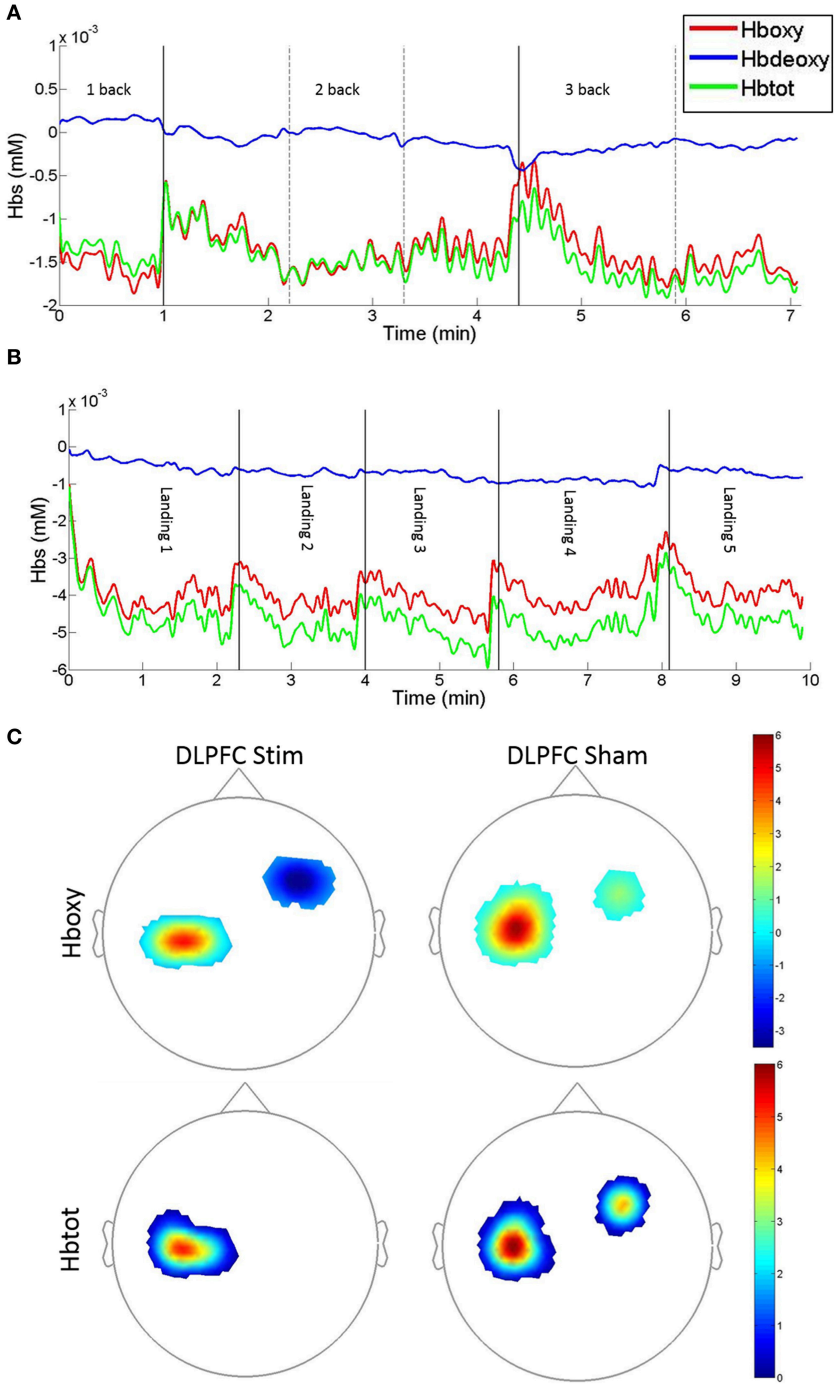


FIGURE 7 | Example fNIRS Hboxy, Hbdeoxy, and Hbtot concentration time-series and group average t-statistic beta maps. (A) N-back concentration time series for DLPFC sham subject #S6 recorded on day 1. Traces denote DLPFC channel #10: source: FC4, detector: F4. **(B)** Easy landing blocked concentration time series for DLPFC stim subject #S1 recorded on day 1. Traces denote DLPFC channel #5: source FFC6h, detector: F4 [vertical solid and dotted lines denote blocks, solid lines indicate 1/2 and 2/3-back block types in **(A)**, and easy landing trials in **(B)**]. An upward displacement in Hboxy and Hbtot concentrations can be seen during pauses between subsequent blocks. **(C)** fNIRS t-statistic beta maps of Day 4 vs. Day 1 Hboxy (top) and Hbtot (bottom) in the Easy landing task. Images are the averages for the DLPFC stim (left) and DLPFC sham (right) groups (Bonferroni corrected $p < 0.0025$, see **Table 2** for the corresponding concentration changes averaged over all channels within M1 and DLPFC).

days 1 to 4 (compared to 37/70, 34/60, 32/50 for DLPFC sham, M1 stim, and M1 sham respectively).

M1 Stimulation

Hboxy

Average Hboxy concentrations across subjects and channels significantly decreased between days 1 and 4 in M1 channels for the M1 stim group, and increased within M1 channels for the M1 sham group (see **Table 2**). Individual channel analysis shows no significant change in Hboxy concentrations from days 1 to 4.

Hbdeoxy

Average Hbdeoxy concentrations across subjects and channels significantly decreased between days 1 and 4 in M1 channels for the M1 stim group, and increased within M1 channels for the M1 sham group (see **Table 2**). Individual channel analysis shows no significant change in Hbdeoxy concentrations from days 1 to 4.

Hbtot

The average Hbtot concentration across subjects and channels significantly increased between days 1 and 4 in M1 channels for the M1 sham group (see **Table 2**). Individual channel analysis shows no significant change in Hbtot concentrations from days 1 to 4.

EEG

Theta (4–7 Hz)

DLPFC stimulation

In each day, significant differences in theta-band power were found between DLPFC stim and sham groups in frontal/central electrodes (**Table 4**). In days 1 and 3, right frontotemporal theta power was higher in DLPFC stim participants. Statistical differences were more broadly distributed in days 2 and 4, encompassing bilateral frontotemporal and midline frontal electrode sites. Comparison of days 1 and 4 revealed a significant increase in midline frontal theta-band power in DLPFC stim, but not DLPFC sham participants (see **Figure 8A** and **Table 4**). Split-plot ANOVA comparing MFT in the easy landing task revealed a significant main effect, with DLPFC stim greater than DLPFC sham [$F_{(1, 12)} = 4.86$, $p = 0.048$]. Additionally, an interaction was found between group and day of training [$F_{(3, 36)} = 4.54$, $p = 0.014$]. Simple-effects comparisons revealed increased MFT in stim compared to sham for only day 4 [day 4: $F_{(1, 12)} = 6.47$, $p = 0.026$]. Simple-effect of day within the DLPFC stim group reached trend-level significance [$F_{(3, 16)} = 3.15$, $p = 0.087$].

M1 stimulation

Theta-band differences between M1 groups during the easy landing task were found only in day 3, and were restricted to central/parietal electrodes (**Table 4**). Broadly-distributed differences in theta-band power were seen between days 1 and 4 in M1 stim participants, but not M1 sham participants (see **Table 4** and **Figure 8A**). No main effects or interactions were found in ANOVA comparing MFT in the easy landing task between M1 groups.

TABLE 4 | Cluster statistics for comparisons of alpha- and theta-band power during the Easy Landing task.

	Electrodes	t^*	p^{**}
THETA (4–7 Hz)			
DLPFC stimulation *Actual vs. Sham* Day 1 Cluster 1	E5, FT8, C4, CP6	2.58	0.024
Day 2 Cluster 1	AF3, F3, E5, FT8, FC1, FC2, E11, T7, Cz, C4, T8, E19, CP2, CP6	3.51	0.020
Day 3 Cluster 1	FT8, E11, T8, CP6	2.99	0.008
Day 4 Cluster 1	AF3, F3, FT8, FC1, FC2, E11, Cz, C4, T8, CP5, CP2, CP6, Pz, P8, Oz	3.81	0.002
Day 4 vs. Day 1 Actual Cluster 1	AF3, F3	3.23	0.044
M1 stimulation *Actual vs. Sham* Day 3 Cluster 1	Cz, CP2, CP5, P7, Pz	2.72	0.030
Day 4 vs. Day 1 Actual Cluster 1	F7, E5, FC5, FC1, T7, C4, CP5, E19, CP2, CP6, P7, Pz, P8, Oz, PO10	3.28	0.002
ALPHA (8–12 Hz)			
DLPFC stimulation *Actual vs. Sham* Day 1 Cluster 1	CP2, CP6, Pz, Oz	2.85	0.030
Day 4 Cluster 1	E5, FT8, Fz, FC2, E11, C4, T8	2.73	0.026
Cluster 2	FT8, E11, T8	3.56	0.040
Cluster 3	AF3, F3, Fz	2.94	0.048
M1 stimulation *Actual vs. Sham* Day 1 Cluster 1	F3, E5, Fz, FC5, FC1, FC2, C4, CP2, CP5, CP6	2.85	0.014
Day 4 vs. Day 1 Actual Cluster 1	CP5, E19, Pz	2.89	0.034
Cluster 2	E11, T8	−2.62	0.050

*Reported t-values are the average t-statistic across all electrodes in a given cluster.
**Reported p-values are corrected for multiple comparisons using cluster-based permutation tests.

Alpha (8–12 Hz)

DLPFC stimulation

Significant differences in alpha-band power were found between DLPFC stim and sham groups in parietal/occipital electrodes (day 1) and frontal/central electrodes (day 3), with greater power in the DLPFC stim group (**Table 4**). No differences in alpha power were found in the comparison of day 1 and 4, for either DLPFC stim or sham groups (see **Table 4** and **Figure 8B**).

M1 stimulation

Alpha-band differences between M1 groups during the easy landing task were found only in day 1, and were broadly distributed over frontal, central, and parietal electrode sites

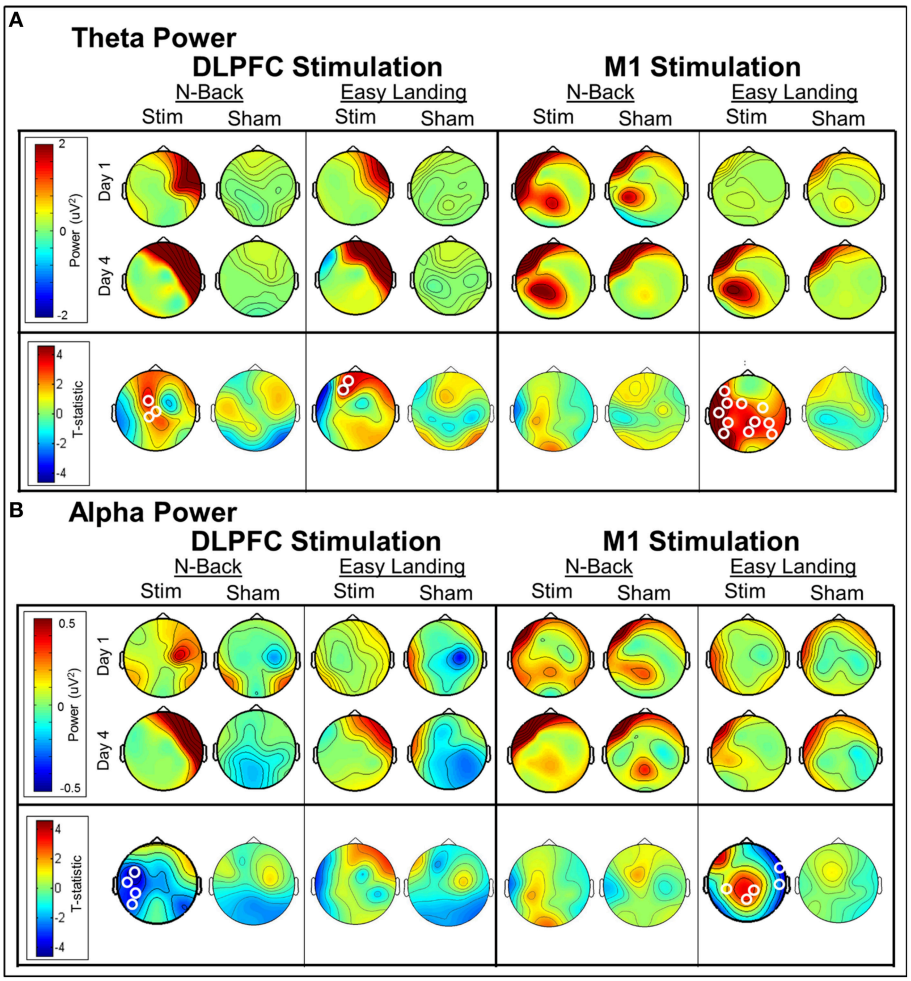


FIGURE 8 | Day 4 vs. day 1 average theta power (A) and alpha power (B) changes per group (t-statistic maps). Significant electrode clusters are depicted by white circles at electrode locations. Increases in midline frontal theta power were seen between days 1 and 4 in both tasks for all groups; however, cluster-level significance between days was found only for the DLPFC stim group. Decreased alpha power was found between days 1 and 4 for DLPFC stim in the N-Back task, where cluster-level significance was found in left temporoparietal electrodes. For M1 stimulation, cluster-level increases in central/parietal alpha and broadly-distributed theta power were found in the Easy Landing task.

(**Table 4**). Two separate clusters of significant differences in alpha-band power were seen between days 1 and 4 for M1 stim participants, but not M1 sham participants (see **Table 4** and **Figure 8B**). The first cluster revealed increased alpha power in the M1 stim group over parietal electrode sites. The second revealed decreased alpha power in right temporal electrodes.

EEG/fNIRS/behavioral correlations

There were positive correlations between change in MFT power (day 4 minus day 1) and both average Hbtot and average Hboxy beta values in DLPFC channels for M1 stim subjects. The direction of this correlation indicates that increased theta from days 1 to 4 is correlated with less reduction of Hboxy/Hbtot from days 1 to 4 in DLPFC fNIRS channels (**Table 5**). There were also strong negative correlations between change in alpha power and both average Hbtot and average Hboxy beta values in M1 channels for M1 stim subjects, indicating that increased parietal alpha power is correlated with reduced fNIRS beta values. No correlations were identified between theta/alpha power and fNIRS beta values for sham groups, there was no correlation between theta and fNIRS beta values at M1 channels, and there was no correlation between alpha power and fNIRS beta values at DLPFC channels (p's > 0.1).

DISCUSSION

Overview

In this study, we measured task-evoked changes in functional neural activity and the modulation of learning from tDCS to the right DLPFC or left M1. Simultaneous fNIRS and EEG measured changes in neural activity as subjects learned to complete flight simulator and n-back training exercises at increasing levels of expertise across four daily consecutive sessions. Assessment of

TABLE 5 | Significant correlations among EEG, fNIRS, and behavioral results in the Easy Landing task, for the M1 stim group.

	r	*p*
SIGNIFICANT EEG/BEHAVIOR CORRELATIONS		
Day 2	0.66	0.038
Online Learning		
MFT × Number of control inputs		
MFT × autopilot displacement	0.85	0.002
MFT × VSDFA	0.84	0.002
MFT × vertical speed variance	0.84	0.002
Offline learning	−0.85	0.002
MFT × autopilot displacement		
MFT × VSDFA	−0.85	0.002
MFT × vertical speed variance	−0.85	0.002
Day 4	−0.70	0.025
Online Learning		
Broadly-Distributed Theta Power × VSDFA		
Central/Parietal Alpha Power × VSDFA	−0.85	0.002
SIGNIFICANT EEG/fNIRS CORRELATIONS (DAY 4–DAY 1)		
MFT × Hbtot	0.82	0.045
MFT × Hboxy	0.81	0.051
Parietal Alpha Power × Hbtot	−0.95	0.005
Parietal Alpha Power × Hboxy	−0.94	0.005

VSDFA, Vertical Speed Deviance from Autopilot.

TABLE 6 | Summary of behavioral and neurophysiological results.

Group	N-Back	G-force	fNIRS	EEG
DLPFC Stim	- Variance	- Variance	- Hboxy and Hbtot in the DLPFC (flight only)	+ MFT power (N-back and flight)
M1 Stim				+ Parietal Alpha power (flight only)

behavioral performances were performed on n-back accuracy, flight metrics of landing performance, as well as for online and offline learning rates associated with practice and skill acquisition. We report that tDCS to the right DLPFC reduced the variability in online learning across individuals in the n-back task, and in g-force on the easy landing task. This was associated with decreased Hboxy and Hbtot in the DLFPC across days for the landing task, and increased MFT power in both the n-back and landing tasks. Additionally, tDCS to the left M1 increased tonic parietal alpha power, which was correlated with changes in Hboxy and Hbtot at M1 fNIRS channels.

Interpretation—Behavior

The observed reduction in group variability in online learning may be attributed to "convergence to the mean" (i.e., increasing online learning rates of low performing individuals and reducing online learning rates of high performing individuals). Subjects may have employed distinct cognitive and behavioral strategies, with correspondingly different brain networks, to complete and learn the n-back task across sessions. tDCS of the right DLPFC may have therefore facilitated the deployment and consolidation of a particular strategy in some subjects, and inhibited certain behaviors in others. The variance in the learning rates did not arise from individual differences of untrained performance, as initial and final performances were similar (see Section Behavioral Results). Furthermore, the results could indicate that all groups reached a ceiling of behavioral performance, or that our measures are under-powered to detect a change in performance statistically, or that a reduction in individual variability produced this observation.

The variability results reported for the easy-landing task were specific to DLPFC stim subjects for the g-force metric [a similar reduction in variance was not seen for the same data in the autopilot displacement (**Figure 6**), the number of control inputs (Supplementary Figure S1), the variability of vertical speed (Supplementary Figure S2), and the vertical speed deviation from autopilot (Supplementary Figure S3)]. Since both the initial and final g-force values were not significantly different across stim and sham groups, the reduction in DLPFC stim group variability implies a similar convergence to the mean phenomenon observed for n-back learning. tDCS of the DLPFC may therefore, facilitate the learning of a smoother landing procedure in subjects who would otherwise consolidated an incorrect landing procedure and increased landing g-forces in subsequent days. Likewise, tDCS of the DLPFC may have hindered some subjects who would have otherwise consolidated a superior landing procedure and decreased landing g-forces in subsequent days.

It should be noted that for the measure of 3D autopilot flight path deviation (Section 3D Autopilot Displacement), it was not readily apparent to subjects when the aircraft deviated from the prescribed flight path of the autopilot; there is no visual field indication that they are deviating from the glide slope, and the Flight Director instrument does not indicate degree of displacement from optimal glide slope. Additionally, for deviation from the autopilots vertical speed (Section Autopilot Vertical Speed Deviation) is possible that, because vertical speed was a peripheral skill required for landing (i.e., non-essential for a successful landing), subjects did not train to maintain a low vertical speed deviation from the reference glide path. As subjects needed only to maintain one constant vertical speed during the landing task, they may have reached maximal capacity to do so beginning from day 1. The combined learning rates and online learning metrics seem to support this view (see Supplementary Table S2, Supplementary Figure S3). Furthermore, low-G Force landings can be performed from a wide range of glide slopes, which can mask large deviations from the "ideal" flight path.

Interpretation—Neurophysiology

We observed an increase in MFT in the DLPFC stim group compared to the DLPFC sham group, as well as experience-related increase in MFT and decrease in central/parietal alpha in DLPFC stim, indicating increased working memory and attention (Klimesch et al., 1997; Jensen and Tesche, 2002; Ishii et al., 2014). Increased theta/alpha band activity in M1 stim compared to M1 sham near the site of stimulation may indicate greater motor cortex excitability (Sauseng et al., 2009). Furthermore, experience-related increases (day 4 vs. 1) in

broad central/parietal theta/alpha in M1 stim during flight tasks implicate greater tactile/proprioceptive monitoring. For example, Baumeister et al. (2008) observed that increased parietal theta during goal-directed learning was associated with increased motor skill performance. Although MFT nor parietal alpha power increases were correlated with behavioral performance increases in this study, it is possible that increases in MFT or parietal alpha may be indirectly associated with cognitive performance enhancement. The significant correlations observed between MFT and online and offline learning of autopilot displacement, vertical speed variance and deviation in the M1 stim group support this hypothesis (**Table 5**).

We observed a decrease in Hboxy and Hbtot in DLPFC channels for the DLPFC stim group in the easy landing task (**Table 2** and results Hboxy and Hbtot). This evidence suggests that tDCS produced more efficient neural activation to consolidate the newly-learned procedural skills as has been previously reported (Wolf et al., 2007; Holland et al., 2011; Ayaz et al., 2012; DiStasio and Francis, 2013). Previous literature from McKendrick et al. (2015) suggest that some, but not all, of these changes may be related to the task performance enhancements associated with tDCS. However, changes in Hbtot concentration may also be related to task reward value (DiStasio and Francis, 2013), the recruitment of additional motor resources (Herff et al., 2013), or a behavioral ceiling effect where low-performing subjects were not able to advance to expert performance levels as shown by Ayaz et al. (2012). Although reward was not explicitly manipulated in the easy landing task, subject's motivations may have played a role based on their prior day's performance. Similarly, the motor resources required for the easy landing task may have changed as subjects learned more advanced motor programs to complete the task. Finally, a ceiling effect could explain the more efficient neural activation, and the trend in meta-learning for stim groups supports this theory (**Figure 5E**).

Hbtot and Hboxy were also significantly correlated with MFT from days 4 to 1 in the easy landing task in the M1 stim group. These results suggest that these separate neurophysiological measures are not totally independent. Future studies should examine the relationships between MFT, Hbtot, and behavioral performance in a larger cohort to determine whether these effects are truly concomitant.

Relation to Prior Investigations of tDCS in Real-World Tasks

To date, there have been few studies in which procedural/real world learning tasks have been tested with a tDCS intervention (Izzetoglu et al., 2014; Nelson et al., 2014), and even fewer with a significant motor component as the focus of performance/training enhancement (Zhu et al., 2015). However, tDCS enhancement of real-world skills has been reported for complex motor control tasks. For example, Beeli et al. (2008) reported that anodal tDCS to either the left or right DLPFC (10/20 EEG site F3 or F4) significantly improved the care of driving style as measures by following distance, average speed and number of errors. Similarly, Sakai et al. (2014) reported that anodal tDCS to the right DLPFC significantly improved car-following and lane-keeping performance in a driving simulator task across days. Finally, Zhu et al. (2015) reported that cathodal tDCS to the left DLPFC suppressed verbal working memory but improved motor learning. The results presented here support these findings, as we observed that tDCS to the right DLPFC reduced online learning variability in higher cognitive measures (e.g., affecting the g-force value of landing by judging multi-modal flight-data in a timely fashion, or n-back accuracy variance) more than those related motor planning or judgment (e.g., flight path deviation see **Table 6**).

Furthermore, real-world skill enhancement from right inferior frontal tDCS has been reported in a perceptual threat detection (Clark et al., 2012; Falcone et al., 2012), and tDCS of the DLPFC has been shown to increase regional cerebral blood oxygenation and behavioral performance in target detection in an air traffic control task (Nelson et al., 2014). The results presented here are indirectly related to these findings as the reduction in behavioral variance we observed from tDCS to the right DLPFC could be attributed to increase in spatial attention, vigilance, or perceptual discrimination (e.g., when to judge an n-back match or the correct time for a nose-flare maneuver during landing). We also observed that tDCS of the right DLPFC decreased Hboxy and Hbtot in DLFPC channels across days in the easy landing task. One possible explanation for the difference reported in cerebral blood oxygenation between the two studies concerns the disparate experimental designs employed. Here, all subjects returned for four consecutive days of testing, regardless of physiological or behavioral measures, whereas Nelson et al. (2014) had subjects return for days 2–4 only if performance and blood flow velocity declined over the course of the first 40-min session.

Relation to Prior Investigations of tDCS in Working Memory

Previous studies have reported evidence that working memory improvements are correlated with the administration of tDCS in diverse contexts (Grafman et al., 1994; Nitsche et al., 2003; Dockery et al., 2009; McKendrick et al., 2015). Specifically, tDCS over DLPFC was associated with acute increases in working memory accuracy (Stagg and Johansen-Berg, 2013; Chhatbar and Feng, 2015; De Putter et al., 2015; Santarnecchi et al., 2015). Although, we observed a reduction in learning rate variance from tDCS to the right DLPFC in the n-back task, but did not find an increase in working memory accuracy for tDCS of either the DLPFC or M1. This discrepancy may be attributed to the adaptive n-back design employed here, the long durations of experimental sessions, and a potential ceiling effect from repeated tDCS and n-back sessions across consecutive days. In addition, the application of tDCS began directly prior to the n-back task (see **Figure 1**) and the effects of stimulation may require more time to produce the reported improvements in n-back accuracy.

Limitations and Future Directions

A goal of this research was to determine if tDCS stimulation would improve training techniques for pilots in a flight simulator.

Such improvements could drastically reduce time and therefore the cost of training a pilot, as it would in any training environment. While our results show decreased variability in training, it is too early to confirm or deny any useful improvements to simulation training until an understanding of the sources and contributing factors to the observed behavioral variance is achieved.

Additional studies must be performed to further investigate n-back accuracy improvement with tDCS by comparing different stimulation montages, stimulation timing, and task paradigms. Because we were unable to parametrically manipulate these parameters in this study, we are unable to determine which of these factors may have led to null effects of tDCS on n-back accuracy. The baseline performance of individuals with differing initial skill levels in n-back and flight tasks are important, and measures of this were limited by the study design employed. In addition, the experimental design employed here (continuous, multiple tasks over 60 min duration) did not provide a sufficient means to control the endogenous brain state of subjects before and throughout the experimental session given the numerous tasks, and instructions and feedback required for subjects to perform them. Thus, subject's diverse experiences and resultant brain states throughout the session may be a significant factor in the interpretation of our findings. For example, the n-back task was performed near the beginning of the stimulation period, while the easy flight landing was performed near the end of the stimulation period. Future studies should examine relationships between tDCS effects and EEG microstates and/or brain metabolic activity.

Some of the null findings in this study were related to exceptionally high within-group variance. One potential method to examine within and across group behavioral variance is to categorize subjects by learning rate bins or perform a cluster analysis of tDCS responders and non-responders. Since the same tDCS protocol may have variable effects across individuals, possibly due to neuroanatomical and neurophysiological differences, and that the same tDCS protocol may produce different effects within an individual over time, due to changes resulting from neural plasticity, the absence of *post-hoc* categorization of subjects likely reduces the statistical power and interpretability of our results (e.g., Supplementary Table S2). Future studies may benefit from real-time assessments and individualized tDCS planning rather than a "one size fits all" approach. While a priori selection or *post-hoc* classification of subjects within experimental groups can control for differences in baseline performance levels, it is not realistic when transferring this technology into real-world training environments.

The high variability between subjects and the need for personalized training becomes more important when we recognize the subject pool for this experiment all fit the western, educated, industrialized, rich and democratic (WEIRD) population. Although this population of subjects for the experiment goal of pilot training was acceptable, we speculate that the inclusion of a wider demographic range of the world populous may produce an even larger variability in behavioral performance. Therefore, a systematic understanding of the sources and contributing factors to the observed behavioral variance is extremely important for the application of tDCS across a wider range of subjects.

CONCLUSIONS

The results presented here underscore the importance of developing the understanding to identify and optimize neurostimulation protocols. Our results suggest that the time course of both online and offline learning is critical for the observed changes in working memory and procedural flight performance. Repeated training sessions reveal time-dependent factors regarding the interaction between tDCS and the learning processes that remain unclear in the literature. Applying such interventions in the real-world will require a much larger investment than initially anticipated in order for the scientific community to measure and catalog the precise behavioral, learning, and neurophysiological changes resulting from each component of procedural skill acquisition. Because there appears to be a differential, region-based effect of neurostimulation interventions, it is critical to determine the optimal targets, stimulation parameters, timing relative to the target behaviors, and synchrony between innate learning processes and strategies and exogenous stimulation for maximally-effective augmentation.

AUTHOR CONTRIBUTIONS

JC, MZ, and MP designed the experiments; JC and DB performed the experiments; JC, BC, and MP analyzed the data; and JC, BC, and MP wrote the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fnhum.2016.00034

REFERENCES

Ayaz, H., Çakır, M. P., Ýzzetoğlu, K., Curtin, A., Shewokis, P., Bunce, S. C., et al. (2012). "Monitoring expertise development during simulated UAV piloting tasks using optical brain imaging," in *Aerospace Conference, 2012 IEEE* (Big Sky, MT: IEEE), 1–11.

Ayaz, H., Onaral, B., Izzetoglu, K., Shewokis, P. A., McKendrick, R., and Parasuraman, R. (2013). Continuous monitoring of brain dynamics with functional near infrared spectroscopy as a tool for neuroergonomic research: empirical examples and a technological development. *Front. Hum. Neurosci.* 7:871. doi: 10.3389/fnhum.2013.00871

Ball, K., Berch, D. B., Helmers, K. F., Jobe, J. B., Leveck, M. D., Marsiske, M., et al. (2002). Effects of cognitive training interventions with older adults: a randomized controlled trial. *JAMA* 288, 2271–2281. doi: 10.1001/jama.288.18.2271

Banissy, M. J., and Muggleton, N. G. (2013). Transcranial direct current stimulation in sports training: potential approaches. *Front. Hum. Neurosci.* 7:129. doi: 10.3389/fnhum.2013.00129

Baumeister, J., Reinecke, K., Liesen, H., and Weiss, M. (2008). Cortical activity of skilled performance in a complex sports related motor task. *Eur. J. Appl. Physiol.* 104, 625–631. doi: 10.1007/s00421-008-0811-x

Beeli, G., Koeneke, S., Gasser, K., and Jancke, L. (2008). Brain stimulation modulates driving behavior. *Behav. Brain Funct.* 4:34. doi: 10.1186/1744-9081-4-34

Been, G., Ngo, T. T., Miller, S. M., and Fitzgerald, P. B. (2007). The use of tDCS and CVS as methods of non-invasive brain stimulation. *Brain Res. Rev.* 56, 346–361. doi: 10.1016/j.brainresrev.2007.08.001

Bell, H. H., and Waag, W. L. (1998). Evaluating the effectiveness of flight simulators for training combat skills: a review. *Int. J. Aviat. Psychol.* 8, 223–242. doi: 10.1207/s15327108ijap0803_4

Bikson, M., Datta, A., and Elwassif, M. (2009). Establishing safety limits for transcranial direct current stimulation. *Clin. Neurophysiol.* 120, 1033–1034. doi: 10.1016/j.clinph.2009.03.018

Blumberg, E. J., Foroughi, C. K., Scheldrup, M. R., Peterson, M. S., Boehm-Davis, D. A., and Parasuraman, R. (2014). Reducing the disruptive effects of interruptions with noninvasive brain stimulation. *Hum. Factors*. 57, 1051–1062. doi: 10.1177/0018720814565189

Boggio, P. S., Castro, L. O., Savagim, E. A., Braite, R., Cruz, V. C., Rocha, R. R., et al. (2006). Enhancement of non-dominant hand motor function by anodal transcranial direct current stimulation. *Neurosci. Lett.* 404, 232–236. doi: 10.1016/j.neulet.2006.05.051

Borghini, G., Astolfi, L., Vecchiato, G., Mattia, D., and Babiloni, F. (2014). Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. *Neurosci. Biobehav. Rev.* 44, 58–75. doi: 10.1016/j.neubiorev.2012.10.003

Braver, T. S., Cohen, J. D., Nystrom, L. E., Jonides, J., Smith, E. E., and Noll, D. C. (1997). A parametric study of prefrontal cortex involvement in human working memory. *Neuroimage* 5, 49–62. doi: 10.1006/nimg.1996.0247

Causse, M., Dehais, F., and Pastor, J. (2011). Executive funcitons and pilot characteristics predict flight simulator performance in general aviation pilots. *Int. J. Aviat. Psychol.* 21, 217–234. doi: 10.1080/10508414.2011.582441

Chhatbar, P. Y., and Feng, W. (2015). Data synthesis in meta-analysis may conclude differently on cognitive effect from transcranial direct current stimulation. *Brain Stimul.* 8, 974–976. doi: 10.1016/j.brs.2015.06.001

Choe, J., Coffman, B. A., Ziegler, M. D., Bhattacharyya, R., and Phillips, M. E. (2015). "tDCS modulates neuronal activity, dynamics and improves procedural learning and skill acquisition," in *Socity for Neuroscience Meeting* (Chicago, IL).

Chrysikou, E. G., Hamilton, R. H., Coslett, H. B., Datta, A., Bikson, M., and Thompson-Schill, S. L. (2013). Noninvasive transcranial direct current stimulation over the left prefrontal cortex facilitates cognitive flexibility in tool use. *Cogn. Neurosci.* 4, 81–89. doi: 10.1080/17588928.2013.768221

Clark, V. P., Coffman, B. A., Mayer, A. R., Weisend, M. P., Lane, T. D., Calhoun, V. D., et al. (2012). TDCS guided using fMRI significantly accelerates learning to identify concealed objects. *Neuroimage* 59, 117–128. doi: 10.1016/j.neuroimage.2010.11.036

Coffman, B. A., Clark, V. P., and Parasuraman, R. (2014). Battery powered thought: enhancement of attention, learning, and memory in healthy adults using transcranial direct current stimulation. *Neuroimage* 85, 895–908. doi: 10.1016/j.neuroimage.2013.07.083

Coffman, B. A., Trumbo, M. C., Flores, R. A., Garcia, C. M., Van Der Merwe, A. J., Wassermann, E. M., et al. (2012b). Impact of tDCS on performance and learning of target detection: interaction with stimulus characteristics and experimental design. *Neuropsychologia* 50, 1594–1602. doi: 10.1016/j.neuropsychologia.2012.03.012

Courtney, S. M., Ungerleider, L. G., Keil, K., and Haxby, J. V. (1996). Object and spatial visual working memory activate separate neural systems in human cortex. *Cereb. Cortex* 6, 39–49. doi: 10.1093/cercor/6.1.39

Cui, X., Bray, S., Bryant, D. M., Glover, G. H., and Reiss, A. L. (2011). A quantitative comparison of NIRS and fMRI across multiple cognitive tasks. *Neuroimage* 54, 2808–2821. doi: 10.1016/j.neuroimage.2010.10.069

Curtis, C. E., and D'Esposito, M. (2003). Persistent activity in the prefrontal cortex during working memory. *Trends Cogn. Sci.* 7, 415–423. doi: 10.1016/S1364-6613(03)00197-9

Datta, A., Baker, J. M., Bikson, M., and Fridriksson, J. (2011). Individualized model predicts brain current flow during transcranial direct-current stimulation treatment in responsive stroke patient. *Brain Stimul.* 4, 169–174. doi: 10.1016/j.brs.2010.11.001

Datta, A., Truong, D., Minhas, P., Parra, L. C., and Bikson, M. (2012). Inter-individual variation during transcranial direct current stimulation and normalization of dose using MRI-derived computational models. *Front. Psychiatry* 3:91. doi: 10.3389/fpsyt.2012.00091

Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21. doi: 10.1016/j.jneumeth.2003.10.009

Delorme, A., Palmer, J., Onton, J., Oostenveld, R., and Makeig, S. (2012). Independent EEG sources are dipolar. *PLoS ONE* 7:e30135. doi: 10.1371/journal.pone.0030135

De Putter, L. M., Vanderhasselt, M. A., Baeken, C., De Raedt, R., and Koster, E. H. (2015). Combining tDCS and working memory training to down regulate state rumination: a single-session double blind sham-controlled trial. *Cognit. Ther. Res.* 39, 754–765. doi: 10.1007/s10608-015-9710-8

Dismukes, R. K. (2008). "Prospective memory in aviation and everyday settings," in *Prospective Memory: Cognitive, Neuroscience, Developmental, and Applied Perspectives*, eds M. Kliegel, M. McDaniel, and G. Einstein (Florence, KY: Psychology Press), 411–428.

DiStasio, M. M., and Francis, J. T. (2013). Use of frontal lobe hemodynamics as reinforcement signals to an adaptive controller. *PLoS ONE* 8:e69541. doi: 10.1371/journal.pone.0069541

Dmochowski, J. P., Datta, A., Bikson, M., Su, Y., and Parra, L. C. (2011). Optimized multi-electrode stimulation increases focality and intensity at target. *J. Neural Eng.* 8:046011. doi: 10.1088/1741-2560/8/4/046011

Dockery, C. A., Hueckel-Weng, R., Birbaumer, N., and Plewnia, C. (2009). Enhancement of planning ability by transcranial direct current stimulation. *J. Neurosci.* 29, 7271–7277. doi: 10.1523/JNEUROSCI.0065-09.2009

Edwards, D., Cortes, M., Datta, A., Minhas, P., Wassermann, E. M., and Bikson, M. (2013). Physiological and modeling evidence for focal transcranial electrical brain stimulation in humans: a basis for high-definition tDCS. *Neuroimage* 74, 266–275. doi: 10.1016/j.neuroimage.2013.01.042

Engle, R. W. (2010). Role of working-memory capacity in cognitive control. *Curr. Anthrop* 51, S17–S26. doi: 10.1086/650572

Essenpreis, M., Elwell, C. E., Cope, M., Van der Zee, P., Arridge, S. R., and Delpy, D. T. (1993). Spectral dependence of temporal point spread functions in human tissues. *Appl. Opt.* 32, 418–425. doi: 10.1364/AO.32.000418

Falcone, B., Coffman, B. A., Clark, V. P., and Parasuraman, R. (2012). Transcranial direct current stimulation augments perceptual sensitivity and 24-hour retention in a complex threat detection task. *PLoS ONE* 7:e34993. doi: 10.1371/journal.pone.0034993

Faria, P., Hallett, M., and Miranda, P. C. (2011). A finite element analysis of the effect of electrode area and inter-electrode distance on the spatial distribution of the current density in tDCS. *J. Neural Eng.* 8:066017. doi: 10.1088/1741-2560/8/6/066017

Flöel, A., Rösser, N., Michka, O., Knecht, S., and Breitenstein, C. (2008). Noninvasive brain stimulation improves language learning. *J. Cogn. Neurosci.* 20, 1415–1422. doi: 10.1162/jocn.2008.20098

Gill, J., and Hamilton, R. (2014). Transfer of working memory performance enhancement induced by transcranial direct current stimulation (tDCS)(P3. 220). *Neurology* 82(10 Suppl.), P3–P220.

Grafman, J., Pascual-Leone, A., Alway, D., Nichelli, P., Gomez-Tortosa, E., and Hallett, M. (1994). Induction of a recall deficit by rapidrate transcranial magnetic stimulation. *Neuroreport* 5, 1157–1160. doi: 10.1097/00001756-199405000-00034

Hauser, T. U., Rotzer, S., Grabner, R. H., Mérillat, S., and Jäncke, L. (2013). Enhancing performance in numerical magnitude processing and mental arithmetic using transcranial Direct Current Stimulation (tDCS). *Front. Hum. Neurosci.* 7:244. doi: 10.3389/fnhum.2013.00244

Hays, R., Jacobs, J., Prince, C., and Salas, E. (1992). Flight simulator training effectiveness: a meta-analysis. *Military Psychol.* 4, 63. doi: 10.1207/s15327876mp0402_1

Heimrath, K., Sandmann, P., Becke, A., Müller, N. G., and Zaehle, T. (2012). Behavioral and electrophysiological effects of transcranial direct current

stimulation of the parietal cortex in a visuo-spatial working memory task. *Front. Psychiatry* 3:56. doi: 10.3389/fpsyt.2012.00056

Herff, C., Heger, D., Fortmann, O., Hennrich, J., Putze, F., and Schultz, T. (2013). Mental workload during n-back task—quantified in the prefrontal cortex using fNIRS. *Front. Hum. Neurosci.* 7:935. doi: 10.3389/fnhum.2013.00935

Holland, R., Leff, A. P., Josephs, O., Galea, J. M., Desikan, M., Price, C. J., et al. (2011). Speech facilitation by left inferior frontal cortex stimulation. *Curr. Biol.* 21, 1403–1407. doi: 10.1016/j.cub.2011.07.021

Horvath, J. C., Forte, J. D., and Carter, O. (2015). Quantitative review finds no evidence of cognitive effects in healthy populations from single-session transcranial direct current stimulation (tDCS). *Brain Stimul.* 8, 535–550. doi: 10.1016/j.brs.2015.01.400

Ishii, R., Canuet, L., Ishihara, T., Aoki, Y., Ikeda, S., Hata, M., et al. (2014). Frontal midline theta rhythm and gamma power changes during focused attention on mental calculation: an MEG beamformer analysis. *Front. Hum. Neurosci.* 8:406. doi: 10.3389/fnhum.2014.00406

Izzetoglu, K., Ayaz, H., Hing, J. T., Shewokis, P. A., Bunce, S. C., Oh, P., et al. (2014). "UAV operators workload assessment by optical brain imaging technology (fNIR)," in *Handbook of Unmanned Aerial Vehicles*, eds K. P. Valavanis and G. J. Vachtsevanos (Dordrecht: Springer Netherlands), 2475–2500.

Jaeggi, S. M., Buschkuehl, M., Jonides, J., and Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proc. Natl. Acad. Sci. U.S.A.* 105, 6829–6833. doi: 10.1073/pnas.0801268105

Jensen, O., and Tesche, C. D. (2002). Frontal theta activity in humans increases with memory load in a working memory task. *Eur. J. Neurosci.* 15, 1395–1399. doi: 10.1046/j.1460-9568.2002.01975.x

Klimesch, W. (1999). EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain Res. Rev.* 29, 169–195. doi: 10.1016/S0165-0173(98)00056-3

Klimesch, W., Doppelmayr, M., Pachinger, T., and Ripper, B. (1997). Brain oscillations and human memory: EEG correlates in the upper alpha and theta band. *Neurosci. Lett.* 238, 9–12. doi: 10.1016/S0304-3940(97)00771-4

Koonce, J. M., and Bramble, W. J. Jr. (1998). Personal computer-based flight training devices. *Int. J. Aviat. Psychol.* 8, 277–292. doi: 10.1207/s15327108ijap0803_7

Lefebvre, S., Thonnard, J. L., Laloux, P., Peeters, A., Jamart, J., and Vandermeeren, Y. (2014). Single session of dual-tDCS transiently improves precision grip and dexterity of the paretic hand after stroke. *Neurorehabil. Neural Repair*. 28, 100–110. doi: 10.1177/1545968313478485

Manenti, R., Brambilla, M., Petesi, M., Ferrari, C., and Cotelli, M. (2013). Enhancing verbal episodic memory in older and young subjects after non-invasive brain stimulation. *Front. Aging Neurosci.* 5:49. doi: 10.3389/fnagi.2013.00049

Martin, D. M., Liu, R., Alonzo, A., Green, M., and Loo, C. K. (2014). Use of transcranial direct current stimulation (tDCS) to enhance cognitive training: effect of timing of stimulation. *Exp. Brain Res.* 232, 3345–3351. doi: 10.1007/s00221-014-4022-x

Maxwell, S. E., and Delaney, H. D. (2004). *Designing Experiments and Analyzing Data: A Model Comparison Perspective*, Vol. 1. Florence, KY: Psychology Press.

McIntire, L. K., McKinley, R. A., Goodyear, C., and Nelson, J. (2014). A comparison of the effects of transcranial direct current stimulation and caffeine on vigilance and cognitive performance during extended wakefulness. *Brain Stimul.* 7, 499–507. doi: 10.1016/j.brs.2014.04.008

McKendrick, R., Ayaz, H., Olmstead, R., and Parasuraman, R. (2014). Enhancing dual-task performance with verbal and spatial working memory training: continuous monitoring of cerebral hemodynamics with NIRS. *Neuroimage* 85, 1014–1026. doi: 10.1016/j.neuroimage.2013.05.103

McKendrick, R., Parasuraman, R., and Ayaz, H. (2015). Wearable functional near infrared spectroscopy (fNIRS) and transcranial direct current stimulation (tDCS): expanding vistas for neurocognitive augmentation. *Front. Syst. Neurosci.* 9:27. doi: 10.3389/fnsys.2015.00027

Mendonca, M. E., Santana, M. B., Baptista, A. F., Datta, A., Bikson, M., Fregni, F., et al. (2011). Transcranial DC stimulation in fibromyalgia: optimized cortical target supported by high-resolution computational models. *J. Pain* 12, 610–617. doi: 10.1016/j.jpain.2010.12.015

Meyers, K. (1979). Drug company employees as research subjects: programs, problems, and ethics. *IRB* 1, 5–6. doi: 10.2307/3564165

Miller, J., Berger, B., and Sauseng, P. (2015). Anodal transcranial direct current stimulation (tDCS) increases frontal-midline theta activity in the human EEG: a preliminary investigation of non-invasive stimulation. *Neurosci. Lett.* 588, 114–119. doi: 10.1016/j.neulet.2015.01.014

Moliadze, V., Antal, A., and Paulus, W. (2010). Electrode-distance dependent after-effects of transcranial direct and random noise stimulation with extracephalic reference electrodes. *Clin. Neurophysiol.* 121, 2165–2171. doi: 10.1016/j.clinph.2010.04.033

Nelson, J. T., McKinley, R. A., Golob, E. J., Warm, J. S., and Parasuraman, R. (2014). Enhancing vigilance in operators with prefrontal cortex transcranial direct current stimulation (tDCS). *Neuroimage* 85, 909–917. doi: 10.1016/j.neuroimage.2012.11.061

Nitsche, M. A., Schauenburg, A., Lang, N., Liebetanz, D., Exner, C., Paulus, W., et al. (2003). Facilitation of implicit motor learning by weak transcranial direct current stimulation of the primary motor cortex in the human. *J. Cogn. Neurosci.* 15, 619–626. doi: 10.1162/089892903321662994

Oostenveld, R., Fries, P., Maris, E., and Schoffelen, J. M. (2011). FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comp. Intel Neurosci.* 2011:156869. doi: 10.1155/2011/156869

Palva, J. M., Monto, S., Kulashekhar, S., and Palva, S. (2010). Neuronal synchrony reveals working memory networks and predicts individual memory capacity. *Proc. Natl. Acad. Sci. U.S.A.* 107, 7580–7585. doi: 10.1073/pnas.0913113107

Pascual-Leone, A., and Hallett, M. (1994). Induction of errors in a delayed response task by repetitive transcranial magnetic stimulation of the dorsolateral prefrontal cortex. *Neuroreport* 5, 2517–2520. doi: 10.1097/00001756-199412000-00028

Phillips, M. E., and Ziegler, M. D. (2014). A neurostimulation-based advanced training system for human performance augmentation. *Brain Stimul.* 7, e11–e12.2013 doi: 10.1016/j.brs.2014.01.041

Reis, J., Fischer, J. T., Prichard, G., Weiller, C., Cohen, L. G., and Fritsch, B. (2015). Time- but not sleep-dependent consolidation of tDCS-enhanced visuomotor skills. *Cereb. Cortex* 25, 109–117. doi: 10.1093/cercor/bht208

Reis, J., Schambra, H. M., Cohen, L. G., Buch, E. R., Fritsch, B., Zarahn, E., et al. (2009). Noninvasive cortical stimulation enhances motor skill acquisition over multiple days through an effect on consolidation. *Proc. Natl. Acad. Sci. U.S.A.* 106, 1590–1595. doi: 10.1073/pnas.0805413106

Richmond, L. L., Wolk, D., Chein, J., and Olson, I. R. (2014). Transcranial direct current stimulation enhances verbal working memory training performance over time and near transfer outcomes. *J. Cogn. Neurosci*. 26, 2443–2454. doi: 10.1162/jocn_a_00657

Robertson, E. M., Pascual-Leone, A., and Miall, R. C. (2004). Current concepts in procedural consolidation. *Nat. Rev. Neurosci.* 5, 576–582. doi: 10.1038/nrn1426

Rosenkopf, L., and Tushman, M. L. (1998). The coevolution of community networks and technology: lessons from the flight simulation industry. *Indust. Corp. Change* 7, 311–346. doi: 10.1093/icc/7.2.311

Rumpf, J. J., Hinselmann, K., Fricke, C., Weise, D., and Classen, J. (2015). P148. Timing of post-training transcranial direct current stimulation to enhance consolidation of motor sequence learning in older people. *Clin. Neurophysiol.* 126, e135–e136. doi: 10.1016/j.clinph.2015.04.219

Ruocco, A. C., Rodrigo, A. H., Lam, J., Di Domenico, S. I., Graves, B., and Ayaz, H. (2014). A problem-solving task specialized for functional neuroimaging: validation of the Scarborough adaptation of the Tower of London (S-TOL) using near-infrared spectroscopy. *Front. Hum. Neurosci.* 8:185. doi: 10.3389/fnhum.2014.00185

Sakai, H., Uchiyama, Y., Tanaka, S., Sugawara, S. K., and Sadato, N. (2014). Prefrontal transcranial direct current stimulation improves fundamental vehicle control abilities. *Behav. Brain Res.* 273, 57–62. doi: 10.1016/j.bbr.2014.07.036

Santarnecchi, E., Brem, A. K., Levenbaum, E., Thompson, T., Kadosh, R. C., and Pascual-Leone, A. (2015). Enhancing cognition using transcranial electrical stimulation. *Curr. Opin. Behav. Sci.* 4, 171–178. doi: 10.1016/j.cobeha.2015.06.003

Sauseng, P., Klimesch, W., Doppelmayr, M., Hanslmayr, S., Schabus, M., and Gruber, W. R. (2004). Theta coupling in the human electroencephalogram during a working memory task. *Neurosci. Lett.* 354, 123–126. doi: 10.1016/j.neulet.2003.10.002

Sauseng, P., Klimesch, W., Gerloff, C., and Hummel, F. C. (2009). Spontaneous locally restricted EEG alpha activity determines cortical excitability in the motor cortex. *Neuropsychologia* 47, 284–288. doi: 10.1016/j.neuropsychologia.2008.07.021

Snedecor, G. W., and Cochran, W. G. (1989). *Statistical Methods*, 8th *Edn*. Iowa City, IA: Iowa State University Press.

Snowball, A., Tachtsidis, I., Popescu, T., Thompson, J., Delazer, M., Zamarian, L., et al. (2013). Long-term enhancement of brain function and cognition using cognitive training and brain stimulation. *Curr. Biol.* 23, 987–992. doi: 10.1016/j.cub.2013.04.045

Sohn, Y. W., and Doane, S. M. (2004). Memory processes of flight situation awareness: interactive roles of working memory capacity, long-term working memory, and expertise. *Hum. Factors* 46, 461–475. doi: 10.1518/hfes.46.3.461.50392

Stagg, C. J., and Johansen-Berg, H. (2013). Studying the effects of transcranial direct-current stimulation in stroke recovery using magnetic resonance imaging. *Front. Hum. Neurosci.* 7:857. doi: 10.3389/fnhum.2013.00857

Teo, F., Hoy, K. E., Daskalakis, Z. J., and Fitzgerald, P. B. (2011). Investigating the role of current strength in tDCS modulation of working memory performance in healthy controls. *Front. Psychiatry* 2:45. doi: 10.3389/fpsyt.2011.00045

Valle, A., Roizenblatt, S., Botte, S., Zaghi, S., Riberto, M., Tufik, S., et al. (2009). Efficacy of anodal transcranial direct current stimulation (tDCS) for the treatment of fibromyalgia: results of a randomized, sham-controlled longitudinal clinical trial. *J. Pain Manag.* 2, 353–361.

Villringer, A., Planck, J., Hock, C., Schleinkofer, L., and Dirnagl, U. (1993). Near infrared spectroscopy (NIRS): a new tool to study hemodynamic changes during activation of brain function in human adults. *Neurosci. Lett.* 154, 101–104. doi: 10.1016/0304-3940(93)90181-J

Williams, B. (2012). *Scenario-Based Training with X-Plane and Microsoft Flight Simulator: Using PC-Based Flight Simulations Based on FAA-Industry Training Standards*. Hoboken, NJ: Wiley and Sons.

