NONLINEAR ANALYSIS IN NEUROSCIENCE AND BEHAVIORAL RESEARCH

EDITED BY: Tobias A. Mattei
PUBLISHED IN: Frontiers in Computational Neuroscience

frontiers Research Topics

frontiers

ISSN 1664-8714
ISBN 978-2-88919-996-9
DOI 10.3389/978-2-88919-996-9

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**

NONLINEAR ANALYSIS IN NEUROSCIENCE AND BEHAVIORAL RESEARCH

Topic Editor:
Tobias A. Mattei, Eastern Maine Medical Center, USA

Three-dimensional phase space representation of a non-linear system with a trajectory involving a strange attractor
Image by Nicolas Desprez. Available under CC BY-SA 3.0 license at: https://en.wikipedia.org/wiki/Attractor#/media/File:Atractor_Poisson_Saturne.jpg

Although nonlinear dynamics have been mastered by physicists and mathematicians for a long time (as most physical systems are inherently nonlinear in nature), the recent successful application of nonlinear methods to modeling and predicting several evolutionary, ecological, physiological, and biochemical processes has generated great interest and enthusiasm among researchers in computational neuroscience and cognitive psychology. Additionally, in the last years it has been demonstrated that nonlinear analysis can be successfully used to model not only basic cellular and molecular data but also complex cognitive processes and behavioral interactions.

The theoretical features of nonlinear systems (such unstable periodic orbits, period-doubling bifurcations and phase space dynamics) have already been successfully applied by several research groups to analyze the behavior of a variety of neuronal and cognitive processes. Additionally the concept of strange attractors has lead to a new understanding of information processing which considers higher cognitive functions (such as language, attention, memory and decision making) as complex systems emerging from the dynamic interaction between parallel streams of information flowing between highly interconnected neuronal clusters organized in a widely distributed circuit and modulated by key central nodes. Furthermore, the paradigm of self-organization derived from the nonlinear dynamics theory has offered an interesting account of the phenomenon of emergence of new complex cognitive structures from random and non-deterministic patterns, similarly to what has been previously observed in nonlinear

studies of fluid dynamics.Finally, the challenges of coupling massive amount of data related to brain function generated from new research fields in experimental neuroscience (such as magnetoencephalography, optogenetics and single-cell intra-operative recordings of neuronal activity) have generated the necessity of new research strategies which incorporate complex pattern analysis as an important feature of their algorithms.

Up to now nonlinear dynamics has already been successfully employed to model both basic single and multiple neurons activity (such as single-cell firing patterns, neural networks synchronization, autonomic activity, electroencephalographic measurements, and noise modulation in the cerebellum), as well as higher cognitive functions and complex psychiatric disorders. Similarly, previous experimental studies have suggested that several cognitive functions can be successfully modeled with basis on the transient activity of large-scale brain networks in the presence of noise. Such studies have demonstrated that it is possible to represent typical decision-making paradigms of neuroeconomics by dynamic models governed by ordinary differential equations with a finite number of possibilities at the decision points and basic heuristic rules which incorporate variable degrees of uncertainty.

This e-book has include frontline research in computational neuroscience and cognitive psychology involving applications of nonlinear analysis, especially regarding the representation and modeling of complex neural and cognitive systems. Several experts teams around the world have provided frontline theoretical and experimental contributions (as well as reviews, perspectives and commentaries) in the fields of nonlinear modeling of cognitive systems, chaotic dynamics in computational neuroscience, fractal analysis of biological brain data, nonlinear dynamics in neural networks research, nonlinear and fuzzy logics in complex neural systems, nonlinear analysis of psychiatric disorders and dynamic modeling of sensorimotor coordination.

Rather than a comprehensive compilation of the possible topics in neuroscience and cognitive research to which non-linear may be used, this e-book intends to provide some illustrative examples of the broad range of fields to which the powerful tools of non-linear analysis can be successfully employed. We sincerely hope that that these articles may stimulate the reader to deepen its interest in the topic of non-linear analysis in neuroscience and cognitive sciences, paving the way for future theoretical and experimental research on this rapidly evolving and promising research field.

Citation: Mattei, T. A., ed. (2016). Nonlinear Analysis in Neuroscience and Behavioral Research. Lausanne: Frontiers Media. doi: 10.3389/978-2-88919-996-9

Table of Contents

frontiers in
COMPUTATIONAL NEUROSCIENCE

EDITORIAL
published: 21 February 2014
doi: 10.3389/fncom.2014.00017

Unveiling complexity: non-linear and fractal analysis in neuroscience and cognitive psychology

Tobias A. Mattei*

Department of Neurological Surgery, The Ohio State University Medical Center, Columbus, OH, USA
**Correspondence: tobias.mattei@osumc.edu*

Edited by:
Misha Tsodyks, Weizmann Institute of Science, Israel

Keywords: non-linear analsyis, complex systems, fractal analysis, cognitive psychology, neurosciences

Although non-linear dynamics has been mastered by physicists and mathematicians for a long time, as most physical systems are inherently non-linear in nature (Kirillov and Dmitry, 2013), the more recent successful application of non-linear and fractal methods to modeling and prediction of several evolutionary, ecologic, genetic, and biochemical processes (Avilés, 1999) has generated great interest and enthusiasm for such type of approach among researchers in neuroscience and cognitive psychology.

After initial works on this emerging field, it became clear that that multiple aspects of brain function as viewed from different perspectives and scales present a nonlinear behavior, with a complex phase space composed of multiple equilibrium points, limit cycles, stability regions, and trajectory flows as well as a dynamics which includes unstable periodic orbits, period-doubling bifurcations, as well as other features typical of chaotic systems (Birbaumer et al., 1995). Moreover it was also demonstrated that non-linear dynamics was able to explain several unique features of the brain such as plasticity and learning (Freeman, 1994).

More recently the concept of strange attractors has lead to a new understanding of information processing in the brain which, instead of the old "localizationist" approaches (Wernicke, 1970), considers higher cognitive functions (such as language, attention, memory and decision-making) as systemic properties which emerge from the dynamic interaction between parallel streams of information flowing between highly interconnected neuronal clusters that are organized in a widely distributed circuit modulated by key central nodes (Mattei, 2013a,b). According to such paradigm, the concept of self-organization has been able to offer a proper account of the phenomenon of evolutionary emergence of new complex cognitive structures from non-deterministic random patterns, similarly to what has been previously observed in nonlinear studies of fluid dynamics (Dixon et al., 2012).

Additionally, the challenges of interpreting massive amounts of information related to brain function generated from emerging research fields in experimental neuroscience (such as functional MRi, magnetoencephalography, optogenetics, and single-cell intra-operative recordings) have generated the necessity of new methods for which incorporate complex pattern analysis as an important feature of their algorithms (Turk-Browne, 2013).

Up to now nonlinear methods have already been successfully employed to describe and model (among many other examples) single-cell firing patterns (Thomas et al., 2013), neural networks synchronization (Yu et al., 2011), autonomic activity (Tseng et al., 2013), electroencephalographic data (Abásolo et al., 2007), noise modulation in the cerebellum (Tokuda et al., 2010), as well as higher cognitive functions and complex psychiatric disorders (Bystritsky et al., 2012). Additionally fractal analysis has been extensively explored not only in the description of the temporal aspects of neuronal dynamics, but also in the evaluation of key structural patterns of cellular organization in both normal and pathological histologic brain samples (Mattei, 2013a,b).