Wolf, M., Ferrari, M., and Quaresima, V. (2007). Progress of near-infrared spectroscopy and topography for brain and muscle clinical applications. *J. Biomed. Opt.* 12, 062104–062104. doi: 10.1117/1.2804899

Xu, Y., Graber, H., and Barbour, R. (2014). "nirsLAB: a computing environment for fNIRS neuroimaging data analysis," in *Biomedical Optics OSA Technical Digest (Online)* (Miami, FL: Optical Society of America), paper BM3A.1.

Yanagisawa, H., Dan, I., Tsuzuki, D., Kato, M., Okamoto, M., Kyutoku, Y., et al. (2010). Acute moderate exercise elicits increased dorsolateral prefrontal activation and improves cognitive performance with Stroop test. *Neuroimage* 50, 1702–1710. doi: 10.1016/j.neuroimage.2009.12.023

Zhu, F. F., Yeung, A. Y., Poolton, J. M., Lee, T. M., Leung, G. K., and Masters, R. S. (2015). Cathodal transcranial direct current stimulation over left dorsolateral prefrontal cortex area promotes implicit motor learning in a golf putting task. *Brain Stimul.* 8, 784–786. doi: 10.1016/j.brs.2015.02.005

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Choe, Coffman, Bergstedt, Ziegler and Phillips. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Human Neuroscience

REVIEW
published: 24 June 2016
doi: 10.3389/fnhum.2016.00284

Does a Combination of Virtual Reality, Neuromodulation and Neuroimaging Provide a Comprehensive Platform for Neurorehabilitation? – A Narrative Review of the Literature

Wei-Peng Teo[1], Makii Muthalib[2,3], Sami Yamin[4,5], Ashlee M. Hendy[6], Kelly Bramstedt[4], Eleftheria Kotsopoulos[4,7], Stephane Perrey[2] and Hasan Ayaz[8,9,10]*

[1] Institute for Physical Activity and Nutrition (IPAN), Deakin University, Burwood, VIC, Australia, [2] EuroMov, University of Montpellier, Montpellier, France, [3] Cognitive Neuroscience Unit, Deakin University, Burwood, VIC, Australia, [4] Liminal Pty Ltd., Melbourne, VIC, Australia, [5] Adult Mental Health, Monash Health, Dandenong, VIC, Australia, [6] School of Exercise and Nutrition Sciences, Deakin University, Burwood, VIC, Australia, [7] Aged Persons Mental Health Service, Monash Health, Cheltenham, VIC, Australia, [8] School of Biomedical Engineering, Science and Health Systems, Drexel University, Philadelphia, PA, USA, [9] Department of Family and Community Health, University of Pennsylvania, Philadelphia, PA, USA, [10] The Division of General Pediatrics, Children's Hospital of Philadelphia, Philadelphia, PA, USA

OPEN ACCESS

Edited by:
Mikhail Lebedev, Duke University, USA

Reviewed by:
Richard B. Reilly, Trinity College Dublin, Ireland
Zoltan Nadasdy, NeuroTexas Institute Research Foundation, USA
Ilya Boristchev, IT Universe, Russia

***Correspondence:**
Wei-Peng Teo
weipeng.teo@deakin.edu.au

Received: *08 December 2015*
Accepted: *25 May 2016*
Published: *24 June 2016*

Citation:
Teo W-P, Muthalib M, Yamin S, Hendy AM, Bramstedt K, Kotsopoulos E, Perrey S and Ayaz H (2016) Does a Combination of Virtual Reality, Neuromodulation and Neuroimaging Provide a Comprehensive Platform for Neurorehabilitation? – A Narrative Review of the Literature. Front. Hum. Neurosci. 10:284. doi: 10.3389/fnhum.2016.00284

In the last decade, virtual reality (VR) training has been used extensively in video games and military training to provide a sense of realism and environmental interaction to its users. More recently, VR training has been explored as a possible adjunct therapy for people with motor and mental health dysfunctions. The concept underlying VR therapy as a treatment for motor and cognitive dysfunction is to improve neuroplasticity of the brain by engaging users in multisensory training. In this review, we discuss the theoretical framework underlying the use of VR as a therapeutic intervention for neurorehabilitation and provide evidence for its use in treating motor and mental disorders such as cerebral palsy, Parkinson's disease, stroke, schizophrenia, anxiety disorders, and other related clinical areas. While this review provides some insights into the efficacy of VR in clinical rehabilitation and its complimentary use with neuroimaging (e.g., fNIRS and EEG) and neuromodulation (e.g., tDCS and rTMS), more research is needed to understand how different clinical conditions are affected by VR therapies (e.g., stimulus presentation, interactivity, control and types of VR). Future studies should consider large, longitudinal randomized controlled trials to determine the true potential of VR therapies in various clinical populations.

Keywords: neurorehabilitation, neuroplasticity, tDCS, fNIRS, EEG, virtual reality therapy

INTRODUCTION

In the last two decades, the application of VR training has become increasingly popular, not only as a means to enhance gaming experiences, but also in the education and healthcare settings to improve learning and rehabilitation outcomes. Particularly in the area of neurorehabilitation, the use of VR technology has shown great promise by providing a sense of realism during training, thereby promoting skill acquisition and retention, and inducing functional recovery (**Figure 1**; for review, see Adamovich et al., 2009).

In the context of neurorehabilitation, VR therapy can be described as a method of brain–computer interaction that involves real-time simulation of an environment, scenario or activity that allows for user interaction and targets multiple senses. In particular, the combination of VR and recent technological advances in robotic and haptic interfaces allow users a seemingly life-like interactional experience in a VE (Jung et al., 2012; Yeh et al., 2014). For example, VR has been used in clinical settings as a training tool for surgeons to learn intricate fine motor skills associated with precision surgery (Wang, 2012; Fang et al., 2014), and as a tool to deliver cognitive-based therapies (Kim et al., 2011; Kandalaft et al., 2013). More complex forms of VR presentation such as augmented VR (whereby VR is superimposed on the actual environment) and immersive VR (first-person interaction in a VR environment) brings the immersive experience to another level with technology such as head-mounted displays (i.e., Oculus® Rift and Microsoft® Hololens) or screens. It is through this naturalistic environment, and allowing for interactive behaviors while being monitored and recorded, that is the primary advantage of implementing VR technology. This means that VR technology can be used to deliver meaningful and relevant stimulation to an individual's nervous system and thereby capitalize on neuroplasticity to promote both cognitive and motor rehabilitation.

In this review, we will discuss the theoretical framework for the use of VR in the context of neurorehabilitation. We will provide evidence for the use of VR in motor rehabilitation for neurological disorders such as PD, CP and stroke and in mood and mental health disorders such as anxiety, PTSD and schizophrenia. We will also review the concurrent use of non-invasive brain stimulation and neuroimaging techniques during VR, discussing how these combined techniques may augment the benefits and complement current VR training protocols.

THEORETICAL FRAMEWORK FOR VR AND LEARNING

Experiential Learning

The most important aspect of using VR is to provide new experiences by allowing users to interact physically and emotionally within a VE that is almost identical to the real world. The combination of physical, mental and emotional interaction encourages active participation and involvement of the user. In this sense, users of VR assimilate knowledge more effectively when they have the freedom to engage in self-directed activities within their learning context. By finding solutions and learning new skills autonomously, users of VR invest mental effort by constructing conceptual models that are both consistent with what they already understand and with the new content that is presented (Garrison and Garrison, 1997). Another key feature of VR training is that it offers users the opportunity to acquire skills in the context where they need to be applied. This results in more meaningful and effective learning, as compared with learning out of context (Nieuwenhuijsen et al., 2006). In physical rehabilitation for example, rehabilitation of fine motor control of the hands and wrists can be "re-trained" by simulating a VE where a stroke patient needs to pour him or herself a glass of water in the kitchen. In this way, patients practice and refine fine motor control of muscles controlling the hands and wrists through manipulating a virtual object that allows the same kind of natural interaction with objects that patients would engage in the real world.

Augmented Feedback: Knowledge of Results and Performance

Another important aspect of VR therapy is the ability to provide augmented feedback to its users. Augmented feedback is additional information provided through any means (e.g., visual, auditory or kinesthetic) that is complimentary to the inherent feedback received via the sensory systems. There is no hard and fast rule as to how or what kinds of information augmented feedback should provide, however, VR therapy offers two vital pieces of information that is essential for learning (Winstein, 1991; Lauber and Keller, 2014); (1) knowledge of performance – information on how the participant performs during movement (i.e., movement sequences, joint angles, force outputs at each phase of movement etc.); (2) knowledge of results – information on the outcome of the performance (i.e., overall quality and quantity of movement). Currently, most commercially available VR games would incorporate visual, auditory and even kinesthetic feedback that can be provided either during or after the game. Very often, these VR games are designed in a manner that users have to maintain or achieve a pre-determined score or level in order for the game to progress. For example, VR applications can provide knowledge of performance throughout gameplay in the form of movement kinematics (i.e., joint angles, velocity, and speed), kinetics (i.e., ground reaction forces and torque) or even the level of activation in specific brain regions during a particular task. In order for users to progress to the next level, users must maintain or exceed a threshold that has been set based on previous trials or specific performance outcomes. Upon task completion, knowledge of results and performance can be provided, allowing both clinicians and users to understand deficiencies in movement patterns that are associated with specific dysfunctional movement outcomes, apply progressions

Abbreviations: BCI, brain computer interface; CP, cerebral palsy; EEG, electroencephalography; fMRI, functional magnetic resonance imaging; fNIRS, functional near-infrared spectroscopy; PD, Parkinson's disease; PTSD, post-traumatic stress disorder; tDCS, transcranial direct current stimulation; VE, virtual environment; VR, virtual reality.

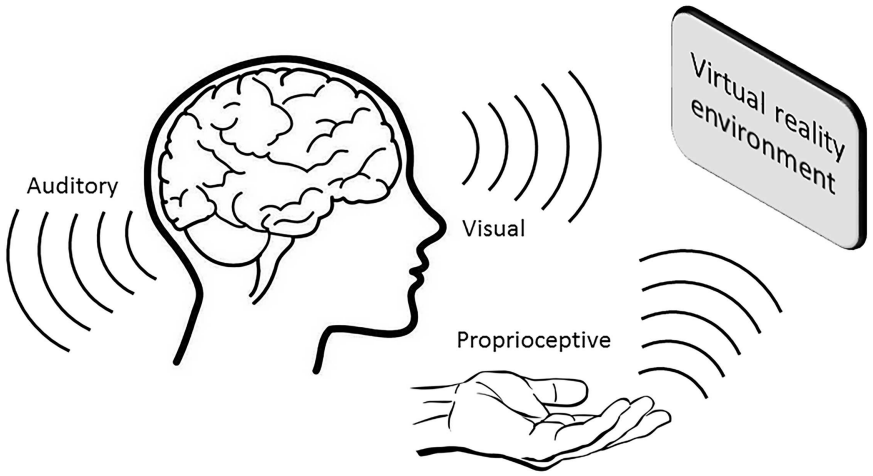


FIGURE 1 | A VR environment induces multisensory feedback that contributes toward greater memory consolidation and retention.

appropriately, and address those deficiencies with a targeted rehabilitation approach.

Observational Learning

Apart from providing feedback that is necessary for learning, another aspect of VR training is the enhancement of observational learning. The basis for learning, or at least its intended outcome, is to mimic or replicate an ideal response that brings about a desired result and induce a lasting change in behavior. In terms of neurorehabilitation, observation of goal-oriented movements or processes provides sensory feedback about the movement, behavior or emotional state, which contributes to learning (Oouchida et al., 2013; Williams and Carnahan, 2014). These observations preferentially activate parts of the brain that are involved with the physical performance itself allowing a motor program to be developed based upon the observed movements (Burke et al., 2010). Training in a VE may facilitate observational learning in four different ways; (1) VR applications can provide an accurate visual representation of the user's body and limb position using motion capture technology; (2) VR applications commonly use an avatar to mimic the movement of users, or conversely, the user could mimic the movements of the avatar; (3) accurate guides or a correct movement pattern can be produced for which users can follow; (4) VR applications can facilitate mental imagery by inducing optimal mood states and instructions for mental imagery.

Motivation

Importantly, the goal-oriented nature of VR tasks may support the maintenance and adherence of neurorehabilitation programs. Unlike traditional therapist-led sessions, where improvements in physical or cognitive function may be subjective or difficult for patients to identify (Van den Broek, 2005), VR programs can provide an objective, quantitative measure of session outcomes and objectives. Furthermore, VR applications can provide both users and clinicians the ability to individualize training programs or alter the progression of a training session based upon the user's personal performance. The capacity to individualize therapy intensity may enhance motivation by allowing users to select practice sessions that are catered to their individual time and need, and more importantly, to manipulate treatment parameters to create optimal learning conditions. Another important consideration for VR to improve motivation is by incorporating competition or co-operation between other players during therapy sessions. Engaging users in a group environment either competing against each other or working in teams promotes an element of enjoyment through increased social interaction, particularly amongst people suffering with similar conditions (Van den Broek, 2005).

EVIDENCE OF VR THERAPY IN MOVEMENT NEUROREHABILITATION

Stroke

The use of motion-controlled VR game consoles, including the Nintendo® Wii and Xbox® Kinect, have been explored as adjuncts to conventional physical therapy (see **Table 1**), specifically for improving upper limb function (Thomson et al., 2014; Laver et al., 2015). VR programs for stroke neurorehabilitation are based on the potential for brain neuroplasticity after neurological injury to support acquisition and retention of new motor skills to recover motor function. The goal of VR therapy in stroke is to apply these motor learning principles for stroke neurorehabilitation, such as providing repetitive, graded intensity, and motivating task-specific training with real time multimodal feedback of movements and performance (Saposnik et al., 2011). Thus, VR systems are designed to enhance conventional therapy by providing a tool to deliver more specific, intensive and enjoyable therapy with real time feedback of performance (Levin et al., 2015).

TABLE 1 | Examples of recent systematic reviews and meta-analyses demonstrating the effects of VR in neurorehabilitation of stroke, PD and CP.

Author and year	Study aims	Studies included and sample (*n*)	Study outcomes	Points of discussion
Stroke				
Laver et al. (2015)	Compared the effects of virtual reality on arm function, walking speed and independence in managing daily activities after stroke versus an alternative intervention or no intervention.	37 studies ($n = 1019$)	12 studies found improved arm function. 4 studies found improved walking speed. 8 studies found slight improvements in activities of daily living.	Low sample size in most studies. Some studies reported pain, headaches or dizziness in small number of participants, but no adverse events overall. Low quality evidence for arm function. Very low quality evidence for walking ability, global motor function and independence in performing daily activities. The quality of the evidence for each outcome was limited due to small numbers of study participants, inconsistent results across studies and poor reporting of study details.
Corbetta et al. (2015)	Compared the effects of VR-based rehabilitation on gait, balance and mobility versus standard therapy.	15 studies ($n = 341$)	Significant improvements in walking speed, balance, and mobility. Significant improvements in mobility if VR training was combined with standard therapy. Insufficient evidence to support to use of combined VR and standard therapy on balance and walking speed.	Substituting some or all of a standard rehabilitation regimen with VR training provides greater benefits in walking speed, balance, and mobility. Although the benefits are small, the cost of administering VR is also small particularly when patient demand is high in a clinic setting.
Luque-Moreno et al. (2015)	Compared the effects of VR interventions on lower extremity rehabilitation.	11 studies ($n = 231$)	High heterogeneity in study designs. Small sample sizes. Mean sample size of 20 per study. Studies were ranked between 4 and 7 points (out of 10) on the PEDro scale.	VR interventions (more than 10 sessions) may have a positive impact on lower limb function. Multimodal approach (i.e., a combination of VR and conventional therapy) may elicit greater results. Adaptability of software seemed to adapt better to patient's requirements, allowing for individualized treatments.
Lohse et al. (2014)	Compared the effects of custom built virtual games and commercially available gaming systems.	26 studies ($n = ?$)	Only 4 studies used commercial games while 20 studies used custom built virtual games. Mean PEDro score for all studies was 5.42 ± 1.6 (out of 10). Methodological limitations of studies include subject, experimenter and therapist blinding, small sample size, and difficulty in determining a dose-response effect. Significant improvements in body function and activity outcomes.	VR intervention improves outcomes compared to conventional therapies. Small samples and few number of studies in commercial games limits the assessment of potential benefits.
Parkinson's disease				
Harris et al. (2015)	Compared the effects of exergaming on static and dynamic balance in older adults and PD.	11 studies ($n = 325$ healthy older adults, 56 PD)	9 studies showed a significant improvement in static balance and postural control in healthy aging individuals. 2 studies found a significant improvement in static balance and postural control people with PD. Studies were ranked between 4 and 8 points (out of 10) on the PEDro scale.	Few studies in PD and small sample size limits the interpretation of the effectiveness of exergaming in PD. Evidence found in this meta-analysis supports the use of exergaming as an adjunctive tool to improve balance and postural control.

(Continued)

TABLE 1 | Continued

Author and year	Study aims	Studies included and sample (*n*)	Study outcomes	Points of discussion
Barry et al. (2014)	Examined the safety, feasibility and effectiveness of exergaming in people with PD.	7 studies ($n = 110$)	Only 2 studies addressed patient safety. No objective measures (such as falls or near falls) or subjective measures (patient's perception) were recorded in any studies. Only 1 study recorded gameplay experience. Good levels of motivation during game play were reported although difficulties with the fast pace and cognitive complexity of some games were raised. Exergaming was found to be just as effective as standard physical therapy for improving clinical measures of balance and cognition even up to 60 days post-intervention.	While the effectiveness and feasibility are often measured, more research is required to establish the safety, particularly in home-based VR therapy. The use of commercial games may be too difficult for some people with PD, and exergames that tailor specifically to the needs and capabilities of patients may be more effective.
Cerebral Palsy				
Dewar et al. (2015)	Systematic review of various interventions to improve postural control in children with CP	45 studies ($n = ?$)	4 studies investigated the use of VR on postural control. 2 studies were rated weak in study conduct while 2 had a strong study design. 3 studies showed improvements in balance, while 2 study showed improvements in walking capacity.	The systematic review provided conflicting evidence of VR on postural control and gait. Due to the preliminary nature of these studies, it is difficult to truly ascertain if indeed the use of VR had any effects on postural control and gait.
Chen et al. (2014)	Examined the effects of virtual gaming on upper extremity function in children with CP	14 studies ($n = 122$)	3 RCTs, 2 cohort studies, 7 case studies and 2 single-subject design studies. For 3 RCTs, no difference was found between VR therapy and conventional therapy. Overall upper extremity function was significantly improved after VR therapy. Strongest effects of VR was shown in younger children, custom-built systems in the home or laboratory setting.	The use of VR may be highly applicable in a pediatric population. Small sample size and the lack of large RCT is a limiting factor in interpreting the results.

PEDro, physiotherapy evidence database (PEDro); RCT, Randomized controlled trials; PD, parkinson's disease; CP, cerebral palsy; VR, virtual reality.

Despite the potential utility of commercial VR game consoles for stroke neurorehabilitation, a number of limitations have been highlighted (Bower et al., 2015): (1) VR games designed for the general population can be too challenging for stroke patients with physical and cognitive deficits; (2) the difficulty levels and control of VR games are often not readily adjustable to rehabilitation targets, and the tasks may lack functional relevance; (3) feedback and scoring provided can be negative and frustrating for the user; (4) current VR games do not include neurological assessment; (5) VR does not integrate multiple environmental factors that connect to motor performance. In response to some of these limitations, there has been an emergence of research and development of modified VR programs specifically designed for stroke neurorehabilitation using adaptable software and hardware components of commercial VR systems (e.g., Kinect system) and guidance from clinicians in their development (Laffont et al., 2014; Bower et al., 2015). These adapted VR systems are progressively optimized with new functions including: (1) allowing automatic adaptation/intensity grading of the activity to the patient's own achievements; (2) allowing the therapist to adapt online the task's characteristics to the patient's needs; (3) allowing multiplayer VR systems via a web-service platform to enhance interactivity; (4) automatic recording of the patient's movements to provide therapists with data describing the quality and quantity of motor function recovery/progression, including the level of compensatory movements (Laffont et al., 2014).

While there is some evidence to suggest that VR may be highly applicable for stroke rehabilitation, the evidence from recent systematic reviews and meta-analyses indicate that current studies are limited by sample size issues and study designs (see **Table 1**). Early VR interventions used commercial applications such as the Nintendo Wii that controls an avatar, however, more recently customized systems have focused on interactive platforms to target activities of everyday living (i.e., reaching and grasping tasks). However, a major challenge with stroke is that no one stroke patient will present with the same motor deficit and therefore an individualized approach to therapy, including VR therapy, is needed. In this sense, future systems must be adaptive and customizable to manage

the heterogenous nature of stroke for patients to gain greater benefits.

Parkinson's Disease

Emerging VR therapies presents as an attractive option for delivering neurorehabilitation therapies to manage the cognitive-motor symptoms in people with PD, as it can be employed at any stage as an adjunct to standard pharmacological (Levodopa therapy) and/or surgical (ablation, deep brain stimulation) treatment (see **Table 1**). This new possibility in the field of neurorehabilitation aims to provide PD patients with a motivating way to perform multiple motor neurorehabilitation exercises with the rationale that the VR system might promote balance training, and cognitive-motor practice. Some commercial VR systems, such as the Nintendo® Wii system using a balance board, has drawn considerable attention from both the research and clinical communities as effective and feasible neurorehabilitation interventions to enhance gait and balance for people with PD (Barry et al., 2014; Harris et al., 2015). More recent studies have implemented custom programming and hardware to their VR systems to specifically improve balance and gait in PD (Mirelman et al., 2011).

It has been demonstrated that a VR neurorehabilitation program of 6–8 weeks involving 40–60 min a day, three times per week appears to be a viable option for significantly improving balance in a clinical population of individuals with PD (Esculier et al., 2012). The intensity/difficulty load of interventions used across existing studies appears a key contributing factor for the discordant findings reported in the literature (Esculier et al., 2012). In addition, activity selection could have contributed to some differential findings among studies. Some studies targeted static slow controlled movements in a closed environment such as the Wii Fit with balance board (dos Santos Mendes et al., 2012; Esculier et al., 2012), while others involve dynamic movements in an open environment such as Wii Sports (Herz et al., 2013).

Despite some evidence for performance improvement in balance, there are still limitations inherent in commercial VR systems that may not directly apply to realistic everyday settings for PD neurorehabilitation. Additionally, the programs are not very scalable, or modifiable, to each individual's needs or progress for all stages of disease. Mirelman et al. (2011, 2013) utilized a custom-made VR system to incorporate virtual obstacles presented on a screen during treadmill walking (18 sessions over 6 weeks). During the gait training they used a novel method (V-TIME) for tracking foot position based on the X-box Kinect technology. Interestingly, Mirelman et al. (2011) observed significant elevated gait speed with and without a cognitive dual-task upon completion of training and 4 weeks post-training. However, this VR gait training protocol confines participants to straight-walking, a gait pattern that is relatively uncommon in real-life environments. Perhaps a more viable approach may be the development of a VR system that may be used in conjunction with activities of daily living. People with PD are known to use visual and/or auditory cues to improve physical performance (Lee et al., 2012), and perhaps the use of augmented VR, via goggles or smart glasses, may be used to provide sensory cues as a feedforward or feedback mechanism to improve physical performance.

Cerebral Palsy

Cerebral Palsy (CP) is the most common pediatric physical disability, thought to affect three to four individuals per 1000 of the population (Aisen et al., 2011; Oskoui et al., 2013) characterized as a spectrum of disorders of motor and postural development that cause limited functionality or dysfunction (Monge Pereira et al., 2015). Studies investigating exercise-based treatments for children with CP has provided growing evidence in the last decades for effectiveness in improving postural control (see **Table 1**). Although effective, traditional physical exercise in the clinical settings consists of repetitive tasks that limits the enthusiasm over regular periodic application.

While the study of VR in children with CP is still at its infancy, Denise Reid at the University of Toronto's Virtual Reality Laboratory (Reid, 2002a,b, 2004) has provided preliminary evidence to support its use. In these studies, children with CP were engaged with VR based exercises for upper extremity and postural control. The self-reported effect of VR on perceived self-efficacy to perform given tasks was tested in an uncontrolled study, before and after intervention (Reid, 2002a). Based on the self-efficacy theory, Reid (2002b) attempted to identify if use of VR could increase the motivation for exercise in children with CP. The pilot study yielded encouraging results for VR use with improvement in perceived performance abilities and satisfaction with performance. In a follow up study on upper-extremity efficiency, improvement was also reported with VR use (Reid, 2002b). Similarly, Reid (2004) later investigated the effect of VR intervention on playfulness and found that VR environments stimulated playfulness in children, specifically the VR tasks that allowed creativity, expression, and choice of activity. You et al. (2005) used fMRI in a case report to investigate cortical reorganization and associated motor function improvement after a VR therapy. Neuroplastic changes were observed in the primary sensorimotor cortex and supplementary motor area following VR therapy, together with enhanced functional motor skills. A later study by Bryanton et al. (2006) compared the VR therapy with conventional exercises in children and found that although children completed more repetitions of the conventional exercises, the range of motion and hold time in stretched position was greater during VR tasks.

While the current research for VR in children is still in the early stages, VR therapies represent a viable option to increase exercise adherence and physical activity as they are both engaging and rewarding particularly in an adolescent population. The process of gamification, one that entails an interactive dynamic storyline and an overall goal, is likely to better capture and retain the attention of children over traditional physical training (Lister et al., 2014). The challenge, particularly in children with CP, will be to incorporate a diversity of activities performed during the game to train a repertoire of fundamental skills so as to further develop their motor and cognitive skills.

EVIDENCE FOR VR THERAPY IN COGNITIVE REHABILITATION AND MENTAL HEALTH

Anxiety, Phobias and Post-traumatic Stress Disorder

Anxiety can be generalized in nature [i.e., generalized anxiety disorder (GAD)], characterized by long-lasting anxiety that is not focused on a specific object, or may be more focal (i.e., phobias) occurring in the presence of, or in anticipation of, a specific object or situation. Preliminary evidence on the use of VR in GAD indicate that a combination of relaxation, controlled exposure and stress inoculation may help patients to cope with various stressors and sources of worry (Gorini et al., 2010; Repetto and Riva, 2011). Additionally, the combination of biofeedback (e.g., heart rate and electro-dermal skin response) may potentially help to identify particular sources of worry and emotion that can be used to modify specific features of the VR environment (Gorini et al., 2010; Repetto and Riva, 2011).

Despite the limited evidence for the use of VR therapy in GAD, there is some support for the use of VR in a range of other anxiety disorders (see **Table 2**) including specific phobias (Cote and Bouchard, 2005; Maskey et al., 2014), panic disorder (Vincelli et al., 2002) and social phobia (Klinger et al., 2004, 2005). Current VR therapies, particularly for phobias, use controlled exposure therapy that allows the patient to experience a sense of presence in an immersive, interactive VE that minimizes avoidance behavior and facilitates emotional involvement. This VE also allows controlled delivery of sensory stimulation via the therapist, for which the patient confronts the feared stimuli in a progressive manner. Another advantage of VR therapy is being able to recreate situations that cannot be re-experienced *in vivo* (i.e., combat situation or terrorist attack). VR therapy may be used as an alternative to imaginal exposure, meaning that patients with PTSD need not rely on internal imagery to visualize an event. A potential limitation in imaginal exposure therapy is that the therapist has no control over, or even knowledge of, what imagery the patient actually evokes (Strosahl and Ascough, 1981). Whereas in the VE, the stimuli presented can be carefully controlled and monitored. As with phobic patients, VR-based exposure therapy may be particularly useful for patients with PTSD for whom avoidance and failure to engage with therapy may hinder the therapeutic process. The efficacy of VR therapy for the treatment of PTSD has predominantly been examined in military populations (Cukor et al., 2009; Goncalves et al., 2012). A systematic review found that VR therapy was just as efficacious as traditional exposure treatment for PTSD (Goncalves et al., 2012). Seven of the 10 studies included in Goncalves's review found that VR environments significantly reduced PTSD symptoms in comparison to control, however, no significant differences in symptoms were observed between VR therapy and traditional exposure treatment.

Whilst the existing literature on the use of VR therapy for phobias, panic disorder, and PTSD were promising, several limitations must be considered (Powers and Emmelkamp, 2008; Meyerbroker and Emmelkamp, 2010; Opris et al., 2012). Meyerbroker and Emmelkamp (2010) noted that VR as a therapeutic tool is difficult to assess as it is often combined with other techniques. This potentially masks any underlying benefits of VR therapy on the patient. Furthermore, most studies do not include behavioral avoidance tasks, which would help to determine how transferable the results are to the real world.

Schizophrenia

As an assessment tool, VR offers the possibility of creating unique environments allowing researchers to better identify and understand specific areas of the brain commonly effected in schizophrenia. It is proposed that binding errors during the memory encoding process are responsible for the episodic memory impairments reported in schizophrenia (Waters et al., 2004). In this sense, VR is able to tease out areas of the brain responsible for binding impairments by providing specific situations or tasks for which patients have to perform. For example, a study by Ledoux et al. (2013) examined contextual binding in schizophrenia using fMRI during a navigation task in a virtual town (i.e., find the grocery store from the school). Their results showed significantly less activation among patients relative to controls in the left middle frontal gyrus, and right and left hippocampi. Ledoux et al. (2013) further suggested that the reduced activation was indicative of context and content not being appropriately linked, therefore affecting the formation of a cognitive map representation in the patient group and eliciting a contextual binding deficit.

As a rehabilitation tool, VR offers a unique potential to expose individuals to controlled rehabilitation environments and allow for interaction within a VE. Indeed, VEs may be perceived as less intimidating for patients as it allows for more gradual increase in task difficulty and may therefore enhance participation with rehabilitation (Rizzo et al., 1998). In particular, VR therapy has been explored as an alternative option to improve cognitive function (Marques et al., 2008) and vocational skills (Tsang and Man, 2013) in schizophrenic patients with some success. However, perhaps one of the most important roles of VR therapy may be to attenuate the deficit in social skills associated with schizophrenia. Traditionally, social skills training using role-play has been effective in remediating these deficits (Benton and Schroeder, 1990), however, role-playing of social skills training are limited in that they require appropriately matched groups, and may produce social anxiety, negative symptoms and poor insight. VR-based techniques offer an alternative to traditional role playing techniques by providing a computer-generated but realistic three-dimensional world and human-like avatars that can provide emotional stimuli. These VR-based techniques may be highly beneficial to re-train conversational skills (i.e., beginning a conversation, breaking silences, and differentiating facial expressions; Ku et al., 2007; Park et al., 2011).

While there is great potential for the role of VR in the treatment of schizophrenia, the evidence for its use remains contentious. Questions still remain if the effects of VR directly affects the condition itself, or perhaps the effects of VR may attenuate other psychiatric comorbidities such as anxiety or

TABLE 2 | Examples of systematic reviews and meta-analyses demonstrated the use of VR in treating PTSD and anxiety disorders.

Author and year	Study aims	Studies included and sample (n)	Study outcomes	Points of discussion
PTSD				
Kuester et al. (2016)	Examined the efficacy of internet-based CBT and expressive writing in people with PTSD vs. waitlist or active controls.	20 studies ($n = 973$ intervention, 805 controls)	Internet-based CBT are showed medium to large effect sizes compared to passive controls, but not against active controls receiving face-to-face CBT with therapist.	Internet-based CBT may be just as beneficial as conventional CBT. Due to large variability in outcome measures of included studies, subgroup analyses was limited.
Goncalves et al. (2012)	A systematic review of the efficacy of VR exposure therapy in the treatment of PTSD vs. waitlist or active controls.	10 studies ($n = ?$)	Patients in VR exposure therapy showed insignificantly better results compared to waitlist controls, but no differences was observed when compared to exposure therapy. Majority of VR exposure therapy used head-mounted displays and customized virtual environment specific to the condition. No difference in dropout rates between VR therapy and conventional CBT.	Preliminary evidence suggests that VR exposure therapy is just as efficacious as conventional CBT. Studies included did not use intent-to-treat analysis or did not state concomitant treatments and/or comorbidities.
Anxiety disorder				
Opris et al. (2012)	Treatment efficacy of VR exposure therapy vs. conventional CBT in anxiety disorders vs. active controls.	23 studies ($n = 608$)	VR therapy was significantly better than waitlist controls. Similar improvements were observed between VR and conventional CBT therapy. Similar improvements in outcome measures were maintained over time in both the VR and conventional CBT groups.	The use of VR exposure therapy may be a viable option. There is a future need to determine the use of VR exposure therapy to other forms of VR therapies targeted at anxiety disorders. No measure of dropout rates in studies reviewed.
Powers and Emmelkamp (2008)	Examined the effects of VR in anxiety disorders vs. waitlist or conventional CBT controls.	13 studies ($n = 397$)	VR therapy was more efficacious that waitlist or active control. Significant improvements were observed in subjective distress, cognitive and behavioral measures, and psychophysiological measures. VR therapy was more effective than *in vivo* exposures. Non-significant trend toward a dose-response relationship was observed between number of sessions and outcome measures.	VR exposure therapy was highly effective in treating anxiety disorders. Behavioral avoidance tests should be administered to assess the impact of treatment on anxiety-provoking situations and generalization to the real world.

CBT, cognitive behavioral Therapy; PTSD, post-traumatic stress disorder; VR, virtual reality.

depression that may trigger visual or auditory hallucinations in sufferers of schizophrenia. As these are still early days for VR therapy in general, there is a need to determine the precise role of VR in treatment therapies for schizophrenia and its limitations.

FUTURE OUTLOOK TO THE MOST PROMISING RESEARCH AVENUES - COMPLEMENTING VR THERAPY USING NON-INVASIVE NEUROMODULATION AND NEUROIMAGING TECHNOLOGIES

This review so far has provided evidence for the use of VR therapy in various clinical populations as a standalone or adjunctive tool with mainstream neurorehabilitation treatment modalities. However, the question remains as to whether the beneficial effects of VR can be augmented via neuromodulation techniques such as tDCS, or if it is possible for a more targeted approach to monitor the effects of VR via non-invasive and portable neuroimaging methods such as fNIRS and EEG.

Augmenting VR Therapy with tDCS

Transcranial direct current stimulation is an emerging non-invasive brain stimulation technique that uses low-intensity constant direct electrical currents to modulate the excitability of cortical neurons and related networks (Nitsche and Paulus, 2000; Lang et al., 2005). By placing either a positive anode or negative cathode electrode over the scalp of the head, tDCS

is able to facilitate (anodal tDCS) or inhibit (cathodal tDCS) excitability of the underlying cortical neurons in a polarity-specific manner. Due to this robust neuromodulatory effect, tDCS has often been used in conjunction either before (offline) or during (online) rehabilitation therapy to improve motor and cognitive performance in healthy and clinical populations (for a recent reviews, see Coffman et al., 2014 and Floel, 2014).

In theory, the application of tDCS with VR therapy to augment neurorehabilitation appear complimentary. A study by Lee and Chun (2014) showed an improvement in stroke-specific clinical measures, manual muscle test and the Korean-modified Barthel Index in subacute stroke patients after 15 sessions of VR therapy with online cathodal tDCS to the unaffected motor cortex compared to sham. Kim et al. (2014) further demonstrated that the addition of online anodal tDCS to the affected motor cortex with VR therapy not only improved upper arm function, but also increased corticospinal excitability in subacute stroke patients.

In contrast to the aforementioned findings, mixed results were reported by Viana et al. (2014) that compared the effects of combining VR with offline anodal tDCS over the affected motor cortex of stroke patients, across 15 1-h VR therapy sessions. While the results showed no statistical differences in stroke-specific clinical measures (i.e., Fugl-Meyer assessment, Wolf motor assessment, and modified Ashworth scale) of upper arm function between patients receiving real tDCS compared to sham, it is important to note that more than 50% of participants receiving anodal tDCS and VR therapy had clinically significant improvements in wrist spasticity following treatment. Based on these limited combined VR and tDCS findings, it can be seen that performing tDCS during the VR therapy is a significant factor for enhancing the effects of VR therapy alone, which is also the case for combining tDCS with neurorehabilitation (Rothwell, 2012). Although combined VR and neuromodulation (tDCS) has been primarily applied in movement disorders, to the best of our knowledge, there are currently no known studies that have investigated this combination in cognitive and mood disorders. Thus therapists that are currently adopting the use of VR therapy in mental and mood disorders can potentially exploit the concurrent use of both VR and tDCS to augment therapy benefits above and beyond VR therapy alone.

It is likely that the combined effects of VR and tDCS is influenced by a combination of several factors, namely (1) general patient characteristics (e.g., brain region affected, and structural/functional reserve) and (2) tDCS parameters including electrode placement (affected or unaffected brain region), polarity (anodal or cathodal) and timing of tDCS application (online or offline). In such circumstances, where the efficacy of combined VR and tDCS interventions are both timing and location dependent (i.e., when and where to stimulate), a method of detecting and monitoring changes to neurophysiological function as patients receive treatment is crucial for optimizing intervention effects. In this regard, neuroimaging methods could be applied to monitor treatment VR progression, which will be discussed in the subsequent section.

Monitoring VR Therapy with Neuroimaging

Neurophysiological changes associated with VR neurorehabilitation can be measured by non-invasive and portable neuroimaging techniques including fNIRS and/or EEG, to ascertain changes in cerebral hemodynamic responses or oscillatory brainwaves, respectively. In particular, the use of fNIRS as a tool to measure online cerebral hemodynamic responses during neurorehabilitation has received attention (for review see Irani et al., 2007; Ferrari and Quaresima, 2012). The use of fNIRS as a neuroimaging method relies on the principle of neurovascular coupling that measures the increase in regional cerebral blood flow (i.e., increase in oxygenated and decrease in deoxygenated hemoglobin) induced by neuronal activation, which is analogous to the blood-oxygenation-level-dependent responses measured by fMRI (Ferrari et al., 2004). Cortical activation measurements by fMRI and fNIRS techniques show highly correlated results in both motor and cognitive tasks (Huppert et al., 2006; Cui et al., 2011; Muthalib et al., 2013). While the application of fNIRS techniques is gaining popularity, EEG has long been used to measure online brain activity during a cognitive or motor task, and in various clinical populations (Bonanni et al., 2008; Gosselin et al., 2011). Of particular importance, EEG is used to detect changes in various brainwaves (i.e., Gamma, Alpha, Beta, Theta, and Delta) which are differentially affected by changes in mood (Huang and Lo, 2009), wakefulness (De Gennaro et al., 2001), neurological diseases (Bonanni et al., 2008), and brain injury (Gosselin et al., 2011). Both fNIRS and EEG have several advantages over fMRI, as they are portable, relatively inexpensive to use, and easy to operate with high temporal resolution. Furthermore, new generation systems are battery operated, wireless and further miniaturized to the size of a smartphone, ideal for ambulatory and untethered measurements consistent with a neuroergonomics approach (Ayaz et al., 2013; de Lissa et al., 2015).

As the use of fNIRS and EEG techniques in VR therapy is still relatively new, most studies to date have focused on healthy individuals (Bayliss and Ballard, 2000; Mingyu et al., 2005; Holper et al., 2010; Seraglia et al., 2011; Basso Moro et al., 2014), with the potential for more studies in clinical populations emerging as the popularity of these portable neuroimaging technologies increases. The current use of fNIRS and EEG in VR therapy has two proposed roles; (1) to monitor and provide augmented feedback regarding regions of cortical activation during therapy and (2) to use fNIRS or EEG as part of a BCI paradigm for therapy. In support of the first role, several studies have investigated the efficacy of fNIRS and EEG to record cortical hemodynamic and oscillatory changes during actual motor tasks and motor imagery in a VR environment. These studies demonstrated the efficacy of fNIRS and EEG to detect task-specific changes in cortical hemodynamics (Holper et al., 2010; Seraglia et al., 2011; Basso Moro et al., 2014) and oscillatory patterns (Bayliss and Ballard, 2000; Mingyu et al., 2005).

The ability of fNIRS and EEG to detect changes in these neurophysiological measures can provide both feedback on



FIGURE 2 | Stroke participants engaged in VR therapy using an X-Box Kinect motion capture system while receiving tDCS.

location and level of activation, for which clinicians and users can use to set intensity and progression of therapy. Furthermore, feedback on cortical activation can also be used to identify areas of hypo- or hyperactivity, which can be modulated using neuromodulatory techniques such as tDCS (see Prospective Integration of Neuromodulation-Neuroimaging with VR Therapy). In support of the second role, identifying cortical areas of activation, patterns and timing of cortical activation associated with various movements or mood states may also be recorded as classifiers for BCI training. Indeed, most BCI studies to date have employed the use of EEG as a measure for cortical activation for which to control a robotic limb or avatar in a VR environment (Lin et al., 2007; Formaggio et al., 2013, 2015). Although relatively new, there are also fNIRS-based BCI approaches demonstrating feasibility for future integration with VR and neurorehabilitation (Sitaram et al., 2007; Hong et al., 2015). Moreover, joint use of fNIRS- and EEG-based BCI approaches have also been demonstrated (Koo et al., 2015; Xuxian et al., 2015). This shows the potential to adopt fNIRS in a similar manner, whereby appropriate cortical hemodynamic responses can be classified to control a robotic or computer interface.

Prospective Integration of Neuromodulation-Neuroimaging with VR Therapy

In the last 5 years, new research suggests that the combination of VR therapy with neuromodulation and neuroimaging techniques may help to improve the effects and delivery of VR therapies. Neuromodulation techniques such as tDCS (**Figure 2**) and neuroimaging methods such as fNIRS (**Figure 3**) and EEG have already been combined and have shown some success (Ang et al., 2015; Dutta et al., 2015; Muthalib et al., 2016). While the use of these techniques in combination with VR is still in its infancy, the available evidence suggests highly complementary effects when combining neuromodulation and neuroimaging with VR therapy.

As both neuroimaging (fNIRS/EEG) and neuromodulation (tDCS) techniques have complimentary capabilities, and they both can be built as wearable and wireless systems, integration of the two presents as a natural avenue for applications in natural environments and real world settings (McKendrick et al., 2015). One potential use is enhancing BCI applications. In its most general form, BCI provides a route for neural output that does not involve the neuromuscular system (Wolpaw et al., 2002). Almost all current non-invasive BCI systems are read-only, that is, brain signals are read directly to the system (via neuroimaging) and the system can provide an output or feedback that relies on sensory input mechanisms and peripheral nervous system to reach back to the brain. A more direct BCI could eliminate the need for sensory input with the use of neuromodulation, and hence provide feedback directly to the brain for a read-write BCI.

Integration of VR therapy for such a read/write BCI hold potential for enhanced or accelerated therapeutical processes in neurorehabilitation. A study by Ang et al. (2015) was one of the first studies that combined VR with both tDCS and EEG to investigate the additive effect of offline anodal tDCS on BCI haptic training with chronic stroke patients. Although this study reported no difference in the Fugl-Meyer assessment and block-and-box test between groups that received real tDCS prior to BCI training, compared to sham, the study did demonstrate an increased state of EEG mu rhythm suppression in the real tDCS group. Thus indicating improved neurophysiological responses in motor preparation. Future studies are necessary to determine



FIGURE 3 | The use of a semi-immersive VR environment and fNIRS system.

the significance of these neurophysiological improvements to clinical outcomes.

Neuroimaging guided tDCS and VR therapy can be applied to neurorehabilitation in general. The optimal location for applying tDCS electrodes to modulate a target cortical region and connected networks is a debated point in the field. Modeling studies of current flow between the tDCS electrodes have provided some guidance to the placement of electrodes to stimulate a specific brain region; however, whether these models predict the actual current flow and polarization of the targeted brain networks is not known. Simultaneous tDCS-neuroimaging could provide a solution to confirming the modeling predictions and/or to guide the direction of current flow between multi-electrode tDCS montages (Ruffini et al., 2014). For example, a motor task could be used to activate a broad network of cortical regions of interest that can be measured using fNIRS/EEG neuroimaging. Once the locations of the tDCS electrodes have been determined using neuroimaging guidance, neuroimaging and tDCS could be simultaneously or independently used to guide and modulate VR therapy. In this scenario, neuroimaging during VR tasks can be used to adapt the intensity based on the level of activation, such as the attentional hub of the dorsolateral prefrontal cortex (DLPFC). The level of DLPFC activation during the initial task would be expected to be high due to the novelty of the activity, however, as the activity progresses the learning of task requirements become automatic and performance improves, less attentional resources would be required, and the level of DLPFC would be expected to decrease (Ayaz et al., 2012; McKendrick et al., 2014). In such a scenario, the VR task can be adapted online to modulate the intensity level and maintain optimal DLPFC activation. Also, if the levels of DLPFC activation remain at high levels and/or performance is stagnant, then tDCS could potentially be applied online to upregulate neuronal networks required to perform the task. Preliminary evidence using a modeling approach to locate tDCS electrodes was provided by McKendrick et al. (2015) using a spatial memory task with concurrent fNIRS and tDCS. They showed that when task performance declined rapidly following baseline, the application of tDCS almost immediately eliminated the performance decrement. Furthermore, they showed that tDCS can modulate the neural activity of specific brain regions near the site of stimulation. However, they cautioned that current models and protocols for determining tDCS montages are lacking, due to complex interactions between stimulation montage, task performance and underlying hemodynamics that are not fully understood. Therefore, additional joint tDCS and fNIRS/EEG studies are required to further unravel these complexities and to better define the pattern of cortical excitation

induced by tDCS during the performance of cognitive and motor tasks.

CURRENT LIMITATIONS OF VR THERAPY

While using a VE offers many unique advantages to traditional treatment and neurorehabilitation approaches, limitations to their efficacy and practicality must be acknowledged. Firstly, larger clinical studies are required to establish the efficacy of using VR in physical and cognitive rehabilitation in different clinical populations. Much of the existing literature report mixed findings from small sample sizes, and often lack appropriate control comparisons (see **Tables 1** and **2**). Secondly, there is little information on the transfer of the training effects of VR into the corresponding physical environment in general, and the VR training parameters associated with optimal transfer to real-world functional improvements are yet to be elucidated. Thirdly, in many clinical populations it is unclear whether advantages of VR over real-world training exist, and if so, precisely what these advantages are (see **Tables 1** and **2** for study limitations). Furthermore, because this literature is extremely vulnerable to selective reporting and Type-I statistical error, there is an inherent bias of publishing results that show correlation between rehabilitation improvements and the application of VR. Therefore to potentially limit any bias in future studies, it may be useful for future studies to adopt a double-blinded protocol for the evaluation of the effectiveness of the use of VR. Lastly, it is important to investigate any unique rehabilitative effects of VR that may be exploited, or whether the benefits of VR can be attributed to the enjoyment of gaming platforms associated with VR themselves (i.e., VR therapies may only present as more effective because they engage and motivate participants throughout their training session, providing increased adherence). While limitations in VR technology exist, the potential for favorable neuroplasticity afforded by such technology undoubtedly warrants further investigation.

CONCLUSION

In summary, this review has discussed the strengths and limitations for the use of VR therapies in motor and mental health neurorehabilitation. The current evidence suggest that a combination of VR and conventional therapies are safe and likely to be more efficacious compared to just traditional or VR therapy alone. However, it is not known if the use of VR therapies can lead to cost-saving benefits (i.e., reduced financial and manpower cost) or even if current commercial or customized systems will be applicable by patients that are living within the community. More importantly, there is a need to elucidate the aspects of VR that are most effective for rehabilitation. While this is not apparent in the current review, future studies should attempt to systematically determine the role of self-projection, sensory feedback or motivation on rehabilitation in relation to specific diseases or impairments. Furthermore, we have discussed the potential for VR therapy to be complemented by other forms of technologies such as neuromodulation (tDCS) and neuroimaging (fNIRS/EEG) in order to augment training benefits of VR, and provide a more targeted approach to neurorehabilitation. Large-scale longitudinal studies will also be required to determine the effects of VR therapy (in combination with tDCS/fNIRS/EEG) and the translation of VR therapy in a non-clinical environment (i.e., home setting).

AUTHOR CONTRIBUTIONS

All authors listed, have made substantial, direct and intellectual contribution to the work, and approved it for publication.

ACKNOWLEDGMENTS

W-PT is supported by an Alfred Deakin Postdoctoral Research Fellowship. MM is supported by a Postdoctoral Research Fellowship of the University of Montpellier.