Finally, recent studies have demonstrated that several cognitive functions can be successfully modeled with basis on the transient activity of large-scale brain networks in the presence of noise (Rabinovich et al., 2008). In fact, it has already been suggested that the observed pervasiveness of the 1/f scaling (also called 1/*f* noise, fractal time, or pink noise) in both neural and cognitive functions may have a very close relationship (if not a causal one) with the phenomenon of metastability of brain states (Kello et al., 2008). Other studies in the emerging field of neuroeconomics have shown that it is possible to represent typical decision-making paradigms by dynamic models governed by ordinary differential equations with a finite number of possibilities at the decision points as well as basic rules to address uncertainty (Holmes et al., 2004).

In this special edition of Frontiers Computational Neuroscience dedicated to the topic of Non-linear and Fractal Analysis in Neuroscience and Cognitive Psychology, special articles from several frontline research groups around the world were carefully selected in order to provide a representative sample of the different research fields in neuroscience and cognitive psychology where non-linear and fractal analysis may be successfully applied.

The selected articles include both classical problems where non-linear method have been traditionally employed (such as EEG data analysis) as well as other new research fields in which non-linear analysis has been shown to be useful not only for modeling normal brain dynamics but also for the diagnosis of neurological and psychiatric disorders, monitoring of their natural history and evaluation of the effects of different therapeutic strategies.

Overall, both theoretical and experimental works in the field seem to demonstrate that the advanced tools of non-linear analysis can much more accurately describe and represent the complexity of brain dynamics than traditional mathematical and computational methods based on linear and deterministic analysis.

Although it seems quite unquestionable that future attempts to model complex brain and cognitive functions will significantly benefit from non-linear methods, the exact cognitive and

neuronal variables that may exhibit a significant chaotic pattern is still an open question. However, taking into account the pervasiveness of non-linear behavior in the brain, which has already been demonstrated by such an extensive literature in so many different fields of neuroscience and cognitive psychology (as well as the remarkable progress that has been achieved by the application of non-linear and fractal analysis in such research areas), maybe the burden of proof should be on the other side. Perhaps the real question to be answered is: Which areas of neuroscience and cognitive psychology would not benefit from the advantages that non-linear and fractal analysis has to offer?

REFERENCES

Abásolo, D., James, C.J., and Hornero, R. (2007). Non-linear analysis of intracranial electroencephalogram recordings with approximate entropy and Lempel-Ziv complexity for epileptic seizure detection. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 2007, 1953–1956. doi: 10.1109/IEMBS.2007.4352700

Avilés, L. (1999). Cooperation and non-linear dynamics: an ecological perspective on the evolution of sociality. *Evolut. Ecol. Res.* 1, 459–477.

Birbaumer, N., Flor, H., Lutzenberger, W., and Elbert, T. (1995). Chaos and order in the human brain. *Electroencephalogr. Clin. Neurophysiol. Suppl.* 44, 450–459.

Bystritsky, A., Nierenberg, A. A., Feusner, J. D., Rabinovich, M. (2012). Computational non-linear dynamical psychiatry: a new methodological paradigm for diagnosis and course of illness. *J. Psychiatr. Res.* 46, 428–435. doi: 10.1016/j.jpsychires.2011.10.013

Dixon, J. A., Holden, J. G., Mirman, D., and Stephen, D. G. (2012). Multifractal dynamics in the emergence of cognitive structure. *Top. Cogn. Sci.* 4, 51–62. doi: 10.1111/j.1756-8765.2011.01162.x

Freeman, W. J. (1994). Role of chaotic dynamics in neural plasticity. *Prog. Brain Res.* 102, 319–33. doi: 10.1016/S0079-6123(08)60549-X

Holmes, P., Shea-Brown, E., Moehlis, J., Bogacz, R., Gao, J., Aston-Jones, G. et al. (2004). Optimal decisions: from neural spikes, through stochastic differential equations, to behavior. *IEICE Trans. Fund. Electron. Commun. Comput. Sci.* 88, 2496–2503. Available online at: http://search.ieice.org/bin/summary.php?id=e88-a_10_2496

Kello, C. T., Anderson, G. G., Holden, J. G., and Van Orden, G. C. (2008). The pervasiveness of 1/f scaling in speech reflects the metastable basis of cognition. *Cogn. Sci.* 32, 1217–1231. doi: 10.1080/03640210801944898

Kirillov, N. O., and Dmitry, E. P. (eds.). (2013). *Nonlinear Physical Systems: Spectral Analysis, Stability and Bifurcations*. Wiley-ISTE. doi: 10.1002/9781118577608. Available online at: http://www.wiley.com/WileyCDA/WileyTitle/productCd-1848214200.html

Mattei, T. A. (2013a). The secret is at the crossways: hodotopic organization and nonlinear dynamics of brain neural networks. *Behav. Brain Sci.* 36, 623–624. discussion: 634–659.

Mattei, T. A. (2013b). Nonlinear (chaotic) dynamics and fractal analysis: new applications to the study of the microvascularity of gliomas. *World Neurosurg.* 79, 4–7. doi: 10.1016/j.wneu.2012.11.047.

Rabinovich, M. I., Huerta, R., Varona, P., and Afraimovich, V. S. (2008). Transient cognitive dynamics, metastability and decision making. *PLoS Comput. Biol.* 4:e1000072. doi: 10.1371/journal.pcbi.1000072

Thomas, P., Straube, A. V., Timmer, J., Fleck, C., and Grima, R. (2013). Signatures of nonlinearity in single cell noise-induced oscillations. *J. Theor. Biol.* 335, 222–234. doi: 10.1016/j.jtbi.2013.06.021

Tokuda, I. T., Han, C. E., Aihara, K., Kawato, M., and Schweighofer, N. (2010). The role of chaotic resonance in cerebellar learning. *Neural Netw.* 23, 836–842. doi: 10.1016/j.neunet.2010.04.006

Turk-Browne, N. B. (2013). Functional interactions as big data in the human brain. *Science* 342, 580–584. doi: 10.1126/science.1238409

Yu, H., Wang, J., Liu, Q., Wen, J., Deng, B., and Wei, X. (2011). Chaotic phase synchronization in a modular neuronal network of small-world subnetworks. *Chaos* 21, 043125. doi: 10.1063/1.3660327

Tseng, L., Tang, S. C., Chang, C. Y., Lin, Y. C., Abbod, M. F., and Shieh, J. S. (2013). Nonlinear and conventional biosignal analyses applied to tilt table test for evaluating autonomic nervous system and autoregulation. *Open Biomed. Eng. J.* 7, 93–99. doi: 10.2174/1874120720130905004

Wernicke, K. (1970). The aphasia symptom-complex: a psychological study on an anatomical basis. *Arch. Neurol.* 22, 280–282. doi: 10.1001/archneur.1970.00480210090013

Received: 05 February 2014; accepted: 05 February 2014; published online: 21 February 2014.

Citation: Mattei TA (2014) Unveiling complexity: non-linear and fractal analysis in neuroscience and cognitive psychology. Front. Comput. Neurosci. ***8****:17. doi: 10.3389/fncom.2014.00017*

This article was submitted to the journal Frontiers in Computational Neuroscience.

frontiers in Computational Neuroscience

ORIGINAL RESEARCH
published: 02 October 2015
doi: 10.3389/fncom.2015.00125

Low-dimensional attractor for neural activity from local field potentials in optogenetic mice

Sorinel A. Oprisan[1], Patrick E. Lynn[2], Tamas Tompa[3,4] and Antonieta Lavin[3]*

[1] Department of Physics and Astronomy, College of Charleston, Charleston, SC, USA, [2] Department of Computer Science, College of Charleston, Charleston, SC, USA, [3] Department of Neuroscience, Medical University of South Carolina, Charleston, SC, USA, [4] Department of Preventive Medicine, Faculty of Healthcare, University of Miskolc, Miskolc, Hungary

We used optogenetic mice to investigate possible nonlinear responses of the medial prefrontal cortex (mPFC) local network to light stimuli delivered by a 473 nm laser through a fiber optics. Every 2 s, a brief 10 ms light pulse was applied and the local field potentials (LFPs) were recorded with a 10 kHz sampling rate. The experiment was repeated 100 times and we only retained and analyzed data from six animals that showed stable and repeatable response to optical stimulations. The presence of nonlinearity in our data was checked using the null hypothesis that the data were linearly correlated in the temporal domain, but were random otherwise. For each trail, 100 surrogate data sets were generated and both time reversal asymmetry and false nearest neighbor (FNN) were used as discriminating statistics for the null hypothesis. We found that nonlinearity is present in all LFP data. The first 0.5 s of each 2 s LFP recording were dominated by the transient response of the networks. For each trial, we used the last 1.5 s of steady activity to measure the phase resetting induced by the brief 10 ms light stimulus. After correcting the LFPs for the effect of phase resetting, additional preprocessing was carried out using dendrograms to identify "similar" groups among LFP trials. We found that the steady dynamics of mPFC in response to light stimuli could be reconstructed in a three-dimensional phase space with topologically similar "8"-shaped attractors across different animals. Our results also open the possibility of designing a low-dimensional model for optical stimulation of the mPFC local network.

Keywords: optogenetics, medial prefrontal cortex, electrophysiology, delay-embedding, nonlinear dynamics

OPEN ACCESS

Edited by:
Tobias Alecio Mattei, Kenmore Mercy Hospital, USA

Reviewed by:
Todd Troyer, University of Texas, USA
Joaquín J. Torres, University of Granada, Spain
Xin Tian, Tianjin Medical University, China

****Correspondence:***
Sorinel A. Oprisan, Department of Physics and Astronomy, College of Charleston, 66 Georege Street, Charleston, SC 29424, USA
oprisans@cofc.edu

Received: *11 June 2015*
Accepted: *18 September 2015*
Published: *02 October 2015*

Citation:
Oprisan SA, Lynn PE, Tompa T and Lavin A (2015) Low-dimensional attractor for neural activity from local field potentials in optogenetic mice. Front. Comput. Neurosci. 9:125. doi: 10.3389/fncom.2015.00125

1. Introduction

Synchronization of neural oscillators across different areas of the brain is involved in memory consolidation, decision-making, and many other cognitive processes (Oprisan and Buhusi, 2014). In humans, sustained theta oscillations were detected when subjects navigated through a virtual maze by memory alone, relative to when they were guided through the maze by arrow cues (Kahana et al., 1999). Also the duration of sustained theta activity is proportional to the length of the maze. However, theta rhythm does not seem to correlate with decision-making processes. The duration of gamma rhythm is proportional to the decision time. Gamma oscillations showed strong coherence across different areas of the brain during associative learning (Miltner et al., 1999). A similar strong coherence in gamma band was found between frontal and parietal cortex during successful

recollection (Burgess and Ali, 2002). Cross-frequency coupling between brain rhythms is essential in organization and consolidation of working memory (Oprisan and Buhusi, 2013). Such a cross-frequency coupling between gamma and theta oscillations is believed to code multiple items in an ordered way in hippocampus where spatial information is represented in different gamma subcycles of a theta cycle (Kirihara et al., 2012; Lisman and Jensen, 2013). It is believed that alpha rhythm suppresses task-irrelevant information, gamma oscillations are essential for memory maintenance, whereas theta rhythms drive the organization of sequentially ordered items (Roux and Uhlhaas, 2014). Synchronization of neural activity is also critical, for example, in encoding and decoding of odor identity and intensity (Stopfer et al., 2003; Broome et al., 2006).

Gamma rhythm involves the reciprocal interaction between interneurons, mainly parvalbumin (PV+) fast spiking interneurons (FS PV+) and principal cells (Traub et al., 1997). The predominant mechanism for neuronal synchronization is the synergistic excitation of glutamatergic pyramidal cells and GABAergic interneurons (Parra et al., 1998; Fujiwara-Tsukamoto and Isomura, 2008).

Nonlinear time series analysis was successfully applied, for example, to extract quantitative features from recordings of brain electrical activity that may serve as diagnostic tools for different pathologies (Jung et al., 2003). In particular, large-scale synchronization of activity among neurons that leads to epileptic processes was extensively investigated with the tools of nonlinear dynamics both for the purpose of early detection of seizures (Jerger et al., 2001; Iasemidis, 2003; Iasemidis et al., 2003; Paivinen et al., 2005) and for the purpose of using the nonlinearity in neural network response to reset the phase of the underlying synchronous activity of large neural populations in order to disrupt the synchrony and re-establish normal activity (Tass, 2003; Greenberg et al., 2010). A series of nonlinear parameters showed significant change during ictal period as compared to the interictal period (Babloyantz and Destexhe, 1986; van der Heyden et al., 1999) and reflect spatiotemporal changes in signal complexity. It was also suggested that differences in therapeutic responsiveness may reflect underlying distinct dynamic changes during epileptic seizure (Jung et al., 2003).

The present study performed nonlinear time series analysis of LFP recordings from PV+ neurons: (1) to determine if nonlinearity is present using time reversal asymmetry and FNN statistics between the original signal and surrogate data; (2) to measure the phase shift (resetting) induced by brief light stimuli, and (3) to compute the delay (lag) time and embedding dimension of LFP data.

We investigated the response of the local neural network in the mPFC activated by light stimuli and determined the number of degrees of freedom necessary for a quantitative, global, description of the steady activity of the network, i.e., long after the light stimulus was switched off. Although each neuron is described by a relatively large number of parameters, using nonlinear dynamics (Oprisan, 2002) it is possible to capture some essential features of the system in a low-dimensional space (Oprisan and Canavier, 2006; Oprisan, 2009). One possible approach to low-dimensional modeling is by using the method of phase resetting, which reduces the complexity of a neural oscillator to a lookup table that relates the phase of the presynaptic stimulus with a reset in the firing phase of the postsynaptic neuron (Oprisan, 2013).

We recently applied delay embedding to investigating the possibility of recovering phase resetting from single-cell recordings (Oprisan and Canavier, 2002; Oprisan et al., 2003). Although techniques for eliminating nonessential degrees of freedom through time scale separation were used extensively (Oprisan and Canavier, 2006; Oprisan, 2009), the novelty of our approach is that we used the phase resetting induced by light stimulus to quickly identify similar activity patterns for the purpose of applying delay embedding technique.