REFERENCES

Adamovich, S. V., Fluet, G. G., Tunik, E., and Merians, A. S. (2009). Sensorimotor training in virtual reality: a review. *NeuroRehabilitation* 25, 29–44. doi: 10.3233/NRE-2009-0497

Aisen, M. L., Kerkovich, D., Mast, J., Mulroy, S., Wren, T. A., Kay, R. M., et al. (2011). Cerebral palsy: clinical care and neurological rehabilitation. *Lancet Neurol.* 10, 844–852. doi: 10.1016/S1474-4422(11)70176-4

Ang, K. K., Guan, C., Phua, K. S., Wang, C., Zhao, L., Teo, W. P., et al. (2015). Facilitating effects of transcranial direct current stimulation on motor imagery brain-computer interface with robotic feedback for stroke rehabilitation. *Arch. Phys. Med. Rehabil.* 96, S79–S87. doi: 10.1016/j.apmr.2014.08.008

Ayaz, H., Onaral, B., Izzetoglu, K., Shewokis, P. A., Mckendrick, R., and Parasuraman, R. (2013). Continuous monitoring of brain dynamics with functional near infrared spectroscopy as a tool for neuroergonomic research: empirical examples and a technological development. *Front. Hum. Neurosci.* 7:871. doi: 10.3389/fnhum.2013.00871

Ayaz, H., Shewokis, P. A., Bunce, S., Izzetoglu, K., Willems, B., and Onaral, B. (2012). Optical brain monitoring for operator training and mental workload assessment. *Neuroimage* 59, 36–47. doi: 10.1016/j.neuroimage.2011.06.023

Barry, G., Galna, B., and Rochester, L. (2014). The role of exergaming in Parkinson's disease rehabilitation: a systematic review of the evidence. *J. Neuroeng. Rehabil.* 11:33. doi: 10.1186/1743-0003-11-33

Basso Moro, S., Bisconti, S., Muthalib, M., Spezialetti, M., Cutini, S., Ferrari, M., et al. (2014). A semi-immersive virtual reality incremental swing balance task activates prefrontal cortex: a functional near-infrared spectroscopy study. *Neuroimage* 85, 451–460. doi: 10.1016/j.neuroimage.2013.05.031

Bayliss, J. D., and Ballard, D. H. (2000). A virtual reality testbed for brain-computer interface research. *IEEE Trans. Rehabil. Eng.* 8, 188–190. doi: 10.1109/86.847811

Benton, M. K., and Schroeder, H. E. (1990). Social skills training with schizophrenics: a meta-analytic evaluation. *J. Consult. Clin. Psychol.* 58, 741–747. doi: 10.1037/0022-006X.58.6.741

Bonanni, L., Thomas, A., Tiraboschi, P., Perfetti, B., Varanese, S., and Onofrj, M. (2008). EEG comparisons in early Alzheimer's disease, dementia with Lewy bodies and Parkinson's disease with dementia patients with a 2-year follow-up. *Brain* 131, 690–705. doi: 10.1093/brain/awm322

Bower, K. J., Louie, J., Landesrocha, Y., Seedy, P., Gorelik, A., and Bernhardt, J. (2015). Clinical feasibility of interactive motion-controlled games for stroke rehabilitation. *J. Neuroeng. Rehabil.* 12:63. doi: 10.1186/s12984-015-0057-x

Bryanton, C., Bosse, J., Brien, M., Mclean, J., Mccormick, A., and Sveistrup, H. (2006). Feasibility, motivation, and selective motor control: virtual reality compared to conventional home exercise in children with cerebral palsy. *Cyberpsychol. Behav.* 9, 123–128. doi: 10.1089/cpb.2006.9.123

Burke, C. J., Tobler, P. N., Baddeley, M., and Schultz, W. (2010). Neural mechanisms of observational learning. *Proc. Natl. Acad. Sci. U.S.A.* 107, 14431–14436. doi: 10.1073/pnas.1003111107

Chen, Y. P., Lee, S. Y., and Howard, A. M. (2014). Effect of virtual reality on upper extremity function in children with cerebral palsy: a meta-analysis. *Pediatr. Phys. Ther.* 26, 289–300. doi: 10.1097/PEP.0000000000000046

Coffman, B. A., Clark, V. P., and Parasuraman, R. (2014). Battery powered thought: enhancement of attention, learning, and memory in healthy adults using transcranial direct current stimulation. *Neuroimage* 85, 895–908. doi: 10.1016/j.neuroimage.2013.07.083

Corbetta, D., Imeri, F., and Gatti, R. (2015). Rehabilitation that incorporates virtual reality is more effective that standard rehabilitation for improving walking speed, balance and mobility after stroke: a systematic review. *J. Physiother.* 61, 117–124. doi: 10.1016/j.jphys.2015.05.017

Cote, S., and Bouchard, S. (2005). Documenting the efficacy of virtual reality exposure with psychophysiological and information processing measures. *Appl. Psychophysiol. Biofeedback* 30, 217–232. doi: 10.1007/s10484-005-6379-x

Cui, X., Bray, S., Bryant, D. M., Glover, G. H., and Reiss, A. L. (2011). A quantitative comparison of NIRS and fMRI across multiple cognitive tasks. *Neuroimage* 54, 2808–2821. doi: 10.1016/j.neuroimage.2010.10.069

Cukor, J., Spitalnick, J., Difede, J., Rizzo, A., and Rothbaum, B. O. (2009). Emerging treatments for PTSD. *Clin. Psychol. Rev.* 29, 715–726. doi: 10.1016/j.cpr.2009.09.001

De Gennaro, L., Ferrara, M., Curcio, G., and Cristiani, R. (2001). Antero-posterior EEG changes during the wakefulness-sleep transition. *Clin. Neurophysiol.* 112, 1901–1911. doi: 10.1016/S1388-2457(01)00649-6

de Lissa, P., Sorensen, S., Badcock, N., Thie, J., and Mcarthur, G. (2015). Measuring the face-sensitive N170 with a gaming EEG system: a validation study. *J. Neurosci. Methods* 253, 47–54. doi: 10.1016/j.jneumeth.2015.05.025

Dewar, R., Love, S., and Johnston, L. M. (2015). Exercise interventions improve postural control in children with cerebral palsy: a systematic review. *Dev. Med. Child Neurol.* 57, 504–520. doi: 10.1111/dmcn.12660

dos Santos Mendes, F. A., Pompeu, J. E., Modenesi Lobo, A., Guedes da Silva, K., Oliveira Tde, P., Peterson Zomignani, A., et al. (2012). Motor learning, retention and transfer after virtual-reality-based training in Parkinson's disease–effect of motor and cognitive demands of games: a longitudinal, controlled clinical study. *Physiotherapy* 98, 217–223. doi: 10.1016/j.physio.2012.06.001

Dutta, A., Jacob, A., Chowdhury, S. R., Das, A., and Nitsche, M. A. (2015). EEG-NIRS based assessment of neurovascular coupling during anodal transcranial direct current stimulation–a stroke case series. *J. Med. Syst.* 39:205. doi: 10.1007/s10916-015-0205-7

Esculier, J. F., Vaudrin, J., Bériault, P., Gagnon, K., and Tremblay, L. E. (2012). Home-based balance training programme using Wii Fit with balance board for Parkinsons's disease: a pilot study. *J. Rehabil. Med.* 44, 144–150. doi: 10.2340/16501977-0922

Fang, T. Y., Wang, P. C., Liu, C. H., Su, M. C., and Yeh, S. C. (2014). Evaluation of a haptics-based virtual reality temporal bone simulator for anatomy and surgery training. *Comput. Methods Programs* 113, 674–681. doi: 10.1016/j.cmpb.2013.11.005

Ferrari, M., Mottola, L., and Quaresima, V. (2004). Principles, techniques, and limitations of near infrared spectroscopy. *Can. J. Appl. Physiol.* 29, 463–487. doi: 10.1139/h04-031

Ferrari, M., and Quaresima, V. (2012). A brief review on the history of human functional near-infrared spectroscopy (fNIRS) development and fields of application. *Neuroimage* 63, 921–935. doi: 10.1016/j.neuroimage.2012.03.049

Floel, A. (2014). tDCS-enhanced motor and cognitive function in neurological diseases. *Neuroimage* 85, 934–947. doi: 10.1016/j.neuroimage.2013.05.098

Formaggio, E., Storti, S. F., Boscolo Galazzo, I., Gandolfi, M., Geroin, C., Smania, N., et al. (2013). Modulation of event-related desynchronization in robot-assisted hand performance: brain oscillatory changes in active, passive and imagined movements. *J. Neuroeng. Rehabil.* 10:24. doi: 10.1186/1743-0003-10-24

Formaggio, E., Storti, S. F., Boscolo Galazzo, I., Gandolfi, M., Geroin, C., Smania, N., et al. (2015). Time-frequency modulation of ERD and EEG coherence in robot-assisted hand performance. *Brain Topogr.* 28, 352–363. doi: 10.1007/s10548-014-0372-8

Garrison, D. R., and Garrison, D. R. (1997). Self-directed learning: toward a comprehensive model. *Adult Educ. Q.* 48, 18–33. doi: 10.1177/074171369704800103

Goncalves, R., Pedrozo, A. L., Coutinho, E. S., Figueira, I., and Ventura, P. (2012). Efficacy of virtual reality exposure therapy in the treatment of PTSD: a systematic review. *PLoS ONE* 7:e48469. doi: 10.1371/journal.pone.0048469

Gorini, A., Pallavicini, F., Algeri, D., Repetto, C., Gaggioli, A., and Riva, G. (2010). Virtual reality in the treatment of generalized anxiety disorders. *Stud. Health Technol. Inform.* 154, 39–43.

Gosselin, N., Bottari, C., Chen, J. K., Petrides, M., Tinawi, S., De Guise, E., et al. (2011). Electrophysiology and functional MRI in post-acute mild traumatic brain injury. *J. Neurotrauma* 28, 329–341. doi: 10.1089/neu.2010.1493

Harris, D. M., Rantalainen, T., Muthalib, M., Johnson, L., and Teo, W. P. (2015). Exergaming as a viable therapeutic tool to improve static and dynamic balance among older adults and people with idiopathic Parkinson's disease: a systematic review and meta-analysis. *Front. Aging Neurosci.* 7:167. doi: 10.3389/fnagi.2015.00167

Herz, N. B., Herz, N. B., Mehta, S. H., Sethi, K. D., Jackson, P., Hall, P., et al. (2013). Nintendo Wii rehabilitation ('Wii-hab') provides benefits in Parkinson's disease. *Parkinsonism. Relat. Disord.* 19, 1039–1042. doi: 10.1016/j.parkreldis.2013.07.014

Holper, L., Muehlemann, T., Scholkmann, F., Eng, K., Kiper, D., and Wolf, M. (2010). Testing the potential of a virtual reality neurorehabilitation system during performance of observation, imagery and imitation of motor actions recorded by wireless functional near-infrared spectroscopy (fNIRS). *J. Neuroeng. Rehabil.* 7:57. doi: 10.1186/1743-0003-7-57

Hong, K. S., Naseer, N., and Kim, Y. H. (2015). Classification of prefrontal and motor cortex signals for three-class fNIRS–BCI. *Neurosci. Lett.* 587, 87–92. doi: 10.1016/j.neulet.2014.12.029

Huang, H. Y., and Lo, P. C. (2009). EEG dynamics of experienced Zen meditation practitioners probed by complexity index and spectral measure. *J. Med. Eng. Technol.* 33, 314–321. doi: 10.1080/03091900802602677

Huppert, T. J., Hoge, R. D., Diamond, S. G., Franceschini, M. A., and Boas, D. A. (2006). A temporal comparison of BOLD, ASL, and NIRS hemodynamic responses to motor stimuli in adult humans. *Neuroimage* 29, 368–382. doi: 10.1016/j.neuroimage.2005.08.065

Irani, F., Platek, S. M., Bunce, S., Ruocco, A. C., and Chute, D. (2007). Functional near infrared spectroscopy (fNIRS): an emerging neuroimaging technology with important applications for the study of brain disorders. *Clin. Neuropsychol.* 21, 9–37. doi: 10.1080/13854040600910018

Jung, E. Y., Park, D. K., Lee, Y. H., Jo, H. S., Lim, Y. S., and Park, R. W. (2012). Evaluation of practical exercises using an intravenous simulator incorporating virtual reality and haptics device technologies. *Nurse Educ. Today* 32, 458–463. doi: 10.1016/j.nedt.2011.05.012

Kandalaft, M. R., Didehbani, N., Krawczyk, D. C., Allen, T. T., and Chapman, S. B. (2013). Virtual reality social cognition training for young adults with high-functioning autism. *J. Autism. Dev. Disord.* 43, 34–44. doi: 10.1007/s10803-012-1544-6

Kim, B. R., Chun, M. H., Kim, L. S., and Park, J. Y. (2011). Effect of virtual reality on cognition in stroke patients. *Ann. Rehabil. Med.* 35, 450–459. doi: 10.5535/arm.2011.35.4.450

Kim, Y. J., Ku, J., Cho, S., Kim, H. J., Cho, Y. K., Lim, T., et al. (2014). Facilitation of corticospinal excitability by virtual reality exercise following anodal transcranial direct current stimulation in healthy volunteers and subacute stroke subjects. *J. Neuroeng. Rehabil.* 11:124. doi: 10.1186/1743-0003-11-124

Klinger, E., Bouchard, S., Legeron, P., Roy, S., Lauer, F., Chemin, I., et al. (2005). Virtual reality therapy versus cognitive behavior therapy for social phobia: a preliminary controlled study. *Cyberpsychol. Behav.* 8, 76–88. doi: 10.1089/cpb.2005.8.76

Klinger, E., Legeron, P., Roy, S., Chemin, I., Lauer, F., and Nugues, P. (2004). Virtual reality exposure in the treatment of social phobia. *Stud. Health Technol. Inform.* 99, 91–119.

Koo, B., Lee, H. G., Nam, Y., Kang, H., Koh, C. S., Shin, H. C., et al. (2015). A hybrid NIRS-EEG system for self-paced brain computer interface with online motor imagery. *J. Neurosci. Methods* 244, 26–32. doi: 10.1016/j.jneumeth.2014.04.016

Ku, J., Han, K., Lee, H. R., Jang, H. J., Kim, K. U., Park, S. H., et al. (2007). VR-based conversation training program for patients with schizophrenia: a preliminary clinical trial. *Cyberpsychol. Behav.* 10, 567–574. doi: 10.1089/cpb.2007.9989

Kuester, A., Niemeyer, H., and Knaevelsrud, C. (2016). Internet-based interventions for posttramatic stress: a meta-analysis of randomized controlled trials. *Clin. Psychol. Rev.* 43, 1–16. doi: 10.1016/j.cpr.2015.11.004

Laffont, I., Bakhti, K., Coroian, F., Van Dokkum, L., Mottet, D., Schweighofer, N., et al. (2014). Innovative technologies applied to sensorimotor rehabilitation after stroke. *Ann. Phys. Rehabil. Med.* 57, 543–551. doi: 10.1016/j.rehab.2014.08.007

Lang, N., Siebner, H. R., Ward, N. S., Lee, L., Nitsche, M. A., Paulus, W., et al. (2005). How does transcranial DC stimulation of the primary motor cortex alter regional neuronal activity in the human brain? *Eur. J. Neurosci.* 22, 495–504. doi: 10.1111/j.1460-9568.2005.04233.x

Lauber, B., and Keller, M. (2014). Improving motor performance: selected aspects of augmented feedback in exercise and health. *Eur. J. Sport Sci.* 14, 36–43. doi: 10.1080/17461391.2012.725104

Laver, K., George, S., Thomas, S., Deutsch, J. E., and Crotty, M. (2015). Virtual reality for stroke rehabilitation: an abridged version of a Cochrane review. *Eur. J. Phys. Rehabil. Med.* 51, 497–506.

Ledoux, A. A., Phillips, J. L., Labelle, A., Smith, A., Bohbot, V. D., and Boyer, P. (2013). Decreased fMRI activity in the hippocampus of patients with schizophrenia compared to healthy control participants, tested on a wayfinding task in a virtual town. *Psychiatry Res.* 211, 47–56. doi: 10.1016/j.pscychresns.2012.10.005

Lee, S. J., and Chun, M. H. (2014). Combination transcranial direct current stimulation and virtual reality therapy for upper extremity training in patients with subacute stroke. *Arch. Phys. Med. Rehabil.* 95, 431–438. doi: 10.1016/j.apmr.2013.10.027

Lee, S. J., Yoo, J. Y., Ryu, J. S., Park, H. K., and Chung, S. J. (2012). The effects of visual and auditory cues on freezing of gait in patients with Parkinson's disease. *Am. J. Phys. Med. Rehabil.* 91, 2–11. doi: 10.1097/PHM.0b013e31823c7507

Levin, M., Magdalon, E. C., Michaelsen, S. M., and Quevedo, A. (2015). Quality of grasping and the role of haptics in a 3D immersive virtual reality environment in individuals with stroke. *IEEE Trans. Neural Syst. Rehabil. Eng.* 23, 1047–1055. doi: 10.1109/TNSRE.2014.2387412

Lin, C. T., Chung, I. F., Ko, L. W., Chen, Y. C., Liang, S. F., and Duann, J. R. (2007). EEG-based assessment of driver cognitive responses in a dynamic virtual-reality driving environment. *IEEE Trans. Biomed. Eng.* 54, 1349–1352. doi: 10.1109/TBME.2007.891164

Lister, C., West, J. H., Cannon, B., Sax, T., and Brodegard, D. (2014). Just a fad. Gamification in health and fitness apps. *JIMR Serious Games* 2:e9. doi: 10.2196/games.3413

Lohse, K. R., Hilderman, C. G., Cheung, K. L., Tatla, S., and Van Der Loos, H. F. (2014). Virtual reality therapy for adults post-stroke: a systematic review and meta-analysis exploring virtual environments and commercial games in therapy. *PLoS ONE* 9:e93318. doi: 10.1371/journal.pone.0093318

Luque-Moreno, C., Ferragut-Garcías, A., Rodríguez-Blanco, C., Heredia-Rizo, A. M., Oliva-Pascual-Vaca, J., Kiper, P., et al. (2015). A decade of progress using virtual reality for poststroke lower extremity rehabilitation: systematic review of the intervention methods. *Biomed. Res. Int.* 2015:7. doi: 10.1155/2015/342529

Marques, A., Queirós, C., and Rocha, N. (2008). "Virtual reality and neuropsychology: a cognitive rehabilitation approach for people with psychiatric disabilities," in *ICDVRAT–Proceedings of 7th International Conference on Disability Virtual Reality and Associated Technologies (Sep 8-11 2008)* (Maia: Citeseer), 39–46.

Maskey, M., Lowry, J., Rodgers, J., Mcconachie, H., and Parr, J. R. (2014). Reducing specific phobia/fear in young people with autism spectrum disorders (ASDs) through a virtual reality environment intervention. *PLoS ONE* 2:e100374. doi: 10.1371/journal.pone.0100374

McKendrick, R., Ayaz, H., Olmstead, R., and Parasuraman, R. (2014). Enhancing dual-task performance with verbal and spatial working memory training: continuous monitoring of cerebral hemodynamics with NIRS. *Neuroimage* 85, 1014–1026. doi: 10.1016/j.neuroimage.2013.05.103

McKendrick, R., Parasuraman, R., and Ayaz, H. (2015). Wearable functional near infrared spectroscopy (fNIRS) and transcranial direct current stimulation (tDCS): expanding vistas for neurocognitive augmentation. *Front. Syst. Neurosci.* 9:27. doi: 10.3389/fnsys.2015.00027

Meyerbroker, K., and Emmelkamp, P. M. (2010). Virtual reality exposure therapy in anxiety disorders: a systematic review of process-and-outcome studies. *Depress. Anxiety* 27, 933–944. doi: 10.1002/da.20734

Mingyu, L., Jue, W., Nan, Y., and Qin, Y. (2005). "Development of EEG biofeedback system based on virtual reality environment," in *Proceedings of the 27th Annual Conference on IEEE Engineering in Medicine and Biology, Shanghai* (Rome: IEEE), 5362–5364. doi: 10.1109/IEMBS.2005.1615693

Mirelman, A., Maidan, I., Herman, T., Deutsch, J. E., Giladi, N., and Hausdorff, J. M. (2011). Virtual reality for gait training: can it induce motor learning to enhance complex walking and reduce fall risk in patients with Parkinson's disease? *J. Gerontol. A Biol. Sci. Med. Sci.* 66, 234–240. doi: 10.1093/gerona/glq201

Mirelman, A., Rochester, L., Reelick, M., Nieuwhof, F., Pelosin, E., Abbruzzese, G., et al. (2013). V-TIME: a treadmill training program augmented by virtual reality to decrease fall risk in older adults: study design of a randomized controlled trial. *BMC Neurol.* 13:15. doi: 10.1186/1471-2377-13-15

Monge Pereira, E., Molina Rueda, F., Alguacil Diego, I. M., Cano de la Cuerda, R., De Mauro, A., and Miangolarra Page, J. C. (2015). Use of virtual reality systems as proprioception method in cerebral palsy: clinical practice guideline. *Neurologia* 30, 550–559. doi: 10.1016/j.nrl.2011.12.004

Muthalib, M., Anwar, A. R., Perrey, S., Dat, M., Galka, A., Wolff, S., et al. (2013). Multimodal integration of fNIRS, fMRI and EEG neuroimaging. *Clin. Neurophysiol.* 124, 2060–2062. doi: 10.1016/j.clinph.2013.03.018

Muthalib, M., Besson, P., Rothwell, J. C., Ward, T. E., and Perrey, A. (2016). Effects of anodal high-definition transcranial direct current stimulation on bilateral sensorimotor cortex activation during sequential finger movements: an fNIRS study. *Adv. Exp. Med. Biol.* 876, 351–359. doi: 10.1007/978-1-4939-3023-4_44

Nieuwenhuijsen, E. R., Zemper, E., Miner, K. R., and Epstein, M. (2006). Health behavior change models and theories: contributions to rehabilitation. *Disabil. Rehabil.* 28, 245–256. doi: 10.1080/09638280500197743

Nitsche, M. A., and Paulus, W. (2000). Excitability changes induced in the human motor cortex by weak transcranial direct current stimulation. *J. Physiol.* 527(Pt 3), 633–639. doi: 10.1111/j.1469-7793.2000.t01-1-00633.x

Oouchida, Y., Suzuki, E., Aizu, N., Takeuchi, N., and Izumi, S. (2013). Application of observational learning in neurorehabilitation. *Int. J. Phys. Med. Rehabil.* 1:5.

Opris, D., Pintea, S., Garcia-Palacios, A., Botella, C., Szamoskozi, S., and David, D. (2012). Virtual reality exposure therapy in anxiety disorders: a quantitative meta-analysis. *Depress. Anxiety* 29, 85–93. doi: 10.1002/da.20910

Oskoui, M., Coutinho, F., Dykeman, J., Jette, N., and Pringsheim, T. (2013). An update on the prevalence of cerebral palsy: a systematic review and meta-analysis. *Dev. Med. Child Neurol.* 55, 509–519. doi: 10.1111/dmcn.12080

Park, K. M., Ku, J., Choi, S. H., Jang, H. J., Park, J. Y., Kim, S. I., et al. (2011). A virtual reality application in role-plays of social skills training for schizophrenia: a randomized, controlled trial. *Psychiatry Res.* 189, 166–172. doi: 10.1016/j.psychres.2011.04.003

Powers, M. B., and Emmelkamp, P. M. (2008). Virtual reality exposure therapy for anxiety disorders: a meta-analysis. *J. Anxiety Disord.* 22, 561–569. doi: 10.1016/j.janxdis.2007.04.006

Reid, D. (2004). The influence of virtual reality on playfulness in children with cerebral palsy: a pilot study. *Occup. Ther. Int.* 11, 131–144. doi: 10.1002/oti.202

Reid, D. T. (2002a). Benefits of a virtual play rehabilitation environment for children with cerebral palsy on perceptions of self-efficacy: a pilot study. *Pediatr. Rehabil.* 5, 141–148. doi: 10.1080/1363849021000039344

Reid, D. T. (2002b). The use of virtual reality to improve upper-extremity efficiency skills in children with cerebral palsy: a pilot study. *Technol. Disabil.* 14, 53–62.

Repetto, C., and Riva, G. (2011). From virtual reality to interreality in the treatment of anxiety disorders. *Neuropsychiatry* 1, 31–43. doi: 10.2196/jmir.3235

Rizzo, A., Buckwalter, G., Neumann, U., Kesselman, C., and Thiebaux, M. (1998). Basic issues in the application of virtual reality for the assessment and rehabilitation of cognitive impairments and functional disabilities. *Cyberpsychol. Behav.* 1, 59–78. doi: 10.1089/cpb.1998.1.59

Rothwell, J. C. (2012). Clinical applications of noninvasive electrical stimulation: problems and potentials. *Clin. EEG Neurosci.* 43, 209–214. doi: 10.1177/1550059412444973

Ruffini, G., Fox, M. D., Ripolles, O., Miranda, P. C., and Pascual-Leone, A. (2014). Optimization of multifocal transcranial current stimulation for weighted cortical pattern targeting from realistic modeling of electric fields. *Neuroimage* 89, 216–225. doi: 10.1016/j.neuroimage.2013.12.002

Saposnik, G., Levin, M., and Outcome Research Canada (SORCan) Working Group (2011). Virtual reality in stroke rehabilitation: a meta-analysis and implications for clinicians. *Stroke* 42, 1380–1386. doi: 10.1161/STROKEAHA.110.605451

Seraglia, B., Gamberini, L., Priftis, K., Scatturin, P., Martinelli, M., and Cutini, S. (2011). An exploratory fNIRS study with immersive virtual reality: a new method for technical implementation. *Front. Hum. Neurosci.* 5:176. doi: 10.3389/fnhum.2011.00176

Sitaram, R., Zhang, H., Guan, C., Thulasidas, M., Hoshi, Y., Ishikawa, A., et al. (2007). Temporal classification of multichannel near-infrared spectroscopy signals of motor imagery for developing a brain-computer interface. *Neuroimage* 34, 1416–1427. doi: 10.1016/j.neuroimage.2006.11.005

Strosahl, K. D., and Ascough, J. C. (1981). Clinical uses of mental imagery: experimental foundations, theoretical misconceptions, and research issues. *Psychol. Bull.* 89, 422–438. doi: 10.1037/0033-2909.89.3.422

Thomson, K., Pollock, A., Bugge, C., and Brady, M. (2014). Commercial gaming devices for stroke upper limb rehabilitation: a systematic review. *Int. J. Stroke* 9, 479–488. doi: 10.1111/ijs.12263

Tsang, M. M., and Man, D. W. (2013). A virtual reality-based vocational training system (VRVTS) for people with schizophrenia in vocational rehabilitation. *Schizophr. Res.* 144, 51–62. doi: 10.1016/j.schres.2012.12.024

Van den Broek, M. D. (2005). Why does neurorehabilitation fail? *J. Head Trauma Rehabil.* 20, 464–473. doi: 10.1097/00001199-200509000-00007

Viana, R. T., Viana, R. T., Laurentino, G. E. C., Souza, R. J. P., Fonseca, J. B., Silva Filho, E. M., et al. (2014). Effects of the addition of transcranial direct current stimulation to virtual reality therapy after stroke: a pilot randomized controlled trial. *NeuroRehabilitation* 34, 437–446. doi: 10.3233/NRE-141065

Vincelli, F., Choi, H., Molinari, E., Wiederhold, B. K., Bouchard, S., and Riva, G. (2002). Virtual reality assisted cognitive behavioral therapy for the treatment of panic disorders with Agoraphobia. *Stud. Health Technol. Inform.* 85, 552–559.

Wang, C. T. (2012). [Research and application of a haptics-based virtual-reality training system for craniomaxillofacial surgery]. *Zhonghua Kou Qiang Yi Xue Za Zhi* 47, 458–462. doi: 10.3760/cma.j.issn.1002-0098.2012.08.003

Waters, F. A., Maybery, M. T., Badcock, J. C., and Michie, P. T. (2004). Context memory and binding in schizophrenia. *Schizophr. Res.* 68, 119–125. doi: 10.1016/S0920-9964(03)00221-4

Williams, C. K., and Carnahan, H. (2014). Motor learning perspectives on haptic training for the upper extremities. *IEEE Trans. Haptics* 7, 240–250. doi: 10.1109/TOH.2013.2297102

Winstein, C. J. (1991). Knowledge of results and motor learning–implications for physical therapy. *Phys. Ther.* 71, 140–149.

Wolpaw, J. R., Birbaumer, N., Mcfarland, D. J., Pfurtscheller, G., and Vaughan, T. M. (2002). Brain-computer interfaces for communication and control. *Clin. Neurophysiol.* 113, 767–791. doi: 10.1016/S1388-2457(02)00057-3

Xuxian, Y., Baolei, X., Changhao, J., Yunfa, F., Zhidong, W., Hongyi, L., et al. (2015). A hybrid BCI based on EEG and fNIRS signals improves the performance of decoding motor imagery of both force and speed of hand clenching. *J. Neural. Eng.* 12, 036004. doi: 10.1088/1741-2560/12/3/036004

Yeh, S. C., Lee, S. H., Chan, R. C., Chen, S., and Rizzo, A. (2014). A virtual reality system integrated with robot-assisted haptics to simulate pinch-grip task: motor ingredients for the assessment in chronic stroke. *NeuroRehabilitation* 35, 435–449. doi: 10.3233/NRE-141134

You, S. H., Jang, S. H., Kim, Y. H., Kwon, Y. H., Barrow, I., and Hallett, M. (2005). Cortical reorganization induced by virtual reality therapy in a child with hemiparetic cerebral palsy. *Dev. Med. Child Neurol.* 47, 628–635. doi: 10.1017/S0012162205001234

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Teo, Muthalib, Yamin, Hendy, Bramstedt, Kotsopoulos, Perrey and Ayaz. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Human Neuroscience

CORRECTION
published: 03 February 2017
doi: 10.3389/fnhum.2017.00053

Corrigendum: Does a Combination of Virtual Reality, Neuromodulation and Neuroimaging Provide a Comprehensive Platform for Neurorehabilitation? – A Narrative Review of the Literature

Wei-Peng Teo[1], Makii Muthalib[2,3], Sami Yamin[4,5], Ashlee M. Hendy[6], Kelly Bramstedt[4], Eleftheria Kotsopoulos[4,7], Stephane Perrey[2] and Hasan Ayaz[8,9,10]*

[1] Institute for Physical Activity and Nutrition (IPAN), Deakin University, Burwood, VIC, Australia, [2] EuroMov, University of Montpellier, Montpellier, France, [3] Cognitive Neuroscience Unit, Deakin University, Burwood, VIC, Australia, [4] Liminal Pty Ltd., Melbourne, VIC, Australia, [5] Adult Mental Health, Monash Health, Dandenong, VIC, Australia, [6] School of Exercise and Nutrition Sciences, Deakin University, Burwood, VIC, Australia, [7] Aged Persons Mental Health Service, Monash Health, Cheltenham, VIC, Australia, [8] School of Biomedical Engineering, Science and Health Systems, Drexel University, Philadelphia, PA, USA, [9] Department of Family and Community Health, University of Pennsylvania, Philadelphia, PA, USA, [10] The Division of General Pediatrics, Children's Hospital of Philadelphia, Philadelphia, PA, USA

OPEN ACCESS

Edited and reviewed by:
Mikhail Lebedev,
Duke University, USA

****Correspondence:***
Wei-Peng Teo
weipeng.teo@deakin.edu.au

Received: *12 January 2017*
Accepted: *24 January 2017*
Published: *03 February 2017*

Citation:
Teo W-P, Muthalib M, Yamin S, Hendy AM, Bramstedt K, Kotsopoulos E, Perrey S and Ayaz H (2017) Corrigendum: Does a Combination of Virtual Reality, Neuromodulation and Neuroimaging Provide a Comprehensive Platform for Neurorehabilitation? – A Narrative Review of the Literature. Front. Hum. Neurosci. 11:53. doi: 10.3389/fnhum.2017.00053

Keywords: neurorehabilitation, neuroplasticity, tDCS, fNIRS, EEG, virtual reality therapy

A corrigendum on

Does a Combination of Virtual Reality, Neuromodulation and Neuroimaging Provide a Comprehensive Platform for Neurorehabilitation? – A Narrative Review of the Literature
by Teo, W.-P., Muthalib, M., Yamin, S., Hendy, A. M., Bramstedt, K., Kotsopoulos, E., et al., (2016). Front. Hum. Neurosci. 10:284. doi: 10.3389/fnhum.2016.00284

In the original article, there was a mistake in the legend for Figure 2 as published. The correct legend appears below. The authors apologize for this error and state that this does not change the scientific conclusions of the article in any way.

Figure 2. Stroke participants engaged in VR therapy using an X-Box Kinect motion capture system, MediMoov by NaturalPad, while receiving tDCS.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Teo, Muthalib, Yamin, Hendy, Bramstedt, Kotsopoulos, Perrey and Ayaz. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 25 May 2016
doi: 10.3389/fnhum.2016.00240

High Working Memory Load Impairs Language Processing during a Simulated Piloting Task: An ERP and Pupillometry Study

Mickaël Causse[1,2]*, *Vsevolod Peysakhovich*[1] and *Eve F. Fabre*[1]

[1] Département Conception et Conduite des Véhicules Aéronautiques et Spatiaux, Institut Supérieur de l'Aéronautique et de l'Espace, Toulouse, France, [2] Ecole de Psychologie, Université Laval, Québec, QC, Canada

Given the important amount of visual and auditory linguistic information that pilots have to process, operating an aircraft generates a high working-memory load (WML). In this context, the ability to focus attention on relevant information and to remain responsive to concurrent stimuli might be altered. Consequently, understanding the effects of WML on the processing of both linguistic targets and distractors is of particular interest in the study of pilot performance. In the present work, participants performed a simplified piloting task in which they had to follow one of three colored aircraft, according to specific written instructions (i.e., the written word for the color corresponding to the color of one of the aircraft) and to ignore either congruent or incongruent concurrent auditory distractors (i.e., a spoken name of color). The WML was manipulated with an n-back sub-task. Participants were instructed to apply the current written instruction in the low WML condition, and the 2-back written instruction in the high WML condition. Electrophysiological results revealed a major effect of WML at behavioral (i.e., decline of piloting performance), electrophysiological, and autonomic levels (i.e., greater pupil diameter). Increased WML consumed resources that could not be allocated to the processing of the linguistic stimuli, as indexed by lower P300/P600 amplitudes. Also, significantly, lower P600 responses were measured in incongruent vs. congruent trials in the low WML condition, showing a higher difficulty reorienting attention toward the written instruction, but this effect was canceled in the high WML condition. This suppression of interference in the high load condition is in line with the engagement/distraction trade-off model. We propose that P300/P600 components could be reliable indicators of WML and that they allow an estimation of its impact on the processing of linguistic stimuli.

Keywords: mental workload evaluation, electroencephalography/event-related potential (EEG/ERPs), pupil size, neuroergonomics, human factors, selective attention, attentional orienting

OPEN ACCESS

Edited by:
Hasan Ayaz,
Drexel University, USA

Reviewed by:
Mario Bonato,
Ghent University, Belgium
Lewis Leewui Chuang,
Max Planck Institute for Biological Cybernetics, Germany

***Correspondence:**
Mickaël Causse
mickael.causse@isae.fr

Received: *07 November 2015*
Accepted: *09 May 2016*
Published: *25 May 2016*

Citation:
Causse M, Peysakhovich V and Fabre EF (2016) High Working Memory Load Impairs Language Processing during a Simulated Piloting Task: An ERP and Pupillometry Study. Front. Hum. Neurosci. 10:240. doi: 10.3389/fnhum.2016.00240

INTRODUCTION

Visual-Auditory Interference

Depending on the current task, the information surrounding us is roughly divided into relevant or irrelevant. Naturally, we tend to ignore the irrelevant information and to privilege that which is relevant. Despite such top-down attentional focus on a primary task, concurrent stimuli can capture human attention, especially when they share common characteristics

with the focal task (Folk and Remington, 1998). As stated by Watkins et al. (2007), although distracting subjects, such attentional capture may be advantageous for survival because even a single stimulus can convey critical information about the environment. The distraction phenomenon has been widely investigated over the last decades (e.g., Parmentier, 2014). According to various authors, distraction may result from three different processing steps (Escera et al., 2000; Berti, 2008, 2013; Horváth et al., 2008). First, a pre-attentive change detection step may occur automatically when novel/deviant stimulus appears in the environment. Second, once the concurrent stimulus is detected, attentional resources may be automatically allocated to it (i.e., involuntary orienting of attention) at the expense of goal-relevant stimuli. Third, if the stimulus is irrelevant to the task, a voluntary reorientation of attentional resources from irrelevant stimulus to relevant stimulus may finally occur. These involuntary and voluntary shifts of attention are assumed to interfere with the processing of the information relevant to the task at hand. Generally, the literature on visual-auditory interference tends to support this three-step model using a wide variety of experimental paradigms such as auditory-visual oddball tasks (Andrés et al., 2006; Boll and Berti, 2009; Bendixen et al., 2010; Parmentier and Andrés, 2010; Ljungberg and Parmentier, 2012), task-irrelevant auditory distractor probes (Scheer et al., 2016), visual-auditory Stroop tasks (Roelofs, 2005; Donohue et al., 2013; Elliott et al., 2014), response competition paradigms (e.g., Lavie and Cox, 1997; Tellinghuisen and Nowak, 2003), and more ecological tasks like decision-making during aircraft landing (e.g., Scannella et al., 2013). Overall, auditory distractors are likely to interfere with the processing of visual targets as longer response times and sometimes a decrement in accuracy are observed (Stuart and Carrasco, 1993; Yuval-Greenberg and Deouell, 2009; Chen and Spence, 2011; Berti, 2013; Donohue et al., 2013).

The electroencephalography (EEG) technique is particularly appropriate for the study of the processing of auditory distractors in that it has a high temporal resolution which enables to observe the different steps of the process (Luck and Kappenman, 2011). The mismatch negativity (MMN) event related potential (ERP), a negative deflection occurring between 150 and 250 ms after stimulus onset, maximal at frontal and central sites, was found to be elicited by novel/deviant auditory distractors; this was interpreted as reflecting the pre-attentive detection of the distractors (e.g., Friedman et al., 2001; Berti, 2013). In addition, the novelty-P3 component, a positive deflection occurring around 300 ms after stimulus onset and highest in the frontal lobes, was found to index the involuntary switch of attention to the distractors (Escera et al., 1998, 2000; Friedman et al., 2001). Finally, the reorienting negativity (RON) component, a later negative deflection occurring around 500 ms after the stimulus onset, maximal at frontal site, was found to index the reorientation of attention back to the task after distraction (Schröger and Wolff, 1998; Schröger et al., 2000; Berti and Schröger, 2001; Wetzel et al., 2004).

The Effect of Working Memory Load on the Processing of Auditory Distractors

Many studies have investigated how the processing of distractors is impacted by both perceptual load (Tellinghuisen and Nowak, 2003; Lavie, 2005; Parks et al., 2011; Lavie et al., 2014; Bonato et al., 2015) and working-memory load (WML; Lavie et al., 2004; Kim et al., 2005; SanMiguel et al., 2008). Lavie et al. (2004) proposed that while an increase in perceptual load may reduce distractor interference, an increase in WML may, on the contrary, increase distractor interference. However, various studies investigating the impact of WML on the processing of both visual targets and auditory distractors found opposite results (SanMiguel et al., 2008; Lv et al., 2010; Sörqvist et al., 2012). In SanMiguel et al.'s (2008) study, participants performed an auditory-visual distraction paradigm. While performing the visual task, participants had to ignore task-irrelevant auditory stimuli (i.e., 20% novel environmental sounds and 80% repetitive standard tones). The WML was also manipulated by an n-back task (Kirchner, 1958). In the low load condition (i.e., 0-back), participants had to decide whether the two digits appearing on screen at the same time were the same or different, while in the high load condition (i.e., 1-back) they had to compare the left digit appearing on the screen with the left digit seen in the previous trial. An increase of response times and a decrease of hit rate showed that participants were distracted by novel sounds. Moreover, behavioral data and an attenuation of the amplitude of the novelty-P3 showed that high WML decreased the distraction effect. In another study, Lv et al. (2010) asked participants to remember the order of three (low load) or seven digits (high load). As in SanMiguel et al. (2008), task-irrelevant auditory stimuli were played during the working memory (WM) task with 80% repetitive standard sounds and 20% novel environmental sounds. Participants responded faster and performed significantly better on the task in the low load condition than in the high load condition. Moreover, lower novelty-P3 amplitudes were found in the high WML condition in comparison to the low WML condition, leading the authors to conclude that high WML decreases the distraction effect. Finally, Sörqvist et al. (2012) measured the auditory brainstem responses (ABR; i.e., a neural signal transmitted by the cochlea to the auditory cortex via the brainstem) of participants completing a visual-verbal version of the n-back task (i.e., low load for 1-back, medium load for 2-back and high load for 3-back). They were presented with a visual sequence of letters and were asked to press the space bar on the computer keyboard when the letter was the same as the letter previously presented n letters back in the sequence. The results of this study demonstrate that a medium increase in WML (i.e., 2-back condition) may disrupt the processing of the distractor (i.e., lower ABR responses) without affecting task performance, while a significant increase of WML (i.e., 3-back condition) may not only result in lower ABR response but also in lower accuracy.

The behavioral and electrophysiological (i.e., ERPs, ABR) results of these three studies tend to confirm that high WML

reduces the distraction effect. The impact of WML on distractors processing may also depend on how WM content (e.g., tones, digits, letters, words, geometric forms, etc.) overlaps task-relevant information. Stroop interference was found to increase when target types overlapped WM content (Kim et al., 2005). This result provides a suitable explanation for the discrepancy in results of studies investigating the impact of WML on auditory distractor processing (i.e., an increase of distraction vs. a decrease of distraction under high WML). However, significantly, all these experiments only used tones as auditory distractors. Linguistic distractors were studied previously by Mayer and Kosson (2004), but as far as we know the impact of WML on the processing of visual task-relevant and auditory task-irrelevant linguistic stimuli has never been investigated.

Visual-Auditory Interference and Load on Working Memory in the Cockpit

Operating an aircraft generates a high WML, pilots have to simultaneously select, process, memorize and retrieve an important amount of information, which requires high multitasking and WM capacity (Konig et al., 2005). Previous studies in flight simulators have demonstrated the critical impact of human WM limitations on piloting performance (Taylor et al., 2000; Causse et al., 2011) and how it is likely to compromise flight safety (Borghini et al., 2014). For instance, high WML was found to affect the ability of the pilots to process ATC verbal instructions (Taylor et al., 2005) and simulated auditory alerts (Dehais et al., 2013; Giraudet et al., 2015). However, pilots find themselves confronted with false or irrelevant information (Belcastro et al., 2016). In order to maintain optimal performance, they sometimes need to insulate themselves from auditory distractors to focus on relevant visual information. For example, they have to ignore irrelevant ATC communications and unjustified warnings (e.g., ground proximity system alarm, Loomis and Porter, 1982) while always maintaining the ability to shift attention to concurrent stimuli to decide whether or not they are task-relevant. However, as previously stated, this ability to detect and process concurrent stimuli might be altered, a situation sometimes referred to as cognitive tunneling (Wickens et al., 2015).

Present Study and Hypothesis

The present study aimed at investigating the impact of high WML on the processing of both linguistic visual targets and auditory distractors. Participants performed a simplified piloting task in which they had to control an aircraft using a joystick. They were instructed to follow one of three different colored aircraft displayed on the left of the computer screen, according to written instructions (i.e., the color corresponding to one of the aircraft) displayed in the center of the screen. Similarly to Donohue et al. (2013), each time a color was displayed on the screen, a concurrent spoken distractor (i.e., a color to be ignored) was simultaneously presented either congruently or incongruently with the written instructions. WML was manipulated via an n-back-like sub-task (i.e., the delay between the displayed instruction and its execution). In the low WML condition, participants were instructed to immediately apply the written instruction. In the high WML condition, they had to apply the 2-back written instruction. We measured piloting performance (i.e., accuracy in following the correct aircraft), ERPs, and pupil diameter.

We expected that accuracy may be higher in the congruent vs. incongruent condition. Also, the piloting task should be cognitively more demanding when the WML is high (i.e., 2-back condition) compared to when it is low (i.e., 0-back condition), thus we expected to observe higher accuracy in the low WML condition compared to the high WML condition (in line with Sörqvist et al., 2012). In addition, given that less attentional resources should be available for processing spoken distractors in the high WML condition, in line with various authors (e.g., SanMiguel et al., 2008; Lv et al., 2010; Sörqvist et al., 2012), incongruence may affect accuracy only in the low WML condition.

As pupil diameter was found to vary according to attentional effort (Smallwood et al., 2011), and task engagement (Gilzenrat et al., 2010), we predicted greater pupil diameter in response to incongruent trials compared to congruent trials (Siegle et al., 2004). This greater pupil diameter may be observed in the low WML condition only. Pupil diameter measurements have been also found to be a reliable psycho-physiological marker of WML (van Gerven et al., 2004; Lisi et al., 2015; Peysakhovich et al., 2015), with larger tonic pupil responses indicating an increase in WML. In line with previous studies, we expected to observe greater pupil diameter in the 2-back condition than in the 0-back condition.

At an electrophysiological level, we expected to observe amplitude modulations of ERP components associated with attention allocation (i.e., novelty-P3). Various studies support the fact that the P3a component and the novelty-P3 component are variations of the same potential (Spencer and Polich, 1999; Simons et al., 2001; Polich and Comerchero, 2003; Polich, 2007). Since no novel auditory distractors were used in the present task, we chose to use the generic term P3a and not novelty-P3 when referring to this component. Based again on previous results showing that an increase in WML may lower the distraction effect (SanMiguel et al., 2008; Lv et al., 2010; Sörqvist et al., 2012), we predicted greater P3a amplitudes in response to incongruent trials than to congruent trials in the low WML condition only, reflecting an involuntary switch of attention to the spoken distractors (e.g., Escera et al., 1998, 2000; Friedman et al., 2001). In addition, spoken distractors seem to lead to involuntary semantic evaluation (Parmentier, 2008; Parmentier et al., 2011; Parmentier and Hebrero, 2013). In order words, spoken distractors are, at least partly, processed. Based on these previous results, we predicted amplitude modulations of ERPs associated with language processing (i.e., N400 component). The N400 component is a negative deflection reaching the scalp around 400 ms after stimulus onset, is highest at central-parietal sites (for a review see Kutas and Federmeier, 2011), and was found to index sematic incongruence processing (Pickering and Schweinberger, 2003). Consequently, we predicted greater N400 amplitudes in response to incongruent trials compared

to congruent trials (Kutas and Hillyard, 1980; Kutas and Federmeier, 2011), indicating the processing of both written instructions and spoken distractors (in line with Donohue et al., 2012). As for the P3a, we may also observe N400 amplitude differences in response to incongruent trials vs. congruent trials in the low WML condition only. Finally, we predicted a general effect of WML with lower P3a/N400 amplitudes in the 2-back condition than in the 0-back condition (SanMiguel et al., 2008; Lv et al., 2010).

MATERIALS AND METHODS

Participants

Participants were 24 healthy volunteers (mean age = 24.6, *SD* ± 1.86), all native French speakers. They were recruited at the Institut Supérieur de l'Aéronautique et de l'Espace (ISAE) and were familiar with the aeronautical domain. All were right-handed as assessed by the Edinburgh Handedness Inventory (Oldfield, 1971), had normal auditory acuity and normal or corrected-to-normal vision. None of the participants reported a prior history of neurological disorder. All participants were informed of their rights and gave written informed consent for participation in the study according to the Helsinki Declaration. The research was carried out fulfilling ethical requirements in accordance with the standard procedures of the University of Toulouse. The experimental protocol was reviewed and approved by a national ethic committee (CEEI/IRB00003888).

Material

The piloting task was displayed on a 22" monitor (1680 × 1250) located at a distance of approximately 70 cm from the participants. The screen luminance and the piloting task were identical in all experimental conditions. As a consequence, no confounding effect of light could have jeopardized pupil measurements. Spoken names of colors (i.e., gray, red, blue, yellow, green) were presented via two stereo speakers, positioned on each side of the computer monitor. They were recorded using a synthetic voice taken from the French Voxygen startup website.[1] Throughout the entire experiment, both pupil diameter variations and EEG signals were recorded (see "Electroencephalography" and "Pupillometry" Sections).

Experiment Design

We used a full factorial design with two within-participant factors: WML (Load: Low, High) and the congruency of task-irrelevant auditory distractor (Congruency: Congruent, Incongruent). Participants performed two blocks of 250 trials each, which only presented either "Low" or "High" load trials. Block order was counter-balanced among participants. The congruency of the task-irrelevant auditory distractor was randomly determined per trial. We computed accuracy for each condition as a percentage of correctly targeted aircraft. We considered that an aircraft was correctly followed if the vertical distance between the user's aircraft and the target one was less than 100 pixels for at least 90% of the trial length.

[1]https://www.voxygen.fr/fr

Task and Stimuli

The task involved controlling an aircraft with a joystick in order to follow one of three possible aircraft that were defined by unique colors. The color name corresponding to the color of the aircraft to target was presented in black ink every 4500 ms in the center of the screen for 1000 ms (i.e., written instructions). In addition, a task-irrelevant auditory distractor (i.e., spoken distractor) was also presented simultaneously for 280 ms (i.e., visual-auditory Stroop paradigm; Roelofs, 2005). These written instructions and spoken distractors created four different trial combinations. In the first combination occurring 10% of the time, the spoken and the written color names (i.e., blue, red, green or yellow) were the same (i.e., congruent trials). In the second combination also occurring 10% of the time, the spoken and the written color names differed from one another (i.e., incongruent trials). In the third combination occurring 10% of the time, the spoken color name did not correspond to any aircraft color (i.e., neutral trials). And finally, in the fourth combination occurring 70% of the time, the spoken color name was "gray" (i.e., standard trials). The neutral and standard trials were not analyzed. These two trial combinations were used to inhibit habituation effects and create a rarity effect toward the congruent/incongruent distractors, respectively.

The three aircraft to target were displayed on the left of the computer screen. **Figure 1** shows the layout of the presentation display. The colors of the three aircraft were randomly chosen for each block among four possible colors: red, blue, yellow and green. The initial horizontal position of the targeted aircraft and the control aircraft were equidistant from the center and the edges of the display. They were respectively positioned 30% from the left/right borders of the screen. In order to create a continuous, engaging, and dynamic interaction with the task, every 50 ms the position of the three aircraft on the left changed, a random shift up to 12 pixels vertically and up to 2 pixels

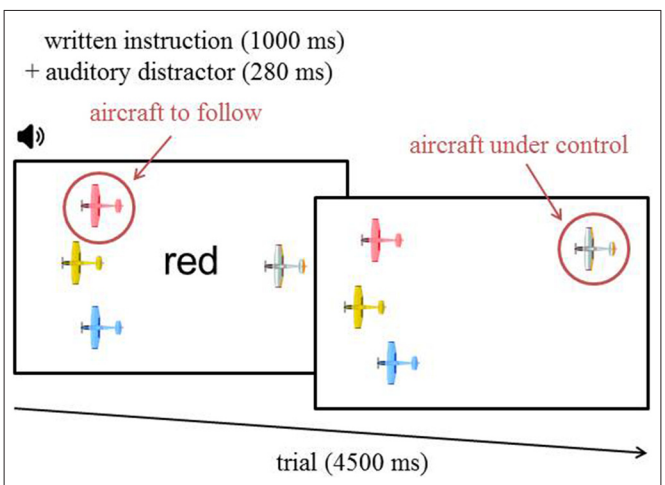


FIGURE 1 | Time course of a task trial. The written instruction was displayed for 1000 ms while the auditory distractor was played for 280 ms. In this particular example, the written target color is "red". Consequently, a congruent auditory distractor would be "red". "Yellow" or "blue" would be incongruent distractors, and "green" would be neutral. The standard distractor was "gray" throughout the whole experiment.

horizontally was applied. The amplitude of this shift was chosen by a moving average filter (of 10th order) of randomly generated numbers (from −1 to 1 to choose the proportion of the greatest authorized shift). A small jitter was also added to the aircraft under control, with a maximum authorized shift up to five pixels vertically and up to one pixel horizontally, so it would be unstable and require continuous control.

Task difficulty was manipulated in terms of WML and the tracking task was designed to be similar to an n-back task. Two difficulty levels corresponded to the delay between the displayed instruction and its execution. Contrary to the classic n-back paradigm in which a participant has to indicate if the current stimulus matches one from n steps earlier in the sequence, our participants had to target the aircraft corresponding to the current written instruction ($n = 0$) in the low WML condition or corresponding to the instruction presented two trials before ($n = 2$) in the high WML condition. After each block, participants filled out the NASA Task Load Index questionnaire (NASA TLX; Hart and Staveland, 1988), see **Figure 2**. This questionnaire provides an evaluation of the subjective mental demand elicited by the task for each level of difficulty.

We did not manipulate piloting complexity *per se*, the control of the aircraft remained constant. Neuroergonomically, this task was designed to recreate an ecological context with an engaging, dynamic and complex situation. Similar to piloting, the participants had to continuously control the trajectory of their aircraft and had to remain responsive to the written and verbal instructions under various WML conditions. In addition, the task also reproduced the multiple conflicting warnings that can confuse crews (Belcastro et al., 2016). Given the voluntary complexity of the task, we did not intend to specifically separate all cognitive processes at each time point. However, the main cognitive abilities engaged during the task were visuospatial (monitoring the position of the aircraft), psychomotor (control of the aircraft) and attentional (toward the instructions).

Procedure

Participants were comfortably seated in an armchair in a sound-dampened experimental room. The room had no windows and the light was kept constant and moderate. After the training session, they were equipped with the EEG electrode cap as well as the electrooculographic electrodes for blink and saccade detection. Eye tracker calibration was performed to record participants' pupil diameter. Participants then performed the four experimental blocks. The instructions were generated so the participant had to change target aircraft every trial. The tracking was continuous during the whole task; when a new instruction was displayed, participants still followed the previous aircraft until they processed the new instruction and switched to the corresponding aircraft. Participants could not repeat aloud the instructions and they were instructed to avoid moving and talking. After each of the four blocks, participants filled out the NASA-TLX.

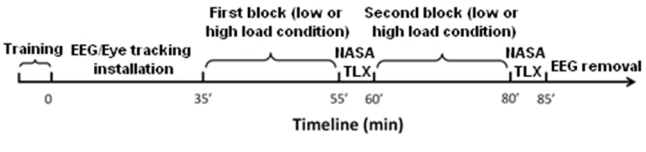


FIGURE 2 | Timeline of the experiment. The whole procedure lasted 85 min, including 5 min training, 30 min to install the EEG sensors, 40 min of EEG recording and 10 min to fill the two NASA TLX questionnaires.