2. Materials and Methods

2.1. Human Search and Animal Research

All procedures were done in accordance to the National Institute of Health guidelines as approved by the Medical University of South Carolina Institutional Animal Care and Use Committee.

2.2. Experimental Protocol

Male PV-Cre mice (B6; 129P2 - Pval$^{btm1(Cre)Arbr/J}$) Jackson Laboratory (Bar Harbor, ME, USA) were infected with the viral vector [AAV2/5. EF1a. DIO. hChR2(H134R) - EYFP. WPRE. hGH, Penn Vector Core, University of Pennsylvania] delivered to the mPFC as described in detail in Dilgen et al. (2013).

Electrophysiological data were recorded using an optrode positioned with a Narishige (Japan) hydraulic microdrive. Extracellular signals were amplified by a Grass amplifier (Grass Technologies, West Warwick, RI, USA), digitized at 10 kHz by a 1401plus data acquisition system, visualized using Spike2 software (Cambridge Electronic Design, LTD., Cambridge, UK) and stored on a PC for offline analysis. Line noise was eliminated by using a HumBug 50/60 Hz Noise Eliminator (Quest Scientific Inc., Canada). The signal was band-pass filtered online between 0.1 and 10 kHz for single- or multi-unit activity, or between 0.1 and 130 Hz for local field potentials (LFP) recordings.

Light stimulation was generated by a 473 nm laser (DPSS Laser System, OEM Laser Systems Inc., East Lansing, MI, USA), controlled via a 1401plus digitizer and Spike2 software (Cambridge Electronic Design LTD., Cambridge, UK). Light pulses were delivered via the 50 μm diameter optical fiber glued to the recording electrode (Thorlabs, Inc., Newton, NJ, USA).

At the top of the recording track the efficacy of optical stimulation was assessed by monitoring single-unit or multi-unit responses to various light pulses (duration 10–250 ms). High firing rate action potentials, low half-width amplitude (presumably from PV-positive interneurons) during the light stimulation, and/or the inhibition of regular spiking units was considered confirmation of optical stimulation of ChR2 expressing PV+ interneurons. The optrode was repositioned along the dorsal ventral axis if no response was found. Upon finding a stable response, filters were changed to record field potentials (0.1–100 Hz). Two different optical stimulations were delivered: (1) a 40 Hz 10-pulse train that lasted 250 ms with 10

ms pulse duration followed by a 15 ms break, and (2) a single pulse with 10 ms duration. In both cases, the recording lasted for 2 s from the beginning of optical stimulus. Local field potential (LFP) activity was monitored for a minimum of 10 min while occasionally stimulating at 40 Hz to ensure the stability of the electrode placement and the ability to induce the oscillation. Additionally, LFP activity was monitored as a tertiary method of assessing anesthesia levels. Several animals were excluded from analysis due to fluctuating levels of LFP activity that resulted from titration of anesthesia levels during the experiment.

3. Data Analysis

For each of the six animals, we analyzed 100 different trials, each with a duration of 2 s measured from the onset of a brief 10 ms stimulus until the next stimulus. For each 2 s long LFP recording, there are two regions of interest: the first approximately 0.5 s that follows the stimulus, which is the transient response of the neural network, and the last 1.5 s of the recording that is the steady activity of the network. The transient response is essential in the subsequent analysis of the steady response since it determines the amount of phase resetting induced by optical stimulus (see Section 3.2 for a detailed description of the procedure employed to determine the phase resetting induced by a light stimulus). The steady activity of the network was investigated to determine if there is any low-dimensional attractor that may explain the observed dynamics.

3.1. Tests for Nonlinearity

Detection of nonlinearity is the first step before any nonlinear analysis. The test is necessary since noisy data and an insufficient number of observations may point to nonlinearity of an otherwise purely stochastic time series (see for example Osborne and Provencale, 1989). There are at least two widely-used methods for testing time series nonlinearities: surrogate data (Theiler et al., 1992; Small, 2005) and bootstrap (Efron, 1982). The most commonly used method to identify time series nonlinearity is a statistical approach based on surrogate data technique. The bootstrap method extracts explicit parametric models from the data (Efron, 1982).

In the following, we will only use the surrogate data method. Testing for nonlinearity with surrogate data requires an appropriate null hypothesis, e.g., that the data are linearly correlated in the temporal domain, but are random otherwise. Once a null hypothesis was selected, surrogate data are generated for the original series by preserving the linear correlations within the original data while destroying any nonlinear structure by randomizing the phases of the Fourier transform of the data (Theiler et al., 1992).

From surrogates, the quantity of interest, e.g., the time reversal asymmetry, is estimated for each realization. Next, a distribution of the estimates is compiled and appropriate statistical tests are carried out with the purpose of determining if the observed data are likely to have been generated by the process set though the null hypothesis. If the selected measure(s) of suspected nonlinearity does not significantly change between the original and the surrogate data, then the null hypothesis is true, otherwise the null hypothesis is rejected.

The number of surrogates to be generated depends on the rate of false rejections of the null hypothesis (Jung et al., 2003). For example, if a significance level of $l = 0.05$ is desired, then at least $n = 1/l = 20$ surrogates need to be generated (Jung et al., 2003; Yuan et al., 2004). A set of values λ_i (with $i = 1, \ldots, n$) of the discriminating statistics is then computed from the surrogates and compared agains the value λ_0 for the original time series. Rejecting the null hypothesis can be done using: (1) rank ordering or significance testing, (2) the average method (Yuan et al., 2004), or (3) the coefficient of variation method (Theiler et al., 1992; Kugiumtzis, 2002; Jung et al., 2003).

In rank ordering, λ_0 must occurs either on the first or on the last place in the ordered list of all values of the discriminating statistics to reject the null hypothesis (see the null hypothesis rejection using FNN Section 4.2).

In the average statistical method, a score γ (sometimes called a Z-score) is derived as follows:

$$\gamma = \left| \frac{\bar{\lambda}}{\lambda_0} - 1 \right|,$$

where $\bar{\lambda} = \frac{1}{n} \sum_{i=1}^{n} \lambda_i$ is the mean value of the discriminating statistics over all surrogates. If the score γ is much less than 1, then the relative discrepancy can be considered negligible. If γ is greater than 1, then the original data and the surrogates are significantly different and the null hypothesis is rejected.

In the coefficient of variation statistical method, a score γ is derived as follows:

$$\gamma = \left| \frac{\bar{\lambda} - \lambda_0}{\sigma_\lambda} \right|, \quad (1)$$

where σ_λ is the standard deviation of the discriminating statistics over all surrogates. If the values λ_i are fairly normally distributed, rejection of the null hypothesis requires a γ- value of about 1.96 at a 95% confidence level (Stam et al., 1998; Jung et al., 2003).

For every trial and every animal we generated $n = 100$ surrogates and used two different discriminating statistics to detect potential nonlinearity in our data. The first γ score was based on the reversibility of the time series. The second discriminating statistics was based on the percentage of false nearest neighbors (see Section 4.2).

A time series is said to be reversible only if its probabilistic properties are invariant with respect to time reversal (Diks et al., 1995). Time irreversibility is a strong signature of nonlinearity (Schreiber and Schmitz, 2000) and rejection of the null hypothesis implies that the time series cannot be described by a linear Gaussian random process (Diks et al., 1995). We used the Tisean function *timerev* to compute the time reversal asymmetry statistics both for the original and the surrogate data (Hegger et al., 1999; Schreiber and Schmitz, 2000). The 100 surrogate data files for each of the 100 trials were generated using Tisean function *surrogate* (Hegger et al., 1999; Schreiber and Schmitz, 2000).

Figure 1A shows one of the original time series (continuous blue line) together with one of its 100 surrogates (dashed red

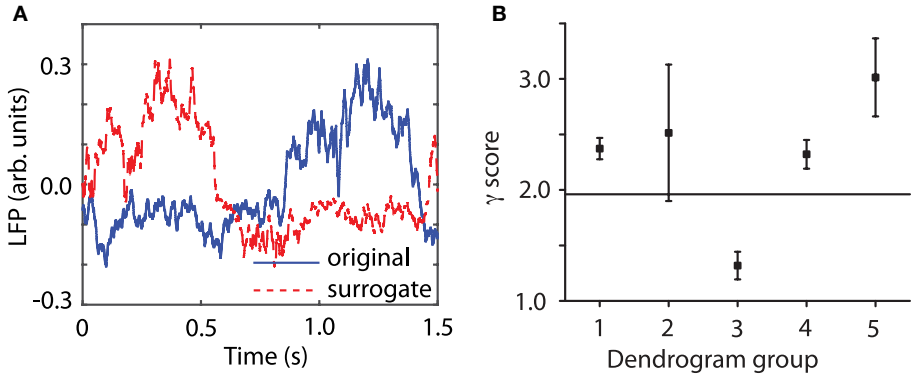


FIGURE 1 | Surrogate data. An ensemble of 100 surrogate data sets similar to the original time series, but consistent with the null hypothesis, were generated using Tisean. **(A)** The original data from a randomly selected trial (blue continuous line) and one of its 100 surrogates (dashed red line) look similar. For each trial, 100 surrogates have been created by a stationary Gaussian linear process using the function *surrogate* of Tisean. **(B)** The discriminating statistic for the original trials and for each of their surrogates over all five groups showed that only the third group does not meet the nonlinearity criterion (the horizontal continuous line) since its γ score is less than 1.96. For all the other four groups the null hypothesis can be rejected.

line). Although the two data sets might look similar, the time reversal asymmetry value for the original data was $\lambda_0 = 0.1893$ and for the surrogate data shown in **Figure 1A** it was $\lambda = 2.4948$. The fact that the surrogates are significantly different from the original data means that, for example, the delay embedding dimension for surrogates is different than for the original data. Indeed, we found that the embedding dimension is higher for surrogates (see **Figure 6C**). It also means that the surrogates do not unfold correctly in the lower-dimensional embedding space of the original data (see Supplementary Materials). We used the coefficient of variation statistical method to compute a γ score from Equation (1). **Figure 1B** shows all γ scores for the first animal. The statistics was computed over groups of original data lumped together based on their "similarity" as determined after correcting for phase resetting induced by the light stimulus (see Section 3.2 below for details) and using the dendrogram (see Section 3.3 below for details). The average γ score of time reversal asymmetry statistics that was computed from individual λ_i- values for each trial in the third group was less than 1.96. Therefore, the null hypothesis that the data had been created by a stationary Gaussian linear process could not be rejected for this group of LFPs. For all the other groups of original data formed out of the 100 trials the γ score was above 1.96 and therefore we rejected the null hypothesis. Although this time reversal asymmetry discriminating statistics seems to exclude the third group of data, we also used the FNN discriminating statistics for all data (see Section 4.2). The FNN reflects the degree of determinism in the original data and therefore serves as a good choice for a discriminating statistic (Hegger et al., 1999; Yuan et al., 2004). Briefly, for the third group of data, which was rejected based on time reversal asymmetry discriminating statistics, we found that the percentage of FNN for all 100 surrogates computed for all trials in the respective group was always larger than for the original data (see **Figure 6C**). Therefore, based on both discriminating statistics, it is likely that nonlinearity is present in all our data.

3.2. Phase Resetting of LFP

LFPs are weighted sums of activities produced by neural oscillators in the proximity of the recording electrode (Ebersole and Pedley, 2003). In order to better understand the effect of a stimulus, such as a brief laser pulse on a neural network, we used a simplified neural oscillator model (see **Figure 2A**) that produced rhythmic activity. We used a Morris-Lecar (ML) model neuron (Morris and Lecar, 1981). When a noise free oscillator with intrinsic firing period P_i (see **Figure 2A**) is perturbed, e.g., by applying a brief rectangular current stimulus, the effect is a transient change in its intrinsic period. For example, a perturbation delivered at phase 0.3, measured from the most recent membrane potential peak, produces a delay of the next peak of activity (continuous blue trace in **Figure 2A**). On the other hand, an identical perturbation delivered to the same free running oscillator at a phase of 0.5 produces a significant advance of the next peak of activity (dashed red trace in **Figure 2A**). As we notice from **Figure 2A**, the cycles after the perturbation return pretty quickly to the intrinsic activity of the cell, i.e., the most significant effect of the perturbation is concentrated during the cycle that contains the perturbation. The induced phase resetting, i.e., the permanent phase shift of post-stimulus activity compared to pre-stimulus phase, depends not only on the strength and duration of the perturbation, but also on its timing (or phase).

One approach often used for reducing the noise is averaging multiple trials. How should a meaningful average be carried out to both reduce the noise and preserve the characteristics of the rhythmic pattern, such as amplitude, phase, and frequency? One possibility is to align all action potentials at stimulus onset and added them up (see the thick black trace in **Figure 2B**) to generate a LFP. In **Figure 2B** we also added a uniform noise to neural oscillator's bias current such that the individual traces are pretty rugged. The effect of noise is especially visible on the dashes and dashed-dotted traces in **Figure 2B** during the slow hyperpolarization. By adding 100 noisy action potential traces produced by resetting the neural oscillator at 100 equally

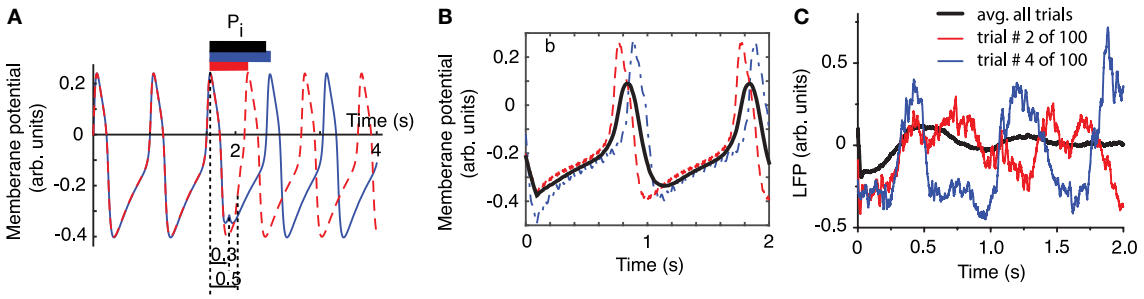


FIGURE 2 | Phase resetting of FLPs. The free-running neural oscillator was perturbed at different phases and, as a result, its phase was reset due to a transient change in the length of the current cycle during which the perturbation was active **(A)**. The intrinsic firing period P_i (see black bar on top of the third cycle that contains the perturbation) was shortened by a perturbation applied at phase $\phi = 0.5$ (see dashed red trace and the corresponding red bar on top of the third cycle). The same perturbation applied at phase $\phi = 0.3$ (measured from the peak of the action potential—see vertical dotted lines) lengthened the current cycle (see continuous blue trace and the corresponding blue bar on top of the third cycle). **(B)** The average membrane potential of 100 noisy traces (thick black line) perturbed at 100 equally spaced phases during the third cycle is less noisy and retains some low frequency oscillations present in all individual traces. All traces were aligned at stimulus onset and only two of them are shown (red dashed and blue dashed-dotted). **(C)** LFP recordings also aligned at laser stimulus onset show an average LFP trace (black thick trace) that is almost noise free and retains some spectral characteristics of its components. At the same time, the shape of the average LFP trace is significantly different from any individual traces.