Electroencephalography

EEG was amplified and recorded with an ActiveTwo BioSemi system (BioSemi, Amsterdam, Netherlands) from 30 Ag/AgCl active electrodes mounted on a cap and placed on the scalp according to the International 10–20 System (FP1, FP2, AF3, AF4, F7, F3, Fz, F4, F8, FC5, FC1, FC2, FC6, CP5, CP1, Cz, CP2, CP6, P7, P3, Pz, P4, P8, T7, T8, PO3, PO4, O1, Oz, O2) plus two sites below the eyes for monitoring eye movements. Analyses were focused on 23 electrodes of interest. Two additional electrodes placed close to Cz, the Common Mode Sense (CMS) active electrode and the Driven Right Leg (DRL) passive electrode, were used to form a feedback loop that maintains the average potential of the participant as close as possible to the AD-box reference potential. Electrode impedance was kept below 5 kΩ for scalp electrodes, and below 10 kΩ for the four eye channels. Skin-electrode contact, obtained using electro-conductive gel, was monitored, keeping voltage offset from the CMS below 25 mV for each measurement site. All the signals were (DC) amplified and digitized continuously with a sampling rate of 512 Hz with an anti-aliasing filter with 3 dB point at 104 Hz (fifth-order sinc filter); no high-pass filtering was applied online. The triggering signals to each word onset were recorded on additional digital channels. EEG data was off-line re-referenced to the average activity of the two mastoids and band-pass filtered (0.1–40 Hz, 12 dB/octave), given that for some participants the low-pass filter was not effective in completely removing the 75 Hz artifact. Epochs were time-locked to instructions onset and extracted in the interval from −200 ms to 800 ms. The 200 ms pre-stimulus baseline was used in all analyses. Given their synchronicity, we could not dissociate the respective contributions of written instructions and auditory distractors on the ERPs. Segments with excessive blinks and/or artifacts (such as excessive muscle activity) were eliminated off-line before data averaging. The lost data (due to artifacts) represented 7%.

Pupillometry

The diameter of participants' left pupil was continuously recorded with a remote SMI RED eye-tracker (SensoMotoric Instruments GmbH, Germany) at a sampling rate of 500 Hz. Before each condition, participants performed a 5-point calibration procedure. The continuous pupillary recordings were cleaned for blink artifacts using linear interpolation, including adjacent 40 ms from each side to avoid eyelid closure artifacts. The data was then filtered with a "two

pass" 9-point filter (low-pass) and segregated into trials by conditions. A trial was validated for the statistical analysis if the time spent blinking during the trial did not exceed 50% (i.e., 2250 ms). This resulted on average in 87% (*SD* = 18%) of validated trials per condition and was not dependent on the condition.

Statistical Analyses

Statistical analyses were performed using Statistica 10 (StatSoft). Differences between the experimental conditions were investigated by using analysis of variance (ANOVA) followed by Tukey's honestly significant difference (HSD) *post hoc* testing.

RESULTS

Performance

A 2 × 2 (Congruency [Congruent; Incongruent] × Load [Low; High]) repeated measures ANOVA revealed a main effect of WML on piloting performance ($F_{(1,23)} = 7.79$, $p < 0.05$, $\eta_p^2 = 0.25$); the participants were better at aircraft targeting in the low WML condition ($M = 85.65$, $SD \pm 13.26$) compared to the high WML condition ($M = 75.55$, $SD \pm 16.65$; see **Figure 3**). On the contrary, we found no effect of congruency ($F_{(1,23)} = 0.08$, $p = 0.78$, neither Load × Congruency interaction ($F_{(1,23)} = 1.18$, $p = 0.29$, $\eta_p^2 = 0.05$) on piloting performance. **Figure 4** shows the temporal evolution of the distance between the user's and the target aircraft for each condition and correct and incorrect trials. This fine-grained analysis confirms that poorer piloting performance in the high WML condition cannot be associated with slow or inaccurate following of the correct aircraft. The decline of piloting performance under high WML was due to incorrect targeting (i.e., the inability to correctly encode and retrieve the aircraft to target). For the incorrect trials, the distance between the target aircraft and the user's was indeed kept constant at about 220 pixels corresponding to a wrong aircraft. On the contrary, the rare incorrect trials under low WML are mainly due to poor following of the correct aircraft (the average distance oscillates just above the 100 pixel threshold).

NASA-TLX Questionnaire

The 2 × 6 (Load [low, high] × NASA-TLX dimensions [Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, Frustration]) repeated measures ANOVA showed a significant effect of load ($F_{(1,23)} = 32.68$, $p < 0.001$, $\eta_p^2 = 0.59$). Participants evaluated the high WML condition as more mentally demanding, see **Figure 5**. The ANOVA also showed a significant effect of the NASA-TLX dimension ($F_{(1,23)} = 8.23$, $p < 0.001$, $\eta_p^2 = 0.26$). The three dimensions with the highest scores were (in descending order): Performance, Effort and Mental Demand. Finally, a Load × NASA-TLX dimensions interaction was found ($F_{(5,15)} = 7.83$, $p < 0.001$, $\eta_p^2 = 0.25$). Most notably, in the high WML condition, the Mental Demand dimension was rated higher on average than all other conditions ($M = 72.50$, $SD = 15.28$). However, this is statistically significant only when compared to Physical Demand and Temporal Demand (HSD: $p < 0.01$ in both comparisons).

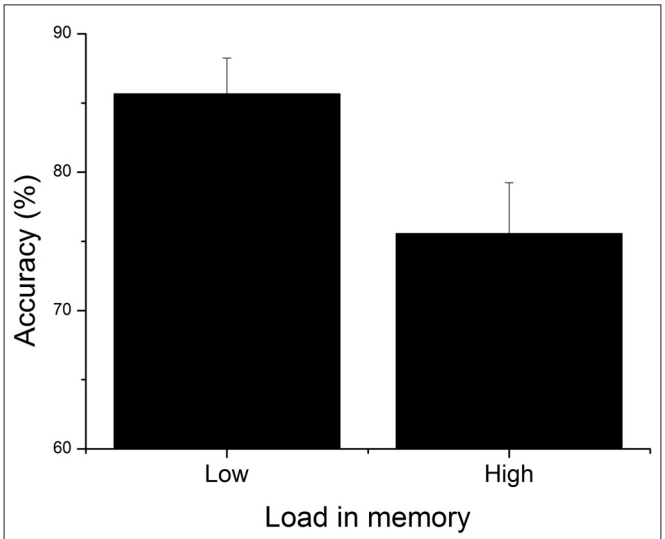


FIGURE 3 | Mean accuracy for low and high load conditions. Error bars indicate SEM.

Electroencephalography

Because of EEG recording issues, data were not available for one participant (corrupted data). First, five 3 × 2 × 2 (Electrode [Fz, Cz, Pz] × Congruency [congruent, incongruent] × Load [low, high]) repeated measure ANOVAs were conducted to assess mean amplitudes of MMN, P3a, P3b, N400 and P600 components on the three midline electrodes. ERPs time windows were determined through both literature and visual analysis of the peak amplitudes. Second, in order to investigate possible topographical differences for these ERPs, the remaining electrodes were collapsed into four regions of interest of five electrodes each (see Siyanova-Chanturia et al., 2012): Left Anterior (AF3, F7, F3, FC1, FC5), Right Anterior (AF4, F4, F8, FC2, FC6), Left Posterior (CP2, CP6, P4, P8, PO4) and Right Posterior (CP1, CP5, P3, P7, PO3). A 4 × 2 × 2 (Region [Left Anterior, Right Anterior, Left Posterior, Right Posterior] × Congruency [congruent, incongruent] × Load [low, high]) ANOVA was conducted. See **Figure 6** for grand average ERP waveforms.

MMN (200–240 ms Time Window)

The MMN amplitude was assessed in terms of the mean amplitude in the 200–240 ms time window. The statistical analysis revealed no significant main effect or interaction ($ps > 0.05$).

P3a (300–330 ms Time Window)

The P3a amplitude was assessed in terms of mean amplitude in the 300–330 ms time window. The analysis revealed a main effect of load ($F_{(1,22)} = 8.13$, $p < 0.01$, $\eta_p^2 = 0.27$), with a greater positivity in the low WML ($M = 3.34$ μV, $SD \pm 6.77$)

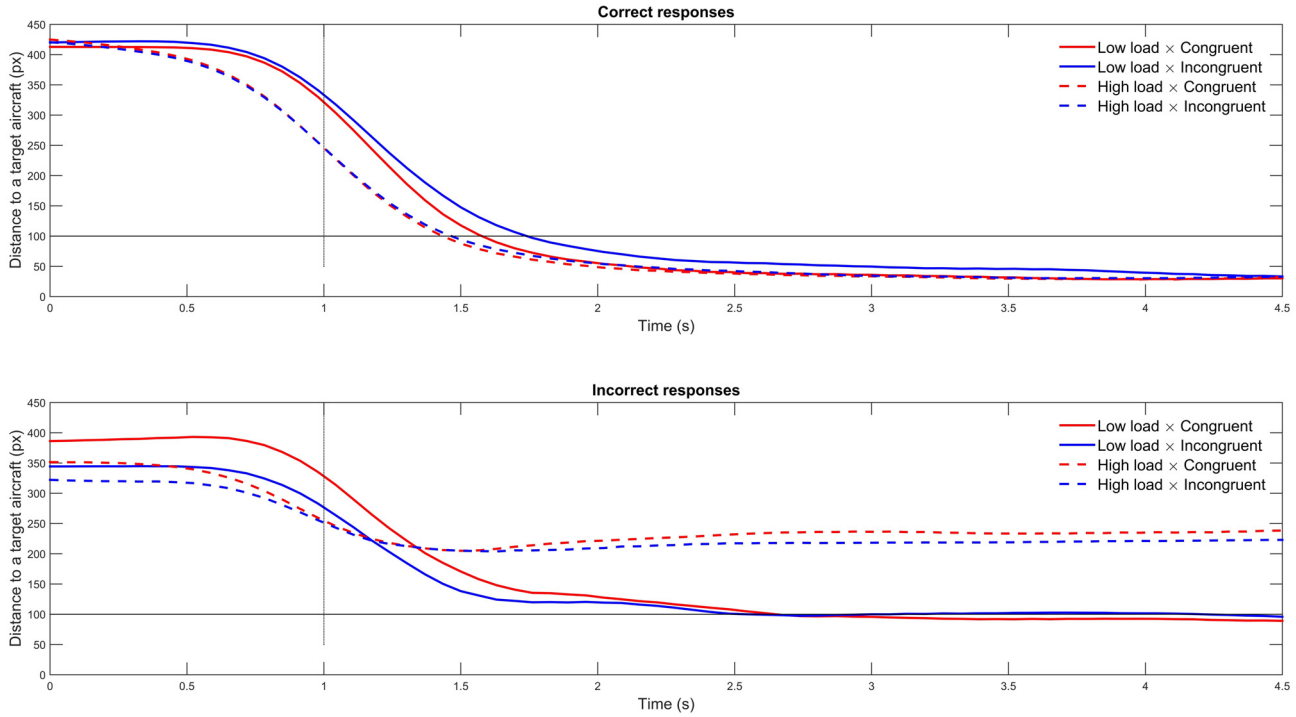


FIGURE 4 | Mean distance to a target aircraft per condition for correct and incorrect trials. The black horizontal line depicts the threshold below which the aircraft was considered as selected. The vertical dashed black line depicts the time when the written instruction disappeared from the computer screen.

than in the high WML condition ($M = 0.10$ μV, $SD \pm 5.92$). The analysis also revealed a significant Electrode × Congruency interaction ($F_{(2,44)} = 3.23$, $p < 0.05$, $\eta_p^2 = 0.13$), with a greater positivity for incongruent trials ($M = 1.21$ μV, $SD \pm 6.64$) than for congruent trials ($M = 0.39$ μV, $SD \pm 7.11$), (HSD: $p < 0.05$) at Fz. No significant differences were found at Cz (HSD: $p = 0.84$) and at Pz (HSD: $p = 0.13$).

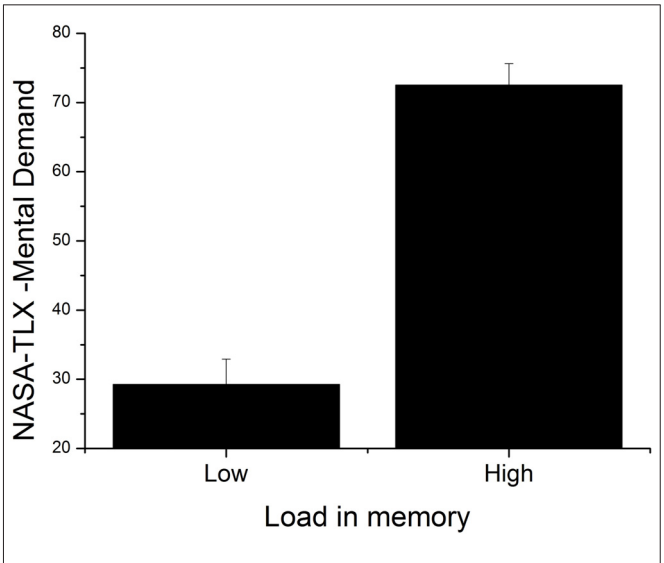


FIGURE 5 | Mean index of perceived mental demand for low and high load conditions. Error bars indicate SEM.

P3b (400–490 ms Time Window)

The P3b amplitude was assessed in terms of the mean amplitude in the 400–490 ms time window. The analysis revealed a main effect of load ($F_{(1,22)} = 5.03$, $p < 0.05$, $\eta_p^2 = 0.19$), with a greater positivity in the low WML condition ($M = 1.27$ μV, $SD \pm 7.89$) than in the high WML condition ($M = -2.03$ μV, $SD \pm 5.24$). The analysis also revealed an Electrode × Congruency interaction ($F_{(2,44)} = 3.52$, $p < 0.05$, $\eta_p^2 = 0.14$), with a greater positivity in response to incongruent trials ($M = -2.08$ μV, $SD \pm 6.28$) compared to congruent trials ($M = -3.22$ μV, $SD \pm 7.76$; HSD: $p = 0.01$) at Fz. No significant differences were found at Cz (HSD: $p = 0.21$) and at Pz (HSD: $p = 0.54$).

N400 (470–540 ms Time Window)

The N400 was assessed in terms of the mean amplitude in the 470–540 ms time window. The analysis revealed a significant Electrode × Load interaction ($F_{(2,44)} = 9.45$, $p < 0.001$, $\eta_p^2 = 0.30$), with greater negativities in the high WML condition (Fz: $M = -3.70$ μV, $SD \pm 6.00$; Cz: $M = -2.91$ μV, $SD \pm 4.93$; Pz: $M = -1.05$ μV, $SD \pm 4.26$) than in the low WML condition (Fz: $M = -2.09$ μV, $SD \pm 7.72$; Cz: $M = -0.36$ μV, $SD \pm 7.52$; Pz: $M = 2.80$ μV, $SD \pm 7.67$; HSD: $ps < 0.001$) on Fz, Cz and Pz. The analysis also revealed a significant Electrode × Congruency interaction ($F_{(2,44)} = 8.95$, $p < 0.001$, $\eta_p^2 = 0.29$), with a

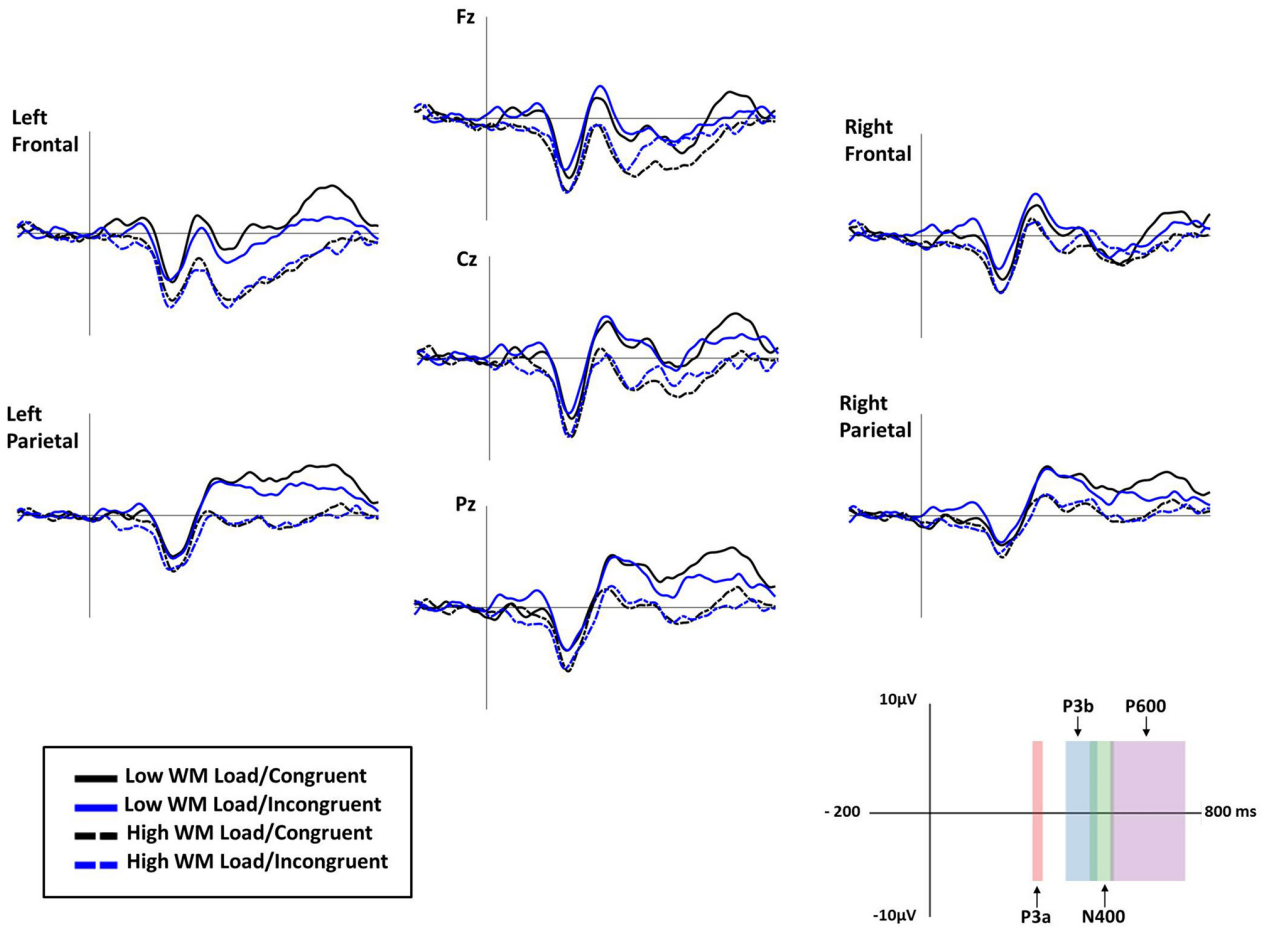


FIGURE 6 | Grand average ERP waveforms for low working memory (WM) load/congruent (black line), low WM load/incongruent (blue line), high WM load/congruent (dashed black line) and high WM load/incongruent trials (dashed blue line) at Fz, Cz, Pz, and left frontal, left parietal, right frontal and right parietal electrode-clusters. In the bottom right corner, the *x*-axis displays the 4 ERP components (vertical bar indicates the onset of both the written instruction and the auditory distractor) and the *y*-axis displays amplitude in microvolts. Negative is plotted down.

greater N400 amplitude for incongruent trials ($M = 0.48$ μV, $SD \pm 6.12$) than for congruent trials ($M = 1.27$ μV, $SD \pm 6.82$) at Pz (HSD: $p < 0.05$). The opposite pattern was found at Fz (incongruent trials: $M = -2.17$ μV, $SD \pm 6.48$; congruent trials: $M = -3.62$ μV, $SD \pm 7.34$; HSD: $p < 0.001$; this apparent contradiction for decreased N400 for incongruent trials in Fz can be explained by the previous more positive P3a on this electrode in this condition). No difference was found at Cz (incongruent trials: $M = -1.33$ μV, $SD \pm 6.16$; congruent trials: $M = -1.94$ μV, $SD \pm 6.78$, (HSD: $p = 0.11$). The topographical analysis revealed no significant effect or interaction.

P600 (530–750 ms Time Window)

The P600 was assessed in terms of the mean amplitude in the 530–750 ms time window. The analysis revealed a main effect of load ($F_{(1,22)} = 8.15$, $p < 0.01$, $\eta_p^2 = 0.27$), with a greater positivity in the low WML condition ($M = 2.44$ μV, $SD \pm 7.08$) than in the high WML condition ($M = -1.02$ μV, $SD \pm 5.18$). A significant Electrode × Congruency interaction ($F_{(2,44)} = 5.53$, $p < 0.01$, $\eta_p^2 = 0.20$), was found with a greater positivity observed for congruent trials ($M = 2.97$ μV, $SD \pm 6.45$) compared to incongruent trials ($M = 1.62$ μV, $SD \pm 5.43$) at Pz (HSD: $p < 0.001$), but not at Fz (HSD: $p = 0.44$) nor at Cz (HSD: $p = 0.75$). The topographical analysis revealed a significant Load × Region interaction ($F_{(3,66)} = 8.83$, $p < 0.001$, $\eta_p^2 = 0.29$), with greater positivity in both left and right posterior regions in the low load condition (respectively: $M = -0.28$ μV, $SD \pm 3.77$; $M = -0.14$ μV, $SD \pm 3.82$) than in the high WML condition (respectively: $M = -1.14$ μV, $SD \pm 7.72$, $p < 0.001$; $M = -1.61$ μV, $SD \pm 8.34$; HSD: $p < 0.001$). The analysis also revealed a second significant Load × Congruency interaction ($F_{(1,22)} = 5.11$, $p < 0.05$, $\eta_p^2 = 0.19$), with a greater positivity in response to congruent trials compared to incongruent trials in the low WML condition (respectively: $M = -0.45$ μV, $SD \pm 6.43$; $M = -1.79$ μV, $SD \pm 8.47$; HSD: $p < 0.05$), but not in the high WML condition (respectively: $M = -0.67$ μV, $SD \pm 3.89$; $M = -0.24$ μV, $SD \pm 4.09$; HSD: $p = 0.44$). Moreover, a greater positivity was also found for incongruent trials in the high WML condition ($M = -0.24$ μV, $SD \pm 4.09$) than in the low WML condition ($M = -1.79$ μV, $SD \pm 8.47$; HSD: $p < 0.05$).

Pupillometry

Two 2 × 2 (Load [low, high]) × Congruency [congruent, incongruent]) repeated measure ANOVAs were carried out on the mean value of 1.5-s recording starting from 1-s post-stimulus for tonic pupil response (absolute diameter) and phasic pupil response (relative dilation). This interval largely includes the peak of pupillary reaction known to appear about 1200–1500 ms post-stimulus (Beatty and Lucero-Wagoner, 2000). We used the mean value of 500 ms pre-stimulus as a baseline value for statistical analyses of the phasic pupil response.

Tonic Pupil Response

The ANOVA showed a significant effect of WML ($F_{(1,23)} = 9.14$, $p < 0.01$, $\eta_p^2 = 0.28$; see **Figure 7**). The high WML condition elicited larger tonic pupil response ($M = 3.76$ mm, $SD \pm 0.11$) compared to the low WML condition ($M = 3.68$ mm, $SD \pm 0.11$). No Congruency ($F_{(1,23)} = 1.82$, $p = 0.19$, $\eta_p^2 = 0.07$), nor Load × Congruency interaction ($F_{(1,23)} = 0.85$, $p = 0.37$, $\eta_p^2 = 0.04$] was found.

Phasic Pupil Response

No significant effects of WML ($F_{(1,23)} = 0.85$, $p = 0.37$, $\eta_p^2 = 0.04$), congruency ($F_{(1,23)} = 0.53$, $p = 0.47$, $\eta_p^2 = 0.02$), or Load × Congruency interaction ($F_{(1,23)} = 0.24$, $p = 0.63$, $\eta_p^2 = 0.01$) were found for phasic pupil response.

DISCUSSION

Operating an aircraft is cognitively demanding and requires high multitasking and WM capacities (Konig et al., 2005). Pilots have to simultaneously process, memorize and retrieve an important amount of visual and auditory information. In addition to this high cognitive load, pilots sometimes have to ignore irrelevant auditory distractors such as background ATC communications and false alarms. Paradoxically, they must remain responsive to unexpected stimuli at all times. Previous studies emphasized that attention-demanding settings that generate a high WML can impair the perception of unexpected/irrelevant stimuli (Berti and Schröger, 2003; Sörqvist et al., 2012). However, the impact of WML on the processing of linguistic material has rarely been tested in an explicit way. In the present study, participants completed a cross-modal version of the Stroop task paradigm (Donohue et al., 2013) adapted to a dynamic piloting task in combination with ERP and pupillary measurements. They were asked to take into account a written target instruction (i.e., the name of a color) and to ignore a concurrent spoken distractor (i.e., also a color). We investigated how WML modulated the processing of both target and distractors and to what extent it affected piloting performance (this latter being dependent on the processing/maintenance/retrieval of written instructions). Overall results revealed a subtle effect of congruency that was observable only at an electrophysiological level, an interaction between congruency and load on P600 amplitude, and a major effect of WML at behavioral, electrophysiological, and autonomic levels.

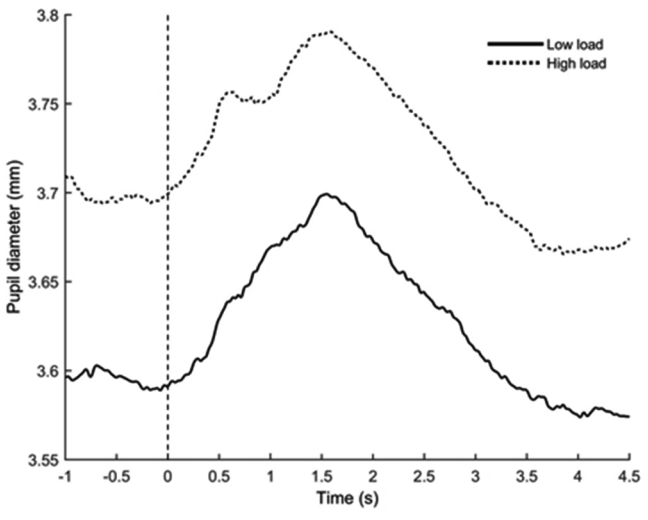


FIGURE 7 | Grand-averaged mean pupil diameter during trials for low load (solid black line) and high load (dotted black line) conditions. The horizontal axis denotes time in seconds (s) and the vertical axis denotes pupil diameter in millimeters (mm). Zero on the timeline indicates the onset of both the written instruction and the auditory distractor.

Impact of the Congruency

At a behavioral level, the results revealed no main effect of the congruency between the written target and the spoken distractor, indicating that the latter does not interfere enough with the processing of the written instruction to affect piloting performance. This absence of interference at a behavioral level may be due to some limitations in the experimental paradigm. First, a recent study showed that distraction is maximal when an auditory distractor is presented 400 ms before the onset of the visual information of interest, but that it is significantly reduced when both stimuli of interest and the distractor are presented at the same time (Donohue et al., 2013). Since in the present study, ERP measurements were performed, it would have been complex to interpret electrophysiological results if the distractor and the target had been presented at different times. Presenting the spoken distractor before the written instruction could have led to a greater distraction effect, observable at the behavioral level. Second, in general, the distraction effect is observable on reaction times rather than accuracy measures. Longer reaction times reflect the penalty yielded by the involuntary orientation of attention to and away from deviant sounds (Parmentier, 2008). Given the continuous control of the aircraft trajectory with the joystick, it was not possible to measure reaction times accurately. We could have captured variations of reaction times if the task allowed such measurements.

However, electrophysiological data demonstrate the increased complexity for processing the written instructions when incongruent spoken distractors were presented simultaneously, with greater P3a and N400 amplitudes in incongruent trials compared to congruent trials. According to Polich (2007), the P3a component may be generated when focal attention on the

task-relevant stimuli is captured by a distractor; and indexes the automatic allocation of attentional resources to the distractor at the expense of goal-relevant stimuli. Greater P3a amplitudes in response to incongruent trials may reflect the recruitment of supplementary attentional resources and the involuntary orienting of attention to the spoken distractors (Escera et al., 1998, 2000; Friedman et al., 2001).

The N400 component was found to index semantic incongruence processing (Kutas and Federmeier, 2011). In line with the literature (Hanslmayr et al., 2008; Kutas and Federmeier, 2011; Donohue et al., 2012), greater N400 amplitudes were found at parietal sites in incongruent trials compared to congruent trials. Some behavioral studies have shown that not only are attentional resources allocated to spoken distractors, but that they also lead to an involuntary semantic evaluation (Parmentier, 2008; Parmentier et al., 2011; Parmentier and Hebrero, 2013). The N400 component may index this involuntary semantic analysis of the spoken distractor. However, as the processing of the spoken distractor was not sufficient to trigger significant distraction observable at a behavioral level, we conclude that the mobilization of additional attentional resources indexed by the P3a component was sufficient to process both the relevant written instruction and the irrelevant spoken distractor.

Impact of the Working-Memory Load

At both behavioral and subjective levels, a higher level of WML was found to decrease the accuracy and was associated with a higher perceived mental demand as shown by the NASA-TLX questionnaire. Moreover, the pupil diameter was also modulated by WML, with greater tonic pupillary responses under high WML compared to low WML, thus objectively confirming the increased task difficulty in the high WML condition.

Given that difficulty levels were generated using an n-back-like task, the tonic pupil diameter indicates the WML that was maintained throughout the block. Thus, when segregated into trials, the tonic reaction is always higher during the high WM condition. Pupil diameter was found to be correlated with WML and task difficulty (Kahneman and Beatty, 1966; Beatty and Lucero-Wagoner, 2000; Karatekin et al., 2007; Causse et al., 2010; Peysakhovich et al., 2015). Greater pupil diameter observed during trials of high WML provides additional evidence that higher WML increases the attentional resources allocated to the task. The present study also demonstrates that it is possible to measure pupil dilation relative to variations in WML in a task that requires natural eye movement if luminosity remains constant.

At an electrophysiological level, an increase in WML was found to affect the allocation of attentional resources (i.e., P3a and P3b components) and the semantic processing of both the written instruction and the spoken distractors (i.e., P600 component). The amplitudes of both the P3a and P3b components were lower in the high load condition than in the low load condition. An increased WM demand in the n-back task was previously shown to enlist attentional resources and processing capacity away from the matching subtask (i.e., the comparison process; Watter et al., 2001) and was associated to reduced P300 amplitudes. For this reason, decreased amplitudes of P3a and P3b in the high WML condition of the present study is interpreted as an "overall" alteration of the ability to orient attention and process environmental stimuli, including the critical written instruction. A classic hypothesis postulates that the P3a component originates from stimulus-driven frontal attention mechanisms when sufficient attentional focus is engaged (Polich, 2007). Recent research has shown that the P3a component indicates information selection within the WM (Berti, 2016). The P3b component originates from temporal-parietal activity when subsequent attentional resources promote context updating operations and memory processing (Knight, 1996; Brázdil et al., 2001; Polich, 2007). The P3b component is also considered to indicate stimulus analysis and response initiation (Verleger et al., 2005). Lower P3a and P3b responses in the 2-back condition indicate the mobilization of resources that cannot be allocated to processing the current written instruction since these resources are utilized by complex operations in WM. An increased N400 negativity was found at parietal sites in the high load condition compared to the low load condition. However, this increased negativity could be attributable to a decrease in amplitude of both P3a and P3b components preceding the N400 component. Therefore, it is difficult to draw definitive conclusions on the impact of load on this component.

Our results also revealed an unpredicted decrease in amplitude of the P600 in high WML compared to the low WML condition. The P600 component is a positive deflection occurring around 500 ms after stimulus onset and is known to reflect language revision processes (Friederici, 1995; Kaan et al., 2000; Papageorgiou et al., 2001). Interestingly, a recent study suggests that the P600 component may also index attention reorientation processes (Sassenhagen and Bornkessel-Schlesewsky, 2015). While the voluntary reorientation of attention was found to be indexed by the RON component (e.g., Schröger and Wolff, 1998; Schröger et al., 2000; Berti and Schröger, 2001; Wetzel et al., 2004), occurring around 400 ms after stimulus onset for simple stimuli (i.e., tones), in the present study, this process step appears to be delayed for linguistic stimuli (i.e., words) and indexed by the P600 component. It is likely that both the voluntary reorientation of attention back to the written instruction and its reanalysis/rechecking were affected in the high WML condition because fewer WM resources were available. This interpretation corresponds with a recent study showing that patients with WM deficits demonstrate lower P600 amplitudes (El-Kholy et al., 2012).

Interaction between Load and Congruency

At a behavioral level, the results of the present study revealed no interaction effect between the WML and target/distractor congruency. Previously mentioned limitations in the experimental paradigm could have contributed to this null effect. At an electrophysiological level, congruency was found to modulate the amplitude of the P600 component only in the

low WML condition, with lower P600 responses in incongruent trials compared to congruent trials. Moreover, lower P600 amplitudes were found in response to incongruent trials in the low WML condition compared to the high WML condition. Taken together, these results are consistent with the idea that the interference effect is reduced as WML increases (e.g., SanMiguel et al., 2008; Lv et al., 2010; Sörqvist et al., 2012; Scharinger et al., 2015). We assume that attentional resources in WM were not sufficient to intensively process auditory distractors in the high WML condition. Consequently, no congruency effect was observed and the voluntary reorientation of attention back to the target instruction was easier. Higher task engagement generated by high WML condition tended to further reduce the effect of distraction observable only at an electrophysiological level. According to Kim et al. (2005), WML can either impair or benefit attentional selection depending on whether it overlaps with target/distractor processing or not. Their results showed that Stroop interference increased when the type of WML overlapped with the type of information required for the task. At the same time, Stroop interference decreased when the type of WML overlapped with distractor processing. In the present study, WML was elicited by exactly the same type of content as the targets/distractors (i.e., verbal stimuli). As a consequence, WML most likely impacted the processing of both written instructions and distractors, as shown by the decline in piloting performance and the mitigation of interference caused by the incongruent distractor, demonstrated by the P600 results.

We found results in line with previous studies (SanMiguel et al., 2008; Lv et al., 2010; Sörqvist et al., 2012) but also apparently in contradiction with the load theory, which predicts that high WML increases distractor interference by impeding inhibitory cognitive control (de Fockert et al., 2001; Lavie et al., 2004; Woodman and Luck, 2004). These contradictory results could possibly be explained by differences in the experimental paradigm. In the studies finding that WML enhances distraction, two independent tasks were combined: a "WM task" and a "selective attention task". In these experiments, a trial consisted in: first, a memorization phase (e.g., memorizing a list of digits), second, a selective attention task (e.g., classifying a written list of famous names such as pop stars or politicians while ignoring distractor faces), and third, a memory probe (e.g., reporting the digit that followed a probe in the memory set presented at the beginning of the trial). However, in the studies in which high WML was found to reduce the distraction effect, including the present study, the WM task and the selective attention task were nested. In other words, when the distraction task and the WM task were concomitant, WML was found to lower distractor interference, while when they were not, WML enhanced distractor interference. Future studies should test this hypothesis by comparing the effect of delayed vs. concomitant WML on distractor processing. Also, in order to get closer to the real piloting situation, future works should use spoken material conveying relevant, neutral or irrelevant information to the piloting task in order to investigate top-down processing modulations associated with spoken information according to their value to the focal task.

CONCLUSION

In the present study, we adapted a visual-auditory version of the Stroop paradigm (Donohue et al., 2013) to a dynamic simulated piloting task in combination with ERPs and pupillary measurements. WML was also manipulated using an n-back task. Electrophysiological results revealed that more attentional resources were mobilized during incongruent trials (i.e., P3a component) and that the incongruence between written instructions and spoken distractors was detected (i.e., N400 component), suggesting that spoken distractors were semantically processed. This result confirms previous behavioral findings showing that not only are attentional resources allocated to spoken distractors but that they also lead to an involuntary semantic evaluation of the latter (Parmentier, 2008; Parmentier et al., 2011; Parmentier and Hebrero, 2013). However, the semantic processing of distractors was not sufficient to impair task accuracy, probably thanks to the mobilization of supplementary attentional resources that enabled participants to process both the target and the incongruent distractor.

Overall, high WML disrupted the processing of both the visual target instruction and the spoken distractors. High WML provoked a decline in task accuracy and increased pupil diameter. At an electrophysiological level, an alteration of the ERPs component was found when the WML was high. In particular we found lower P3a/P3b responses indexing the mobilization of resources by the WM task that could not be allocated to orient the attention and process environmental stimuli, including the critical written instruction. We also found lower P600 responses, showing the impairment of voluntary reorientation of attention back to the processing of written instruction, thus altering the reanalysis/rechecking process. In addition, lower P600 responses in incongruent trials than to congruent trials were significant in the low WML condition only, indexing an easier voluntary reorientation of attention back to the target instruction because interference was reduced in the high WML condition. Our electrophysiological results can be related to a recent study (Scheer et al., 2016) that support a three-stage distraction model with ERPs that reflect the post-sensory detection of the task-irrelevant stimulus, engagement, and re-orientation back to the relevant task. They showed that the difficulty of a steering task not only diminished the amplitudes of early P3, late P3 but also the re-orientation negativity (RON) to the steering task (reorientation being rather indexed by P600 component in our study). Our results are also consistent with theories such as enhancing inhibitory control (Scharinger et al., 2015) and the task engagement/distraction trade-off model (Sörqvist and Rönnberg, 2014) with the idea that an higher cognitive engagement in a task can diminish the distractibility and responsiveness to additional stimuli. From an operational point of view, we confirm that high WML can compromise the ability of pilots to process, maintain, and execute ATC verbal instructions (2005) and to react to critical auditory alerts (Giraudet et al., 2015). We also demonstrate that P300 and P600 components are good candidates to detect variations in WM demand and that they allow estimation of its impact on the processing of linguistic stimuli.

AUTHOR CONTRIBUTIONS

MC: designed the experiment, conducted data analysis, interpreted the data and wrote the manuscript; EFF: administered the experiment, conducted data analysis, interpreted the data and wrote the manuscript; VP: designed the experiment, developed the experimental task, conducted data analysis and wrote the manuscript.

ACKNOWLEDGMENTS

Authors would like to thank Marine Gonzalez and Louise Giraudet for their precious help with data collection. The authors would also like to thank the two reviewers for their helpful comments. We are also grateful to Zarrin Chua and Joseph Shea for proofreading earlier versions of the article.

REFERENCES

Andrés, P., Parmentier, F. B., and Escera, C. (2006). The effect of age on involuntary capture of attention by irrelevant sounds: a test of the frontal hypothesis of aging. *Neuropsychologia* 44, 2564–2568. doi: 10.1016/j.neuropsychologia.2006.05.005

Beatty, J., and Lucero-Wagoner, B. (2000). The pupillary system. *Handb. Psychophysiol.* 2, 142–162.

Belcastro, C. M., Foster, J. V., Newman, R. L., Groff, L., Crider, D. A., and Klyde, D. H. (2016). "Aircraft loss of control: problem analysis for the development and validation of technology solutions," *Paper presented at the AIAA Guidance, Navigation and Control Conference,* San Diego, CA.

Bendixen, A., Grimm, S., Deouell, L. Y., Wetzel, N., Mädebach, A., and Schröger, E. (2010). The time-course of auditory and visual distraction effects in a new crossmodal paradigm. *Neuropsychologia* 48, 2130–2139. doi: 10.1016/j.neuropsychologia.2010.04.004

Berti, S. (2008). Cognitive control after distraction: event-related brain potentials (ERPs) dissociate between different processes of attentional allocation. *Psychophysiology* 45, 608–620. doi: 10.1111/j.1469-8986.2008.00660.x

Berti, S. (2013). The role of auditory transient and deviance processing in distraction of task performance: a combined behavioral and event-related brain potential study. *Front. Hum. Neurosci.* 7:352. doi: 10.3389/fnhum.2013.00352

Berti, S. (2016). Switching attention within working memory is reflected in the p3a component of the human event-related brain potential. [original research]. *Front. Hum. Neurosci.* 9:701. doi: 10.3389/fnhum.2015.00701

Berti, S., and Schröger, E. (2001). A comparison of auditory and visual distraction effects: behavioral and event-related indices. *Brain Res. Cogn. Brain Res.* 10, 265–273. doi: 10.1016/s0926-6410(00)00044-6

Berti, S., and Schröger, E. (2003). Working memory controls involuntary attention switching: evidence from an auditory distraction paradigm. *Eur. J. Neurosci.* 17, 1119–1122. doi: 10.1046/j.1460-9568.2003.02527.x

Boll, S., and Berti, S. (2009). Distraction of task-relevant information processing by irrelevant changes in auditory, visual and bimodal stimulus features: a behavioral and event-related potential study. *Psychophysiology* 46, 645–654. doi: 10.1111/j.1469-8986.2009.00803.x

Bonato, M., Spironelli, C., Lisi, M., Priftis, K., and Zorzi, M. (2015). Effects of multimodal load on spatial monitoring as revealed by ERPs. *PLoS One* 10:e0136719. doi: 10.1371/journal.pone.0136719

Borghini, G., Astolfi, L., Vecchiato, G., Mattia, D., and Babiloni, F. (2014). Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. *Neurosci. Biobehav. Rev.* 44, 58–75. doi: 10.1016/j.neubiorev.2012.10.003

Brázdil, M., Rektor, I., Daniel, P., Dufek, M., and Jurák, P. (2001). Intracerebral event-related potentials to subthreshold target stimuli. *Clin. Neurophysiol.* 112, 650–661. doi: 10.1016/s1388-2457(01)00463-1

Causse, M., Dehais, F., and Pastor, J. (2011). Executive functions and pilot characteristics predict flight simulator performance in general aviation pilots. *Int. J. Aviat. Psychol.* 21, 217–234. doi: 10.1080/10508414.2011.582441

Causse, M., Sénard, J.-M., Démonet, J. F., and Pastor, J. (2010). Monitoring cognitive and emotional processes through pupil and cardiac response during dynamic versus logical task. *Appl. Psychophysiol. Biofeedback* 35, 115–123. doi: 10.1007/s10484-009-9115-0

Chen, Y.-C., and Spence, C. (2011). Crossmodal semantic priming by naturalistic sounds and spoken words enhances visual sensitivity. *J. Exp. Psychol. Hum. Percept. Perform.* 37, 1554–1568. doi: 10.1037/a0024329

de Fockert, J. W., Rees, G., Frith, C. D., and Lavie, N. (2001). The role of working memory in visual selective attention. *Science* 291, 1803–1806. doi: 10.1126/science.1056496

Dehais, F., Causse, M., Vachon, F., Régis, N., Menant, E., and Tremblay, S. (2013). Failure to detect critical auditory alerts in the cockpit evidence for inattentional deafness. *Hum. Factors* 56, 631–644. doi: 10.1177/0018720813510735

Donohue, S. E., Appelbaum, L. G., Park, C. J., Roberts, K. C., and Woldorff, M. G. (2013). Cross-modal stimulus conflict: the behavioral effects of stimulus input timing in a visual-auditory Stroop task. *PLoS One* 8:e62802. doi: 10.1371/journal.pone.0062802

Donohue, S. E., Liotti, M., Perez, R. III, and Woldorff, M. G. (2012). Is conflict monitoring supramodal? Spatiotemporal dynamics of cognitive control processes in an auditory Stroop task. *Cogn. Affect. Behav. Neurosci.* 12, 1–15. doi: 10.3758/s13415-011-0060-z

El-Kholy, O. A., Ramadan, M., El-Sheikh, M., and Ali, M. (2012). Impairment in working memory in multiple sclerosis. *Egypt. J. Psychiatry* 33, 117–125. doi: 10.7123/01.ejp.0000414294.51794.7d

Elliott, E. M., Morey, C. C., Morey, R. D., Eaves, S. D., Shelton, J. T., and Lutfi-Proctor, D. A. (2014). The role of modality: auditory and visual distractors in stroop interference. *J. Cogn. Psychol.* 26, 15–26. doi: 10.1080/20445911.2013.859133

Escera, C., Alho, K., Schröger, E., and Winkler, I. (2000). Involuntary attention and distractibility as evaluated with event-related brain potentials. *Audiol. Neurotol.* 5, 151–166. doi: 10.1159/000013877

Escera, C., Alho, K., Winkler, I., and Näätänen, R. (1998). Neural mechanisms of involuntary attention to acoustic novelty and change. *J. Cogn. Neurosci.* 10, 590–604. doi: 10.1162/089892998562997

Folk, C. L., and Remington, R. (1998). Selectivity in distraction by irrelevant featural singletons: evidence for two forms of attentional capture. *J. Exp. Psychol. Hum. Percept. Perform.* 24, 847–858. doi: 10.1037/0096-1523.24.3.847

Friederici, A. D. (1995). The time course of syntactic activation during language processing: a model based on neuropsychological and neurophysiological data. *Brain Lang.* 50, 259–281. doi: 10.1006/brln.1995.1048

Friedman, D., Cycowicz, Y. M., and Gaeta, H. (2001). The novelty P3: an event-related brain potential (ERP) sign of the brain's evaluation of novelty. *Neurosci. Biobehav. Rev.* 25, 355–373. doi: 10.1016/s0149-7634(01)00019-7

Gilzenrat, M. S., Nieuwenhuis, S., Jepma, M., and Cohen, J. D. (2010). Pupil diameter tracks changes in control state predicted by the adaptive gain theory of locus coeruleus function. *Cogn. Affect. Behav. Neurosci.* 10, 252–269. doi: 10.3758/CABN.10.2.252

Giraudet, L., St-Louis, M.-E., Scannella, S., and Causse, M. (2015). P300 event-related potential as an indicator of inattentional deafness? *PLoS One* 10:e0118556. doi: 10.1371/journal.pone.0118556

Hanslmayr, S., Pastötter, B., Bäuml, K.-H., Gruber, S., Wimber, M., and Klimesch, W. (2008). The electrophysiological dynamics of interference during the Stroop task. *J. Cogn. Neurosci.* 20, 215–225. doi: 10.1162/jocn.2008.20020

Hart, S., and Staveland, L. (1988). "Development of NASA-TLX (Task Load Index): results of empirical and theoretical research," in *Human Mental Workload*, (Vol. 1) eds P. A. Hancock and N. Meshkati (Amsterdam: North Holland Press), 139–183.

Horváth, J., Winkler, I., and Bendixen, A. (2008). Do N1/MMN, P3a and RON form a strongly coupled chain reflecting the three stages of auditory distraction? *Biol. Psychol.* 79, 139–147. doi: 10.1016/j.biopsycho.2008.04.001

Kaan, E., Harris, A., Gibson, E., and Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Lang. Cogn. Process.* 15, 159–201. doi: 10.1080/016909600386084

Kahneman, D., and Beatty, J. (1966). Pupil diameter and load on memory. *Science* 154, 1583–1585. doi: 10.1126/science.154.3756.1583

Karatekin, C., Marcus, D. J., and Couperus, J. W. (2007). Regulation of cognitive resources during sustained attention and working memory in 10-year-olds and adults. *Psychophysiology* 44, 128–144. doi: 10.1111/j.1469-8986.2006.00477.x

Kim, S.-Y., Kim, M.-S., and Chun, M. M. (2005). Concurrent working memory load can reduce distraction. *Proc. Natl. Acad. Sci. U S A* 102, 16524–16529. doi: 10.1073/pnas.0505454102

Kirchner, W. K. (1958). Age differences in short-term retention of rapidly changing information. *J. Exp. Psychol.* 55, 352–358. doi: 10.1037/h0043688

Knight, R. T. (1996). Contribution of human hippocampal region to novelty detection. *Nature* 383, 256–259. doi: 10.1038/383256a0

Konig, C. J., Buhner, M., and Murling, G. (2005). Working memory, fluid intelligence and attention are predictors of multitasking performance, but polychronicity and extraversion are not. *Hum. Perf.* 18, 243–266. doi: 10.1207/s15327043hup1803_3

Kutas, M., and Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event related brain potential (ERP). *Annu. Rev. Psychol.* 62, 621–647. doi: 10.1146/annurev.psych.093008.131123

Kutas, M., and Hillyard, S. A. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science* 207, 203–205. doi: 10.1126/science.7350657

Lavie, N. (2005). Distracted and confused?: selective attention under load. *Trends Cogn. Sci.* 9, 75–82. doi: 10.1016/j.tics.2004.12.004

Lavie, N., Beck, D. M., and Konstantinou, N. (2014). Blinded by the load: attention, awareness and the role of perceptual load. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 369:20130205. doi: 10.1098/rstb.2013.0205

Lavie, N., and Cox, S. (1997). On the efficiency of visual selective attention: efficient visual search leads to inefficient distractor rejection. *Psychol. Sci.* 8, 395–396. doi: 10.1111/j.1467-9280.1997.tb00432.x

Lavie, N., Hirst, A., de Fockert, J. W., and Viding, E. (2004). Load theory of selective attention and cognitive control. *J. Exp. Psychol. Gen.* 133, 339–354. doi: 10.1037/0096-3445.133.3.339

Lisi, M., Bonato, M., and Zorzi, M. (2015). Pupil dilation reveals top-down attentional load during spatial monitoring. *Biol. Psychol.* 112, 39–45. doi: 10.1016/j.biopsycho.2015.10.002

Ljungberg, J. K., and Parmentier, F. B. (2012). Cross-modal distraction by deviance. *Exp. Psychol.* 59, 355–363. doi: 10.1027/1618-3169/a000164

Loomis, J. P., and Porter, R. F. (1982). The performance of warning systems in avoiding controlled-flight-into-terrain (CFIT) accidents. *Aviat. Space Environ. Med.* 53, 1085–1090.

Luck, S. J., and Kappenman, E. S. (2011). *The Oxford Handbook of Event-Related Potential Components.* New York, NY: Oxford University Press.