spaced phases we produced a smooth average (see the thick black trace in **Figure 2B**). Therefore, on the positive side, we could use a (weighted) sum of noisy traces to reduce the noise in our data. The other positive outcome is that the (weighted) sum retains some of the characteristics of the individual traces, such as the intrinsic firing frequency. However, we also notice form **Figure 2B** that the shape of the (weighted) average is quite different from any of its constituents, which raises the question: is this averaging procedure the right ways of computing a (weighted) average from individual trials? Based on **Figures 2A,B**, we can conclude that the mismatch between the average (black thick line) and the individual trials (blue and red traces) is due to the fact that the periodically delivered stimulus found the background oscillatory activity of the neuron at different phases, therefore, produced different phases resettings. Without correcting for the stimulus induced phase resetting effect on each trial we lose the phase and amplitude information by simply adding all individual traces. We noticed the same effects when attempting to remove the noise in out LFP data be averaging all trials aligned at the onset of the light stimulus (see **Figure 2C**). As a result, whenever performing an averaging of noisy rhythmic patterns for the purpose of reducing the noise, first the individual traces must be corrected for the phase resetting induced by the external stimulus.

After dropping the 0.5 s transient, we noticed that even very similar LFP traces, such as those shown in **Figure 3A**, do not overlap perfectly due to the phase resetting (or the permanent phase shift) induced by light stimuli that arrived at different phases of the LFP activity.

In order to correct the LFP recordings for the phase resetting induced by the brief laser pulse, we performed a circular shift of each LFP trace with respect to one, arbitrarily selected, trace that was considered as a "reference" LFP. The phase resetting maximized the coefficient of correlation between any trial and the arbitrary "reference" (see **Figure 3B**). As a result of the circular shift, the coefficient of correlation increased significantly from an average of 0.0143 ± 0.055 (red trace in **Figure 3C**) to 0.5854 ± 0.1383 (blue trace in **Figure 3C**). Additionally, the root-mean-square (rms) error, i.e., the Euclidian norm of the difference between each 1.5 s long trial and the "reference" trial, was computed (see **Figure 3D**). The rms error before circularly shifting the trials was 13.4 ± 2.9. By circularly shifting the trials to remove the effect of phase resetting induced by the light stimulus, we were able to decrease the rms error to 8.5 ± 1.8 (see green curve with squares in **Figure 2D**).

3.3. Dendrograms of Phase Shifted LFPs

The circular shift performed in the previous section with the purpose of maximizing the coefficient of correlation between any trial and an arbitrary "reference" helps correctly defining the relative phase of trials with respect to each other. Another helpful step in the process of automatic data classification before attempting a delay embedding reconstruction was to separate the trials in "similar"-looking groups. Since we were interested in finding out if there is any attractor of network's steady activity, it is expected that phases space traces of different trials would remain close to each other at all times. This implies that individual recordings present some "similarities" that could be detected using the dendrograms, e.g., for the purpose of separating clean data from artifacts (due to malfunction of laser trigger, etc.) We used dendrograms to find the similarity trees of all 1.5 s long, phase-corrected, trials that allowed us to further decrease the rms error to an arbitrarily selected "reference" from the same group (see blue solid circles in **Figure 3D**). The dendrogram in **Figure 4A** used the Euclidian distance to measure similarities between the phase-shifted LFP trials.

The dendrogram could be used, for example, to separated groups of trials based on an arbitrary selection of the cutoff distance along dendrogram's trees. For example, by selecting a cluster distance larger than 40 (see **Figure 4A**) all 100 trials belong to just one group. As already discussed, lumping all trails

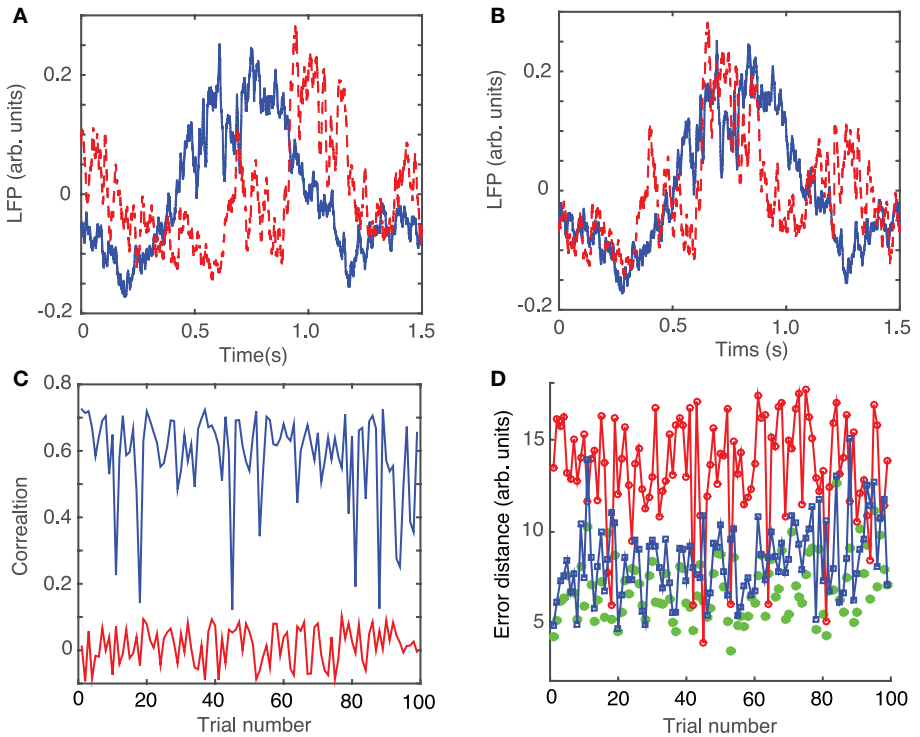


FIGURE 3 | Phase resetting correction of LFPs. Steady LFP activity recorded 0.5 s after the 10 ms stimulus switched off look similar **(A)** and can be better overlapped by an appropriate circular shifting **(B)** that removes the waveform phase shift due to phase resetting induced by the same light stimulus arriving at different phases during the ongoing rhythm. **(C)** Without phase shifting to correct for the phase resetting, the correlations between the waveforms of different trials with respect to an arbitrarily selected "reference" trial are relatively small at an average of 0.0143 ± 0.055 (red trace). A significant improvement in pair correlation between trials occurs after appropriately shifting the waveforms to maximize the correlation coefficient (blue trace) with an average correlation of 0.5854 ± 0.1383. **(D)** Similar to correlation, the root-mean-square error between a trial and the corresponding "reference" trial significantly decreases. The rms error decreases form 13.4 ± 2.9 for correlation between trials without phase resetting correction (red line), to 8.5 ± 2.1 after phase-shifting all trials to correct for phase resetting (blue line), to 6.9 ± 1.8 for phase-shifted dendrogram-based correlation (green solid circles).

in one group may inadvertently lump together low-dimensional attractors with data affected by various equipment malfunctions. Such an approach would make the task of identifying any phase space attractor to which all trajectories remain close at all times more computationally intensive. By decreasing the cluster distance threshold, we could form two groups or more. In the following, we used a cutoff cluster distance close to 20 and obtained five dendrogram-based groups (see the shaded rectangles in **Figure 4A**). The plots of the LFPs for each of the first three groups (**Figure 4B**) show pretty similar waveforms and quite different form the last two groups of the dendrogram (see **Figure 4C**). Therefore, it may be easier to visually identify an attractor (if one exists) by looking at reconstructed attractor of individual trials from the same group, for example by comparing traces from group 1 against each other (see **Figure 7B1**). The same is true when comparing trials from group 5 against each other (see **Figure 7B5**). It is unlikely that we would be able to find any trials from group 1 that remain close to any trials from group 5, a fact that we learned during data preprocessing stage using dendrograms.

The same numerical procedure was applied to all data from six animals of which we only show one detailed example.

4. Delay Embedding Method

Given the complexity of a single pyramidal neuron and the intricacy of synaptic coupling in the mPFT cortex (Schnitzler and Gross, 2005), we would expect a rather high-dimensional delay embedding for our LFP recordings.

In electrophysiology, we record the membrane potential time series, which is just one of many independent variables required for a full characterization of neural network activity. Even though we have direct access to only one variable of the $d-$dimensional dynamical system, i.e., the light-activated local network, it is still possible to faithfully recover, or reconstruct, the phase space dynamics through delay embedding method (Abarbanel, 1996; Kantz and Schreiber, 1997; Schuster and Just, 2005; Kralemann et al., 2008). For a time series $x_i = x(i\Delta t)$ with $i = 1, 2, \ldots, N$ where N is the number of data points and Δt is the (uniform) sampling time, a $d-$dimensional embedding vector is defined as

$$x_i = (x_i, x_{i+n}, \ldots, x_{i+(d-1)n}),$$

where $\tau = n\Delta t$ is the delay, or lag, time (Packard et al., 1980; Takens, 1981).

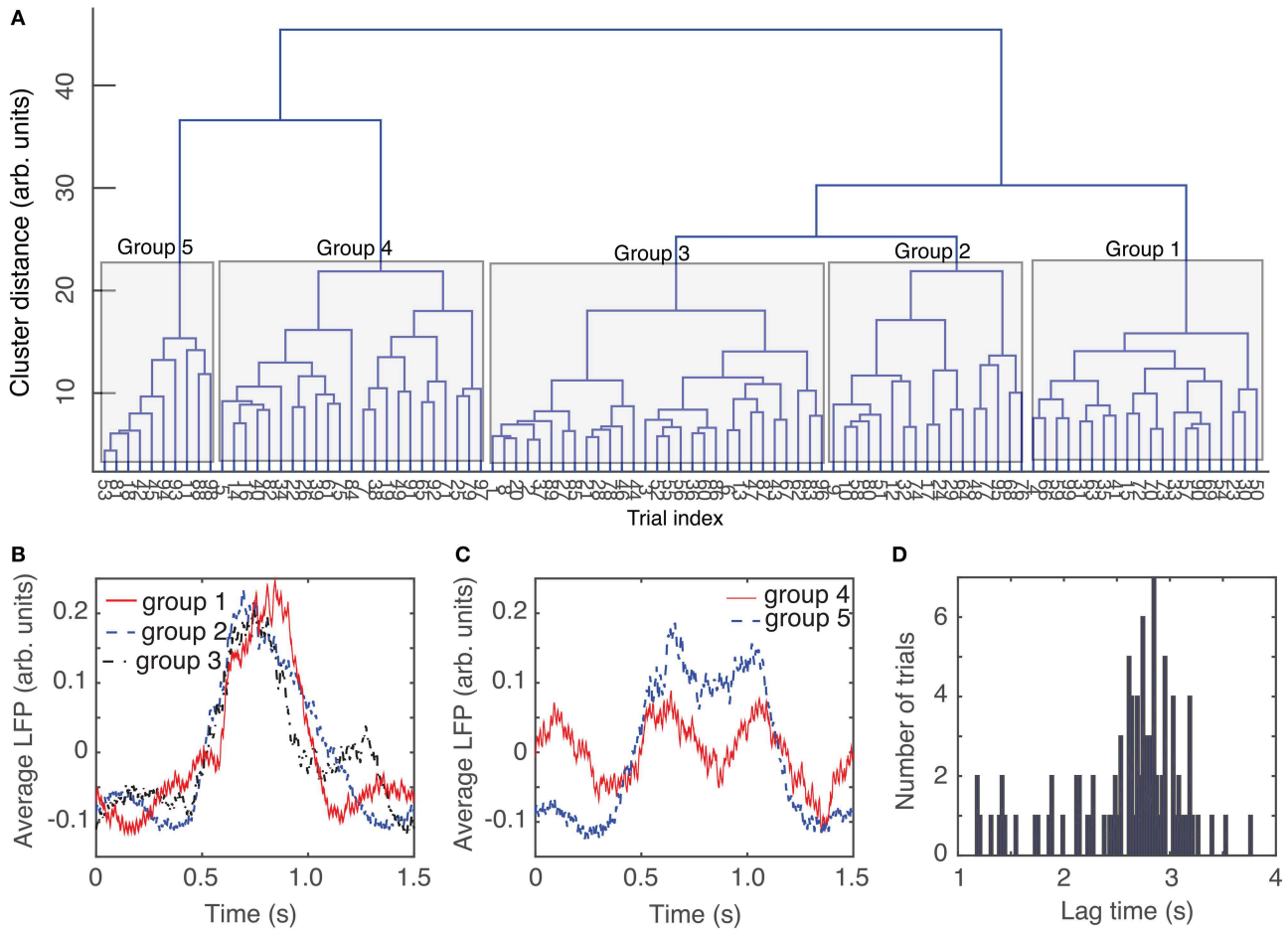


FIGURE 4 | Dendrogram-based grouping of similar LFPs. Even after maximizing the correlation between trials by phase-shifting them to correct for phase resetting effect, the distance to an average LFP "template" can be further improved by grouping LFP waveforms according to their Euclidian-based dendrogram **(A)**. Depending on the cutoff threshold for the distance along a tree, we can have a very coarse representation with only one group (for distance larger than 40) or five groups for distance around 20 (see the shaded areas that mark different groups). As the dendrogram suggests, the average LFPs for the first three groups are quite similar **(B)**. The average LFP for the last two groups are also very similar **(C)**, but quite different from the previous three groups. The resulting rms error of a trial with respect to its corresponding group average is definitely an improvement over the simple averaging of all trials (see **Figure 3D**). Even though the group average could be pretty close to capturing features of individual LFP trials form the respective group, the delay time for phase space reconstruction has a wide range of values **(D)** and there is no obvious group correlation.