Lv, J.-Y., Wang, T., Qiu, J., Feng, S.-H., Tu, S., and Wei, D.-T. (2010). The electrophysiological effect of working memory load on involuntary attention in an auditory-visual distraction paradigm: an ERP study. *Exp. Brain Res.* 205, 81–86. doi: 10.1007/s00221-010-2360-x

Mayer, A. R., and Kosson, D. (2004). The effects of auditory and visual linguistic distractors on target localization. *Neuropsychology* 18, 248–257. doi: 10.1037/0894-4105.18.2.248

Oldfield, R. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9, 97–113. doi: 10.1016/0028-3932(71)90067-4

Papageorgiou, C., Liappas, I., Asvestas, P., Vasios, C., Matsopoulos, G. K., Nikolaou, C., et al. (2001). Abnormal P600 in heroin addicts with prolonged abstinence elicited during a working memory test. *Neuroreport* 12, 1773–1778. doi: 10.1097/00001756-200106130-00051

Parks, N. A., Hilimire, M. R., and Corballis, P. M. (2011). Steady-state signatures of visual perceptual load, multimodal distractor filtering and neural competition. *J. Cogn. Neurosci.* 23, 1113–1124. doi: 10.1162/jocn.2010.21460

Parmentier, F. B. (2008). Towards a cognitive model of distraction by auditory novelty: the role of involuntary attention capture and semantic processing. *Cognition* 109, 345–362. doi: 10.1016/j.cognition.2008.09.005

Parmentier, F. B. (2014). The cognitive determinants of behavioral distraction by deviant auditory stimuli: a review. *Psychol. Res.* 78, 321–338. doi: 10.1007/s00426-013-0534-4

Parmentier, F. B., and Andrés, P. (2010). The involuntary capture of attention by sound. *Exp. Psychol.* 57, 68–76. doi: 10.1027/1618-3169/a000009

Parmentier, F. B., and Hebrero, M. (2013). Cognitive control of involuntary distraction by deviant sounds. *J. Exp. Psychol. Learn. Mem. Cogn.* 39, 1635–1641. doi: 10.1037/a0032421

Parmentier, F. B., Turner, J., and Elsley, J. V. (2011). Distraction by auditory novelty. The course and aftermath of novelty and semantic effects. *Exp. Psychol.* 58, 92–101. doi: 10.1027/1618-3169/a000072

Peysakhovich, V., Causse, M., Scannella, S., and Dehais, F. (2015). Frequency analysis of a task-evoked pupillary response: luminance-independent measure of mental effort. *Int. J. Psychophysiol.* 97, 30–37. doi: 10.1016/j.ijpsycho.2015.04.019

Pickering, E. C., and Schweinberger, S. R. (2003). N200, N250r and N400 event-related brain potentials reveal three loci of repetition priming for familiar names. *J. Exp. Psychol. Learn. Mem. Cogn.* 29, 1298–1311. doi: 10.1037/0278-7393.29.6.1298

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clin. Neurophysiol.* 118, 2128–2148. doi: 10.1016/j.clinph.2007.04.019

Polich, J., and Comerchero, M. D. (2003). P3a from visual stimuli: typicality, task and topography. *Brain Topogr.* 15, 141–152. doi: 10.1023/A:1022637732495

Roelofs, A. (2005). The visual-auditory color-word Stroop asymmetry and its time course. *Mem. Cognit.* 33, 1325–1336. doi: 10.3758/bf03193365

SanMiguel, I., Corral, M.-J., and Escera, C. (2008). When loading working memory reduces distraction: behavioral and electrophysiological evidence from an auditory-visual distraction paradigm. *J. Cogn. Neurosci.* 20, 1131–1145. doi: 10.1162/jocn.2008.20078

Sassenhagen, J., and Bornkessel-Schlesewsky, I. (2015). The P600 as a correlate of ventral attention network reorientation. *Cortex* 66, A3–A20. doi: 10.1016/j.cortex.2014.12.019

Scannella, S., Causse, M., Chauveau, N., Pastor, J., and Dehais, F. (2013). Effects of the audiovisual conflict on auditory early processes. *Int. J. Psychophysiol.* 89, 115–122. doi: 10.1016/j.ijpsycho.2013.06.009

Scharinger, C., Soutschek, A., Schubert, T., and Gerjets, P. (2015). When flanker meets the n-back: what EEG and pupil dilation data reveal about the interplay between the two central-executive working memory functions inhibition and updating. [Research Support, Non-U.S. Gov't]. *Psychophysiology* 52, 1293–1304. doi: 10.1111/psyp.12500

Scheer, M., Bülthoff, H. H., and Chuang, L. L. (2016). Steering demands diminish the early-P3, late-P3 and RON components of the event-related potential of task-irrelevant environmental sounds. *Front. Hum. Neurosci.* 10:73. doi: 10.3389/fnhum.2016.00073

Schröger, E., Giard, M.-H., and Wolff, C. (2000). Auditory distraction: event-related potential and behavioral indices. *Clin. Neurophysiol.* 111, 1450–1460. doi: 10.1016/s1388-2457(00)00337-0

Schröger, E., and Wolff, C. (1998). Attentional orienting and reorienting is indicated by human event-related brain potentials. *Neuroreport* 9, 3355–3358. doi: 10.1097/00001756-199810260-00003

Siegle, G. J., Steinhauer, S. R., and Thase, M. E. (2004). Pupillary assessment and computational modeling of the Stroop task in depression. *Int. J. Psychophysiol.* 52, 63–76. doi: 10.1016/j.ijpsycho.2003.12.010

Simons, R. F., Graham, F. K., Miles, M. A., and Chen, X. (2001). On the relationship of P3a and the Novelty-P3. *Biol. Psychol.* 56, 207–218. doi: 10.1016/s0301-0511(01)00078-3

Siyanova-Chanturia, A., Pesciarelli, F., and Cacciari, C. (2012). The electrophysiological underpinnings of processing gender stereotypes in language. *PLoS One* 7:e48712. doi: 10.1371/journal.pone.0048712

Smallwood, J., Brown, K. S., Tipper, C., Giesbrecht, B., Franklin, M. S., Mrazek, M. D., et al. (2011). Pupillometric evidence for the decoupling of attention from perceptual input during offline thought. *PLoS One* 6:e18298. doi: 10.1371/journal.pone.0018298

Sörqvist, P., and Rönnberg, J. (2014). Individual differences in distractibility: an update and a model. *Psych J.* 3, 42–57. doi: 10.1002/pchj.47

Sörqvist, P., Stenfelt, S., and Rönnberg, J. (2012). Working memory capacity and visual-verbal cognitive load modulate auditory-sensory gating in the brainstem: toward a unified view of attention. *J. Cogn. Neurosci.* 24, 2147–2154. doi: 10.1162/jocn_a_00275

Spencer, K. M., and Polich, J. (1999). Poststimulus EEG spectral analysis and P300: attention, task and probability. *Psychophysiology* 36, 220–232. doi: 10.1111/1469-8986.3620220

Stuart, D. M., and Carrasco, M. (1993). Semantic component of a cross-modal Stroop-like task. *Am. J. Psychol.* 106, 383–405. doi: 10.2307/1423183

Taylor, J., O'Hara, R., Mumenthaler, M., Rosen, A., and Yesavage, J. (2005). Cognitive ability, expertise and age differences in following air-traffic control instructions. *Psychol. Aging* 20, 117–133. doi: 10.1037/0882-7974.20.1.117

Taylor, J., O'Hara, R., Mumenthaler, M., and Yesavage, J. (2000). Relationship of CogScreen-AE to flight simulator performance and pilot age. *Aviat. Space Environ. Med.* 71, 373–380.

Tellinghuisen, D. J., and Nowak, E. J. (2003). The inability to ignore auditory distractors as a function of visual task perceptual load. *Percept. Psychophys.* 65, 817–828. doi: 10.3758/bf03194817

van Gerven, P. W., Paas, F., Van Merriënboer, J. J., and Schmidt, H. G. (2004). Memory load and the cognitive pupillary response in aging. *Psychophysiology* 41, 167–174. doi: 10.1111/j.1469-8986.2003.00148.x

Verleger, R., Jaśkowski, P., and Wascher, E. (2005). Evidence for an integrative role of P3b in linking reaction to perception. *J. Psychophysiol.* 19, 165–181. doi: 10.1027/0269-8803.19.3.165

Watkins, S., Dalton, P., Lavie, N., and Rees, G. (2007). Brain mechanisms mediating auditory attentional capture in humans. *Cereb. Cortex* 17, 1694–1700. doi: 10.1093/cercor/bhl080

Watter, S., Geffen, G. M., and Geffen, L. B. (2001). The n-back as a dual-task: P300 morphology under divided attention. *Psychophysiology* 38, 998–1003. doi: 10.1111/1469-8986.3860998

Wetzel, N., Berti, S., Widmann, A., and Schröger, E. (2004). Distraction and reorientation in children: a behavioral and ERP study. *Neuroreport* 15, 1355–1358. doi: 10.1097/01.wnr.0000129858.40478.be

Wickens, C. D., Hollands, J. G., Banbury, S., and Parasuraman, R. (2015). *Engineering Psychology and Human Performance.* London and New York: Psychology Press.

Woodman, G. F., and Luck, S. J. (2004). Visual search is slowed when visuospatial working memory is occupied. *Psychon. Bull. Rev.* 11, 269–274. doi: 10.3758/bf03196569

Yuval-Greenberg, S., and Deouell, L. Y. (2009). The dog's meow: asymmetrical interaction in cross-modal object recognition. *Exp. Brain Res.* 193, 603–614. doi: 10.1007/s00221-008-1664-6

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Causse, Peysakhovich and Fabre. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution and reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 14 October 2015
doi: 10.3389/fnhum.2015.00526

The impact of expert visual guidance on trainee visual search strategy, visual attention and motor skills

*Daniel R. Leff[1†], David R. C. James[1†], Felipe Orihuela-Espina[1,2], Ka-Wai Kwok[1], Loi Wah Sun[1], George Mylonas[1], Thanos Athanasiou[1], Ara W. Darzi[1] and Guang-Zhong Yang[1]**

[1] Hamlyn Centre for Robotic Surgery, Imperial College London, London, UK, [2] National Institute for Astrophysics, Optics and Electronics (INAOE), Tonantzintla, Mexico

Minimally invasive and robotic surgery changes the capacity for surgical mentors to guide their trainees with the control customary to open surgery. This neuroergonomic study aims to assess a "Collaborative Gaze Channel" (CGC); which detects trainer gaze-behavior and displays the point of regard to the trainee. A randomized crossover study was conducted in which twenty subjects performed a simulated robotic surgical task necessitating collaboration either with verbal (control condition) or visual guidance with CGC (study condition). Trainee occipito-parietal (O-P) cortical function was assessed with optical topography (OT) and gaze-behavior was evaluated using video-oculography. Performance during gaze-assistance was significantly superior [biopsy number: (mean ± SD): control = 5.6 ± 1.8 vs. CGC = 6.6 ± 2.0; $p < 0.05$] and was associated with significantly lower O-P cortical activity [ΔHbO_2 mMol × cm [median (IQR)] control = 2.5 (12.0) vs. CGC 0.63 (11.2), $p < 0.001$]. A random effect model (REM) confirmed the association between guidance mode and O-P excitation. Network cost and global efficiency were not significantly influenced by guidance mode. A gaze channel enhances performance, modulates visual search, and alleviates the burden in brain centers subserving visual attention and does not induce changes in the trainee's O-P functional network observable with the current OT technique. The results imply that through visual guidance, attentional resources may be liberated, potentially improving the capability of trainees to attend to other safety critical events during the procedure.

Keywords: functional near infrared spectroscopy, optical topography, neuroergonomics, graph theory, collaborative gaze, visual attention, skills assessment, mentoring

OPEN ACCESS

Edited by:
Klaus Gramann, Berlin Institute of Technology, Germany

Reviewed by:
Peter König, University of Osnabrück, Germany
Frederic Dehais, Institut Supérieur de l'Aéronautique et de l'Espace, France

***Correspondence:**
Guang-Zhong Yang, Hamlyn Centre for Robotic Surgery, Imperial College London, Level 4, Bessemer Building, South Kensington Campus, London, SW7 2AZ, UK
g.z.yang@imperial.ac.uk

† These authors have contributed equally to this work.

Received: *28 July 2015*
Accepted: *10 September 2015*
Published: *14 October 2015*

Citation:
Leff DR, James DRC, Orihuela-Espina F, Kwok K-W, Sun LW, Mylonas G, Athanasiou T, Darzi AW and Yang G-Z (2015) The impact of expert visual guidance on trainee visual search strategy, visual attention and motor skills. Front. Hum. Neurosci. 9:526. doi: 10.3389/fnhum.2015.00526

Highlights

1. A randomized crossover study assessing the impact of trainer visual guidance upon trainee visual cognition, occipito-parietal (O-P) brain function and technical performance.
2. Visual guidance is associated with enhanced gaze behavior, improved technical accuracy and attenuated activity across O-P cortices.
3. Parameters of network performance such as cost and global efficiency are not detrimentally effected by visual guidance.

Introduction

In high-risk industry, collaboration between operators is integral to performing goal-orientated tasks successfully (e.g., pilots, air-traffic controller, surgeons, etc). Regarding surgery, collaboration is necessary between surgeons and their assistant(s), theatre nurse(s) and occasionally members of allied specialties. Recent developments in technologies for robotic surgery such as dual console systems (e.g., da Vinci® Si) enable two surgeons to operate simultaneously, facilitating both high-level co-operation and mentorship as well as potentially streamlining the operators' cognitive resources towards improved safety. However, in this scenario, it is important that communication between both surgeons is effective to enable a seamless flow of information between the two operators and ensure an efficient workflow. Similarly, excellent communication facilitates technical skills training in surgery. During "open" surgery, expert trainers' employ a variety of methods for communication with trainees that include a combination of verbal instruction, physical pointing or actual demonstration(s). However, during robotic minimally invasive surgery (MIS), there may be circumstances in which the trainee or collaborating surgeon is using both instruments simultaneously within the operative field of view, constraining the trainer/master surgeon and rendering them reliant solely on verbal communication.

Within MIS and robotic surgery, techniques exist such as telestration that aid information transfer between surgeons and/or between trainer and trainee. Telestration allows information to be "*drawn*" onto a monitor at a remote site by the surgeon guiding the procedure. This information is then displayed on the operator's screen with the aim of guiding performance and may be undertaken either remotely or locally (Ferguson and Stack, 2010). Remote guidance or telementoring enables surgeons to be guided by a mentor at a location remote from the operation. This form of instruction has been applied to better enable regional experts to guide surgeons at local centers and to provide assistance and mentoring from surgical experts in other countries (Micali et al., 2000; Schlachta et al., 2010).

There has been interest in the role that gaze behavior may have in improving the flow of communication between collaborating subjects. For example, it has been demonstrated that shared gaze during visual collaboration enables a more efficient search strategy when compared to verbal collaboration alone (Brennan et al., 2008). Therefore, it is anticipated that observing a guiding surgeon's point of regard instead of, or in conjunction with their verbal instruction(s) will significantly improve the performance of the operating surgeon by providing supplementary cues critical to task success. Based on this concept, a new system referred to as "*collaborative gaze control*" (CGC) was developed to enable an operating surgeon to be directed by visual guidance as opposed to or in conjunction with verbal instruction(s) from an expert (Kwok et al., 2012). With CGC enabled, the trainer's gaze behavior is extracted in real-time. Their point of regard is subsequently relayed to the trainee's screen, which may be in a remote location. Therefore, the trainee's operative manoeuvres can be directed more precisely, potentially obviating the dependence on verbal instruction(s). Importantly, in manipulating target salience, visual search is modulated leading to enhanced behavioral performance (Avraham et al., 2008).

More recently, there is evidence that workload can be inferred from saccadic eye movements (Tokuda et al., 2011), pupillary responses (Zheng et al., 2015) and blink frequency (Zheng et al., 2012). Challenging, effortful visual search results in greater visual cortical (V1) excitation (Kojima and Suzuki, 2010). Evaluating the impact that technological manipulation of visual search has on an operator's cortical function helps to determine whether performance enhancement is offset by the need for greater attentional demands at brain level. This is encompassed by "*neuroergonomics*" which concerns the investigation of the brain behavior at work (Parasuraman, 2003), a paradigm that has been applied to surgery in order to investigate how recruited brain regions may be modulated by novel performance-enhancing tools (James et al., 2010b, 2013).

In order to examine this effect, functional Near Infrared Spectroscopy (fNIRS) a non-invasive neuroimaging modality is utilized to measure task-evoked fluctuations in oxygenated and deoxygenated hemoglobin (HbO_2 and HHb respectively) within cortical tissues that reflects the magnitude of cortical activation (Jöbsis, 1977). This is based upon the principle that neuronal activity and the associated increased metabolic demand within the brain leads to local hemodynamic changes, so termed *"neurovascular coupling"* (Roy and Sherrington, 1890). Unlike functional magnetic resonance imaging (fMRI), fNIRS is relatively resistant to motion artifact and can be used in conjunction with ferromagnetic instruments and has been successfully applied to monitor the cortical responses in surgeons (Leff et al., 2008a,b,c; Ohuchida et al., 2009; James et al., 2011, 2013). Broadly, these studies highlight the importance of the prefrontal cortex (PFC) in supporting "cognitive phases" of skill learning (Leff et al., 2008a), evolution in PFC excitation with technical skills training (Leff et al., 2008c), and relative PFC redundancy amongst expert surgeons (Ohuchida et al., 2009). More recently, investigators have demonstrated the impact of the type of learning (e.g., implicit vs. explicit) and the influence of technology to stabilize performance and enhance neuronal efficiency amongst surgeons (Zhu et al., 2011; James et al., 2013).

Functional brain connectivity captured in coherence or cross-correlation between different brain regions can be used to investigate efficiency in brain networks (Zhu et al., 2011; James et al., 2013). Graph Theory, a popular method for interrogating brain networks, can model the organization, development and function of complex networks (Sporns et al., 2004; Bullmore and Sporns, 2009; Sporns, 2011) and has been successfully employed to networks derived from fNIRS data (Niu et al., 2012; James et al., 2013). In this regard, studies investigating graph topology such as the number of connections, cost and efficiency have demonstrated associations between task performance and brain network efficiency or cost-efficiency (Bassett et al., 2009). Despite the above, there have been no studies investigating the influence of varying trainer/mentor guidance on brain function or network architectures amongst trainees.

The aim of this paper is to investigate the influence of a gaze channel on changes in visual search strategies, technical performance, and brain behavior in a group of task naïve subjects being instructed to perform simulated biopsy using robotic MIS. Therefore, it is anticipated that compared to verbal guidance technical procedural skills may be superior during gaze-assistance owing to the improved perceptual flow of information to the trainee. The primary hypothesis is that increased target saliency will lead to a *"bottom-up"* search strategy, reflected in a more focused pattern of V1 activation and a reduction in the need for recruitment of extra-striatal visual association areas. Conversely, verbal communication (gold standard) is anticipated to lead to a more effortful *"top down"* visual search strategy, necessitating recruitment of additional cortical regions outside V1, manifest as greater excitation in centers of visual attention. The secondary hypothesis is that collaborative gaze may facilitate the flow of information transfer in the visual-parietal network manifest as reduced network costs, improved efficiency and reduced network burden.

Materials and Methods

Subjects

The study was carried out in accordance with the recommendations of the Local Regional Research Ethics Committee (LREC 05/Q0403/142) with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. Following ethical approval a randomized control trial was conducted in which 20 subjects (1 female) were recruited from Imperial College London (mean age, years $\pm$ SD = 28.9 $\pm$ 1.5). Left-handed subjects and those with a history of neuropsychiatric illness or previous exposure to the task were excluded (Orihuela-Espina et al., 2010). Subjects were included on the basis that they were task naïve. The task was performed under both guidance conditions (order randomized) such that subjects served as their own controls and bias associated with learning or ordering effects was minimized.

Task Paradigm

The robotic surgical task entailed the subject (*"trainee"*) and an expert (*"trainer"*) collaborating in taking virtual biopsies from a simulated gastric mucosa in a shared surgical environment as depicted in **Figure 1**. Haptic manipulators (Phantom, Omni, SensAble Technologies, USA) were used to control robotic graspers in the virtual scene. The task necessitated the trainee take a virtual biopsy and pass the specimen to the guiding trainer. Both the trainee's and the trainer's graspers were visible within the same field of view with the former located inferiorly and the latter superiorly as depicted in **Figure 1** (panels i–iv). Within the operative field, seven nodules were visible to the trainee. The choice of nodule for biopsy was randomly determined and this selection was available only to the trainer. Therefore, the appropriate biopsy site had to be conveyed to the trainee either visually or verbally by the trainer. Once the biopsy was taken by the trainee, the specimen was passed towards the trainer's graspers and when successfully transferred to the trainer, it disappeared from the field of view. This process was repeated as many times as possible during the allotted task periods.

Prior to commencing the study, all subjects received a standardized period of task familiarization. All subjects performed the simulated biopsy task under verbal (control) and visual instruction (CGC; Kwok et al., 2012). The order was randomized (random number generator) in order to control for learning effects. Regarding the control task, the location of the biopsy site was described by the trainer using verbal instructions. With CGC enabled, a portable eyetracker ($\times$50 eyetracker Tobii Technologies, Sweden) situated beneath the trainer's monitor detected their fixation point and conveyed this to the trainee's screen as a cross. Therefore, with CGC enabled, the trainer's target selection would be conveyed to the trainee. For each condition (verbal and CGC) a block design experiment was employed comprising a baseline rest period (30 s) followed by five task blocks each of which comprised alternating episodes of simulated nodule biopsy (30 s) and inter-trial rest periods (30 s). During rest periods, subjects were asked to remain still with their eyes open regarding a black screen on the task monitor. Within functional neuroimaging experiments, block design paradigms have the advantage of allowing the hemodynamic response to return to baseline between each session, therefore providing reliable indices of task-evoked cortical activity. Furthermore, the block design allows task data to be averaged, increasing the signal to noise ratio.

Cortical Activity

Brain activation was assessed using a commercially available 24-channel Optical topography (OT) system (ETG-4000, Hitachi Medical Corp., Japan). Sixteen optodes (8 emitters and 8 detectors) were positioned in a 4 $\times$ 4 array over the O-P cortices as displayed in **Figure 1**. A "channel" represents a banana-shaped volume of cortex where changes in absorption of near infrared light from the optode emitters are interpreted as changes in HbO_2 and HHb. The array was centered on "Oz" of the International 10–20 system (Jurcak et al., 2007) with the intention of capturing activation within the visual cortex. Cortical data was subject to both manual and automated data integrity checks (Orihuela-Espina et al., 2010) to identify and eliminate data contaminated with noise, optode movement and saturation-related artifacts (i.e., apparent non-recordings and "mirroring"). Since both ambient light and near infra-red light from eye-tracking systems have the potential to influence OT data (Orihuela-Espina et al., 2010), laboratory lights were dimmed and the probes were shielded using a combination of external fixation tapes and shower cap.

Technical Performance

The number of nodules that the trainee was able to successfully biopsy and transfer to the trainer's graspers across the task period and the trainee's instrument pathlength (metres) were recorded and used as objective metrics of technical performance. This was preferred to restricting the overall number of moves towards calculating time/nodule biopsied, and helped to ensure that subjects were focusing on the task quality and not the procedural time, or perceiving the number of movements.

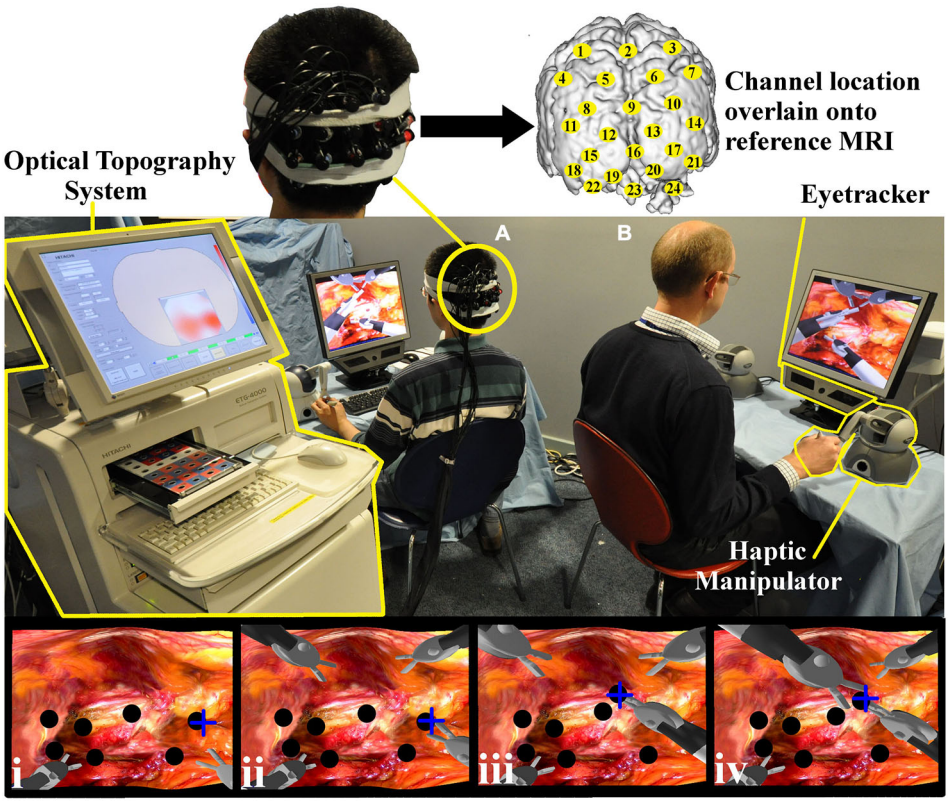


FIGURE 1 | Experimental task set up. Both the trainee **(A)** and trainer **(B)** control the virtual instruments, each with two haptic manipulators (Phantom Omni, SensAble Tech, USA). The trainer's right hand manipulator is highlighted (yellow). Gaze behavior is detected with portable eyetracker (X50 eyetracker, Tobii Technologies, Sweden) situated below both monitors (trainer eyetracker highlighted yellow). An Optical topography (OT) system (ETG-4000, Hitachi Medical Corp. Japan) positioned outside the trainee's field of view (left, highlighted) records cortical hemodynamic data from 24 cortical loci (channels). Appropriate channel locations (yellow circles) are understood by projecting 3D positional data onto a T1 weighted MRI image (upper subplot). The lowermost row of channels was centered on Oz of the International 10–10 system (Jurcak et al., 2007). Task images can be appreciated on trainer and trainee monitors and sample screen shots are represented in which the trainee's instruments are located inferiorly (i–iv). With the collaborative gaze channel (CGC) enabled, the trainee regards the blue cross indicating the intended biopsy target (i). The trainee then grasps the nodule (black circle) (ii) and passes it to the trainer's instrument (iii–iv). With the channel disabled, the trainee performs identical maneouvres but only with verbal instructions from the trainer.

Gaze Behavior

Subject and trainer gaze behavior was recorded throughout the study with portable eyetracking technology (×50 eyetracker, Tobii Technologies, Sweden) situated beneath the task monitor (as displayed in **Figure 1**). The gaze behavior of the trainer was interrogated to derive their fixation point in order to display this as a cross on the trainee's monitor thereby facilitating gaze-guidance in CGC (study condition). The trainer's fixation point was not visible to the trainee during episodes of verbal guidance (control condition). The trainee's fixation points were recorded to determine the time taken, termed "gaze latency" (GL, seconds), to fixate on the same area of the surgical scene as the expert.

Heart Rate Monitoring

A portable band electrocardiogram (Bioharness v2.3.0.5; Zephyr Technology Limited, USA) was used to acquire continuous heart rate data, from which heart rate variability (HRV) was derived and used to infer subject stress (Task Force of the European Society of Cardiology the North American Society of Pacing Electrophysiology, 1996).

Data Analysis

Cortical Hemodynamics

Cortical hemodynamic data and network graph econometrics were observed to be non-Gaussian and therefore analyzed using non-parametric tests of significance. Channel-wise cortical activation was determined as a task-evoked statistically significant increase in HbO_2 coupled to a significant decrease in HHb from baseline rest (Wilcoxon Rank Sign, $p < 0.05$). For each channel of data and hemoglobin species a variable ΔHb was computed (Hb task–Hb rest). To investigate the influence of the mode of guidance (CGC vs. control) and stress on cortical hemodynamics (i.e., ΔHbO_2 and ΔHHb) random effects models (REM) were generated (Intercooled Stata, v10.0 for windows, Stata Corporation, USA).

Cortical hemodynamic data was subsequently used to construct a task-evoked network of the 24 channels using graph theory (Bullmore and Sporns, 2009). A 24 × 24 bidimensional cross-correlation matrix was constructed by cross-correlating data between all channels, as previously described (James et al., 2013). This matrix represents the strength of functional associations within the network of 24 channels. Comparisons between graphs of different functional networks are potentially sensitive to the method used for thresholding, for which an optimal solution does not yet exist (van Wijk et al., 2010). Therefore, to evaluate the active network, the matrix was pruned to eliminate "inactive" graph nodes. This approach renders a network for each subject during each task condition.

Econometric data from these networks was then calculated to derive: (a) the number of network connections; (b) the maximum global efficiency (Achard and Bullmore, 2007); (c) the normalized cost (Achard and Bullmore, 2007); and (d) the task-induced "*network burden*" (James et al., 2010a). Network economy is defined as efficiency minus cost (Achard and Bullmore, 2007). The network burden is defined here as—economy which equates to "cost-efficiency". If a network is economical the cost-efficiency is high and accordingly the network burden is low. Network measures were also compared between the study and control groups using REM analysis to determine whether the mode of guidance (CGC vs. control) significantly influenced network econometrics. Statistical significance was set at $p = 0.05$.

Performance and Gaze Behavior

The number of nodules biopsied by each subject during the allotted task time and the instrument pathlength (metres) were determined. GL (seconds) was derived from the eye-tracking data stream. Behavioral performance and GL data was observed to be Gaussian and therefore analyzed using paired *t*-tests. These data were subsequently incorporated into the REM analysis in order to assess whether the guidance mode (control vs. CGC) was a predictor of performance accuracy and efficiency in visual search.

Heart Rate Analysis

HRV as calculated by the standard deviation of the R to R interval (SD_{RR}) was derived from the HR data stream (Task Force of the European Society of Cardiology the North American Society of Pacing Electrophysiology, 1996). The SD_{RR} decreases under stress and was incorporated into the REM analysis, to exclude any potential confounding effect that differences in HRV or changes in mean HR may exert on changes in cortical hemodynamics. Furthermore, HRV was utilized to determine which mode of guidance (verbal vs. CGC) trainee's found the most stressful by undertaking a univariate random effects analysis ($p = 0.05$).

Results

Technical Performance

Biopsy number and instrument pathlength was analyzed to determine whether CGC improved trainees' technical performance. As illustrated in **Figure 2A**, gaze-guidance

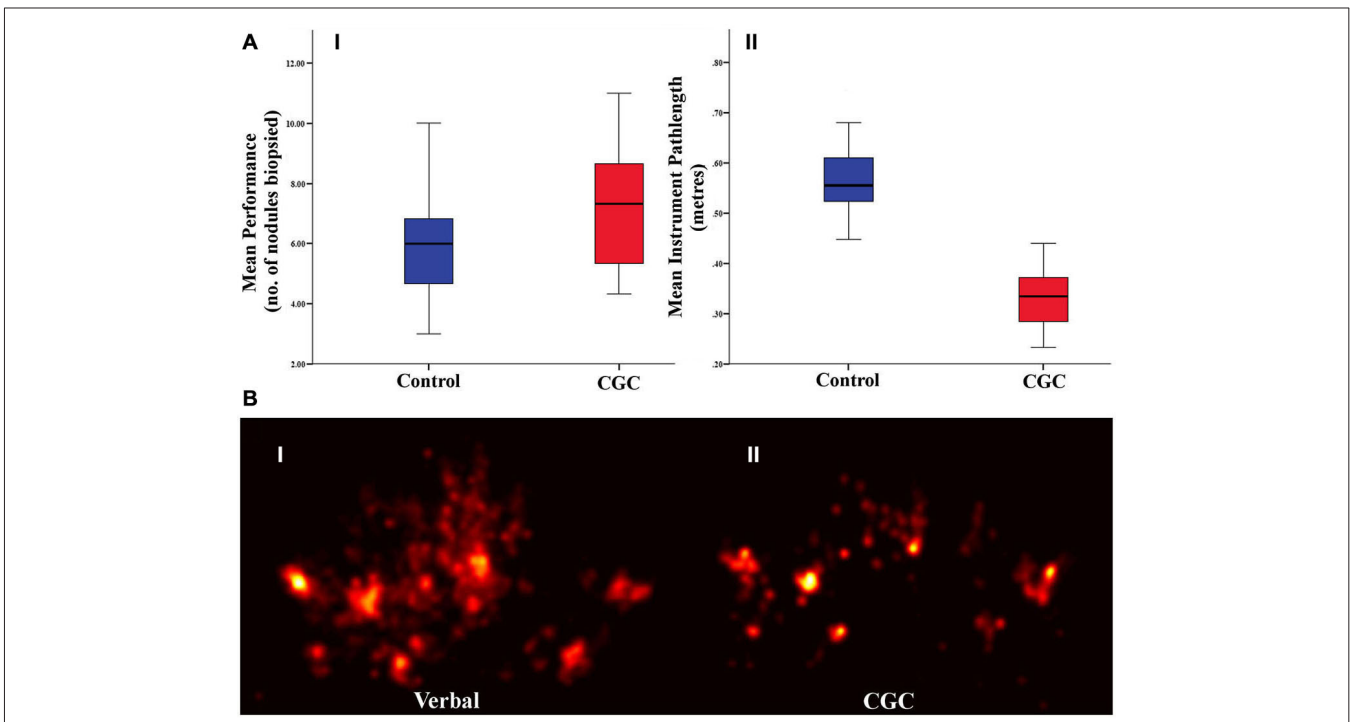


FIGURE 2 | (A) Technical performance as indexed by the number of biopsies retrieved (I) and instrument path length (II). Box plots indicate mean and error bars represent 95% confidence interval. **(B)** Gaze plots from a representative subject under control (I) and gaze guidance (II) demonstrate more focussed fixations during gaze-assistance.

TABLE 1 | The influence of guidance mode on technical performance, visual search behavior, changes in cortical hemodynamics, network topological properties and systemic effects.

Outcome variable	Control condition (Mean ± SD)	CGC condition (Mean ± SD)	*t*-value	*p*-value
Technical lerformance				
Biopsy number	5.6 ± 1.8	6.6 ± 2.0	−3.394	**0.003**
Instrument path length (m)	0.6 ± 0.1	0.3 ± 0.7	11.765	***0.000***
Visual search				
Gaze latency (s)	1.4 ± 0.3	0.8 ± 0.2	7.292	***0.000***
Outcome variable	**Control condition median (IQR)**	**CGC condition median (IQR)**	***z*-value**	***p*-value**
Cortical hemodynamics				
ΔHbO_2 (mMol × cm)	2.5 (12.0)	0.6 (11.2)	−4.049	***0.000***
ΔHHb (mMol × cm)	−1.4 (5.0)	−1.0 (4.5)	−1.098	0.272
ΔHbT (mMol × cm)	3.6 (13)	1.1 (11.6)	−6.064	***0.000***
Cortical network				
Normalized cost (a.u.)	0.10 (0.13)	0.19 (0.43)	−0.722	0.470
Global efficiency (a.u.)	0.03 (0.05)	0.02 (0.08)	−0.220	0.826
Network burden (a.u.)	0.09 (0.14)	0.18 (0.46)	−0.847	0.397
Network edges (a.u.)	56.0 (304.0)	81.0 (120.0)	−0.589	0.556
Systemic effect				
Heart rate ($beatsmin^{-1}$)	71.2 (10.0)	73.4 (8.1)	−0.392	0.695
SD_{NN}	57.7 (42.0)	47.2 (36.9)	−0.784	0.433

$p < 0.05$ = bold, $p < 0.001$ = bold italic.

under the influence of CGC resulted in enhanced technical performance. **Table 1** highlights the differences in technical performance according to the mode of guidance. With gaze-assistance, trainees' biopsied a significantly greater number of nodules [biopsy number (mean ± SD): control = 5.6 ± 1.8 vs. CGC = 6.6 ± 2.0, $p < 0.05$] using significantly shorter instrument pathlength (metres) [mean ± SD: control = 0.6 ± 0.1 vs. CGC = 0.3 ± 0.7, $p < 0.001$]. This implies that trainees were faster, more productive and used virtual instruments more economically when operating from the CGC mode.

Gaze Behavior

GL which represents the temporal delay between trainer and trainee gaze fixation was analyzed to determine whether gaze guidance streamlined trainee visual search. **Figure 2B** depicts the visual search pattern acquired from a representative trainee under both guidance conditions. It is apparent that whilst operating under gaze guidance, trainee fixations appear to be more localized to the nodule to be biopsied. GL was significantly shorter in CGC mode [GL seconds (mean ± SD): control = 1.4 ± 0.3 vs. CGC = 0.8 ± 0.2, $p < 0.001$]. This suggests that gaze assistance manifests as more rapid fixation on the appropriate target nodule to be biopsied.

Cortical Activation

Cortical hemodynamic change was analyzed to compare trainee brain responses between verbal and gaze-assisted modes of operation, with the hypothesis that verbal guidance would induce higher amplitude and spatially broader O-P hemodynamic changes. Topograms of a representative subject depicting the

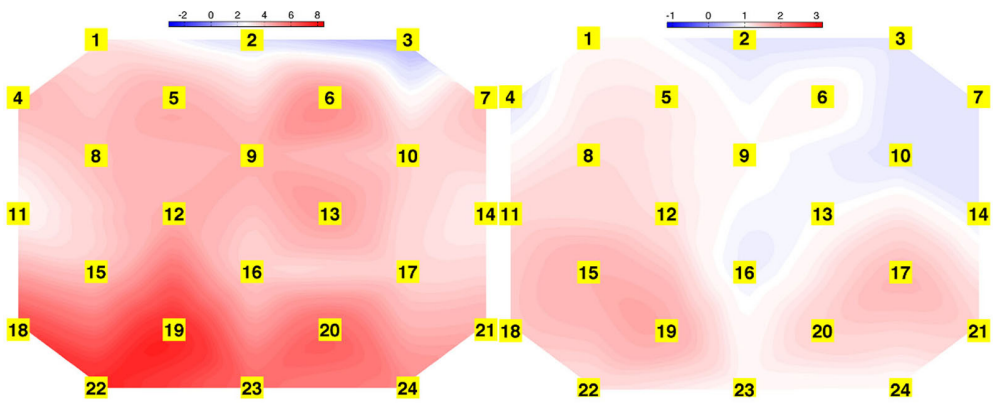


FIGURE 3 | Topograms derived from task averaged HbO_2 response of a representative subject for verbal (left) and gaze guidance (right) conditions, depicting spatially broader task-evoked oxygenated hemoglobin change during verbal guidance.

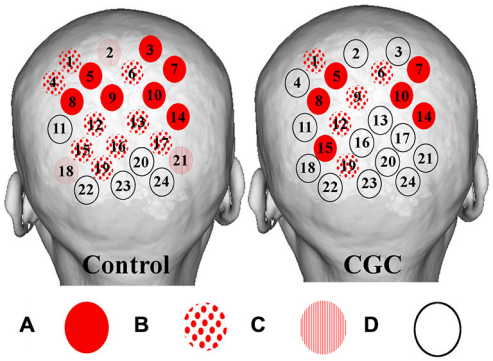


FIGURE 4 | Figure depicting group averaged (O-P) channel activation for verbal (left) and gaze guidance (right). Magnitude of statistical changes in cortical hemodynamics reflect intensity of brain activation as follows: **(A)** statistically significant ($p < 0.05$) increase in HbO_2 coupled to statistically significant ($p < 0.05$) decrease in HHb (red circles); **(B)** increase HbO_2 and decrease HHb with one species reaching statistical significance, $p < 0.05$ (spots); **(C)** increase HbO_2 and decrease HHb with neither species reaching statistical significance (stripes); and **(D)** no coupled increase HbO_2 and decrease HHb (clear circles). Verbal guidance resulted in a greater number of activating channels (control vs. CGC = 19/24 vs. 11/24).

average change in HbO_2 overlying the O-P cortices are displayed in **Figure 3**. **Table 1**, depicts cortical hemodynamic change as ΔHbO_2 (mMol × cm) averaged across the O-P cortices for both verbal and gaze-guidance. Cortical hemodynamic change evoked by verbal guidance was more diffuse as illustrated in **Figure 4** (CGC: 11/24 channels active vs. verbal: 19/24 channels active), more likely to involve bilateral parietal as well as bilateral visual cortices and was greater in magnitude than the response evoked by gaze guidance (ΔHbO_2 mMol × cm [median (IQR)]: control = 2.5 (12.0) vs. CGC = 0.63 (11.2), $p < 0.001$; ΔHbT mMol × cm [median (IQR)]: control = 3.6. (13.0) vs. CGC = 1.1 (11.6), $p < 0.001$). Overall, this data supports the primary hypothesis that training in CGC mode evokes an attenuated O-P brain response. The mode of guidance did not significantly influence the magnitude of ΔHHb [ΔHHb mMol × cm [median (IQR)]: control = −1.4 (5.0) vs. CGC = −1.0 (4.5), $p = 0.27$]. Similarly, as highlighted in **Table 2**, REM analysis revealed that guidance mode was a predictor of ΔHbO_2 ($p < 0.001$) but not of ΔHHb ($p = 0.19$).

Cortical Networks

Graph theoretical econometric data were computed and compared between guidance modes with the hypothesis that the performance of functional network in CGC mode would be associated with less cost and greater efficiency. **Figure 5** depicts the activated cortical network under control and CGC conditions for a representative subject. **Table 1** represent results of econometric analysis delineating the number of cortical connections, normalized cost, maximum global efficiency and cognitive burden. Differences in these network topological properties between modes guidance did not reach statistical threshold. Additionally, even when subject-level clustering was considered (**Table 2**) guidance mode was not found to predict network properties (e.g., cost, efficiency, etc). This suggests that CGC does not induce changes in the trainee's O-P functional network observable with the current OT technique.

Heart Rate Data

HR and SD_{RR} were monitored to determine the influence of guidance mode on stress-related change in systemic responses (**Table 1**). Between-condition differences in HR and SD_{RR} were not statistically significant [Median HR (IQR): control = 71.2 (10.0) vs. CGC = 73.4 (8.1) $p = 0.70$; Median SD_{RR} (IQR): control = 57.7 (42.0) vs. CGC = 47.2 (36.9), $p = 0.43$). Additionally, upon REM analysis, neither HR nor SD_{RR} were observed to be predictors for changes in cortical hemodynamics.

Harms

No harms occurred in the study.

Discussion

In this study, performance on a simulated surgical task has been improved by modulating the manner in which collaborating surgeons interact with one another. Communicating through collaborative gaze-driven control leads to a greater number of successful biopsies and a reduction in instrument path length, the latter being a measure of dexterity previously shown to reflect skill level in laparoscopic and open surgery (Bann et al., 2003; Xeroulis et al., 2009). The foundation for this improvement appears to be a change in visual search strategy manifest as a reduced GL indicating that with gaze-assistance, trainee fixation points more rapidly reach those of the expert. This was accompanied by an amelioration of cortical excitation across primary visual centers in the brain, but without an appreciable difference in O-P network costs or burden.

The current paper offers a potential mechanistic explanation for improvements observed in novices' performance when training under the influence of expert visual cues (Wilson et al., 2011; Chetwood et al., 2012). Experienced operators are known to utilize more effective gaze-strategies than novices, characterized by fixating on relevant target locations and adopting optimal psychomotor control (Wilson et al., 2011). Unlike novices who learn mapping rules by switching their point of regard between tool and target, experts utilize a target locking strategy and rarely need to check tool locations (Leong et al., 2008). As demonstrated by Wilson et al. (2011), novices trained to observe and then "mimic" the more focused gaze patterns of experts improve their laparoscopic performance and multi-tasking capabilities more than novices trained to observe expert performance without the benefit of expert gaze-cues. Similarly, Chetwood et al. (2012) observed improved completion times and reduced errors in novices guided by expert gaze vs. expert verbal instructions. However, unlike the current experiment, the aforementioned studies were not designed to explain the foundation for improved performance owing to gaze guidance, resulting instead in speculation regarding adaptation in visual

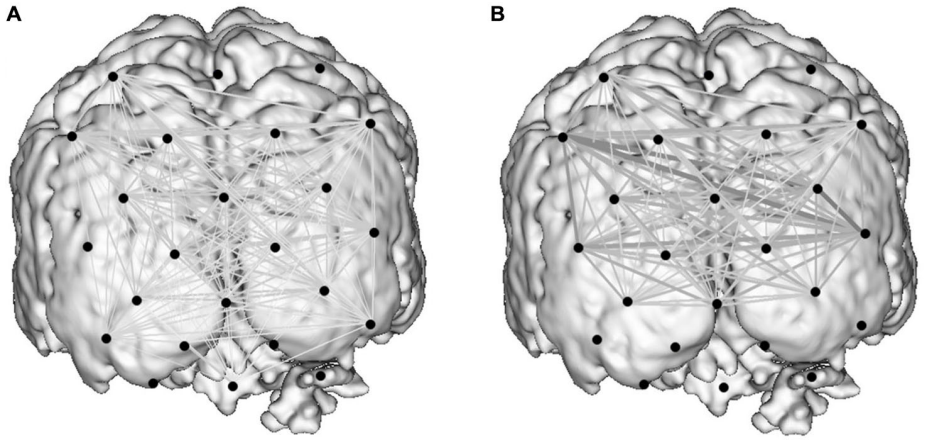


FIGURE 5 | Activity-guided cortical networks for a representative subject during the control condition (A) and study condition (B). Approximate channel locations (black circles) are overlain onto reference MRI atlas. The strength of functional associations between nodes in the network is represented by the boldness of network edges.

cognitive function. Here, improved performance as a result of expert gaze guidance is understood as a reduction in visual activation and hence attentional demand on the visual cortex. This is in line with studies demonstrating learning related plasticity in activation maps implying attenuation of attentional resources associated with training and expertise (Dayan and Cohen, 2011). By manipulating the visual behavior of novices in a way that they align more closely with those of experts it is conceivable that novices may bypass the early "cognitive" phases of visual-motor learning (Fitts and Posner, 1967). This notwithstanding confirming that the gaze behavior of trainees operating under gaze guidance was characterized by less random saccadic activity and was indeed more "expert" cannot be confirmed using GL alone and would necessitate a more elaborate analysis of eye-tracking data such as using exploit/explore ratio (Dehais et al., 2015) or visual entropy (Di Nocera et al., 2007).

There is evidence from functional neuroimaging studies that streamlined visual search strategies lead to reduced activation in the visual cortex (Kojima and Suzuki, 2010). For example, Kojima and Suzuki (2010) observed greater hemodynamic responses in fNIRS channels centred on the visual cortex during more effortful search strategies. However, it must be acknowledged that the introduction of a target feature into the surgical scene might be anticipated to increase visual attention owing to changes in visual saliency. This is relevant since the eye-tracking derived fixation point of the expert was projected to trainee as a visually salient target. Interestingly, shifts in visual attention secondary to manipulations in visual saliency as a result of gaze-guidance (i.e., the trainer's fixation point) did not manifest as greater activation in the visual cortex when compared to verbal instruction. Rather, the resultant visual search is potentially streamlined from a *"top-down"* to *"bottom-up"* strategy (van der Stigchel et al., 2009; Theeuwes, 2010). Specifically, if a target markedly differs from its background, it

TABLE 2 | Results of univariate random effect models (REM), evaluating the influence of the independent variable (mode of guidance) on dependent variables including performance, changes in cortical hemodynamics, cortical network metrics, heart rate (HR) and heart rate variability (HRV).

Dependent variable	Coefficient	*S.E.*	$p > z$	95% *C.I.*
Biopsy number	0.090	0.040	**0.025**	0.011 to −0.168
Instrument pathlength (m)	−3.20	0.312	***0.000***	−3.808 to −2.586
Gaze latency (s)	−0.761	0.118	***0.000***	−0.992 to −0.530
ΔHbO_2 (mMol × cm)	−1.294	0.326	***0.000***	−1.933 to −0.654
ΔHHb (mMol × cm)	−0.198	0.151	0.188	−0.494 to −0.097
ΔHbT (mMol × cm)	−1.094	0.303	***0.000***	−1.689 to −0.500
No. of connections	−0.000	0.000	0.754	−0.002 to −0.001
Normalized cost (a.u.)	−0.036	0.101	0.720	−0.234 to −0.161
Network burden (a.u.)	−0.333	0.097	0.732	−0.244 to −0.157
Global efficiency (a.u.)	−0.106	0.282	0.706	−0.659 to −0.446
Mean HR ($beatsmin^{-1}$)	0.008	0.009	0.353	−0.009 to −0.025
SD_{NN}	−0.002	0.003	0.595	−0.008 to −0.004

($p < 0.05$ = bold, $p < 0.001$ = bold italic).

is visually salient and is more likely to be detected by a *"bottom-up"* search strategy guided by the saliency of the scene, whereas if a target requires greater cognitive input to be identified, a *"top down"* search ensues which is dependent on the PFC and parietal cortex (PC; van der Stigchel et al., 2009; Theeuwes, 2010). Bottom up saliency is not coded in the primary visual cortex (Betz et al., 2013), and this mode results in search simplification leading to a reduction in activity in visual association areas (Kojima and Suzuki, 2010). Enhanced saliency through visual guidance may parallel visual processing of natural stimuli (Einhäuser and König, 2010), whereby responses in V1 cells are optimally sparse (Vinje and Gallant, 2000). In the current study, this effect has been observed as a reduction in O-P cortical hemodynamic changes with comparatively fewer channels reaching statistical threshold for activation.

Parietal cortical activity is also associated with oculomotor intention and attention and may be important in planning eye movements (Kanwisher and Wojciulik, 2000). Verbal guidance may result in demanding visual search since it necessitates that auditory information be explicitly processed and translated into visual-spatial co-ordinates to understand the desired target's location, and parietal lobe activation has been shown to be important in spatial integration (Molholm et al., 2006). Conversely, gaze-guidance protocols may share many similarities with implicit learning protocols (Wilson et al., 2011). Implicit learning, a form of unconscious, incidental and procedural knowledge demands fewer attentional resources than explicit learning, a form of conscious, intentional or declarative knowledge. Implicit motor learning has been shown to reduce non-essential co-activation or connectivity between verbal-analytic and motor planning regions during laparoscopic performance (Zhu et al., 2011).

Here, as well as investigating connectivity (i.e., correlations), network topology has been explored with graph theory, which provides a powerful method for quantitatively describing the topology of brain connectivity (He and Evans, 2010). Graph theory has been utilized to interrogate cortical networks in both pathological and non-pathological brains (Achard and Bullmore, 2007; Bassett et al., 2009), and allows network parameters such as cost and efficiency to be determined (Bullmore and Bassett, 2011). Presently, graph theory was applied to experimental data in order to further appreciate the impact of a "gaze-channel" on functional brain networks. From the active network analysis (i.e., that which retains only activated nodes), it is evident that compared to verbal-guidance, gaze-assistance does not lead to significant differences in O-P network topologies, therefore disproving the secondary hypothesis. Therefore, our conclusion is that collaborative gaze exerts a positive effect on technical skills, alleviates burden on the visual cortices, and yet critically does not significantly alter performance of the functional O-P network.

Intuitively verbal instructions about target location are time consuming to deliver, more complex to interpret and harder to translate into the "visual" workspace, ultimately relying therefore on greater cognitive work as evidenced by enhanced task performance when visual guidance is employed (Chetwood et al., 2012). We suspect that gaze assistance makes the flow of information between the trainer and trainee more seamless by increasing the perceptual fidelity of the instruction given. Extrapolating this effect to the *in vivo* setting, a reduction in the attentional demands necessary to execute a procedure may manifest as a liberation of resources to devote to other safety critical aspects of clinical care (e.g., reacting to unexpected events, multitask decision making, planning operative steps, *etc.*). Future studies may capitalize on a framework that enables combined analysis of brain responses, visual behavior and HRV to improve the detection of changes in workload as has been demonstrated in pilots (Duratin et al., 2014). Furthermore, although not specifically investigated within the confines of this study, it is feasible that in using visual guidance the need to verbalize the intended target is bypassed and as such the trainer can focus on supplementary aspects of the procedure. For example, if the site of suture placement is already determined and displayed visually, a trainer can then focus verbal instruction on the technical aspects of suturing manoeuvres required to achieve accurate tissue apposition.

Conclusion

To summarize, this study demonstrates that capitalizing on visual behavior enhances communication between collaborating surgeons, and improves operator performance. This may be achieved through a *bottom up* allocation of resources within the visual cortex of the surgeon being instructed. It is plausible that trainees instructed in this fashion will be better able to devote neural resources to other safety critical aspects of the procedure. In investigating these hypotheses, fNIRS technology is well placed to make an impact, as it overcomes the limitations of traditional scanning environments (Cutini et al., 2012). However, future validation of graph theory measures for fNIRS connectivity analysis will necessitate comparison against models of anticipated responses and structural connectivity as have been observed using other neuroimaging technologies such as fMRI (van den Heuvel et al., 2009; Zhang et al., 2010). Critically, demonstration of correspondence between predicted and observed patterns of functional connectivity would support the feasibility and validity of fNIRS-derived connectivity measures.

Author Contributions

Study design and protocols were conceived by DRCJ, DRL, FO-E, LWS, K-WK, GM, G-ZY, and AWD. Data collection was performed by DRCJ, DRL, FO-E, LWS and K-WK. Data analysis was performed by DRCJ, FO-E, DRL, K-WK, LWS, GM and TA. The manuscript was written by DRCJ, DRL and FO-E and final critical editing was performed by DRL, LWS, GM, TA, G-ZY and AWD.

Funding

This work was funded in part by research grants from the Academy of Medical Sciences (Lecturer Starter Grant) and Cancer Research UK (Academic Lecturership).

References

Achard, S., and Bullmore, E. (2007). Efficiency and cost of economical brain functional networks. *PLoS Comput. Biol.* 3:e17. doi: 10.1371/journal.pcbi.0030017

Avraham, T., Yeshurun, Y., and Lindenbaum, M. (2008). Predicting visual search performance by quantifying stimulus similarities. *J. Vis.* 8:9. doi: 10.1167/8.4.9

Bann, S. D., Khan, M. S., and Darzi, A. W. (2003). Measurement of surgical dexterity using motion analysis of simple bench tasks. *World. J. Surg.* 27, 390–394. doi: 10.1007/s00268-002-6769-7

Bassett, D. S., Bullmore, E. T., Meyer-Lindenberg, A., Apud, J. A., Wienberger, D. R., and Coppola, R. (2009). Cogntive fitness of cost-efficient functional networks. *Proc. Natl. Acad. Sci. U S A* 106, 11747–11752. doi: 10.1073/pnas.0903641106

Betz, T., Wilming, N., Bogler, C., Haynes, J. D., and König, P. (2013). Dissociation between saliency signals and activity in the early visual cortex. *J. Vis.* 13:6. doi: 10.1167/13.14.6

Brennan, S. E., Chen, X., Dickinson, C. A., Neider, M. B., and Zelinsky, G. J. (2008). Coordinating cognition: the costs and benefits of shared gaze during collaborative search. *Cognition.* 106, 1465–1477. doi: 10.1016/j.cognition.2007.05.012

Bullmore, E., and Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat. Rev. Neurosci.* 10, 186–198. doi: 10.1038/nrn2575

Bullmore, E. T., and Bassett, D. S. (2011). Brain graphs: graphical models of the human brain connectome. *Ann. Rev. Clin. Psychol.* 7, 113–140. doi: 10.1146/annurev-clinpsy-040510-143934

Cutini, S., Basso Moro, S., and Bisconti, S. (2012). Functional near infrared optical imaging in cognitive neuroscience: an introductory review. *J. Near Infrared Spectrosc.* 20, 75–92. doi: 10.1255/jnirs.969

Chetwood, A. S., Kwok, K. W., Sun, L. W., Mylonas, G. P., Clark, J., Darzi, A., et al. (2012). Collaborative eye tracking: a potential training tool in laparoscopic surgery. *Surg. Endosc.* 26, 2003–2009. doi: 10.1007/s00464-011-2143-x

Dayan, E., and Cohen, L. G. (2011). Neuroplasticity subserving motor skill learning. *Neuron.* 72, 443–445. doi: 10.1016/j.neuron.2011.10.008

Dehais, F., Peysakhovich, V., Scanella, S., Fongue, J., and Gateau, T. (2015). ""Automation suprise" in aviation: real-time solutions," in *CHI'15 Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul: Republic of Korea), 2525–2534.

Di Nocera, F., Camilli, M., and Terenzi, M. (2007). A random glance at the Flight Deck: Pilots' scanning strategies and real-time assessment of mental workload. *J. Cogn. Eng. Decis. Mak.* 1, 271–285. doi: 10.1518/155534307X255627

Duratin, G., Gagnon, J. F., Tremblay, S., and Dehais, F. (2014). Using near infrared spectroscopy and heart rate variability to detect mental overload. *Behav. Brain Res* 259, 16–23. doi: 10.1016/j.bbr.2013.10.042

Einhäuser, W., and König, P. (2010). Getting real-sensory processing of natural stimuli. *Curr. Opin. Neurobiol.* 20, 389–395. doi: 10.1016/j.conb.2010.03.010

Ferguson, S., and Stack, B. C. (2010). Addition of a telestrator for teaching during video-based procedures. *Otolaryngol. Head Neck Surg.* 143, 159–160. doi: 10.1016/j.otohns.2009.12.043

Fitts, P. M., and Posner, M. I. (1967). *Learning and Skilled Performance in Human Performance.* Belmont, CA: Brooks/Cole.

He, Y., and Evans, A. (2010). Graph theoretical modeling of brain connectivity. *Curr. Opin. Neurol.* 23, 341–350. doi: 10.1097/wco.0b013e32833aa567

James, D. R., Leff, D. R., Orihuela-Espina, F., Kwok, K. W., Mylonas, G. P., Athanasiou, T., et al. (2013). Enhanced frontoparietal network architectures following "gaze-contingent" versus "free-hand" motor learning. *Neuroimage.* 64, 267–276. doi: 10.1016/j.neuroimage.2012.08.056

James, D. R. C., Orihuela-Espina, F., Leff, D. R., Sodergren, M. H., Athanasiou, T., Darzi, A. W., et al. (2011). The ergonomics of Natural Orifice Translumenal Endoscopic Surgery (NOTES) in terms of performance, stress and cognitive behaviour. *Surgery.* 149, 525–533. doi: 10.1016/j.surg.2010.11.019

James, D. R. C., Orihuela-Espina, F., Leff, D. R., Mylonas, G. P., Kwok, K. W., Darzi, A. W., et al. (2010a). Cognitive burden estimation for visuomotor learning with fNIRS. *Med. Image Comput. Comput. Assist. Interv.* 13, 319–326. doi: 10.1007/978-3-642-15711-0_40

James, D. R. C., Orihuela-Espina, F., Leff, D. R., Mylonas, G. P., Kwok, K. W., Darzi, A. W., et al. (2010b). Neuroergonomic assessment of the robotic enhancement of surgery. *Surg. Endosc.* 24, 192–269.

Jöbsis, F. F. (1977). Noninvasive, infrared monitoring of cerebral and myocardial oxygen sufficiency and circulatory parameters. *Science.* 198, 1264–1267. doi: 10.1126/science.929199

Jurcak, V., Tsuzuki, D., and Dan, I. (2007). 10/20, 10/10 and 10/5 systems revisited: their validity as relative head-surface-based positioning systems. *Neuroimage.* 34, 1600–1611. doi: 10.1016/j.neuroimage.2006.09.024

Kanwisher, N., and Wojciulik, E. (2000). Visual attention: insights from brain imaging. *Nat. Rev. Neurosci.* 1, 91–100. doi: 10.1038/35039043

Kojima, H., and Suzuki, T. (2010). Hemodynamic change in occipital lobe during visual search: visual attention allocation measured with NIRS. *Neuropsychologia.* 48, 349–352. doi: 10.1016/j.neuropsychologia.2009.09.028

Kwok, K. W., Sun, L. W., Mylonas, G. P., James, D. R., Orihuela-Espina, F., and Yang, G. Z. (2012). Collaborative gaze channelling for improved cooperation during robotic assisted surgery. *Ann. Biomed. Eng.* 40, 2156–2167. doi: 10.1007/s10439-012-0578-4

Leff, D. R., Elwell, C. E., Orihuela-Espina, F., Atallah, L., Delpy, D. T., Darzi, A. W., et al. (2008a). Changes in prefrontal cortical behaviour depend upon familiarity on a bimanual co-ordination task: an fNIRS study. *Neuroimage.* 39, 805–813. doi: 10.1016/j.neuroimage.2007.09.032

Leff, D. R., Leong, J. J., Aggarwal, R., Yang, G. Z., and Darzi, A. (2008b). Could variations in technical skills acquisition in surgery be explained by differences in cortical plasticity? *Ann. Surg.* 247, 540–543. doi: 10.1097/sla.0b013e31815fa42e

Leff, D. R., Orihuela-Espina, F., Atallah, L., Athanasiou, T., Leong, J. J., Darzi, A. W., et al. (2008c). Functional prefrontal reorganization accompanies learning-associated refinements in surgery: a manifold embedding approach. *Comput. Aided Surg.* 13, 325–339. doi: 10.3109/10929080802531482

Leong, J. H., Atallah, L., Mylonas, G. P., Leff, D. R., Emery, R. J., Darzi, A. W., et al. (2008). "Investigation of partial directed coherence for hand-eye coordination in laparoscopic training," in *Lecture Notes in Computer Science.* eds T. Dohi, I. Sakuma, and H. Liao (Berlin Heidelberg: Springer-Verlag), 270–278.

Micali, S., Virgili, G., Vannozzi, E., Grassi, N., Jarrett, T. W., Bauer, J. J., et al. (2000). Feasibility of telementoring between Baltimore (USA) and Rome (Italy): the first five cases. *J. Endourol.* 14, 493–496. doi: 10.1089/end.2000.14.493

Molholm, S., Sehatpour, P., Mehta, A. D., Shpaner, M., Gomez-Ramirez, M., Ortigue, S., et al. (2006). Audio-visual multisensory integration in superior parietal lobule revealed by human intracranial recordings. *J. Neurophysiol.* 96, 721–729. doi: 10.1152/jn.00285.2006

Niu, H., Wang, J., Zhao, T., Shu, N., and He, Y. (2012). Revealing topological organization of human brain functional networks with resting-state functional near infrared spectroscopy. *PLoS One* 7:e45771. doi: 10.1371/journal.pone.0045771

Ohuchida, K., Kenmotsu, H., Yamamoto, A., Sawada, K., Hayami, T., Morooka, K., et al. (2009). The frontal cortex is activated during learning of endoscopic procedures. *Surg. Endosc.* 2310, 2296–2301. doi: 10.1007/s00464-008-0316-z

Orihuela-Espina, F., Leff, D. R., James, D. R. C., Darzi, A. W., and Yang, G. Z. (2010). Quality control and assurance in functional near infrared spectroscopy (fNIRS) experimentation. *Phys. Med. Biol.* 55, 1–24. doi: 10.1088/0031-9155/55/13/009

Parasuraman, R. (2003). Neuroergonomics: research and practice. *Theor. Issues Erg.* 4, 5–20. doi: 10.1080/14639220210199753

Roy, C. S., and Sherrington, C. S. (1890). On the regulation of the blood-supply of the brain. *J. Physiol.* 11, 85–158. doi: 10.1113/jphysiol.1890.sp000321

Schlachta, C. M., Lefebvre, K. L., Sorsdahl, A. K., and Jayaraman, S. (2010). Mentoring and telementoring leads to effective incorporation of laparoscopic colon surgery. *Surg. Endosc.* 24, 841–844. doi: 10.1007/s00464-009-0674-1

Sporns, O., Chialvo, D. R., Kaiser, M., and Hilgetag, C. C. (2004). Organisation, development and function of complex brain networks. *Trends Cogn. Sci.* 8, 418–425. doi: 10.1016/j.tics.2004.07.008

Sporns, O. (2011). *Networks of the Brain.* Cambridge: MIT Press.

Task Force of the European Society of Cardiology the North American Society of Pacing Electrophysiology (1996). Heart rate variability. Standards

of measurement, physiological interpretation and clinical use. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. *Eur. Heart J.* 17, 354–381. doi: 10.1161/01.cir.93.5.1043

Theeuwes, J. (2010). Top-down and bottom-up control of visual selection. *Acta Psychologica.* 135, 77–99. doi: 10.1016/j.actpsy.2010.02.006

Tokuda, S., Obinata, G., Palmer, E., and Chaparro, A. (2011). Estimation of mental workload using saccadic eye movements in a free-viewing task. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 2011, 4523–4529. doi: 10.1109/iembs.2011.6091121

van den Heuvel, M. P., Mandl, R. C., Kahn, R. S. and Hulshoff Pol, H. E. (2009). Functionally linked resting-state networks reflect the underlying structural connectivity architecture of the human brain. *Hum. Brain Mapp.* 30, 3127–3141. doi: 10.1002/hbm.20737

van der Stigchel, S., Belopolsky, A. V., Peters, J. C., Wijnen, J. G., Meeter, M., and Theeuwes, J. (2009). The limits of top-down control of visual attention. *Acta. Psychologica.* 132, 201–212. doi: 10.1016/j.actpsy.2009.07.001

van Wijk, B. C., Stam, C. J., and Daffertshofer, A. (2010). Comparing brain networks of different size and connectivity density using graph theory. *PLoS One* 5:e13701. doi: 10.1371/journal.pone.0013701

Vinje, W. E., and Gallant, J. L. (2000). Sparse coding and decorrelation in primary visual cortex during natural vision. *Science.* 287, 1273–1276. doi: 10.1126/science.287.5456.1273

Wilson, M. R., Vine, S. J., Bright, E., Masters, R. S., Defriend, D., and McGrath, J. S. (2011). Gaze training enhances laparoscopic technical skill acquisition and multi-tasking performance: a randomized, controlled study. *Surgi. Endosc.* 25, 3731–3739. doi: 10.1007/s00464-011-1802-2

Xeroulis, G., Dubrowski, A., and Leslie, K. (2009). Simulation in laparoscopic surgery: a concurrent validity study for FLS. *Surg. Endosc.* 23, 161–165. doi: 10.1007/s00464-008-0120-9

Zhang, D., Snyder, A. Z., Shimony, J. S., Fox, M. D., and Raichle, M. E. (2010). Noninvasive functional and structural connectivity mapping of the human thalamocortical system. *Cereb. Cortex.* 20, 1187–1194. doi: 10.1093/cercor/bhp182

Zheng, B., Jiang, X., Tien, G., Mengeghetti, A., Panton, O. N., and Atkins, M. S. (2012). Workload assessment of surgeons: correlation between NASA TLX and blinks. *Surg. Endosc.* 26, 2746–2750. doi: 10.1007/s00464-012-p2268-6

Zheng, B., Jiang, X., and Atkins, M. S. (2015). Detection of changes in surgical difficulty: Evidence from pupil responses. *Surg. Innov.* doi: 10.1177/1553350615573582 [Epub ahead of print].

Zhu, F. F., Poolton, J. M., Wilson, M. R., Hu, Y., Maxwell, J. P., Masters, R. S., et al. (2011). Implicit motor learning promotes neural efficiency during laparoscopy. *Surg. Endosc.* 25, 2950–2955. doi: 10.1007/s00464-011-1647-8

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Leff, James, Orihuela-Espina, Kwok, Sun, Mylonas, Athanasiou, Darzi and Yang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution and reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 06 July 2016
doi: 10.3389/fnhum.2016.00344

The Role of Cognitive and Perceptual Loads in Inattentional Deafness

Mickaël Causse[1,2]*, *Jean-Paul Imbert*[3], *Louise Giraudet*[1], *Christophe Jouffrais*[4] and *Sébastien Tremblay*[2]

[1] Département Conception et Conduite des Véhicules Aéronautiques et Spatiaux, Institut Supérieur de l'Aéronautique et de l'Espace (ISAE), Toulouse, France, [2] School of Psychology, Co-Dot Laboratory, Université Laval, Québec, QC, Canada, [3] Laboratoire d'Informatique Interactive (LII), École Nationale de l'Aviation Civile (ENAC), Toulouse, France, [4] Centre National de la Recherche Scientifique (CRNS) and Université de Toulouse, IRIT, Toulouse, France

The current study examines the role of cognitive and perceptual loads in inattentional deafness (the failure to perceive an auditory stimulus) and the possibility to predict this phenomenon with ocular measurements. Twenty participants performed Air Traffic Control (ATC) scenarios—in the Laby ATC-like microworld—guiding one (low cognitive load) or two (high cognitive load) aircraft while responding to visual notifications related to 7 (low perceptual load) or 21 (high perceptual load) peripheral aircraft. At the same time, participants were played standard tones which they had to ignore (probability = 0.80), or deviant tones (probability = 0.20) which they had to report. Behavioral results showed that 28.76% of alarms were not reported in the low cognitive load condition and up to 46.21% in the high cognitive load condition. On the contrary, perceptual load had no impact on the inattentional deafness rate. Finally, the mean pupil diameter of the fixations that preceded the target tones was significantly lower in the trials in which the participants did not report the tones, likely showing a momentary lapse of sustained attention, which in turn was associated to the occurrence of inattentional deafness.

OPEN ACCESS

Edited by:
Klaus Gramann,
Berlin Institute of Technology, Germany

Reviewed by:
Peter König,
University of Osnabrück, Germany
Massimo Grassi,
University of Padova, Italy

***Correspondence:**
Mickaël Causse
mickael.causse@isae.fr

Received: *15 December 2015*
Accepted: *21 June 2016*
Published: *06 July 2016*

Citation:
Causse M, Imbert J-P, Giraudet L, Jouffrais C and Tremblay S (2016) The Role of Cognitive and Perceptual Loads in Inattentional Deafness. Front. Hum. Neurosci. 10:344. doi: 10.3389/fnhum.2016.00344

Keywords: inattentional deafness, cognitive load, perceptual load, pupil diameter, neuroergonomics

INTRODUCTION

The Air Traffic Control (ATC) environment involves supervisory control of emergency response, and security surveillance. Air traffic controllers must deal with dynamic and cognitively demanding tasks: guiding aircraft through a controlled airspace and optimizing trajectories whilst adhering to minimum distance and altitude separation minima requirement. This task must be completed in the face of temporal pressure, stress, and high-risk decision-making situations. Several research tried to identify the characteristics of the ATC environment that create cognitive demand (e.g., Manning et al., 2002; Loft et al., 2007). Manning et al. (2001) showed that these characteristics include, among others, the total number of aircraft controlled, the number of aircraft changing altitude, and the total conflict alert displayed. Other studies revealed that the dynamic density of the airspace at a given moment accounts for approximately half the variance in workload (Laudeman et al., 1998; Kopardekar and Magyarits, 2003). Although task demand has a strong relationship with workload, this relationship depends on the ATC operator capacity to select priorities and manage its cognitive resources (Loft et al., 2007).

The auditory channel is an essential means for air traffic controllers to exchange information with pilots and other controllers through radio and phone communications. They must also be vigilant and responsive to the occurrence of auditory alarms such as ground collision avoidance alerts or area infringement warnings that have been increasingly integrated into ATC workstations (Cabrera et al., 2005). Given that the auditory modality provides information without

requiring head/gaze movements (Edworthy et al., 1991), it is particularly suitable for the transmission of alerts and warnings in emergency situations because perception is not dependent on the direction of gaze at a particular moment (Harris, 2011). However, research in the field of aviation has provided ample evidence that individuals can still remain unaware of unexpected task-relevant and often safety-critical auditory stimuli if deeply involved in demanding tasks (Dehais et al., 2012, 2014; Giraudet et al., 2015b).

Several studies support the notion of a central bottleneck of attention processing (Jolicoeur, 1999; Arnell and Larson, 2002; Lavie, 2005; Dux et al., 2006; Raveh and Lavie, 2015; Wahn and König, 2015) but other works propose modality-specific restrictions of attention (Duncan et al., 1997; Talsma et al., 2006; Martens et al., 2010; Keitel et al., 2013). In accordance with the first view, Tombu et al. (2011) proposed a central attentional bottleneck that includes the inferior frontal junction, superior medial frontal cortex, and bilateral insula that temporally limits cognitive processes such as perceptual encoding or decision-making. In contrast, other studies show support for modality specific limitations by demonstrating that attentional capacity between modalities is greater than attentional capacity within the same modality (Talsma et al., 2006). Furthermore, Martens et al. (2010) showed that an attentional blink is produced only when targets are both presented within the same modality (auditory or visual) but not cross-modally, thus favoring the idea of a modality-specific sensory system rather than a central amodal system. From a theoretical viewpoint, multiple resource theory (Wickens, 1980) posits that there are multiple, independent pools of resources and that tasks that share the same limited resource would interfere with each other but would not affect other tasks that require a different type of resource. For example, Kim et al. (2005), showed that Stroop interference increased when the type of working memory (WM) load overlapped with the type of information required for the target task. At the same time, Stroop interference decreased when the type of WM load overlapped with distractor processing. Beyond this debate between central vs. modality-specific attentional limitations, many studies show that WM load also affects the ability to process visual or auditory environmental stimuli. For example, Sörqvist et al. (2012) demonstrated that brain response to an irrelevant sound decreased as a function of central WM load, induced by a visual-verbal version of the n-back task. In the same way, it has been shown that manipulating the task load of the primary task reduced markedly the sensitivity to auditory distractors during a duration-discrimination task (Berti and Schröger, 2003).

Given the evidence for both sides of the amodal vs. modality-specific debate on attentional capacity, we might postulate the existence of both central limitations in the control of attention and executive control (Rossi et al., 2009), with a key role of the prefrontal cortex (Asplund et al., 2010) and higher-order multisensory cortices (Calvert and Thesen, 2004), and additional capacity limits in modality-specific sensory brain areas (Talsma and Kok, 2001). Such a hypothesis is supported by Vachon and Tremblay (2008) using an attentional blink paradigm. Their results tend to support the idea that attentional limitations are due to a mixture of both modality-specific and amodal resource constraints. Based on the results of Berti and Schröger (2003) and Sörqvist et al. (2012) showing the adverse effect of WM load, as well as similar fundamental works (Wood and Cowan, 1995; Spence and Read, 2003; Lavie et al., 2004; Hughes et al., 2013) and observations in flight and ATC simulators (Dehais et al., 2012, 2014; Giraudet et al., 2015a,b) indicating that a high cognitive load context can lead to the neglect of auditory alerts, we may reasonably postulate that the risk of missed alarms is quite important in complex activities such as ATC.

The high cognitive and perceptual loads typical of ATC operations may consume most of attentional resources, thus reducing the remaining attentional capacity for processing unexpected stimuli such as auditory alarms. This failure to perceive auditory stimuli has been called inattentional deafness (Macdonald and Lavie, 2011; Koreimann et al., 2014). Given the potential impact of inattentional deafness in safety-critical occupations, it is important to understand the factors that promote this phenomenon and to be able to detect its occurrence. When no visual feedback from the operator is available, or when the alarm is triggered by a system, it is almost impossible to interpret human reactions. However, recent studies have found electro-encephalographic indicators of the occurrence of inattentional deafness with diminution of the amplitude of the P300 evoked potential (Giraudet et al., 2015a,b). These results are promising since they allow an offline analysis to test alarm designs and to evaluate the conditions favoring inattentional deafness. However, the online detection of inattentional deafness with ERP is complex under ecological conditions given the low signal-to-noise ratio of the event-related EEG activity. A more robust way for detecting inattentional deafness online is still to be developed, but the ability to predict its occurrence using a physiological measure has excellent potential. With the visual modality monopolizing most of attentional resources, we suggest that recording eye movements while operators are exposed to alarms can inform about their auditory capacity in real time, particularly if they are displaying inattentional deafness. Eye-tracking has already proven very useful for interface design and for usability tests (Goldberg and Kotval, 1999). Several behavioral ocular metrics such as the number of fixations and their duration, the scanpath direction and length, or the switching rate between areas of interest can provide a non-invasive measure of cognitive activity (for a review see, Jacob and Karn, 2003). Evidence suggests that when the eye is free to move, fixation location is strongly correlated with where attention is focused (Findlay and Gilchrist, 2003). But while eye tracking is known to reflect visual cognition, it is uncertain whether ocular behavior could reflect further mental processes beyond basic visual encoding of task-relevant information. Also, the pupil diameter is a classic measure to index cognitive activity and is generally higher in context of high mental workload (Kahneman and Beatty, 1966; Palinko et al., 2010; Peysakhovich et al., 2015) or when the level of vigilance is high (Beatty, 1982). For example, Beatty (1982) showed that vigilance decrement was associated to decreased amplitude of the phasic task-evoked pupillary response during an auditory vigilance task, while tonic or baseline pupillary diameter exhibited no such relationship.

Inattentional deafness is generally studied by varying perceptual load (Koreimann et al., 2009, 2014; Macdonald and Lavie, 2011; Molloy et al., 2015; Raveh and Lavie, 2015), while the effects of variations in mental workload (central demand) are less well investigated (Giraudet et al., 2015b). Importantly, no studies have examined and compared the impact of these two loads on the ability to perceive auditory stimuli. The present study had two main objectives: to further understand how cognitive and perceptual loads impact auditory detection sensitivity, and to assess the possibility of eye-movements and pupil diameter to predict the occurrence of inattentional deafness. Twenty participants performed a realistic ATC simulation task called Laby (Imbert et al., 2014a) while an auditory oddball task was presented. Participants had to react to deviant tones (simulating auditory alarms) by button pressing, as an indicator of their detection of the sound. We separately examined the impact of cognitive and perceptual loads on auditory detection sensitivity with a 2 × 2 factorial design. The cognitive load varied with the number of central aircraft to control, and the perceptual load with the number of peripheral aircraft to monitor. In a previous study also using Laby, we demonstrated that the cerebral response (P300) to deviant auditory tones was diminished when the visual design of Laby was poor (Giraudet et al., 2015a). Also, this study showed that approximatively 6% of the deviant tones were missed in the high cognitive load condition with the poor visual design. To further understand the factors that promote the occurrence of inattentional deafness, in the present study we intended to increase the inattentional deafness rate by using a more engaging and complex version of the Laby. Inducing a high level of missed alarms would enable comparison between the ocular behavior of missed and reported alarms. We hypothesized that: (1) the high cognitive and perceptual load conditions should generate more missed alerts than the low cognitive and perceptual load conditions; (2) increased cognitive and perceptual load should impact ocular measurements; and (3) ocular measurements may predict the occurrence of inattentional deafness.

MATERIALS AND METHODS

Participants

Twenty participants, all students of Université Laval were recruited for this study (Mean age = 23.5 years, Standard Deviation (SD) = 4.2). None had a history of neurological disease, psychiatric disturbance, substance abuse, or took psychoactive medications. They all received full information on the experimental protocol, signed an informed consent and received compensation for their participation in the study.

Experimental Design

We used a 2 × 2 factorial design crossing two independent variables, cognitive and perceptual loads. The cognitive load was manipulated by the number of central aircraft in the corridor. The low cognitive load condition was the first half of the scenarios, with one aircraft to guide. The high cognitive load condition was the second half of the scenarios, with two aircraft to guide. The perceptual load was manipulated by the number of peripheral aircraft around the corridor (between 5 and 21). The perceptual load was unique for each scenario and the order in which low and high perceptual load scenarios was performed was counterbalanced across participants.

The Laby Microworld and The Auditory Oddball Task

The ATC Task

The Laby microworld is a functional simulation of ATC, developed to create and evaluate new designs for controller's visualization. It is built on the main task of guiding aircraft around a route shown on the center part of the screen (light green path). For the first half of the Laby scenario, there was only one aircraft to monitor. In order to increase the main task demand, at the beginning of the second half of the scenario, a second aircraft entered the corridor and participants had to guide both aircraft along the route (**Figure 1**).

In order to maintain the central aircraft within the corridor or to follow altitude instructions, participants had to regularly modify their heading and altitude, using drop-down menus (**Figures 2A,B**).

In addition to the central aircraft, participants had to monitor a set of static aircraft (5 in the low perceptual load condition, 21 in the high perceptual load condition) located around the main aircraft corridor (**Figure 1**). "Color-Blink" visual notifications were displayed in or around the radar label located in the vicinity of these peripheral aircraft (**Figure 3**). Color-Blink uses colored text with the word "ALRT" which switches from white to red (see **Figures 3A,B**). It is used in ATC operational radar visualization for high-priority short-term conflict alerts. The Laby interface design is similar to operational radar visualization, and has been used in a previous study comparing the performance of several visual notifications in peripheral vision (Imbert et al., 2014b). Compared to other enhanced designs, the Color-Blink notification was found to be less salient and had a lower detection rate among controllers (see, Imbert et al., 2014a). We thus selected the Color-Blink notification to increase the overall monitoring effort in the present study. In another study also with Laby (Giraudet et al., 2015a), we showed that a high cognitive load condition of the Laby was associated with 6% unreported ton. Also, as we intended to increase the inattentional deafness rate with a more engaging and complex versions of the Laby in the present study, two modifications were performed. In the present study, there was two aircraft to guide in the high cognitive load condition (one in the previous study). In addition, contrary to the previous studies in which the heading indications to give to the aircraft was already computed by the system and just had to be selected by the participants in a drop-down menus, in the present study, the participants had to mentally calculate the various heading that the aircraft should follow to turn and stay in the corridor. An orange heading indicator was displayed on the top left corner in order to help participants to transform direction into heading values in degree.

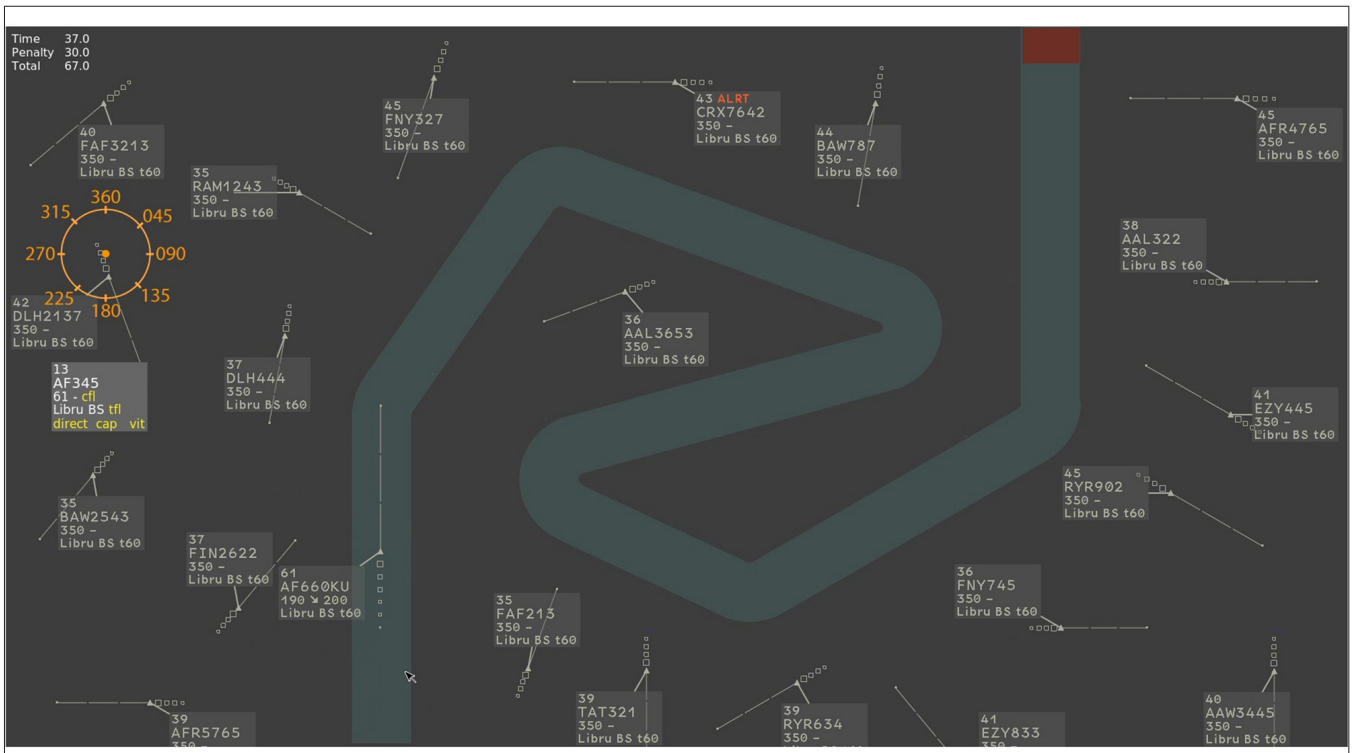


FIGURE 1 | Screenshot of the Laby microworld simulation. An example with 21 static peripheral aircraft positioned around the corridor. The central aircraft navigates through the corridor.

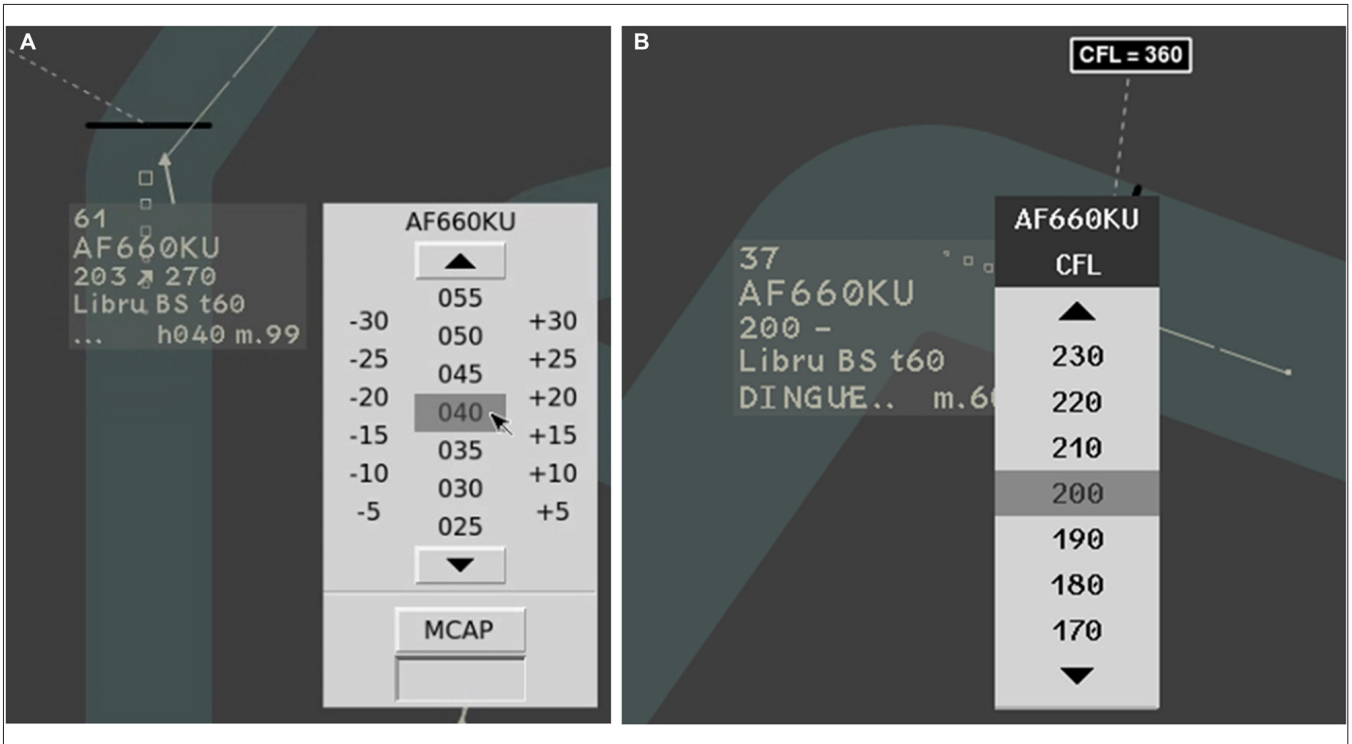


FIGURE 2 | Zoom on the Laby interface. (A) The menu used to select the heading of the central aircraft. **(B)** The menu used to select the altitude of the central aircraft. The menus appeared when clicking on the radar label.

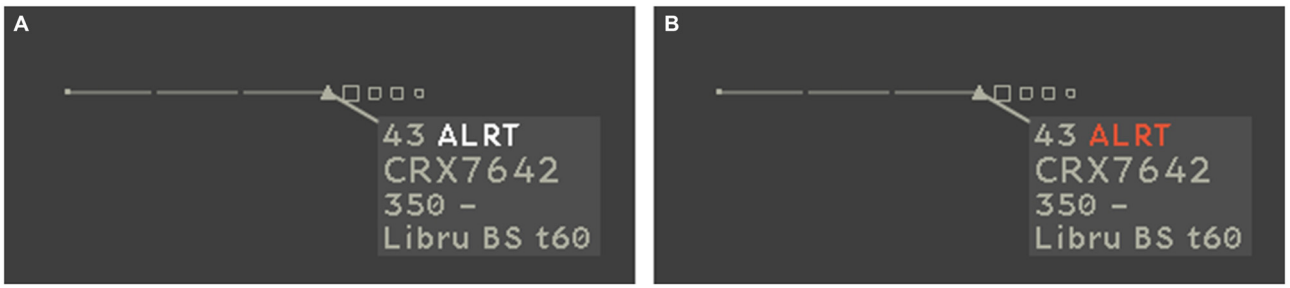


FIGURE 3 | The visual notification Color-Blink is inspired by operational warnings triggered in Air Traffic Control (ATC) radar screens when minimum separation between aircraft is lost. The text ALRT switches from white **(A)** to red **(B)** at a rate of 200 ms white/800 ms red.

Visual notifications were randomly displayed in the radar label located in the vicinity of these peripheral aircraft. Only one notification was issued at a time. The notification disappeared as soon as the participant clicked on the aircraft. If the participant did not react within a given time (5 s), the notification disappeared. Thirty-four visual notifications were displayed in each scenario.

A score was displayed on the top left of the screen. The score decreased for the following three reasons: first, when a participant led an aircraft outside the corridor; second, when he/she gave an incorrect altitude instruction; third, when he/she failed to click on a peripheral notification in the time limit. A deviation in the assigned route resulted in the aircraft crossing the border and initiating a visual alert in the center of the screen. An error in the altitude instructions resulted in the aircraft maintaining its trajectory, with no alert, and continued control. The score aimed to engage the participant in the ATC-like simulation in order to avoid them paying attention to the auditory alarm detection task only. The score was not considered in the analysis. The simulation ended as soon as the first aircraft reached the arrival area (colored red), at the end of the corridor.

Auditory Oddball to Simulate an Alarm Detection Task

In parallel to the ATC task, participants had to perform an auditory alarm detection task. Standard pure tones (1000 Hz, 52.5 dB SPL, 500 ms long, probability = 0.8) and deviant pure tones (2000 Hz, 52.5 dB SPL, 500 ms long, probability = 0.2) were randomly played. The tones were not representative of the auditory alerts recently integrated in ATC operations, their frequencies were chosen from previous works (Giraudet et al., 2014, 2015a,b). The mean time window between successive tones was 4.2 s. Participants were told to consider the deviant tones as auditory warnings and to report them as fast as possible by pressing a specific button. The auditory oddball detection task had no impact on the score. The number of auditory alarms (10) was the same in the four experimental conditions. There were ten tones in the first half (with one main aircraft) and ten tones in the second half (two aircraft). In order to increase the sound detection task difficulty, A 42 dB white noise was played continuously during each ATC scenario and the oddball control task. A control condition was also performed by the participant. They only had to react to the deviant auditory tones of the oddball while fixating a cross at the screen.

Procedure

The whole procedure lasted about 1 h. First, participants were seated comfortably at 60 cm from the 30″ screen in a sound-attenuated room with their right hand on the computer mouse and their left hand on the auditory alarm button. Second, they completed a training phase of 5 min to familiarize with the Laby microworld software, i.e., enter correctly path and altitude instructions by the drop-down menus, acknowledge visual notifications, and report deviant tones. After the training, the eye tracker was calibrated and participants completed the four ATC scenarios. Between scenarios, the eye tracker was recalibrated. Finally, participants performed the auditory oddball control task.

Eye Tracking Measurements and Data Processing

Continuous eye tracking was performed with a Tobii T1750 during the four ATC scenarios. The signal was recorded at a sampling rate of 300 Hz. For all eye movement analyses, the threshold to detect a fixation was set at 100 ms and the fixation field corresponded to a circle with a 30-pixel radius (equivalent to 1.15° of visual angle when seated at a distance of 50 cm). The position of both eyes on the screen was recorded. Data analysis was performed using MATLAB 7.1 (The Mathworks). Heatmaps visualizations of the distribution of fixations were generated using the open source software Open Gaze and Mouse Analyzer (OGAMA; Vosskühler et al., 2008).

Statistical Analysis

The impact of cognitive and perceptual loads on the accuracy to the central aircraft guiding task, peripheral notifications detection rate and missed auditory stimuli (rare tones) were analyzed. We also calculated the mean fixation duration on each of the four whole scenario and the mean duration of the fixation time that preceded the onset of a deviant tone (time-locked analysis) as well as the mean pupil diameter of this fixation (averaged on both eyes). Statistical analyses were performed using Statistica 10 (StatSoft©). Differences between the experimental conditions were investigated with the use of within-subjects analysis of variance (ANOVA) followed by

post hoc testing (Tukey's honestly significant difference, Tukey HSD). We finally computed a multivariate logistic regression in order to further determine the variables that predicted inattentional deafness.

RESULTS

Effects of Cognitive and Perceptual Loads on Performance to The ATC Task

We examined if the performance in the ACT task depended on cognitive and perceptual loads, see **Figure 4**. The 2 × 2 (cognitive load × perceptual load) repeated measures ANOVA showed no significant effect of the cognitive and perceptual loads on the accuracy to the central aircraft guiding task (respectively, $F_{(1,19)} = 0.86$, $p > 0.05$, $\eta_p^2 = 0.04$; $F_{(1,19)} = 0.64$, $p > 0.05$, $\eta_p^2 = 0.03$). The interaction term was not significant ($F_{(1,19)} = 0.37$, $p > 0.05$, $\eta_p^2 = 0.01$). Regarding the peripheral notifications detection rate, the 2 × 2 (cognitive load × perceptual load) repeated measures ANOVA showed a significant effect of the cognitive load, with a lower performance in the high cognitive load condition ($F_{(1,19)} = 37.45$, $p < 0.001$, $\eta_p^2 = 0.66$). The perceptual load had a near significant impact ($F_{(1,19)} = 3.74$, $p = 0.06$, $\eta_p^2 = 0.16$). The interaction term was not significant ($F_{(1,19)} = 0.00$, $p > 0.05$, $\eta_p^2 = 0.00$).

Effects of Cognitive and Perceptual Loads on Ocular Behavior

The analysis of the dispersion of fixations across the Laby interface is shown in **Figure 5**. In the high cognitive load condition, there is an increase in the overall time spent fixating the central part of the Laby interface where the central aircraft are moving. Also, the time spent fixating peripheral aircraft is increased in the high perceptual load condition.

We analyzed the extent to which the overall fixation time, averaged across each whole condition duration, were affected by cognitive and perceptual loads (**Figure 6**). The 2 × 2 (cognitive load × perceptual load) showed a significant effect of cognitive load on fixation duration ($F_{(1,19)} = 7.69$, $p < 0.05$, $\eta_p^2 = 0.29$). The effect of the perceptual load was not significant ($F_{(1,19)} = 0.00$, $p > 0.05$, $\eta_p^2 = 0.00$) neither the interaction term ($F_{(1,19)} = 0.41$, $p > 0.05$, $\eta_p^2 = 0.02$). Overall average fixation durations were approximatively 420 ms (see **Figure 6**), which is consistent with a previous study using the same eye tracker during a simulated combat control system microworld. In this latter work, participants demonstrated average fixation durations above 300 ms in several experimental conditions (Hodgetts et al., 2015).

Fixation durations before a saccade have been shown to be modulated by the relative angle of the saccade (see, Wilming et al., 2013). The alternation between the two central planes in the high load condition could lead to systematic differences in the angle between subsequent saccades in comparison to the low load condition with only one central plane. This difference in angle by means of saccadic momentum can in turn lead to differences in fixation duration. Consequently, we compared the average angle between two saccades in the low vs. high cognitive load condition in order to examine a possible effect of momentum on fixation times. This analysis revealed that the mean angle slightly increased with increased cognitive load (low cognitive load = 84.64° (SD = 0.81); high cognitive load = 85.95° (SD = 1.08)) but the analysis did not reach the significance threshold ($F_{(1,19)} = 2.24$, $p > 0.05$, $\eta_p^2 = 0.10$). Saccadic momentum cannot explain by itself the variations of fixation times across the two cognitive load conditions.

Effects of Cognitive and Perceptual Loads on the Inattentional Deafness Rate

The control condition revealed that the inattentional deafness rate (missed alerts = 1-hit rate) was extremely low, with 2% (SD = 5.93) of missed alert. As a matter of fact, two participants omitted a few deviant tones whereas the 18 others reacted to 100% of the deviant tones. This result confirms that the

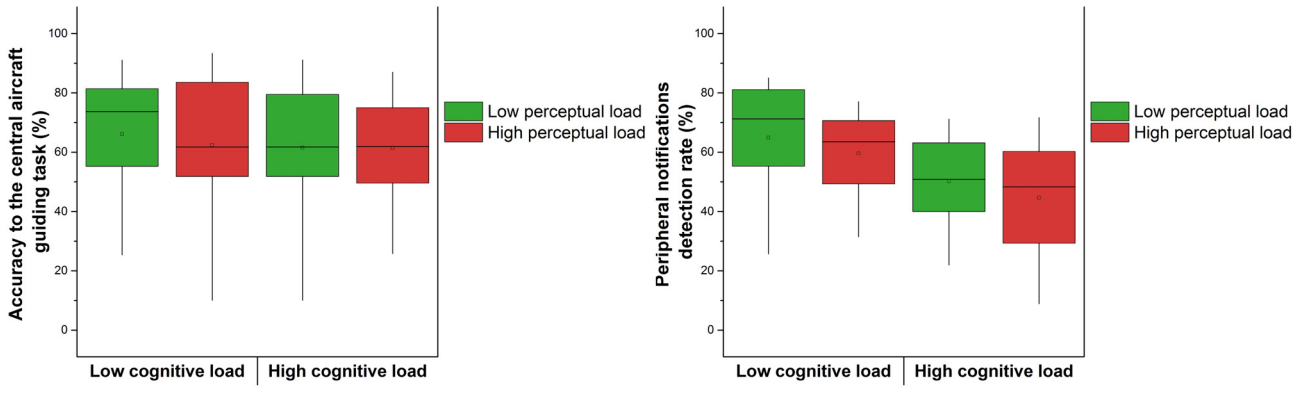


FIGURE 4 | Correct responses to the central aircraft guiding sub-task according to the levels of cognitive and perceptual loads. Validation of the peripheral aircraft sub-task according to the levels of cognitive and perceptual load. The square in the center of the boxes represent the mean, the horizontal line in the center of the boxes represent the 50th percentile (median), the end of the boxes represent the 25th and 75th percentiles, and the whiskers represent the 5th and 95th percentiles.

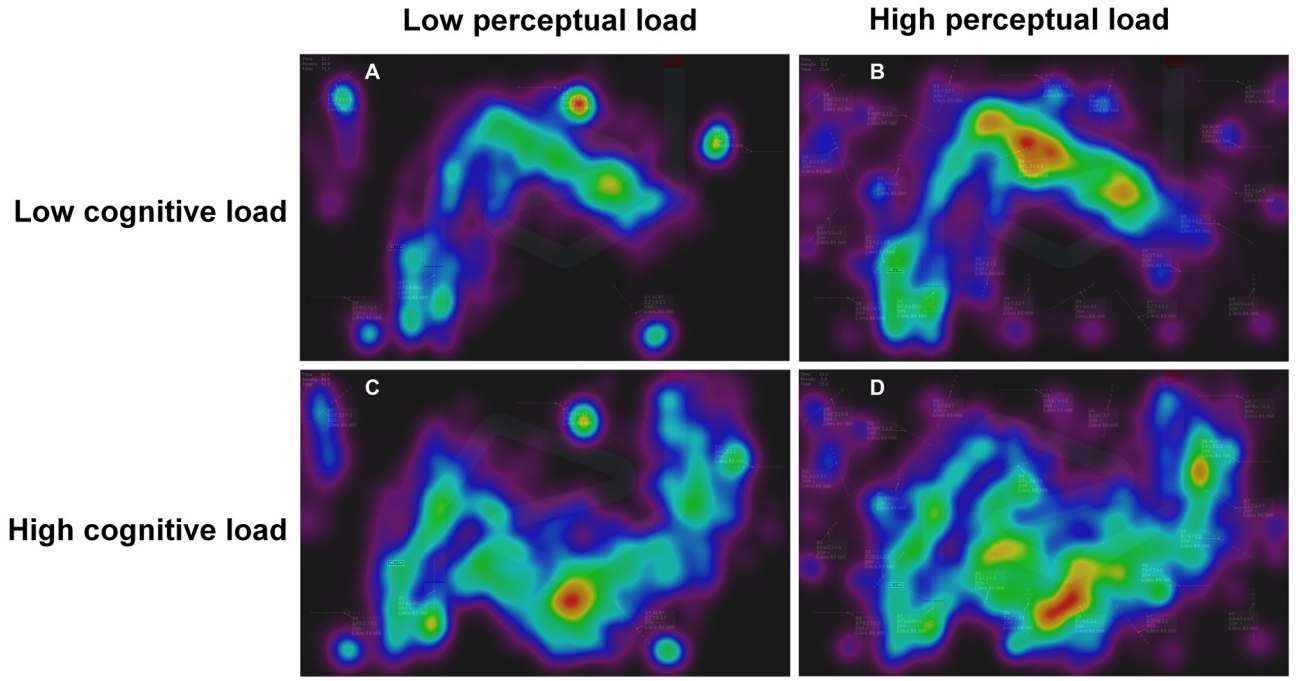


FIGURE 5 | Heatmap visualizations of the distribution of fixations on the Laby interface. (A) Low cognitive load/low perceptual load; **(B)** low cognitive load/high perceptual load; **(C)** high cognitive load/low perceptual load; **(D)** high cognitive load/high perceptual load.

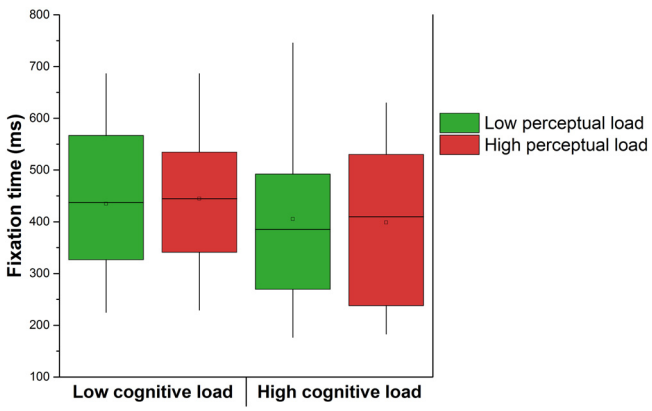


FIGURE 6 | Effect of the cognitive and perceptual loads on the fixation time averaged across the whole condition duration. The square in the center of the boxes represent the mean, the horizontal line in the center of the boxes represent the 50th percentile (median), the end of the boxes represent the 25th and 75th percentiles, and the whiskers represent the 5th and 95th percentiles.

tones were clearly perceptible despite the continuous white noise. We then examined if the inattentional deafness rate depended on cognitive and perceptual loads. The 2 × 2 (cognitive load × perceptual load) repeated measures ANOVA showed a significant effect of the cognitive load on the percentage of missed auditory alarms with an increased percentage of missed auditory stimuli in the high cognitive load condition ($F_{(1,19)} = 24.49$, $p < 0.001$, $\eta_p^2 = 0.56$), see **Figure 7**. The perceptual load had no significant impact ($F_{(1,19)} = 0.10$, $p > 0.05$, $\eta_p^2 = 0.00$). The interaction term was not significant ($F_{(1,19)} = 0.49$, $p > 0.05$, $\eta_p^2 = 0.02$). We finally examined if the specificity (true negative rate rate) and the discriminability index (d′ = Z(hit rate) − Z(false alarm rate); Stanislaw and Todorov, 1999) depended on cognitive and perceptual loads. A first 2 × 2 (cognitive load × perceptual load) repeated measures ANOVA revealed no significant main effect of cognitive load ($F_{(1,19)} = 1.23$, $p > 0.05$, $\eta_p^2 = 0.06$) and perceptual load ($F_{(1,19)} = 0.07$, $p > 0.05$, $\eta_p^2 = 0.00$) neither interaction effect ($F_{(1,19)} = 0.22$, $p > 0.05$, $\eta_p^2 = 0.01$) on the specificity. A second 2 × 2 (cognitive load × perceptual load) repeated measures ANOVA revealed significant main effect of cognitive load ($F_{(1,19)} = 18.88$, $p < 0.001$, $\eta_p^2 = 0.50$) but no effect of perceptual load ($F_{(1,19)} = 0.00$, $p > 0.05$, $\eta_p^2 = 0.00$) neither interaction ($F_{(1,19)} = 3.30$, $p > 0.05$, $\eta_p^2 = 0.14$) on the discriminability index. In summary, the cognitive load had a specific impact on the sensibility (hit rate), which increased the number of missed alerts. This is illustrated in **Figure 7**, the cognitive load had a specific impact on the true positive rate, not on the false positive rate. The d′ variations are only due to this effect of cognitive load on the true positive rate. We finally estimated if the variations in loads impacted the reaction time to alerts. In all four experimental conditions, mean reaction times were markedly below the mean time available to respond between two tones (i.e., 4.2 s), $M = 1.67$ s in low cognitive load condition; 1.53 s in high cognitive load condition; 1.65 s in low perceptual load condition; 1.55 s in the high perceptual load condition. The 2 × 2 repeated measures ANOVA showed no significant effect of

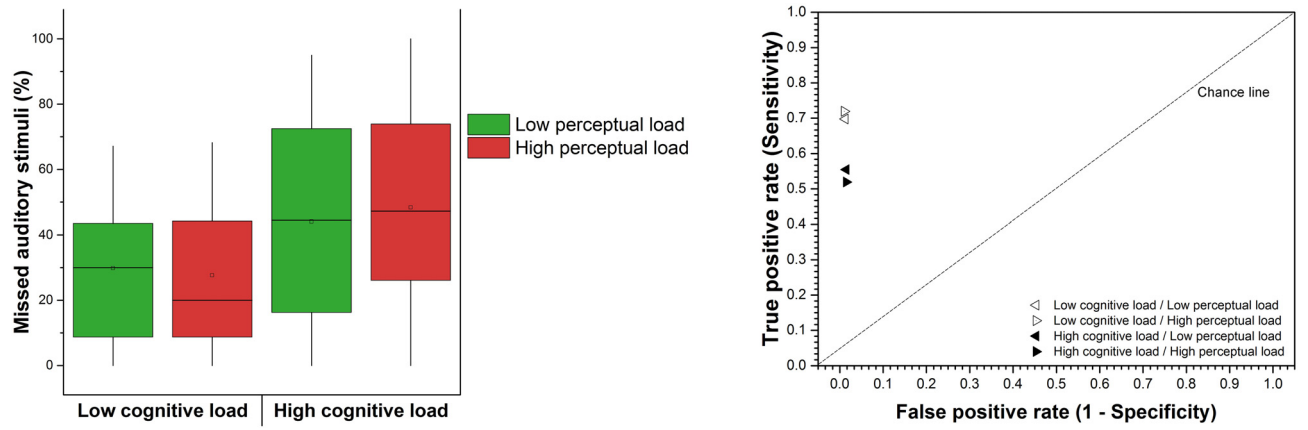


FIGURE 7 | Left: mean percentage of missed auditory stimuli according to the levels of cognitive and perceptual loads (referred to as inattentional deafness rate). The square in the center of the boxes represent the mean, the horizontal line in the center of the boxes represent the 50th percentile (median), the end of the boxes represent the 25th and 75th percentiles, and the whiskers represent the 5th and 95th percentiles. **Right:** true positive rate vs. false positive rate according to the levels of cognitive and perceptual loads.

cognitive ($F_{(1,19)} = 1.02$, $p > 0.05$, $\eta_p^2 = 0.05$) and perceptual load ($F_{(1,19)} = 1.03$, $p > 0.05$, $\eta_p^2 = 0.05$) neither significant interaction ($F_{(1,19)} = 0.06$, $p > 0.05$, $\eta_p^2 = 0.00$) on reaction times.

Multivariate Logistic Regression Analysis

In order to further investigate the factors that promote inattentional deafness, we performed a multivariate logistic regression analysis stratified by trial (trials with deviant tones only) with all participants grouped together (number of trials = 800). We used the occurrence of inattentional deafness (yes/no) as binary dependent variable and cognitive and perceptual load were introduced as categorical independent variables. Furthermore, the mean duration of the fixation that occurred just before the occurrence of the rare tone and the mean pupil diameter during this fixation were used as a continuous variable. The Wald chi-square p-values confirmed that the cognitive load was a significant predictor of the occurrence of inattentional deafness (Wald statistic = 14.38, $p < 0.001$) while perceptual load was not (Wald statistic = 0.77, $p > 0.05$). In addition, the regression also showed that pupil diameter of the fixation that preceded the rare sound was significantly lower in the trials in which the participants did not react to the target tones (Wald statistic = 18.66, $p < 0.001$) (see **Figure 8**). Finally, the duration of the previous fixation was not predictive (Wald statistic = 0.20, $p > 0.05$).

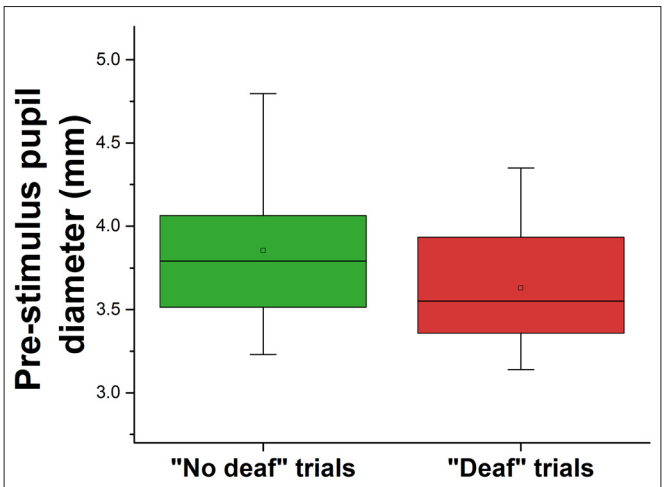


FIGURE 8 | Pre-stimulus pupil diameter (fixation $n-1$) for hit and missed auditory stimuli. The square in the center of the boxes represent the mean, the horizontal line in the center of the boxes represent the 50th percentile (median), the end of the boxes represent the 25th and 75th percentiles, and the whiskers represent the 5th and 95th percentiles.

DISCUSSION

Summary of Results

Our results showed that a high level of cognitive load, manipulated by the number of planes to guide, significantly increased the inattentional deafness rate whereas the perceptual load, manipulated by the number of peripheral aircraft to monitor, had no significant impact. The cognitive load also impacted ocular behavior with lower fixation time in the high load condition, while perceptual load had no significant effect. Finally, logistic regressions showed that the mean pupil diameter of the fixation that preceded the onset of the tones predicted inattentional deafness.

Effects of Cognitive and Perceptual Loads

Participants missed 28.76% of alarms in the low cognitive load condition (irrespective of the perceptual load) compared to 46.21% in the high cognitive load condition. This strikingly high rate of missed alerts cannot be attributed to sensorial difficulties as only 2% of alerts were missed in the control condition, in which the participant completed the tone detection

task only (only 2 out of 20 participants missed any alerts). In the high cognitive load condition, participants engaged in greater mental effort by shifting their attention from one plane to the other, and also had a higher workload due to the need to calculate/modify heading and to change altitude parameters more often. These factors increased the chance of experiencing inattentional deafness.

Interestingly, perceptual load did not increase the missed alarm rate. In order to disentangle cognitive from perceptual load as much as possible, the high perceptual load did not generate a supplementary cognitive effort as the number of peripheral aircraft was increased while the number of associated peripheral alerts remained constant. Indeed, as demonstrated by Manning et al. (2001), the total conflict alert displayed contribute to increase cognitive effort. Inattentional deafness seems to be produced when an individual is engaged and monopolized in a task rather than when the individual is gazing more passively at visual information. One might argue that the additional number of peripheral aircraft was simply ignored by the participant which may explain this lack of effect of the high perceptual load condition. However, the heatmap illustrating the distribution of fixations clearly showed that fixations on the peripheral aircraft increased in the high perceptual load condition.

As demonstrated by the heatmaps and by a decline in the detection rate of peripheral notifications, the high load condition resulted in an important focus on the central aircraft, a behavior that can be compared to attentional tunneling (Wickens and Alexander, 2009; Régis et al., 2014). In general, fixation duration is known to reflect the attention (Findlay and Gilchrist, 2003) and mental effort (De Rivecourt et al., 2008) of an observer. In this last study on simulated flights, De Rivecourt et al. (2008) showed that momentary altitude changes can result in increased mean fixation duration. Variation of the fixation duration should be considered as task dependent: both shorter and longer fixations may indicate an increase in workload, and in particular shorter fixations indicated higher workload and increased temporal pressure in our study. This strong engagement of cognitive resources seemed to contribute to create a momentarily "deafness" to auditory stimuli.

One might also argue that this high inattentional deafness rate was due to an insufficient time window to report the alarm (i.e., mean time window = 4.2 s). Yet, in all four experimental conditions, mean reaction times were markedly below the mean time available to respond between two tones (around 1 s), and the reaction times did not significantly vary across the four experimental conditions. Even if we cannot completely eliminate the idea that a relatively small number of deviant tones were not reported because participants reacted too late, these two results tend to exclude this explanation as a major contributor of variation in the inattentional deafness rate with increased cognitive load. The analysis of d′ confirmed that the decline in the number of reported alerts in the high cognitive load condition is associated with a loss of sensitivity to deviant tones, and not due to an effect on the ability to discriminate the two tones. In this latter situation, the number of false alarms would have likely increased.

Importantly the inattentional deafness rate of the present study was considerably higher than in previous research using Laby (Giraudet et al., 2015a), whereby the percentage of unreported tones was 6% in the high cognitive load condition. To further understand the factors that promote the occurrence of inattentional deafness, the present study had employed two modifications to create a more engaging and complex version of the Laby task. First, there were two aircraft to guide in the high cognitive load condition whereas only one was displayed in the previous study. Second, in the current study participants had to mentally calculate the various headings that the aircraft should follow to turn and stay within the corridor, whereas previously these were pre-calculated by the system and just required selection from a drop-down menu. These modifications lead to a considerable rise in the incidence of inattentional deafness. It must be noted that in both studies, the importance of reporting the sounds was emphasized and that the time between the two tones was identical in both. The mental calculation of heading was undoubtedly a key factor in this increase of inattentional deafness as even in the low load condition, in which only one aircraft was displayed, the inattentional deafness rate was greater than three times that observed in the previous study in which the heading was given by the system.

Lower Pupil Diameter Predicts Inattentional Deafness

The multivariate logistic regression confirmed that cognitive load significantly predicted the occurrence of inattentional deafness. Most importantly, the regression also revealed that pupil diameter was lower during the fixation that preceded the onset of the target tones in the "deaf" trials. This result is counterintuitive as inattentional deafness was indubitably increased by the high cognitive load context, which is supposed in turn to increase the pupil diameter (Kahneman and Beatty, 1966; Palinko et al., 2010; Peysakhovich et al., 2015). Yet, as previously mentioned, Beatty (1982) showed that vigilance decrement was associated to decreased amplitude of the phasic task-evoked pupillary response during an auditory vigilance task, while tonic or baseline pupillary diameter exhibited no such relationship. In addition, a very recent study (Unsworth and Robison, 2016) indicated that pupillary diameter can index lapses of sustained attention. They showed that compared to focused states, inattentive and mind-wandering states are associated with lower pretrial baseline pupil responses and that distracted states are associated with larger pretrial baseline pupil diameter. These results support the notion that pupil diameter is sensitive to different types of lapses of attention, which is consistent with theories of locus coeruleus norepinephrine (LC-NE) functioning. In our study, despite a context of sustained high cognitive load, momentary lapses of sustained attention may have occurred, which could explain the relationship between lower phasic pupil diameter and inattentional deafness occurrence. This assumption can be also related to a past study that revealed that information overload resulted in a leveling of the dilation pattern, which suggested a momentary suspension of processing effort (Peavler, 1974).

Our results demonstrating an effect of cognitive load but not of perceptual load on inattentional deafness, are somewhat contradictory to a study by Macdonald and Lavie (2011) in which participants were engaged in a visual discrimination task of a cross shape. In the low visual load condition, this discrimination was made according to the line color and in the high visual-load condition, participants had to discriminate subtle line length. One brief pure tone was presented simultaneously at the final trial onset. Failures to notice the presence of this tone reached a rate of 79% in the high-visual-load condition, significantly greater than in the low-load condition. We could postulate that the type of perceptual load manipulated by the authors likely generated an indirect increase in mental effort and task engagement due to the comparison process of the line length. For example, Fierro et al. (2000) showed that the line-length comparison process engages the parietal cortex, indicating that spatial cognition is also taxed in such a task. Also, the paradigm used by Macdonald and Lavie (2011) was quite different as only one pure sound was presented in the study while 10 target tones per condition were presented in the current study. We believe that our paradigm is closer to a real life context or complex activity such as ATC in which the auditory environment is composed of a mixture of different sounds that can be repeated several times. As our paradigm can be related to "change deafness" studies, for example in which a subtle change between two voices is unnoticed (Vitevitch, 2003), a future study would look to reproduce the same paradigm but with only deviant tone.

Conclusion

The present study suggests that inattentional deafness is promoted by cognitive load rather than by a "passive" perceptual load that does not generate a supplementary amount of work. A strong engagement of cognitive resources in a given task can momentarily render one "deaf" to auditory stimuli. In our study, the key factor that promoted inattentional deafness was most likely the cognitive load generated by the mental calculation of heading and by the numerous tasks to conduct. This result confirmed previous studies showing that inattentional deafness drastically increases in the context of high cognitive load (Giraudet et al., 2015b), which can have serious consequences in safety-critical occupations like ATC. Finally, the mean pupil diameter of the period that just preceded the rare sound onset was significantly lower in the trials in which the participants did not react to the target tones, likely showing a momentary lapse of sustained attention, which in turn promoted the occurrence of inattentional deafness.

ETHICS STATEMENT

The study was approved by the National Scientific Research Ethics Committee in Paris (CEEI/IRB00003888).

AUTHOR CONTRIBUTIONS

LG, J-PI and MC designed the experiment. LG conducted the experiments. LG, MC, J-PI analyzed and interpreted the data. MC, LG, ST and CJ wrote the article.

FUNDING

Financial support was provided by a Discovery grant from the National Sciences and Engineering Research Council of Canada (NSERC), awarded to Sébastien Tremblay [Grant number CG073877], by the Institut Supérieur de l'Aéronautique et de l'Espace to in the form of an operating grant to Mickaël Causse, and also by the Direction Générale de l'Armement in the form of a scholarship to Louise Giraudet.

ACKNOWLEDGMENTS

We would like to thank Danny Lebel for his technical support, and Marie-Kim Côté for her help running the experiment. We are also grateful to Zarrin Chua and Helen Hodgetts for proofreading earlier versions of the article.

REFERENCES

Arnell, K. M., and Larson, J. M. (2002). Cross-modality attentional blinks without preparatory task-set switching. *Psychon. Bull. Rev.* 9, 497–506. doi: 10.3758/bf03196305

Asplund, C. L., Todd, J. J., Snyder, A. P., and Marois, R. (2010). A central role for the lateral prefrontal cortex in goal-directed and stimulus-driven attention. *Nat. Neurosci.* 13, 507–512. doi: 10.1038/nn.2509

Beatty, J. (1982). Phasic not tonic pupillary responses vary with auditory vigilance performance. *Psychophysiology* 19, 167–172. doi: 10.1111/j.1469-8986.1982.tb02540.x

Berti, S., and Schröger, E. (2003). Working memory controls involuntary attention switching: evidence from an auditory distraction paradigm. *Eur. J. Neurosci.* 17, 1119–1122. doi: 10.1046/j.1460-9568.2003.02527.x

Cabrera, D., Ferguson, S., and Laing, G. (2005). "Development of auditory alerts for air traffic control consoles," in *Proceedings 119th Audio Engineering Society Convention* (New York, NY).

Calvert, G. A., and Thesen, T. (2004). Multisensory integration: methodological approaches and emerging principles in the human brain. *J. Physiol. Paris* 98, 191–205. doi: 10.1016/j.jphysparis.2004.03.018

De Rivecourt, M., Kuperus, M., Post, W., and Mulder, L. (2008). Cardiovascular and eye activity measures as indices for momentary changes in mental effort during simulated flight. *Ergonomics* 51, 1295–1319. doi: 10.1080/00140130802120267

Dehais, F., Causse, M., Régis, N., Menant, E., Labedan, P., Vachon, F., et al. (2012). "Missing critical auditory alarms in aeronautics: evidence for inattentional deafness?" in *Proceedings of the Human Factors and Ergonomics Society 56th Annual Meeting* (Boston), 1639–1643. doi: 10.1177/1071181312561328

Dehais, F., Causse, M., Vachon, F., Régis, N., Menant, E., and Tremblay, S. (2014). Failure to detect critical auditory alerts in the cockpit evidence for inattentional deafness. *Hum. Factors* 56, 631–644. doi: 10.1177/0018720813510735

Duncan, J., Martens, S., and Ward, R. (1997). Restricted attentional capacity within but not between sensory modalities. *Nature* 387, 808–810. doi: 10.1038/42947

Dux, P. E., Ivanoff, J., Asplund, C. L., and Marois, R. (2006). Isolation of a central bottleneck of information processing with time-resolved fMRI. *Neuron* 52, 1109–1120. doi: 10.1016/j.neuron.2006.11.009

Edworthy, J., Loxley, S., and Dennis, I. (1991). Improving auditory warning design: relationship between warning sound parameters and perceived urgency. *Hum. Factors* 33, 205–231.

Fierro, B., Brighina, F., Oliveri, M., Piazza, A., La Bua, V., Buffa, D., et al. (2000). Contralateral neglect induced by right posterior parietal rTMS in healthy subjects. *Neuroreport* 11, 1519–1521. doi: 10.1097/00001756-200005150-00030

Findlay, J. M., and Gilchrist, I. D. (2003). "Visual selection, covert attention and eye movements," in *Active Vision: The Psychology of Looking and Seeing Chapter 3*, (New York, NY: Oxford University Press), 35–54.

Giraudet, L., Berenger, M., Imbert, J.-P., Tremblay, S., and Causse, M. (2014). "Inattentional deafness in simulated air traffic control tasks: a behavioral and P300 analysis," in *AHFE 2014, 5th International Conference on Applied Human Factors and Ergonomics* (Kraków, Poland).

Giraudet, L., Imbert, J.-P., Bérenger, M., Tremblay, S., and Causse, M. (2015a). The neuroergonomic evaluation of human machine interface design in air traffic control using behavioral and EEG/ERP measures. *Behav. Brain Res.* 294, 246–253. doi: 10.1016/j.bbr.2015.07.041

Giraudet, L., St-Louis, M.-E., Scannella, S., and Causse, M. (2015b). P300 event-related potential as an indicator of inattentional deafness? *PLoS One* 10:e0118556. doi: 10.1371/journal.pone.0118556

Goldberg, J. H., and Kotval, X. P. (1999). Computer interface evaluation using eye movements: methods and constructs. *Int. J. Ind. Ergon.* 24, 631–645. doi: 10.1016/s0169-8141(98)00068-7

Harris, D. (2011). *Human Performance on the Flight Deck*. Farnham: Ashgate Publishing, Ltd.

Hodgetts, H. M., Tremblay, S., Vallières, B. R., and Vachon, F. (2015). Decision support and vulnerability to interruption in a dynamic multitasking environment *Int. J. Hum. Comput. Stud.* 79, 106–117. doi: 10.1016/j.ijhcs.2015.01.009

Hughes, R. W., Hurlstone, M. J., Marsh, J. E., Vachon, F., and Jones, D. M. (2013). Cognitive control of auditory distraction: impact of task difficulty, foreknowledge and working memory capacity supports duplex-mechanism account. *J. Exp. Psychol. Hum. Percept. Perform.* 39, 539–553. doi: 10.1037/a0029064

Imbert, J.-P., Hodgetts, H. M., Parise, R., Vachon, F., Dehais, F., and Tremblay, S. (2014a). Attentional costs and failures in air traffic control notifications. *Ergonomics* 57, 1817–1832. doi: 10.1080/00140139.2014.952680

Imbert, J.-P., Hodgetts, H. M., Parise, R., Vachon, F., and Tremblay, S. (2014b). "The LABY microworld a platform for research, training and system engineering," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Chicago), 27–31.

Jacob, R., and Karn, K. S. (2003). "Eye tracking in human-computer interaction and usability research: ready to deliver the promises," in *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research*, eds J. Hyona, R. Radach and H. Deubel (Oxford), 573–605.

Jolicoeur, P. (1999). Restricted attentional capacity between sensory modalities. *Psychon. Bull. Rev.* 6, 87–92. doi: 10.3758/bf03210813

Kahneman, D., and Beatty, J. (1966). Pupil diameter and load on memory. *Science* 154, 1583–1585. doi: 10.1126/science.154.3756.1583

Keitel, C., Maess, B., Schröger, E., and Müller, M. M. (2013). Early visual and auditory processing rely on modality-specific attentional resources. *Neuroimage* 70, 240–249. doi: 10.1016/j.neuroimage.2012.12.046

Kim, S.-Y., Kim, M.-S., and Chun, M. M. (2005). Concurrent working memory load can reduce distraction. *Proc. Natl. Acad. Sci. U S A* 102, 16524–16529. doi: 10.1073/pnas.0505454102

Kopardekar, P., and Magyarits, S. (2003). *Measurement and Prediction of Dynamic Density*. 5th Edn. Budapest, Hungary: Eurocontrol/FAA ATM R and D Seminar.

Koreimann, S., Gula, B., and Vitouch, O. (2009). Inattentional deafness under dynamic musical conditions. *Psychol. Res.* 78, 304–312. doi: 10.1007/s00426-014-0552-x

Koreimann, S., Gula, B., and Vitouch, O. (2014). Inattentional deafness in music. *Psychol. Res.* 78, 304–312. doi: 10.1007/s00426-014-0552-x

Laudeman, I. V., Shelden, S., Branstrom, R., and Brasil, C. L. (1998). Dynamic Density: An Air Traffic Management Metric. *Technical Report NASA/TM-1998-112226. Moffett Field, CA: National Aeronautics and Space Administration, Ames Research Center*.

Lavie, N. (2005). Distracted and confused? selective attention under load. *Trends Cogn. Sci.* 9, 75–82. doi: 10.1016/j.tics.2004.12.004

Lavie, N., Hirst, A., de Fockert, J. W., and Viding, E. (2004). Load theory of selective attention and cognitive control. *J. Exp. Psychol. Gen.* 133, 339–354. doi: 10.1037/0096-3445.133.3.339

Loft, S., Sanderson, P., Neal, A., and Mooij, M. (2007). Modeling and predicting mental workload in en route air traffic control: critical review and broader implications. *Hum. Factors* 49, 376–399. doi: 10.1518/001872007x197017

Macdonald, J. S., and Lavie, N. (2011). Visual perceptual load induces inattentional deafness. *Atten. Percept. Psychophys.* 73, 1780–1789. doi: 10.3758/s13414-011-0144-4

Manning, C. A., Mills, S. H., Fox, C., Pfleider, E., and Mogilka, H. J. (2001). *Investigating the Validity of Performance and Objective Workload Evaluation Research (POWER)*. (Report No. DOT/FAA/AM-01/10). Washington, DC: FAA Office of Aerospace Medicine.

Manning, C., Mills, S., Fox, C., Pfleiderer, E., and Mogilka, H. (2002). The relationship between air traffic control communication events and measures of controller taskload and workload. *Air Traffic Control Qual.* 10, 69–83.

Martens, S., Kandula, M., and Duncan, J. (2010). Restricted attentional capacity within but not between sensory modalities: an individual differences approach. *PLoS One* 5:e15280. doi: 10.1371/journal.pone.0015280

Molloy, K., Griffiths, T. D., Chait, M., and Lavie, N. (2015). Inattentional deafness: visual load leads to time-specific suppression of auditory evoked responses. *J. Neurosci.* 35, 16046–16054. doi: 10.1523/JNEUROSCI.2931-15.2015

Palinko, O., Kun, A. L., Shyrokov, A., and Heeman, P. (2010). "Estimating cognitive load using remote eye tracking in a driving simulator," in *Proceedings of the 2010 symposium on eye-tracking research and applications* (Austin).

Peavler, W. S. (1974). Pupil size, information overload, and performance differences. *Psychophysiology* 11, 559–566. doi: 10.1111/j.1469-8986.1974.tb01114.x

Peysakhovich, V., Causse, M., Scannella, S., and Dehais, F. (2015). Frequency analysis of a task-evoked pupillary response: luminance-independent measure of mental effort. *Int. J. Psychophysiol.* 97, 30–37. doi: 10.1016/j.ijpsycho.2015.04.019

Raveh, D., and Lavie, N. (2015). Load-induced inattentional deafness. *Atten. Percept. Psychophys.* 77, 483–492. doi: 10.3758/s13414-014-0776-2

Régis, N., Dehais, F., Rachelson, E., Thooris, C., Pizziol, S., Causse, M., et al. (2014). Formal detection of attentional tunneling in human operator-automation interactions. *IEEE Trans. Hum. Mach. Syst.* 44, 326–336. doi: 10.1109/thms.2014.2307258

Rossi, A. F., Pessoa, L., Desimone, R., and Ungerleider, L. G. (2009). The prefrontal cortex and the executive control of attention. *Exp. Brain Res.* 192, 489–497. doi: 10.1007/s00221-008-1642-z

Sörqvist, P., Stenfelt, S., and Rönnberg, J. (2012). Working memory capacity and visual-verbal cognitive load modulate auditory-sensory gating in the brainstem: toward a unified view of attention. *J. Cogn. Neurosci.* 24, 2147–2154. doi: 10.1162/jocn_a_00275

Spence, C., and Read, L. (2003). Speech shadowing while driving on the difficulty of splitting attention between eye and ear. *Psychol. Sci.* 14, 251–256. doi: 10.1111/1467-9280.02439

Stanislaw, H., and Todorov, N. (1999). Calculation of signal detection theory measures. *Behav. Res. Methods Instrum. Comput.* 31, 137–149. doi: 10.3758/bf03207704

Talsma, D., Doty, T. J., Strowd, R., and Woldorff, M. G. (2006). Attentional capacity for processing concurrent stimuli is larger across sensory modalities than within a modality. *Psychophysiology* 43, 541–549. doi: 10.1111/j.1469-8986.2006.00452.x

Talsma, D., and Kok, A. (2001). Nonspatial intermodal selective attention is mediated by sensory brain areas: evidence from event-related potentials. *Psychophysiology* 38, 736–751. doi: 10.1111/1469-8986.3850736

Tombu, M. N., Asplund, C. L., Dux, P. E., Godwin, D., Martin, J. W., and Marois, R. (2011). A unified attentional bottleneck in the human brain. *Proc. Natl. Acad. Sci. U S A* 108, 13426–13431. doi: 10.1073/pnas.1103583108

Unsworth, N., and Robison, M. K. (2016). Pupillary correlates of lapses of sustained attention. *Cogn. Affect. Behav. Neurosci.* doi: 10.3758/s13415-016-0417-4 [Epub ahead of print].

Vachon, F., and Tremblay, S. (2008). Modality-specific and amodal sources of interference in the attentional blink. *Percept. Psychophys.* 70, 1000–1015. doi: 10.3758/pp.70.6.1000

Vitevitch, M. S. (2003). Change deafness: the inability to detect changes between two voices. *J. Exp. Psychol. Hum. Percept. Perform.* 29, 333–342. doi: 10.1037/0096-1523.29.2.333

Vosskühler, A., Nordmeier, V., Kuchinke, L., and Jacobs, A. M. (2008). OGAMA (Open Gaze and Mouse Analyzer): open-source software designed to analyze eye and mouse movements in slideshow study designs. *Behav. Res. Methods* 40, 1150–1162. doi: 10.3758/BRM.40.4.1150

Wahn, B., and König, P. (2015). Audition and vision share spatial attentional resources, yet attentional load does not disrupt audiovisual integration. *Front. Psychol.* 6:1084. doi: 10.3389/fpsyg.2015.01084

Wickens, C. D. (1980). "The structure of attentional resources," in *Attention and Performance VIII* (Vol. 8), ed. R. Nickerson (Hillsdale, NJ: Erlbaum), 239–257.

Wickens, C. D., and Alexander, A. L. (2009). Attentional tunneling and task management in synthetic vision displays. *Int. J. Aviat. Psychol.* 19, 182–199. doi: 10.1080/10508410902766549

Wilming, N., Harst, S., Schmidt, N., and König, P. (2013). Saccadic momentum and facilitation of return saccades contribute to an optimal foraging strategy. *PLoS Comput. Biol.* 9:e1002871. doi: 10.1371/journal.pcbi.1002871

Wood, N. L., and Cowan, N. (1995). The cocktail party phenomenon revisited: attention and memory in the classic selective listening procedure of Cherry (1953). *J. Exp. Psychol. Gen.* 124, 243–262. doi: 10.1037/0096-3445.124.3.243

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Causse, Imbert, Giraudet, Jouffrais and Tremblay. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution and reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Human Neuroscience

REVIEW
published: 30 June 2016
doi: 10.3389/fnhum.2016.00290

From Trust in Automation to Decision Neuroscience: Applying Cognitive Neuroscience Methods to Understand and Improve Interaction Decisions Involved in Human Automation Interaction

Kim Drnec[1], Amar R. Marathe[1], Jamie R. Lukos[2] and Jason S. Metcalfe[1]*

[1] Human Research and Engineering Directorate, U.S. Army Research Laboratory, Aberdeen, MD, USA, [2] Advanced Concepts and Applied Research Branch, Space and Naval Warfare Systems Center Pacific, San Diego, CA, USA

Human automation interaction (HAI) systems have thus far failed to live up to expectations mainly because human users do not always interact with the automation appropriately. Trust in automation (TiA) has been considered a central influence on the way a human user interacts with an automation; if TiA is too high there will be overuse, if TiA is too low there will be disuse. However, even though extensive research into TiA has identified specific HAI behaviors, or trust outcomes, a unique mapping between trust states and trust outcomes has yet to be clearly identified. Interaction behaviors have been intensely studied in the domain of HAI and TiA and this has led to a reframing of the issues of problems with HAI in terms of reliance and compliance. We find the behaviorally defined terms reliance and compliance to be useful in their functionality for application in real-world situations. However, we note that once an inappropriate interaction behavior has occurred it is too late to mitigate it. We therefore take a step back and look at the interaction decision that precedes the behavior. We note that the decision neuroscience community has revealed that decisions are fairly stereotyped processes accompanied by measurable psychophysiological correlates. Two literatures were therefore reviewed. TiA literature was extensively reviewed in order to understand the relationship between TiA and trust outcomes, as well as to identify gaps in current knowledge. We note that an interaction decision precedes an interaction behavior and believe that we can leverage knowledge of the psychophysiological correlates of decisions to improve joint system performance. As we believe that understanding the interaction decision will be critical to the eventual mitigation of inappropriate interaction behavior, we reviewed the decision making literature and provide a synopsis of the state of the art understanding of the decision process from a decision neuroscience perspective. We forward hypotheses based on this understanding that could shape a research path toward the ability to mitigate interaction behavior in the real world.

Keywords: trust in automation, interaction decisions, decision making, human automation interaction, neuroergonomics

OPEN ACCESS

Edited by:
Klaus Gramann, Berlin Institute of Technology, Germany

Reviewed by:
Agnieszka Wykowska, Ludwig-Maximilians-Universität, Germany
Dietrich Manzey, Technische Universitaet Berlin, Germany

****Correspondence:***
Kim Drnec
kdrnec@gmail.com;
kim.a.drnec2.ctr@mail.mil

Received: *07 November 2015*
Accepted: *30 May 2016*
Published: *30 June 2016*

Citation:
Drnec K, Marathe AR, Lukos JR and Metcalfe JS (2016) From Trust in Automation to Decision Neuroscience: Applying Cognitive Neuroscience Methods to Understand and Improve Interaction Decisions Involved in Human Automation Interaction. Front. Hum. Neurosci. 10:290. doi: 10.3389/fnhum.2016.00290

INTRODUCTION

The purpose of this review is to address a largely unexplored aspect of human automation interaction (HAI); that is, the human decision that leads to interaction behavior, traditionally considered a manifestation of the user's level of Trust in Automation (TiA). The extension of this concept has been that, if HAI is to be actively managed in joint human-automation systems, one must calibrate the TiA of the user so that decisions about interactions with automation are appropriate. Further, it has been considered that if one could measure instantaneous levels of TiA, inappropriate interaction decisions could be predicted and mitigated. Research interest in HAI systems is motivated in large part because of observations that even the most advanced HAI systems have not yet fully realized the ultimate vision of both safe and seamless integration of the human into the system that would lead to improved task performance. Specifically, successful applications of automation within task spaces involving human operators have not yet been realized without simultaneous definition of significant context-specific design constraints that delineate human and automation responsibilities. Such constraints may improve focused aspects of performance, but also increase the risk in other ways, particularly in circumstances and moments involving handoff of control authority, and these constraints limit more generalized application of HAI concepts and methods, particularly in terms of improving joint system efficiency (Parasuraman and Riley, 1997; Dekker and Woods, 2002; Dzindolet et al., 2003; Jamieson and Vicente, 2005; Parasuraman and Manzey, 2010).

Decades of human factors research have resulted in an understanding of what factors affect TiA, but as of yet, it remains unclear how specific levels of TiA translate into specific human decisions regarding interaction with a given automation. This knowledge gap may exist because human behavior and joint system performance can be thought of as the result of a combination of many factors only one of which is TiA (Hancock et al., 2011; Schaefer et al., 2014). Research aimed at predicting interaction behaviors has previously met with some success, particularly with respect to decision aid systems (Bliss et al., 1995; Meyer and Bitan, 2002; Meyer et al., 2014) and automated driving aids (Kumagai et al., 2003; Gold et al., 2015; Terai et al., 2015). We consider such results to suggest that understanding the interaction behavior may be a more fruitful and immediate route toward active, online mitigation of problems thought to arise from mis-calibrated TiA. This idea is developed with the appreciation that interaction behaviors result from decisions about how and when to interact, and any individual interaction decision may or may not be motivated by a change in TiA.

Our specific proposal is that, as much as behavior is a key to managing HAI, understanding the process of decision-making in the context of HAI is critical for understanding and predicting interaction behaviors. It is an important step that is implicitly necessary for eventual online mitigation of inappropriate interaction behavior. This is especially applicable in our discussion inasmuch as we believe that TiA reflects changing degrees of perceived risk and uncertainty and is an instance of value based decision making in dynamic contexts. We further suggest that physiological correlates of value based decisions could be measured and leveraged to provide valuable data that may increase the likelihood of predicting a consequent interaction. To develop the connection between TiA and value based decision making, the discussion begins by reviewing extant human factors literature to demonstrate that, while TiA is one of many important factors influencing HAI performance, it is ultimately the interaction behavior that is of interest. It is then argued that this behavior, if intentional, is the result of a decision, and thus understanding the decision process leading to the behavior may facilitate near- to medium-term solutions for active mitigation strategies while understanding of the nuances and complexities of TiA continues to evolve over the long term. The discussion then turns towards a synthesis of selected cognitive neuroscience literature that focuses particularly on value based decision making. We conclude with future research directions that would be necessary to enable decision based monitoring and prediction of interaction behaviors and the eventual development of active mitigations for the types of HAI problems currently believed to be brought about by mis-calibrated TiA.

THE IMPORTANCE OF TiA IN JOINT SYSTEM PERFORMANCE

The term automation, or automated system, as used here is best defined by Parasuraman et al. (2000) as a "machine execution of functions". This definition includes automation with capabilities as diverse as controlling a sophisticated cockpit system or as simple as an automated coffee maker. Because automation is not yet fully "intelligent" it has no agency for adapting to unexpected circumstances, and therefore often requires the supervision and/or occasional intervention of humans. Part of this supervisory role requires that there be HAIs, but these interactions need to be appropriate, or the joint system performance will suffer. Decades of HAI system research have indicated that appropriate interactions are the result of decisions subsequent to calibrated TiA. An established conceptual model of factors influencing HAI performance is provided in **Figure 1**. As the model suggests, TiA has traditionally been considered to be the critical component driving human user decisions about interactions such as intervening in an automated task. Given the importance that TiA has been accorded to overall joint system performance, in this section we provide a brief review of important aspects of TiA and its dynamics. We aim to highlight the complex relationship between TiA and human user behavior, and implications arising from this relationship that imply that even if moment to moment levels of TiA were to be measured, it is unclear how such information could be leveraged to predict an interaction behavior.

Early theories about the construct of TiA were developed from the psychological construct of interpersonal trust, and they posited that calibrated TiA was critical for successful HAI system performance (Sheridan, 1980; Sheridan and Hennessy, 1984). There are aspects of interpersonal trust that are analogous to TiA,

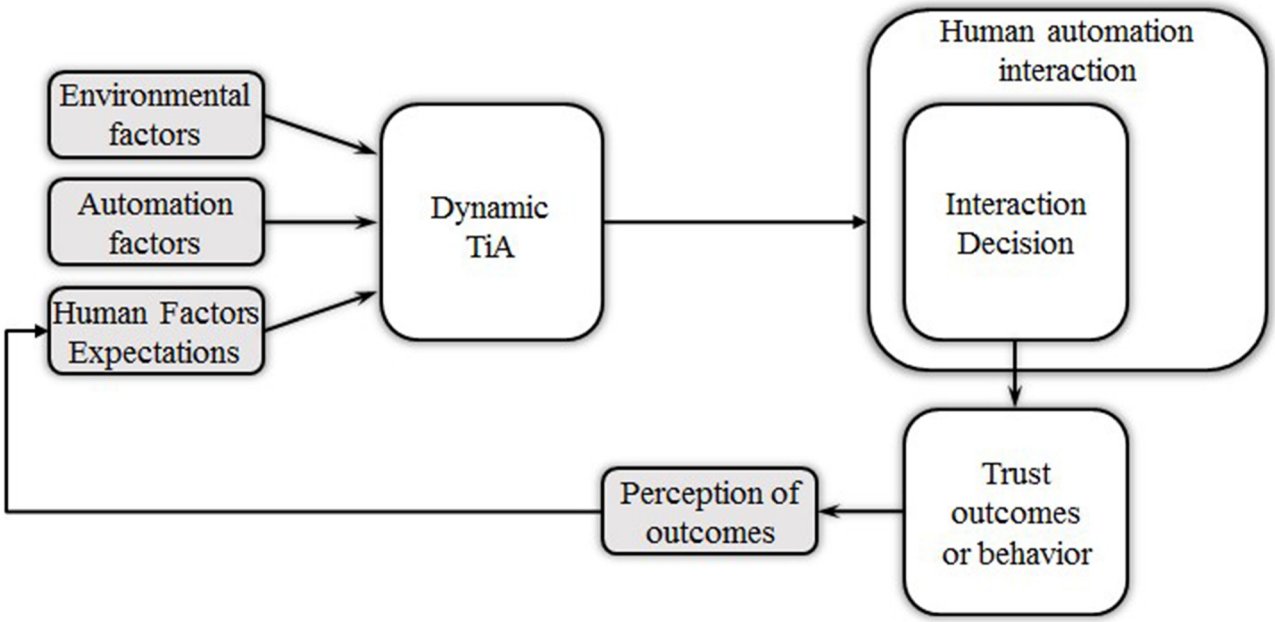


FIGURE 1 | A conceptual organization of trust and human user-automation interaction (Adapted from Hancock et al., 2011 with permission from Sage Publishing). This article focuses on interaction decisions that are part of the overall human automation interaction (HAI).

in particular that there needs to be a sense of risk or vulnerability on the part of the trustor for trust to develop (Lee and Moray, 1994; Muir, 1994; Corritore et al., 2003; Lee and See, 2004; Evans and Krueger, 2011). However, it has been debated whether the two constructs are homologous (Madhavan and Wiegmann, 2007), and so trust as it specifically applies to automation became a central point of interest in human factors research aimed at improving joint system performance (Lee and Moray, 1992; Muir, 1994; Muir and Moray, 1996; Lee and See, 2004). Myriad definitions of TiA imply that it is the result of a feeling of trustworthiness towards the automation such that a human user can depend on the automation to perform the task for which it was designed. It is worth noting that if the consequence of the task to the human user is small, if TiA develops at all, its level becomes irrelevant because the outcome of the joint system fails to be important. Therefore, much like interpersonal trust (Lee and Moray, 1994; Muir, 1994), TiA develops in the face of a sense of risk. In these situations, TiA then develops and shows dynamic changes from the ongoing comparison of the expectations about the automation's behavior and observations by the human user about the automation's performance weighted heavily on the risk borne by the human user (Sheridan and Hennessy, 1984; Muir, 1994; Muir and Moray, 1996).

Determinants and Dynamics of Trust in Automation: Expectations and Observations

One of the first explicit theories of TiA (Muir, 1994) stated that appropriate levels of TiA would develop if three expectations were met during the course of automation interaction. These expectations are technical competence, persistence, and fiduciary responsibility, but they play differential roles in TiA development and dynamics throughout the course of automation use. For example, perceptions of competence might be more important in the early stages of automation use than later in time. The expectation of technical competence is the expectation that the automation will accurately and successfully perform the functions for which it was designed. Persistence, perhaps here better conceived of as predictability, relates to the issue of reliability in that an automation that performs in a particular manner now will be expected to perform in a same or similar manner when it encounters similar circumstances in the future. Finally, fiduciary responsibility addresses the notion that a given human user will hold expectations of an automation of a particular type that will impact role allocation. That is, the human user will expect that the automation will necessarily be responsible for its designed functions as they understand them and thus fewer personal resources need to be allocated to carrying out those functions. The importance of these expectations related to TiA dynamics differ depending on the stage of the interaction with the automation.

When first presented with an automated system there is limited information available for the human user to observe, and thus little with which to evaluate the trustworthiness of the automation. Some key elements that significantly affect early levels of TiA include initial expectations borne from biases toward automations in general, and initial observations about the design of the automation (Muir and Moray, 1996; Nass et al., 1996; Dzindolet et al., 2002; Lee and See, 2004; Parasuraman and Miller, 2004; Miller, 2005; Merritt and Ilgen, 2008; Merritt, 2011; Merritt et al., 2012; Pak et al., 2012). After

having been introduced to an automation, human users tend to explore different strategies for subsequent interaction (Lee and Moray, 1992) and thus learn more about the automation's behavior. This experimentation arguably helps the human user gauge competence, which emerges as one of the most important predictors of TiA at this early stage. However, it is worth noting that for various reasons, human users are notoriously poor at making accurate judgments of competence (Sheridan and Hennessy, 1984; Lee and Moray, 1992; Dzindolet et al., 2002, 2003; Madhavan et al., 2006; Verberne et al., 2012; Merritt et al., 2014). Once automation competence has been judged, whether correct or not, the most important factor driving levels of TiA is persistence or predictability of performance over time (Lee and Moray, 1992). Persistence of performance is important enough that as long as errors are predictable and the automation error rate is at a consistent rate of approximately 30% or less, most human users will decide to continue to use and benefit from the automation (Parasuraman et al., 2000; Wickens and Dixon, 2007; Wang et al., 2009). As levels of TiA dynamically change throughout the course of observations about the automation's behavior, theory posits that interaction decisions, and consequent behaviors, should reflect the extant level of TiA. If there is too much or too little TiA, as it goes, a human user may decide to overuse or underuse the automation, respectively. Specific patterns of behaviors resulting from decisions about how to interact with the automation have been well documented and are commonly referred to as trust outcomes as they are believed to directly reflect certain levels of TiA.

Trust Outcomes and their Relationship to TiA

The trust outcomes most commonly discussed are misuse and disuse and are described in detail by Parasuraman and Riley (1997). Misuse refers to instances when the automation is used without undo skepticism, tending to result in overuse (Parasuraman and Riley, 1997; Bahner et al., 2008; Parasuraman and Manzey, 2010). Misuse has two related causes; automation bias and complacency (Manzey et al., 2006; Parasuraman and Manzey, 2010). They are related in that they both result in a lack of monitoring where lack of attention plays a central role (Parasuraman and Manzey, 2010). Automation bias arises through the mere presence of an automated system, possibly because humans demonstrate a tendency to choose the route of least cognitive effort, making it easier, or at least preferable, to accept that feedback from an automation as correct (Dzindolet et al., 1999, 2001; Skitka et al., 1999, 2000; Wang et al., 2008; Parasuraman and Manzey, 2010; Goddard et al., 2014; Mosier and Skitka, 1996). Complacency, less well understood, can be said to occur when monitoring is less than optimal and joint system performance suffers (Parasuraman and Manzey, 2010). However, both automation bias and complacency tend to increase in cases of high workload and high consequence environments wherein users often make conscious decisions to rely on even imperfect automation (Dixon et al., 2007; Wickens and Dixon, 2007). Disuse describes a continuum that spans from the user underutilizing the automation to entirely abandoning the automation in favor of a manual mode. Disuse tends to occur if a human user has a high expectation of automation performance and then observes unexpected errors or has more self confidence in her ability to perform the task than confidence in the efficacy of the automation for the same task (Lee and Moray, 1994; Parasuraman and Riley, 1997; Moray et al., 2000; Dzindolet et al., 2003).

Although trust outcomes have been well defined, a synthesis of the literature, and recent experimental evidence (Wiczorek and Manzey, 2010; Chancey et al., 2015) indicate a far more complex relationship between TiA and trust outcomes than is implied in the above discussion, and this implies that predicting interactions based on extant TiA levels is problematic. Such complex interactions involve perceived risk, self-confidence, workload, and even personality type (Lee and Moray, 1994; Muir, 1994; Parasuraman and Riley, 1997; Lee and See, 2004; Merritt and Ilgen, 2008; Hancock et al., 2011; Schaefer et al., 2012; Merritt et al., 2014). For instance, human users have been documented as reporting a high level of TiA and then, paradoxically choosing a manual operation mode, demonstrating disuse (Lee and Moray, 1992). Conversely, even when low TiA has been reported, human users may misuse even a poorly competent automation, particularly under high workload conditions (Daly, 2002; Biros et al., 2004). Clearly levels of TiA do not map uniquely onto trust outcomes, regardless of how they are represented (i.e., attention, intervention rate, etc.). Therefore, they are not predictive of the way a human user will decide to interact with an automation, limiting the use of measuring TiA for real world applications to improve HAI. We suggest that this is because TiA is far more complex than may be useful for those with more immediate concerns regarding actively managing HAI. However, it is important to note that when TiA is studied it is the interaction behavior that is of interest most often. Therefore, regardless of the manageability of trust, what might be learned if we focus more simply and exclusively on the behavior?

TiA as Predictable Behavior

The present discussion is not the first to offer that a shift in focus from trust to behavior is well justified. In fact, a number of researchers in this domain have re-framed the problem space of TiA into one of reliance and compliance, which are defined exclusively in terms of observable behavior, and are not intended to imply specific psychological cause such as trust (Meyer, 2001, 2004; Parasuraman et al., 2008; Rice, 2009; Meyer et al., 2014). Indeed, there is a non-unique mapping of reliance and compliance to traditional trust outcomes such that an observation of inappropriate amounts of either may alternately signal disuse or misuse and possibly motivate conflicting interpretations of TiA. We find these behaviorally defined terms to be useful in their functionality for application in real world situations. That is, objectively defined and observable behaviors are especially valuable for the purposes of modeling and prediction because they obviate the need for drawing inferences to and making assumptions about manifestations of more subjectively defined constructs, such as TiA, automation bias or complacency that are difficult to measure objectively and

thus unsuitable for use in attempts at active optimization and/or mitigation.

Reliance is the tendency of the human user to accept the lack of an alarm, alert, warning, or prompt as a true reflection of the state of the world (Lee and Moray, 1994; Singh et al., 1997; Parasuraman et al., 2000; Yeh and Wickens, 2000; Moray, 2003; Dixon et al., 2006). That is, in the absence of an alarm or warning, the human user accepts, often tacitly, that all is well and there is no reason for possible intervention. Compliance, on the other hand, is defined when the user responds to, putatively agrees with, and ultimately takes the action specified by an alarm or recommendation from the automation (Meyer, 2001, 2004). Though reliance and compliance are often discussed in terms of optimal behavior, too much of either in the wrong context is detrimental to system performance. For instance, if an alarm is absent and the human user assumes that no circumstances warranting an alarm exist and thus fails to monitor the automation over significant time, he or she is at risk of over-reliance and the consequences thereof. Conversely, over-compliance occurs when the human accepts all suggestions from the automation (when present) without confirming their validity.

Beyond observation of general behavioral patterns, the greater benefit of defining compliance and reliance behaviors has been in providing an avenue towards greater precision in understanding the factors that affect automation use during HAI, which might eventually lead to prediction of an interaction. For instance, some have observed that reliance and compliance are differentially affected by error type, i.e., false alarms vs. misses in target detection tasks (Meyer, 2004; Rice and Geels, 2010; Wiczorek and Manzey, 2014), and by the predictive value of the alarm (Meyer and Bitan, 2002; Manzey et al., 2014). If a human user observes frequent failures to trigger alarms, the frequency of monitoring the automation will increase, thereby reducing reliance on the automated agent (Masalonis and Parasuraman, 1999; Bagheri and Jamieson, 2004; Meyer, 2004; Madhavan and Wiegmann, 2007; Parasuraman and Manzey, 2010; Geels-Blair et al., 2013). Compliance, however, is degraded by higher rates of false alarms. In particular, when higher rates of false alarms are observed, users tend to consume critical time and attentional resources to verify alarms before choosing a response. Further progress in this line of inquiry has resulted in more general characterization of how interaction behaviors change with the positive and negative predictive value of an alarm. Positive predictive value is derived from a Bayesian calculation of the likely existence of a hazard given an alarm and, likewise, the probability of an alarm given no existing hazard (Meyer et al., 2014). Negative predictive value is calculated similarly, but in the absence of an alarm. Therefore, positive predictive value decreases as false alarm rate goes up and negative predictive value decreases with more frequent misses (no alarm in the presence of a hazard). Interaction behavior is thus differentially affected by changing positive vs. negative predictive value. Positive predictive value has been shown to have strong effects on reliance, but only for values less than 0.75 (Meyer et al., 2014), where values below this threshold have been associated with excessive time spent monitoring the automation. Research in HAI domains has thus befitted considerably from the use of these narrowly and objectively defined behavioral terms.

We advocate here for shifting research towards more clearly defined behaviors and the factors that affect them because of how this shift creates important opportunities for systematic research into HAI. The domain of application for such a shifted focus would include contexts where TiA may be involved, at least inasmuch as TiA reflects assessments of the relative value of specific behavioral options defined in terms of probable risk versus reward. We argue that such behavior-based understandings are important for progress on multiple levels from phenomenology to predictive modeling. The extant work discussed above has provided an essential corpus of knowledge regarding the relationship between automation performance characteristics (i.e., error rate, type, and predictive value) and human user interaction behaviors. However, we also suggest that in order to be useful down the road for real-world mitigation of inappropriate interactions, this shift from trust to behavior does not go quite far enough for two important reasons. First, to mitigate a potentially detrimental interaction behavior in a dynamic context, prediction is necessary. This is because a behavior that has already occurred cannot be changed and the consequences are likely to be too immediate to offset in *post hoc* fashion. Moreover, the predictive power required must occur on a time-scale that allows a reasonable opportunity to enact a mitigation when an inappropriate behavior is expected. Second, the current understanding of reliance and compliance is tied to automation design; an automation that frequently misses events reduces reliance, and an automation that frequently produces false alarms reduces compliance. This understanding, then, usefully provides an improved framework for HAI, but has yet to account for variability in individual instances of HAI. Therefore, the predictive power of the current understanding of interaction behaviors based on population averages remains limited to overall design strategies whereas we are interested in building towards eventual prediction and mitigation of reliance and compliance at the level of individual instances of interaction behavior. In order to improve the ability to predict an interaction behavior we thus believe it is necessary to consider not only the effects of automation design on interaction behaviors such as compliance and reliance, but also the individual internal phenomena that precedes the behavior.

INTERACTION BEHAVIOR REFRAMED AS A DECISION TO INTERACT WITH AN AUTOMATION

Before an intentional behavior occurs, the human user must make a decision as to which among a limited array of options will be selected. Here we argue that research concerned with improving HAI would benefit greatly from studying the decisions that precede interaction behaviors. Such an approach satisfies the need for focus on individual interactions in a manner that affords prediction on a time-scale that is useful for active mitigation. We define such decisions as interaction decisions, given as specific to the intention to interact with an automation.

While TiA has often been considered to motivate interaction decisions, the richness of the decision process itself, as well as accompanying stereotypical psychophysiological indicators thereof, has not been thoroughly investigated as a source of information that could be applied to the prediction of a consequent interaction behavior. Our starting point in this pursuit is to understand the underlying psychological and physiological processes of decision making, with a particular focus on value based decision making. This understanding can provide a cornerstone for the advancement of scientifically based hypotheses about how interaction behaviors may eventually be predicted for the sake of active mitigation. Predicting decision outcomes, or the interaction behavior in a real-world HAI context, is of course not trivial, but laboratory-based research in decision neuroscience has established decision making as a reasonably stereotyped process with clear behavioral and physiological precursors. Further, the ability to predict decision outcomes has been pursued by both the cognitive neuroscience (Soon et al., 2008; Haynes, 2011; Perez et al., 2015) and brain computer interface (Musallam et al., 2004) communities. Indeed, attempts at predicting some types of decision outcome behaviors have already met with success in a laboratory environment, possibly because the specifics of a decision process in the brain begin even before there is conscious awareness of the impending decision (Soon et al., 2008, 2013; Haynes, 2011; Perez et al., 2015). Some of these studies have been criticized because there lacks a sense of risk or value to the decision maker in a controlled experiment, and therefore, the assumption is that the decision outcomes that are being predicted are trivial (Gold and Shadlen, 2007; Lavazza and De Caro, 2010).

The lack of risk, value, or reward in these controlled laboratory environments is in contrast to interaction decisions that inherently involve some type of personal risk or reward. For example, over relying on an automation can compromise joint system performance, and therefore causes degradations in joint system performance. Thus, we are chiefly concerned with decisions that are based on expected value and risk, or value based decisions (Rangel et al., 2008; Wallis, 2012). Value based decisions are particularly relevant to the HAI context because it is often required that a human user continuously weigh the expected personal value of allowing the automation to complete the task versus performing it manually or, rather, whether to comply with the recommendation of an automated system. Importantly, this assessment and subsequent judgment of value to the user must be made against the backdrop of risk that the decision may compromise joint system performance. Thus, through the common elements of risk, reward, and expected value, we believe interaction decisions during HAI to be an instance of value based decision making. We believe that understanding the value based decision process is important to improving HAI and, therefore, we briefly discuss results of value based decision making research as it relates to HAI in order to support hypotheses forwarded in the discussion, aimed at establishing a research path that will allow the eventual prediction of HAI interaction behaviors.

The Importance of Considering the Decision Process

An important argument in favor of studying the decision process in order to improve HAI is that significant efforts in cognitive neuroscience have revealed decision making as fairly stereotyped, and therefore, a potentially predictable process. Moreover, this body of research has identified a number of psychophysiological correlates that unfold in advance of, and during a decision. Critically, these correlates are measurable, and therefore useful for understanding the decision process, at least in laboratory settings. Some have observed that these correlates unfold in predictable ways through defined cognitive stages, and therefore measuring them has potential use for active mitigations of inappropriate interaction decisions and behavior. This approach is fundamentally different than attempting to measure and calibrate TiA because the psychophysiological correlates of a decision are measurable whereas the construct of TiA is yet to be defined in a way that is equally useful for active monitoring. In general, many cognitive neuroscientists model decisions as comprising three cognitive stages (Fellows, 2004; Bogacz, 2007). However, five cognitive processes, some analogous to stages of the cognitive models of general decisions, have been described in value based decision making (Rangel et al., 2008) and are therefore relevant to our discussion of interaction decisions. These processes, which are not discrete stages *per se*, are: (1) representation of the problem, i.e., identification of alternative choices, and of internal and external states that affect the value of the choices; (2) evaluation of gathered evidence that allows the assignment of a value to the alternatives; (3) comparison of these values in order to make a decision; (4) accumulation of the comparative value for each alternative and making the decision; and (5) generation of prediction errors that provide feedback in order for learning to occur. The psychophysiological processes that unfold within the first four processes will be discussed as they relate to HAI contexts such as risk and reward. The fifth process, generating feedback on the decision has been studied in the context of learning, and may be useful for later development of adaptive mitigation strategies, but is beyond the scope of the current review (Nieuwenhuis et al., 2005; Christie and Tata, 2009; Cohen et al., 2011; van de Vijver et al., 2011). We note that we discuss these processes sequentially mainly for organizational purposes, however, during the decision process they may overlap or even occur in parallel (Rangel et al., 2008).

The Value Based Decision Process

In order for the need to identify alternatives to arise, there must be some recognition of the need for a decision; in a sense it is the motivation to perform a task (Gold and Shadlen, 2007). Decisions must be initiated by either salient external or internal stimuli. These stimuli will often produce an orienting response (Sokolov et al., 2002; Glimcher and Rustichini, 2004; Delgado et al., 2005), characterized in humans by a measurable increase in tonic skin conductance (SC) levels and a decrease in heart rate variability (Figner and Murphy, 2011). In an HAI domain, relevant stimuli typically include those specifying alerts from the automation, acute changes in environment,

or internal feelings that the current behavior is inappropriate (typically seen as an error-related potential in the brain or a gradual shift in peripheral physiology). Once the need for a decision has been established, however, decision alternatives are identified. As alternatives are identified, in the case of interaction decisions, the human user will also identify, if not consciously, a representation of internal and external states (Rangel et al., 2008). These representations play an important part during the process of assigning values to individual alternatives. For example, a human user is more likely to take control from the automation if they detect that the automation is malfunctioning and they perceive an associated risk. The neural basis of this early stage in the decision process is not well understood. For example, it is unclear how the brain decides which alternatives should be considered, and if there is a functional limit to the number that can be assessed at one time (Rangel et al., 2008). Nevertheless, such questions are important for determining how to leverage physiological indicators into models of decision making during HAI.

Once the possible alternatives are identified evidence for or against each alternative must be evaluated in order to make an optimal decision. In the case of interaction decisions, which due to the presence of risk are analogous to value based decisions, it has been hypothesized by some (Rangel et al., 2008; Glimcher and Fehr, 2013) Reading hidden intentions in the human brain cognitive valuation systems that the brain might use; the Pavlovian, Habitual, and Goal directed. We believe the goal directed system to be most relevant to our discussion because the goal directed system assigns values to potential actions by calculating action-outcome associations from previous experience and comparing this value to the perceived rewards associated with possible outcomes of the decision (Rangel et al., 2008). In the goal directed valuation system the value assigned to a piece of evidence is equivalent to the potential value of the alternative it supports, with the value assigned to an alternative being equal to the expected reward of the action. In the context of HAI, an important research question would be whether the probability of success is greater by relying on the automation or not and, moreover, how that probability scales with perceived risk to determine the direction of a given interaction decision.

When a person valuates a piece of evidence they will do so by observing the relevant data (e.g., visual scanning, sound or other stimulus), consulting their memory, and integrating this against a backdrop of expectations (Mulder et al., 2014). In the case of visual evidence, gaze fixation is thought to support evidence evaluation such that evidence about the value of an alternative is sampled at each fixation (Krajbich et al., 2010; Krajbich and Rangel, 2011). Memory consultation causes a person to compare past decision outcomes with the available alternatives. The brain creates a prediction error that would represent the difference between the expected value of choosing current decision alternatives from the value that has been experienced in the past by choosing alternatives that are similar in nature (Hare et al., 2008). For example, consider a human user who has previously experienced aberrant behavior from an automation, but there has been no decrement in joint system performance, and joint system performance continues to remain better than what would be expected from only one agent performing the task. Even in risky environments such as in the battlefield, or during a search and rescue operation, the experienced user is more likely to rely on an automation (Lyons and Stokes, 2012) than a user who has not experienced the aberrant behavior because the experienced user has realized the value in relying on the automation despite the probability of an error.

At a cellular level, data have suggested that the cognitive evaluation of evidence is supported by neural "evaluators" that store dynamic estimates of which decision alternative is supported by the evidence. For instance, studies using fMRI in risk reward scenarios have identified two candidate neural evaluators; the amygdala and ventral striatum. A reward based fMRI study indicated that the amygdala evaluates the cost or risk of acting on an alternative (Yacubian et al., 2006; Basten et al., 2010). In the same fMRI study the ventral striatum was implicated in the formation of representations of the expected value or reward of an alternative (Yacubian et al., 2006; Kable and Glimcher, 2007; Rangel et al., 2008; Basten et al., 2010; Lim et al., 2011). Other authors, however, have found that in addition to the amygdala and the ventral striatum that the lateral orbitofrontal cortex and the medial orbital frontal cortex also act as neural evaluators for risk and reward, respectively (Hare et al., 2008; Rangel et al., 2008; Rangel and Hare, 2010). The neural substrates that have been observed support value based decision making processes are detailed in **Table 1**.

The neural evaluators, then, form representations of the risks and rewards for each alternative; the benefit of relying or complying with an automation, as opposed to choosing to complete the task manually. As the risks and rewards of a potential interaction behavior are processed by the amygdala and ventral striatum, the value of these representations must be assessed relative to each other; they must be compared. Neural correlates of this third process involved in value based decision making, comparison of the values assigned to the evidence, have been observed in fMRI studies. That is, value based comparison has been suggested as supported by activation in the ventral medial prefrontal cortex (vmPFC; Chib et al., 2009; Gläscher et al., 2009; Basten et al., 2010), whereas the "comparator" function in perceptual decisions has been associated with increased activity in the dorsolateral prefrontal cortex (dlPFC; Basten et al., 2010; Philiastides et al., 2011). While the evidence comparison process unfolds, some have hypothesized that the comparative value, also known as the decision variable, is accumulated in the lateral intraparietal cortex (LIP) until a decision threshold is reached, bringing about a decision (Platt and Glimcher, 1999; Kiani and Shadlen, 2009; Mulder et al., 2014). Evidence to support this hypothesis has mainly been shown in primate studies of single cell recordings during cued saccade trials (Platt and Glimcher, 1999; Platt, 2002). However, there has been evidence from fMRI studies that the human parietal cortex is also involved in accumulating the decision variable (Ploran et al., 2007; Heekeren et al., 2008). It is interesting to note that the temporal integration of activity in the frontal-parietal regions, which are considered to be involved in comparing and accumulating compared value

TABLE 1 | Neural substrates and their putative role in decision making.

Neural substrate	Role in decision making	Reference
Amygdala	Processes/computes the value of negative stimuli	Yacubian et al. (2006) and Basten et al. (2010)
Ventral striatum	Processes/computes the value of positive stimuli	Yacubian et al. (2006), Basten et al. (2010) and Lim et al. (2011)
Ventral medial prefrontal cortex (vmPFC)	Calculates the difference of value signals from amygdala and ventral striatum in value based decisions	Basten et al. (2010) and Philiastides et al. (2011)
Dorsolateral prefrontal cortex (dlPFC)	Calculates the difference of signals from amygdala and ventral striatum in perceptual decisions	Basten et al. (2010) and Philiastides et al. (2011)
Lateral intraparietal cortex (LIP)	Accumulates and integrates the value of evidence processed by the vmPFC (evidence largely from monkeys)	Platt and Glimcher (1999); Platt (2002); Basten et al. (2010) and Rorie et al. (2010)
	A cortical area involved in gaze fixation, saccade, and attention, underlying evidence accumulation	Coe et al. (2002) and Goldberg et al. (2006)

signals, has been observed as preceeding the conscious decision to act (Gold and Shadlen, 2007; Soon et al., 2013; Perez et al., 2015).

The putative involvement of the parietal cortex in decision making is noteworthy because of its central role in the process. For example, gaze fixation, critical for evidence evaluation (Poole and Ball, 2006) in visually based decisions, is controlled by the LIP in monkeys (Coe et al., 2002). This region forms a "salience map" for the oculomotor system to saccade to a target, or maintain gaze fixation on a target (Goldberg et al., 2006). The LIP then, not only plays a role in accumulating the comparative value of the evidence as discussed above, but is critical for its initial evaluation. Brain computer interface research has also found that the medial intraparietal cortex in monkeys forms representations of the value of an alternative that has been encoded in the vmPFC (Musallam et al., 2004), such that the intent of the monkey to choose one alternative over an other can be decoded from intracellular electrodes. Although much evidence of the importance of the parietal cortex during decision making, and especially value based decision making, has come from primate research, there is evidence that analogs in the human parietal cortex are also central to decision making (Ploran et al., 2007; Heekeren et al., 2008).

Despite the hypothesized causal role of integrated frontal-parietal activity in the conscious decision to act, there are no measurable psychophysiological variables that allow an accurate determination of the exact time that a decision threshold is reached. However, psychophysiological correlates occurring hundreds of milliseconds before a conscious decision involving risk and reward (Cohen et al., 2009), inherent in value based decision making, have been identified. For instance, the readiness potential, a slow negativity in scalp recording of cortical activity precedes fully endogenous decisions by a few hundred milliseconds (Libet, 1993). Even more proximal to the decision spectral correlates have been observed. In a paradigm involving playing a competitive game against a computer, spectral decomposition of scalp-recorded EEG led to the finding that the decision process was accompanied by a general shift in power between lower bands (delta, 1–4 Hz and theta, 5–7 Hz) to higher frequency bands (alpha, 8–12 Hz and beta, 13–35 Hz), as well as a broadband increase in cross-trial phase coherence at about 220 ms post stimulus (Cohen and Donner, 2013). Similar indications were found during complex real world choice tasks and a two-choice forced-decision paradigm. In these cases, significant correlations of increased power were seen in delta, theta, beta, and gamma (36 + Hz) bands of EEG activity approximately 250–500 ms post-stimulus (Guggisberg et al., 2007; Davis et al., 2011). In decisions involving risk, risk is represented by an asymmetry in the alpha band such that there is an increased alpha power in the right frontal region (Gianotti et al., 2009). These spectral correlates of value based decision making are measurable in real time and available for current application outside a laboratory, which is encouraging in the context of improving HAI. However, these scalp recorded spectral correlates occur only hundreds of milliseconds before an interaction decision and therefore have limited use because they are so temporally proximal to the behavior itself.

The proximity of these value based decision correlates to the actual decision may be discouraging in the context of predicting and mitigating interaction behavior. Nonetheless, research efforts in decision neuroscience and in brain computer interface have found in fMRI studies that the correlates of an outcome of a decision to move at a time chosen by the subject are measurable up to 7 s before the conscious awareness of the decision is reached. Moreover, through analysis of these correlates, the intended goal of the decision can be decoded before conscious awareness of it arises (Haynes et al., 2007; Soon et al., 2008, 2013; Haynes, 2011; Perez et al., 2015). In two fMRI studies (Haynes, 2011; Soon et al., 2013) subjects were asked to decide at will when to either press a button on their left or right side, or add or multiply a set of numbers, and then to report when they were consciously aware of the decision. Spatial pattern analysis of the blood oxygen level dependent signal, a measure of neural activation in fMRI studies, revealed that the frontal polar cortex appeared to encode the intentions of the subjects before they reported having made the decision. In a driving study using implanted EEG electrodes in human epilepsy patients a modulation of gamma power in the posterior lateral cortex predicted whether the subjects would turn left or right at an intersection before they consciously made the decision

(Perez et al., 2015). These studies made use of technology such as fMRI that as yet is not available for real world applications, unlike EEG, because of the need for the subject to lie still in the large, importable fMRI equipment. Further, although EEG is portable, it cannot directly measure activity in deep cortical and subcortical areas, and these are exactly the areas that showed activity prior to a conscious decision. However, the results are encouraging for making use of the interaction decision process to improve HAI and, moreover, can be leveraged into research aimed at developing models that capture relations between cortical and subcortical brain activations during value based decision making. Identification of such relations is an important avenue for future research aimed at active mitigation during HAI.

Indeed, decision neuroscience and brain computer interface research have facilitated the development of precise understandings of decision making that could facilitate the development of methods for identifying an interaction decision within the contextual space of HAI. Efforts in decision neuroscience as well as in more applied domains, such as neuroergonomics, have shown that there are clear and measurable behavioral and psychophysiological correlates (fMRI activation patterns, EEG, SC, gaze fixation, heart rate, etc.) of component processes that are antecedent to the decision. In addition, decision neuroscience has begun to provide an understanding of the underlying cortical and sub-cortical processes involved in decision making. While many of these processes have as yet only been identified by fMRI, their understanding will allow meaningful hypotheses to be advanced about measures that can be recorded in real time.

DISCUSSION

HAI systems have as yet to live up to their expectations, and one critical reason is that human users often make inappropriate decisions about how and when to interact with an automation. These interaction decisions have traditionally been considered to be motivated by extant levels of TiA. Therefore, if TiA can be measured, it is expected that it can be managed and inappropriate interaction behaviors could be mitigated. Given that substantive theory, it was appreciated that TiA is an important construct that undoubtedly affects human user interaction behavior, and hence we reviewed and synthesized the TiA literature. From that exercise, we observed that the relationship between TiA and human behavior is complex and not fully understood. Further, relevant to immediate real world applications for improving HAI system performance, TiA cannot be readily measured, and even if it were measurable in real time it is unclear how certain levels of TiA map onto specific interaction behaviors. By contrast, specific behaviors such as reliance and compliance are readily observed and measured in real time and do not have the confounding effect of inferring psychological causality. Such cause-agnostic variables are particularly attractive in HAI research aimed at defining concrete methods for improving joint system performance, both in terms of initial system design as well as for ultimate real-time applications.

Reframing the problem space of HAI and TiA as a problem of behavior, rather than of TiA, has been successful in allowing general predictions of interaction behavior based on knowledge of system design given specific environmental and internal conditions such as increased risk or increased workload, respectively. For example, knowing that an automation is prone to false alarms will allow the general prediction that a human user will often fail to comply with alerts. This understanding allows a system designer to set thresholds for alarms that are appropriate to intended use. For example, in high risk environments it may be better to set an alarm threshold low so that critical cases are not missed. This predictive ability has been significant in designing systems, but the knowledge is unlikely to allow active mitigation of interaction behavior on an individual basis in real time application, an implicit goal of the aims of HAI focused research. This is the case for two important reasons. First, we argue that focusing attention on interaction behavior does not go far enough because once the behavior has occurred it is too late to mitigate it in an *post hoc* fashion in a timely manner. Second, behaviorally based predictions for automation use are by definition general because the predictive ability has been achieved through extensive observation of how human user behavior is affected by system design. The behaviorally based predictions, however, do not take into account individual variation and dynamic changes in environment. Therefore, they are unlikely to apply to individuals on a case by case basis. For example, it has been shown that a human user may continue to rely on an automation that persistently commits errors of omission, or "misses", due to automation independent reasons such as increasing workload. Therefore, an approach is needed that considers individual cognitive and behavioral aspects, as well as ensuring that there is time to not only mitigate behavior, but allow the prediction of the likelihood of an interaction behavior.

We note that antecedent to the interaction behavior is a decision, which has been characterized by the decision neuroscience community as fairly stereotyped and accompanied by measurable psychophysiological correlates. These properties of decisions suggest predictability, and importantly, as decisions are individual in nature, these properties also imply the likelihood of behavioral prediction at an individual level. We thus believe that understanding the interaction decision is a useful approach to improving HAI, and that with future research that the decision correlates can be leveraged to predict the likelihood of an individual's impending interaction behavior. This approach not only satisfies the problems just discussed, but takes into consideration the human and environmental variability that is found in real world situations in ways that research focused on reliance and compliance has yet to achieve. While it is true that many of the psychophysiological correlates of value based decisions are only measurable with fMRI, we consider that the discoveries afforded by fMRI studies provide a solid basis to form specific scientific hypotheses to guide future research aimed at understanding the interaction decision and consequent interaction behavior. We understand that the fruit of decision neuroscience research might be applied to any domain where it would be advantageous for one decision outcome over another. However, our domain of interest is HAI, and therefore, our focus

is in leveraging what is known about value based decision making to understand interaction decisions in the hopes of eventually predicting the likelihood of one decision over another. This goal will require future research, and we begin by forwarding hypotheses to guide research efforts.

One of our first assumptions in this review is that interaction decisions are in fact a case of value based decisions, and that assumption guides our first hypothesis; interaction decisions are a special instance of value based decisions, and therefore the neural correlates accompanying value based decisions will be observable during interaction decisions. One of the first avenues of research is to demonstrate that value based decisions and interaction decisions are analogous in that the sense of risk and reward are inherent in both. This could be achieved by measuring bilateral frontal alpha power during an interaction decision to look for the characteristic asymmetry found in situations entailing risk. Further research to support this hypothesis should necessarily include observing the predicted vmPFC-parietal activation found in value based decision making, during an interaction decision. The strength of this evidence, if found, could be enhanced if concurrent with the parietal-vmPFC activation there is significantly less activation of the dlPFC.

If our first hypothesis is confirmed, we can begin to make more specific hypotheses. Our second hypothesis relates to the fact that there is little understanding of the first stage of value based decision making, the observation of alternatives and representation of internal and external states. Understanding this stage could be particularly important for mitigating inappropriate behavior because it is also accompanied by physiological changes (SC, decreased heart rate variability), which are readily observed. We believe that future research should be aimed at revealing neural correlates of this stage. We hypothesize that activity in the frontal polar cortex and in areas of the parietal cortex during, or just prior to a conscious decision to interact with an automation will occur along with or just prior to the physiological correlates. Evidence previously discussed, that activation patterns in the area of the frontal polar cortex in humans, and in the area of the mid-parietal in monkeys, can be decoded to reveal behavioral intention supports this hypothesis. Should this hypothesis be confirmed, it would add significant evidence that the approach of focusing on interaction decisions will provide an improved method to mitigate interaction behaviors on an individual level. For example, consider the fact that users tend to rely on an automation in the face of risk as demonstrated by traditional behavioral research. However, if this interaction behavior is inappropriate, but the decision to rely can be decoded, there is a chance to mitigate the inappropriate interaction behavior.

While not hypotheses, we believe that future research should also be focused on understanding the psychophysiological basis for, or correlates of, the interaction behaviors of relying or complying on an automation. One first step should include finding the psychophysiological correlates of the demonstrated tendencies of users to rely or comply with automations in circumstances such as workload and risk. For example, what psychophysiological processes drive a human user to perhaps over comply with an automation in conditions of increased cognitive workload, and are there measurable correlates that would suggest that this is the likely interaction behavior? Conversely, what psychophysiological correlates can be found, apart from alpha asymmetry, that appear to suggest the human user is perceiving increased risk, and therefore more likely to over rely on an automation? In order to mitigate disuse, potential psychophysiological correlates, such as particular levels of heart rate variability, SC and EEG power require future study.

Finally, though on a longer horizon, we suggest that it might ultimately be feasible to leverage the understanding of value based decisions in behavioral mitigations aimed at improving HAI system performance. Logically, management of a particular interaction between a human and an automated system requires a minimum of three elements. First, it is essential to understand the capabilities and vulnerabilities of both the particular operator and the particular automation as well as how these may vary under different task and contextual constraints. With such knowledge, one may be able to infer an optimal strategy for allocation of control or decision authority, such as has been done with behaviorally based predictions. Second, though the behavior of automated controls is relatively predictable with knowledge of how its control system was designed, establishing likelihood of human behaviors is a much more challenging task. Therefore, it is also critical to develop methods for prediction of likely changes in operator behavior on a time-scale that leaves room for active intervention through the understanding of the interaction decision. Third, an understanding of how to influence the decision process of humans in principled ways is necessary to ultimately define appropriate systems of actuation when inappropriate behaviors are expected. Of these three elements, it seems that the second may be the most challenging. This is because it is relatively trivial to establish baseline operational or performance characteristics of both humans and automated systems and it is already known that human behaviors and perceptions are subject to influence by a variety of factors, including workload, display properties, transparency, and may be amenable to influence by other task and contextual factors. However, predicting impending behavioral choices is particularly challenging because this requires methods to develop advance insight into the unfolding of the decision process that has largely been studied through the use of fMRI. Here, we offer one way of addressing this; by the application of modern techniques from cognitive neuroscience and psychophysiology.

CONCLUSION

The main purpose of this review is to explore the gap between the understanding of TiA and the actual human user interaction behavior which does not appear to have a clear mapping from TiA levels. We argue in this article that, in addition to understanding the influence of changing levels of TiA, understanding the antecedent decision of the human user's interaction behavior is critical for improving HAI system performance. Decisions have not been explicitly studied in the context of HAI and TiA specifically, but due to the

importance of these interaction decisions we reviewed decision making literature and summarized findings that provide a basic understanding of the psychophysiological processes involved in decision making. We are particularly interested in value based decision making because, just as in the case of TiA, if there is no risk, the behavior ceases to be important. While the value based decision process is not yet fully understood as it relates to interaction behaviors, there is a significant understanding of the underlying psychophysiological processes and correlates. This knowledge can be used to advance hypotheses that define a research path aimed at achieving mitigation of human user interaction behaviors.

AUTHOR CONTRIBUTIONS

KD: completed literature review and steered review towards current domain; main contributing author. JSM: final approval authority, guided literature review, advanced concepts germane to content, significant contribution to authorship. ARM: advised in fundamental concepts regarding BCI, HCI, and HAI; also reviewed, edited, and approved successive steps in the process. JRL: provided expertise in the cognitive neuroscience of trust in automation; also reviewed, edited and approved successive review and writing processes.

ACKNOWLEDGMENTS

This research was supported by the U.S. Office of the Secretary of Defense through the Autonomy Research Pilot Initiative (MIPR DWAM31168) as well as by an appointment to the US Army Research Laboratory Postdoctoral Fellowship program administered by the Oak Ridge Associated Universities through a cooperative agreement with the US Army Research Laboratory. We would also like to acknowledge Dr. Kristin Schaefer for her insights and guidance regarding understanding the complex set of human factors that influence trust in automation. The authors are also very grateful for the feedback and guidance of two reviewers whose suggestions have proven rather valuable in improving the quality of the final work.

REFERENCES

Bagheri, N., and Jamieson, G. A. (2004). "Considering subjective trust and monitoring behavior in assessing automation-induced complacency," in *Human Performance, Situation Awareness and Automation: Current Research and Trends*, eds D. A. Vicenzi, M. Moulousa, and P. A. Hancock (Mahwah, NJ: Erlbaum), 54–59.

Bahner, J. E., Hüper, A.-D., and Manzey, D. (2008). Misuse of automated decision aids: complacency, automation bias and the impact of training experience. *Int. J. Hum. Comput. Stud.* 66, 688–699. doi: 10.1016/j.ijhcs.2008.06.001

Basten, U., Biele, G., Heekeren, H. R., and Fiebach, C. J. (2010). How the brain integrates costs and benefits during decision making. *Proc. Natl. Acad. Sci. U S A* 107, 21767–21772. doi: 10.1073/pnas.0908104107

Biros, D., Daly, M., and Gunsch, G. (2004). The influence of task load and automation trust on deception detection. *Group Decis. Negot.* 13, 173–189. doi: 10.1023/b:grup.0000021840.85686.57

Bliss, J. P., Gilson, R. D., and Deaton, J. E. (1995). Human probability matching behaviour in response to alarms of varying reliability. *Ergonomics* 38, 2300–2312. doi: 10.1080/00140139508925269

Bogacz, R. (2007). Optimal decision-making theories: linking neurobiology with behaviour. *Trends Cogn. Sci.* 11, 118–125. doi: 10.1016/j.tics.2006.12.006

Chancey, E. T., Bliss, J. P., Proaps, A. B., and Madhavan, P. (2015). The role of trust as a mediator between system characteristics and response behaviors. *Hum. Factors* 57, 947–958. doi: 10.1177/0018720815582261

Chib, V. S., Rangel, A., Shimojo, S., and O'Doherty, J. P. (2009). Evidence for a common representation of decision values for dissimilar goods in human ventromedial prefrontal cortex. *J. Neurosci.* 29, 12315–12320. doi: 10.1523/JNEUROSCI.2575-09.2009

Christie, G. J., and Tata, M. S. (2009). Right frontal cortex generates reward-related theta-band oscillatory activity. *Neuroimage* 48, 415–422. doi: 10.1016/j.neuroimage.2009.06.076

Coe, B., Tomihara, K., Matsuzawa, M., and Hikosaka, O. (2002). Visual and anticipatory bias in three cortical eye fields of the monkey during an adaptive decision-making task. *J. Neurosci.* 22, 5081–5090.

Cohen, M. X., and Donner, T. H. (2013). Midfrontal conflict-related theta-band power reflects neural oscillations that predict behavior. *J. Neurophysiol.* 110, 2752–2763. doi: 10.1152/jn.00479.2013

Cohen, M. X., Elger, C. E., and Fell, J. (2009). Oscillatory activity and phase-amplitude coupling in the human medial frontal cortex during decision making. *J. Cogn. Neurosci.* 21, 390–402. doi: 10.1162/jocn.2008.21020

Cohen, M. X., Wilmes, K. A., and van de Vijver, I. (2011). Cortical electrophysiological network dynamics of feedback learning. *Trends Cogn. Sci.* 15, 558–566. doi: 10.1016/j.tics.2011.10.004

Corritore, L., Kracher, B., and Wiedenbeck, S. (2003). On-line trust: concepts, evolving themes, a model. *Int. J. Hum. Comput. Stud.* 58, 737–758. doi: 10.1016/s1071-5819(03)00041-7

Daly, M. A. (2002). *Task Load and Automation Use in an Uncertain Environment.* No. AFIT/GAQ/ENV/02M-05. Wright-Patterson AFB, OH: Air Force Institute of Technology, School of Engineering and Management.

Davis, C. E., Hauf, J. D., Wu, D. Q., and Everhart, D. E. (2011). Brain function with complex decision making using electroencephalography. *Int. J. Psychophysiol.* 79, 175–183. doi: 10.1016/j.ijpsycho.2010.10.004

Dekker, S. W., and Woods, D. D. (2002). Maba-maba or abracadabra? Progress on human-automation co-ordination. *Cogn. Technol. Work* 4, 240–244. doi: 10.1007/s101110200022

Delgado, M. R., Frank, R. H., and Phelps, E. A. (2005). Perceptions of moral character modulate the neural systems of reward during the trust game. *Nat. Neurosci.* 8, 1611–1618. doi: 10.1038/nn1575

Dixon, S. R., Wickens, C. D., and McCarley, J. S. (2006). "How do automation false alarms and misses affect operator compliance and reliance?," in Paper presented at the *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting* (Champaign, IL), 25–29.

Dixon, S. R., Wickens, C. D., and McCarley, J. S. (2007). On the independence of compliance and reliance: are automation false alarms worse than misses? *Hum. Factors* 49, 564–572. doi: 10.1518/001872007x215656

Dzindolet, M. T., Pierce, L. G., Beck, H. P., Dawe, L. A. and Anderson, B. W. (2001). Predicting misuse and disuse of combat identification systems. *Mil. Psychol.* 13, 147–164. doi: 10.1207/s15327876mp1303_2

Dzindolet, M. T., Pierce, L. G., Beck, H. P., and Dawe, L. A. (1999). "Misuse and disuse of automated aids.," in *Proceedings of the 43rd Annual Meeting of the Human Factors and Ergonomics Society* (Santa Monica, CA), 339–343.

Dzindolet, M. T., Pierce, L. G., Beck, H. P., and Dawe, L. A. (2002). The perceived utility of human and automated aids in a visual detection task. *Hum. Factors* 44, 79–94. doi: 10.1518/0018720024494856

Dzindolet, M., Peterson, S., Pomranky, T., and Beck, P. (2003). The role of trust in automation reliance. *Int. J. Hum. Comput. Stud.* 58, 697–718. doi: 10.1016/s1071-5819(03)00038-7

Evans, A. M., and Krueger, J. I. (2011). Elements of trust: risk and perspective-taking. *J. Exp. Soc. Psychol.* 47, 171–177. doi: 10.1016/j.jesp.2010.08.007

Fellows, L. K. (2004). The cognitive neuroscience of human decision making: a review and conceptual framework. *Behav. Cogn. Neurosci. Rev.* 3, 159–172. doi: 10.1177/1534582304273251

Figner, B., and Murphy, R. O. (2011). "Using skin conductance in judgment and decision making research," in *A Handbook of Process Tracing Methods for Decision Research*, eds M. Schulte-Mecklenbeck, A. Kuehberger, and R. Ranyard (New York, NY: Psychology Press), 163–184.

Geels-Blair, K., Rice, S., and Schwark, J. (2013). Using system-wide trust theory to reveal the contagion effects of automation false alarms and misses on compliance and reliance in a simulated aviation task. *Int. J. Aviat. Psychol.* 23, 245–266. doi: 10.1080/10508414.2013.799355

Gianotti, L. R., Knoch, D., Faber, P. L., Lehmann, D., Pascual-Marqui, R. D., Diezi, C., et al. (2009). Tonic activity level in the right prefrontal cortex predicts individuals' risk taking. *Psychol. Sci.* 20, 33–38. doi: 10.1111/j.1467-9280.2008.02260.x

Gläscher, J., Hampton, A. N., and O'Doherty, J. P. (2009). Determining a role for ventromedial prefrontal cortex in encoding action-based value signals during reward-related decision making. *Cereb. Cortex* 19, 483–495. doi: 10.1093/cercor/bhn098

Glimcher, P. W., and Fehr, E. (2013). Neuroeconomics: Decision Making and the Brain. (New York, NY: Academic Press).

Glimcher, P. W., and Rustichini, A. (2004). Neuroeconomics: the consilience of brain and decision. *Science* 306, 447–452. doi: 10.1126/science.1102566

Goddard, K., Roudsari, A., and Wyatt, J. C. (2014). Automation bias: empirical results assessing influencing factors. *Int. J. Med. Inform.* 83, 368–375. doi: 10.1016/j.ijmedinf.2014.01.001

Gold, C., Körber, M., Hohenberger, C., Lechner, D., and Bengler, K. (2015). Trust in automation-Before and after the experience of take-over scenarios in a highly automated vehicle. *Procedia Manuf.* 3, 3025–3032. doi: 10.1016/j.promfg.2015.07.847

Gold, J. I., and Shadlen, M. N. (2007). The neural basis of decision making. *Annu. Rev. Neurosci.* 30, 535–574. doi: 10.1146/annurev.neuro.29.051605.113038

Goldberg, M. E., Bisley, J. W., Powell, K. D., and Gottlieb, J. (2006). Saccades, salience and attention: the role of the lateral intraparietal area in visual behavior. *Prog. Brain Res.* 155, 157–175. doi: 10.1016/s0079-6123(06)55010-1

Guggisberg, A. G., Dalal, S. S., Findlay, A. M., and Nagarajan, S. S. (2007). High-frequency oscillations in distributed neural networks reveal the dynamics of human decision making. *Front. Hum. Neurosci.* 1:14. doi: 10.3389/neuro.09.014.2007

Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y., de Visser, E. J., and Parasuraman, R. (2011). A meta-analysis of factors affecting trust in human-robot interaction. *Hum. Factors* 53, 517–527. doi: 10.1177/0018720811417254

Hare, T. A., O'Doherty, J., Camerer, C. F., Schultz, W., and Rangel, A. (2008). Dissociating the role of the orbitofrontal cortex and the striatum in the computation of goal values and prediction errors. *J. Neurosci.* 28, 5623–5630. doi: 10.1523/JNEUROSCI.1309-08.2008

Haynes, J. D. (2011). Decoding and predicting intentions. *Ann. N Y Acad. Sci.* 1224, 9–21. doi: 10.1111/j.1749-6632.2011.05994.x

Haynes, J. D., Sakai, K., Rees, G., Gilbert, S., Frith, C., and Passingham, R. E.. (2007). Reading hidden intentions in the human brain. *Curr. Biol.* 17, 323–328. doi: 10.1016/j.cub.2006.11.072

Heekeren, H. R., Marrett, S., and Ungerleider, L. G. (2008). The neural systems that mediate human perceptual decision making. *Nat. Rev. Neurosci.* 9, 467–479. doi: 10.1038/nrn2374

Jamieson, G. A., and Vicente, K. J. (2005). Designing effective human-automation-plant interfaces: a control-theoretic perspective. *Hum. Factors* 47, 12–34. doi: 10.1518/0018720053653820

Kable, J. W., and Glimcher, P. W. (2007). The neural correlates of subjective value during intertemporal choice. *Nat. Neurosci.* 10, 1625–1633. doi: 10.1038/nn2007

Kiani, R., and Shadlen, M. N. (2009). Representation of confidence associated with a decision by neurons in the parietal cortex. *Science* 324, 759–764. doi: 10.1126/science.1169405

Krajbich, I., Armel, C., and Rangel, A. (2010). Visual fixations and the computation and comparison of value in simple choice. *Nat. Neurosci.* 13, 1292–1298. doi: 10.1038/nn.2635

Krajbich, I., and Rangel, A. (2011). Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions. *Proc. Natl. Acad. Sci. U S A* 108, 13852–13857. doi: 10.1073/pnas.1101328108

Kumagai, T., Sakaguchi, Y., Owuwa, M., and Akamatsu, M. (2003). "Prediction of driving behavior through probabilistic inference.," in Paper presented at the *Proceedings of the Eighth International Conference On Engineering Applications of Neural Networks (EANN'03)* (Torremolinos, Malaga, Spain), 117–123.

Lavazza, A., and De Caro, M. (2010). Not so fast. On some bold neuroscientific claims concerning human agency. *Neuroethics* 3, 23–41. doi: 10.1007/s12152-009-9053-9

Lee, J., and Moray, N. (1992). Trust, control strategies and allocation of function in human-machine systems. *Ergonomics* 35, 1243–1270. doi: 10.1080/00140139208967392

Lee, J. D., and Moray, N. (1994). Trust, self-confidence and operators' adaptation to automation. *Int. J. Hum. Comput. Stud.* 40, 153–184. doi: 10.1006/ijhc.1994.1007

Lee, J. D., and See, K. A. (2004). Trust in automation: designing for appropriate reliance. *Hum. Factors* 46, 50–80. doi: 10.1518/hfes.46.1.50.30392

Libet, B. (1993). Unconscious cerebral initiative and the role of conscious will in voluntary action. . *Neurophysiol. Conscious.*, (Boston: Birkhäuser), 269–306. doi: 10.1007/978-1-4612-0355-1_16

Lim, S.-L., O'Doherty, J. P., and Rangel, A. (2011). The decision value computations in the vmPFC and striatum use a relative value code that is guided by visual attention. *J. Neurosci.* 31, 13214–13223. doi: 10.1523/JNEUROSCI.1246-11.2011

Lyons, J. B., and Stokes, C. K. (2012). Human-human reliance in the context of automation. *Hum. Factors* 54, 112–121. doi: 10.1177/0018720811427034

Madhavan, P., and Wiegmann, D. (2007). Similarities and differences between human-human and human-automation trust: an integrative review. *Theor. Issues Ergon. Sci.* 8, 277–301. doi: 10.1080/14639220500337708

Madhavan, P., Wiegmann, D. A., and Lacson, F. C. (2006). Automation failures on tasks easily performed by operators undermine trust in automated aids. *Hum. Factors* 48, 241–256. doi: 10.1518/001872006777724408

Manzey, D., Bahner, J. E., and Hueper, A.-D. (2006). Misuse of automated aids in process control: complacency, automation bias and possible training interventions. *Proc. Hum. Factors Ergon. Soc. Annu. Meeting*, 50, 220–224. doi: 10.1177/154193120605000303

Manzey, D., Gérard, N., and Wiczorek, R. (2014). Decision-making and response strategies in interaction with alarms: the impact of alarm reliability, availability of alarm validity information and workload. *Ergonomics* 57, 1833–1855. doi: 10.1080/00140139.2014.957732

Masalonis, A. J., and Parasuraman, R. (1999). Trust as a construct for evaluation of automated aids: past and future theory and research. *Proc. Hum. Factors Ergon. Soc. Annu. Meeting*, 43, 184–187. doi: 10.1177/154193129904300312

Merritt, S. M. (2011). Affective processes in human automation interactions. *Hum. Factors* 53, 356–370. doi: 10.1177/0018720811411912

Merritt, S. M., Heimbaugh, H., LaChapell, J., and Lee, D. (2012). I trust it, but I don't know why effects of implicit attitudes toward automation on trust in an automated system. *Hum. Factors* 55, 520–534. doi: 10.1177/0018720812465081

Merritt, S. M., Huber, K., LaChapell-Unnerstall, J., and Lee, D. (2014). *Continuous Calibration of Trust in Automated Systems.* St. Louis, MO: University of Missouri-St. Louis.

Merritt, S. M., and Ilgen, D. R. (2008). Not all trust is created equal: dispositional and history-based trust in human-automation interactions. *Hum. Factors* 50, 194–210. doi: 10.1518/001872008x288574

Meyer, J. (2001). Effects of warning validity and proximity on responses to warnings. *Hum. Factors* 43, 563–572. doi: 10.1518/001872001775870395

Meyer, J. (2004). Conceptual issues in the study of dynamic hazard warnings. *Hum. Factors* 46, 196–204. doi: 10.1518/hfes.46.2.196.37335

Meyer, J., and Bitan, Y. (2002). Why better operators receive worse warnings. *Hum. Factors* 44, 343–353. doi: 10.1518/0018720024497754

Meyer, J., Wiczorek, R., and Günzler, T. (2014). Measures of reliance and compliance in aided visual scanning. *Hum. Factors* 56, 840–849. doi: 10.1177/0018720813512865

Miller, C. A. (2005). "Trust in adaptive automation: the role of etiquette in tuning trust via analogic and affective methods," in Paper Presented at the

Proceedings of the 1st International Conference on Augmented Cognition (Las Vegas, NV).

Moray, N. (2003). Monitoring, complacency, scepticism and eutactic behaviour. *Int. J. Ind. Ergon.* 31, 175–178. doi: 10.1016/s0169-8141(02)00194-4

Moray, N., Inagaki, T., and Itoh, M. (2000). Adaptive automation, trust and self-confidence in fault management of time-critical tasks. *J. Exp. Psychol. Appl.* 6, 44–58. doi: 10.1037/1076-898x.6.1.44

Mosier, K. L., and Skitka, L. J. (1996). Human decision makers and automated decision aids: made for each other. *Autom. Hum. Perform. Theory Appl.* 40, 201–220.

Muir, B. M. (1994). Trust in automation: Part I. Theoretical issues in the study of trust and human intervention in automated systems. *Ergonomics* 37, 1905–1922. doi: 10.1080/00140139408964957

Muir, B. M., and Moray, N. (1996). Trust in automation. Part II. Experimental studies of trust and human intervention in a process control simulation. *Ergonomics* 39, 429–460. doi: 10.1080/00140139608964474

Mulder, M., van Maanen, L., and Forstmann, B. (2014). Perceptual decision neurosciences-a model-based review. *Neuroscience* 277, 872–884. doi: 10.1016/j.neuroscience.2014.07.031

Musallam, S., Corneil, B., Greger, B., Scherberger, H., and Andersen, R. (2004). Cognitive control signals for neural prosthetics. *Science* 305, 258–262. doi: 10.1126/science.1097938

Nass, C., Fogg, B., and Moon, Y. (1996). Can computers be teammates? *Int. J. Hum. Comput. Stud.* 45, 669–678. doi: 10.1006/ijhc.1996.0073

Nieuwenhuis, S., Aston-Jones, G., and Cohen, J. D. (2005). Decision making, the P3, and the locus coeruleus-norepinephrine system. *Psychol. Bull.* 131, 510–532. doi: 10.1037/0033-2909.131.4.510

Pak, R., Fink, N., Price, M., Bass, B., and Sturre, L. (2012). Decision support aids with anthropomorphic characteristics influence trust and performance in younger and older adults. *Ergonomics* 55, 1059–1072. doi: 10.1080/00140139.2012.691554

Parasuraman, R., and Manzey, D. (2010). Complacency and bias in human use of automation: an attentional integration. *Hum. Factors* 52, 381–410. doi: 10.1177/0018720810376055

Parasuraman, R., and Miller, C. A. (2004). Trust and etiquette in high-criticality automated systems. *Commun. ACM* 47, 51–55. doi: 10.1145/975817.975844

Parasuraman, R., and Riley, V. (1997). Humans and automation; use misuse, disuse, abuse. *Hum. Factors* 39, 230–253. doi: 10.1518/001872097778543886

Parasuraman, R., Sheridan, T. B., and Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Trans. Syst. Man Cybern. A Syst. Hum.* 30, 286–297. doi: 10.1109/3468.844354

Parasuraman, R., Sheridan, T. B., and Wickens, C. D. (2008). Situation awareness, mental workload and trust in automation: viable, empirically supported cognitive engineering constructs. *J. Cogn. Eng. Decis. Mak.* 2, 140–160. doi: 10.1518/155534308x284417

Perez, O., Mukamel, R., Tankus, A., Rosenblatt, J. D., Yeshurun, Y., and Fried, I. (2015). Preconscious prediction of a driver's decision using intracranial recordings. *J. Cogn. Neurosci.* 27, 1492–1502. doi: 10.1162/jocn_a1004_00799

Philiastides, M. G., Auksztulewicz, R., Heekeren, H. R., and Blankenburg, F. (2011). Causal role of dorsolateral prefrontal cortex in human perceptual decision making. *Curr. Biol.* 21, 980–983. doi: 10.1016/j.cub.2011.04.034

Platt, M. L. (2002). Neural correlates of decisions. *Curr. Opin. Neurobiol.* 12, 141–148. doi: 10.1016/s0959-4388(02)00302-1

Platt, M. L., and Glimcher, P. W. (1999). Neural correlates of decision variables in parietal cortex. *Nature* 400, 233–238. doi: 10.1038/22268

Ploran, E. J., Nelson, S. M., Velanova, K., Donaldson, D. I., Petersen, S. E., and Wheeler, M. E. (2007). Evidence accumulation and the moment of recognition: dissociating perceptual recognition processes using fMRI. *J. Neurosci.* 27, 11912–11924. doi: 10.1523/jneurosci.3522-07.2007

Poole, A., and Ball, L. J. (2006). Eye tracking in HCI and usability research. *Encyclopedia Hum. Comput. Interact.* 1, 211–219. doi: 10.4018/978-1-59140-562-7.ch034

Rangel, A., Camerer, C., and Montague, P. R. (2008). A framework for studying the neurobiology of value-based decision making. *Nat. Rev. Neurosci.* 9, 545–556. doi: 10.1038/nrn2357

Rangel, A., and Hare, T. (2010). Neural computations associated with goal-directed choice. *Curr. Opin. Neurobiol.* 20, 262–270. doi: 10.1016/j.conb.2010.03.001

Rice, S. (2009). Examining single-and multiple-process theories of trust in automation. *J. Gen. Psychol.* 136, 303–322. doi: 10.3200/GENP.136.3.303–322

Rice, S., and Geels, K. (2010). Using system-wide trust theory to make predictions about dependence on four diagnostic aids. *J. Gen. Psychol.* 137, 362–375. doi: 10.1080/00221309.2010.499397

Rorie, A. E., Gao, J., McClelland, J. L., and Newsome, W. T. (2010). Integration of sensory and reward information during perceptual decision-making in lateral intraparietal cortex (LIP) of the macaque monkey. *PLoS One* 5:e9308. doi: 10.1371/journal.pone.0009308

Schaefer, K. E., Billings, D. R., Szalma, J. L., Adams, J. K., Sanders, T. L., Chen, J. Y., et al. (2014). *A Meta-Analysis of Factors Influencing the Development of Trust in Automation: Implications for Human-Robot Interaction.* No. ARL-TR-6984. Aberdeen, MD: Army Research Laboratory's, Human Research and Engineering Directorate.

Schaefer, K. E., Sanders, T. L., Yordon, R. E., Billings, D. R., and Hancock, P. (2012). "Classification of robot form: factors predicting perceived trustworthiness," in Paper Presented at the *Proceedings of the Human Factors and Ergonomics Society 56th Annual Meeting* (Boston, MA).

Sheridan, T. (1980). "Theory of man-machine interaction as related to computerized automation," in *Paper Presented at the Proceedings of the Sixth Annual Advanced Control Conference*, (Cambridge, MA: MIT Press).

Sheridan, T. B., and Hennessy, R. T. (1984). *Research and Modeling of Supervisory Control Behavior. Report of a Workshop.* Washington, DC: National Research Council, Committee on Human Factors.

Singh, I. L., Molloy, R., and Parasuraman, R. (1997). Automation-induced monitoring inefficiency: role of display location. *Int. J. Hum. Comput. Stud.* 46, 17–30. doi: 10.1006/ijhc.1996.0081

Skitka, L. J., Mosier, K. L., and Burdick, M. (1999). Does automation bias decision-making? *Int. J. Hum. Comput. Stud.* 51, 991–1006. doi: 10.1006/ijhc.1999.0252

Skitka, L. J., Mosier, K. L., Burdick, M., and Rosenblatt, B. (2000). Automation bias and errors: are crews better than individuals? *Int. J. Aviat. Psychol.* 10, 85–97. doi: 10.1207/s15327108ijap1001_5

Sokolov, E., Spinks, J., Naatanen, R., and Risto-Ktttubebm, G. (2002). *The Orienting Response in Information Processing.* Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Soon, C. S., Brass, M., Heinze, H.-J., and Haynes, J.-D. (2008). Unconscious determinants of free decisions in the human brain. *Nat. Neurosci.* 11, 543–545. doi: 10.1038/nn.2112

Soon, C. S., He, A. H., Bode, S., and Haynes, J.-D. (2013). Predicting free choices for abstract intentions. *Proc. Natl. Acad. Sci. U S A* 110, 6217–6222. doi: 10.1073/pnas.1212218110

Terai, H., Okuda, H., Hitomi, K., Bando, T., Miyajima, C., Hirayama, Y., et al. (2015). An experimental study on the difference in drivers' decision making behavior during manual and supported driving. *Procedia Manuf.* 3, 3136–3141. doi: 10.1016/j.promfg.2015.07.974

van de Vijver, I., Ridderinkhof, K. R., and Cohen, M. X. (2011). Frontal oscillatory dynamics predict feedback learning and action adjustment. *J. Cogn. Neurosci.* 23, 4106–4121. doi: 10.1162/jocn_a_00110

Verberne, F. M. F., Ham, J., and Midden, C. J. H. (2012). Trust in smart systems: sharing driving goals and giving information to increase trustworthiness and acceptability of smart systems in cars. *Hum. Factors* 54, 799–810. doi: 10.1177/0018720812443825

Wallis, J. D. (2012). Cross-species studies of orbitofrontal cortex and value-based decision-making. *Nat. Neurosci.* 15, 13–19. doi: 10.1038/nn.2956

Wang, L., Jamieson, G. A., and Hollands, J. G. (2008). "Selecting methods for the analysis of reliance on automation," in Paper Presented at the *Proceedings of the Human Factors and Ergonomics Society 52nd Annual Meeting* (Santa Monica, CA), 287–291.

Wang, L., Jamieson, G. A., and Hollands, J. G. (2009). Trust and reliance on an automated combat identification system. *Hum. Factors* 51, 281–291. doi: 10.1177/0018720809338842

Wickens, C. D., and Dixon, S. R. (2007). The benefits of imperfect diagnostic automation: a synthesis of the literature. *Theor. Issues Ergon. Sci.* 8, 201–212. doi: 10.1080/14639220500370105

Wiczorek, R., and Manzey, D. (2010). "Is operators' compliance with alarm systems a product of rational consideration?," in Paper presented at the *Proceedings of the Human Factors and Ergonomics Society 54th Annual Meeting* (San Francisco, CA), pp. 1722–1726

Wiczorek, R. and Manzey, D. (2014). Supporting attention allocation in multitask environments: effects of likelihood alarm systems on trust, behavior, and performance. *Hum. Factors* 56, 1209–1221. doi: 10.1177/0018720814528534

Yacubian, J., Gläscher, J., Schroeder, K., Sommer, T., Braus, D. F., and Büchel, C. (2006). Dissociable systems for gain-and loss-related value predictions and errors of prediction in the human brain. *J. Neurosci.* 26, 9530–9537. doi: 10.1523/jneurosci.2915-06.2006

Yeh, M., and Wickens, C. D. (2000). Effects of cue reliability, realism and interactivity on biases of attention and trust in augmented reality. *Proc Hum. Factors Ergon. Soc. Annu. Meeting*, 44, 511–514. doi: 10.1177/154193120004400507

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer DM and handling Editor declared their shared affiliation, and the handling Editor states that the process nevertheless met the standards of a fair and objective review.

Copyright © 2016 Drnec, Marathe, Lukos and Metcalfe. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution and reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read, for greatest visibility

COLLABORATIVE PEER-REVIEW
Designed to be rigorous – yet also collaborative, fair and constructive

FAST PUBLICATION
Average 85 days from submission to publication (across all journals)

COPYRIGHT TO AUTHORS
No limit to article distribution and re-use

TRANSPARENT
Editors and reviewers acknowledged by name on published articles

SUPPORT
By our Swiss-based editorial team

IMPACT METRICS
Advanced metrics track your article's impact

GLOBAL SPREAD
5'100'000+ monthly article views and downloads

LOOP RESEARCH NETWORK
Our network increases readership for your article

Frontiers
EPFL Innovation Park, Building I • 1015 Lausanne • Switzerland
Tel +41 21 510 17 00 • Fax +41 21 510 17 01 • info@frontiersin.org
www.frontiersin.org

Find us on