Two parameters are essential for a correct delay embedding reconstruction of the phase space: the lag time τ and the embedding dimension d_E. The delay, or lag, time τ is the time interval between successive components of the embedded vector. Although we assumed that the same delay time applies to each component of the embedded vector, the delay embedding method also allows for different delays along different directions of the phase space (Vlachos and Kugiumtzis, 2010).

4.1. Lag Time

The quality of phase space reconstruction is affected, among other factors, by the amount of noise, the length of the time series, and the choice of the delay time. For example, a too small delay time τ leads to embedded vector with highly correlated, or indistinguishable, components. Geometrically, this means the all trajectories are near the diagonal of the embedding space and the attractor has a dimension close to one irrespective of its complexity. To avoid such *redundancy*, the delay time τ should be large enough to make the components of the embedded vector independent of each other. However, a too large delay time completely de-correlates the components of the embedded vector. Geometrically, this means that phase space points fill the entire embedding space randomly and the attractor has a dimension close to the embedding space dimension. Although there is no universal method for selecting the "right" delay time, in practice we use a few different approaches to avoid both the *redundancy* due to a too short delay time and the *irrelevance* due to a too large delay time (Casdagli et al., 1991).

One of the methods often used for estimating the lag time τ is the *autocorrelation* of the time series. Although researchers agree that autocorrelation could provide a good estimation of

the time lag, there is no consensus regarding the specifics. For example, Zeng et al. (1991) considered that τ is the time at which the autocorrelation decays to e^{-1}, Schiff and Chang (1992) considered the first time when the autocorrelation is not significantly different from zero, Schuster (Schuster and Just, 2005) suggested using the first zero of autocorrelation function to ensure linear independence of the coordinates, King et al. (1987) considered the time of the first inflection of the autocorrelation, and Holzfuss and Mayer-Kress (1986) considered the first time the autocorrelation reaches a minimum.

In addition to autocorrelation, Fraser and Swinney (1986) suggested using the first local minimum of *the average mutual information* (AMI) to estimate the time lag. Their method measures the mutual dependence between x_i and x_{i+n} with variable lag time $n\Delta t$ (see also Kantz and Schreiber, 1997; Hegger et al., 1999).

Additionally, the total time spanned (Broomhead and King, 1986) by each embedded vector, i.e., $t_w = (d-1)\tau$, is a significant measure of potential crossover between temporal correlation that could induce spurious spatial, or geometrical, correlation between phase space points (Theiler, 1990).

4.2. Embedding Dimension

The embedding dimension was selected based on Takes's theorem (Takens, 1981) that ensured a faithful reconstruction of a $d-$dimensional attractor in an embedding space with at most $2d+1$ dimensions. For a dissipative system, Hausdoff dimension could be estimated from a time series and used as the dimension of the attractor (Holzfuss and Mayer-Kress, 1986; Kennel et al., 1992; Provenzale et al., 1992). Good estimators of Hausdorf's dimension are the correlation dimension (Grassberger and Procaccia, 1983) or the Lyapunov dimension (Kaplan and Yorke, 1979). Once the range ($d \leq d_E \leq 2d+1$) of embedding dimensions is known, additional tests could determine the optimum embedding dimension d_E.

Kennel et al. (1992) introduced the false nearest neighbors (FNN) procedure to obtain the optimum embedding dimension (see also Kennel et al., 1992; Hegger et al., 1999; Sen et al., 2007). The idea behind FNN approach is to estimate the number of points in the neighborhood of every given point for a fixed embedding dimension. High dimensional attractors projected onto a too low dimensional embedding space show a significant number of false neighbors, i.e., phase space points that look close to each other although in the true attractor space they are far apart. The FFN method compares the Euclidian distance R_d between two neighbors x_i and x_j computed in a $d-$dimensional space against the distance R_{d+1} in a $(d+1)-$dimensional embedding space (Kennel et al., 1992). If the ratio of relative distances between neighbors in the two embedding spaces, i.e., $f = \sqrt{\frac{R_{d+1}^2 - R_d^2}{R_d^2}}$, is larger than a predefined value then the two points x_i and x_j are false neighbors, i.e., the points are neighbors because of a too low projection and not because of the true dynamics. The ratio f is usually set between 1.5 and 15 (Kennel et al., 1992; Abarbanel, 1996; Kantz and Schreiber, 1997). Additionally, if the distance R_{d+1} is larger than the coefficient of variation $\sigma/\bar{x}$ of the data then the two points are false neighbors. The reason is that σ is a measure of the size of the attractor and two points that are false neighbor will be indeed stretched to the extremities of the attractor in dimension $d+1$. Abarbanel (1996) found that for many nonlinear systems the value of f approaches 15, but the range is quite wide from 9 to 17 (Konstantinou, 2002). By successively computing the fraction of FNNs in different embedding dimensions, it is possible to estimate an optimum embedding. Some algorithms that takes into account the temporal window $t_w = (d-1)\tau$ spanned by the embedded vectors allow simultaneous estimation of both embedding dimension and lag time (see Stefánsson et al., 1997).

5. Results

5.1. Experimental Data

Since we were interested in uncovering any possible attractor of phase space trajectories, we only considered the last 1.5 s of each 2 s long recording. We first performed a phase shift of every 1.5 s long LFP recording to correct for the phase resetting due to light stimulus (see **Figure 3B** for two similar-looking LFT traces that were phase-shifted with respect to each other to maximize the correlation coefficient and correct for the phase resetting effect).

5.2. Lag Time

As described in Section 4.1, we used two different approaches to estimating the lag time τ: (1) the autocorrelation function (Casdagli et al., 1991), and (2) the AMI method (Fraser and Swinney, 1986). The first zero crossing of the autocorrelation function is the time τ beyond which $x(t+\tau)$ is completely de-correlated from $x(t)$. However, the first zero crossing of the autocorrelation function takes into account only linear correlations of the data (Abarbanel, 1996). The first minimum of the nonlinear autocorrelation function called *Average Mutual Information* (AMI) (Fraser and Swinney, 1986) is considered a more suitable choice since this is the time when $x(t+\tau)$ adds maximum information to the knowledge we have from $x(t)$ (Kantz and Schreiber, 1997). In most practical applications the two methods are used together and they usually give similar estimations of the lag time.

We computed the lag times for individual trials (see **Figure 4D** for the distribution of all lag times for animal # 1) and also for group averages (see **Table 1**). Although only the autocorrelation-based lag time are shown both in **Figure 4D** and **Table 1**, the AMI-based lag time values (not shown) were within 10% of those obtained with the autocorrelation.

TABLE 1 | Estimated lag times.

Mouse #	Avg.	Std.	Group 1	Group 2	Group 3	Group 4	Group 5
1	2599	542	1504	2962	2886	2578	2721
2	3150	885	3128	2982	4337	3390	3401
3	1759	483	2297	1814	1812	2203	
4	2645	708	3394	2924	2611	3332	
5	1842	708	1501	1722	1717	2286	
6	1661	594	1518	1767	1583	1736	1374

In **Table 1**, the second column (called Avg.) and the third columns (called Std.) represent the average, respectively, the standard deviation of the corresponding lag time distributions, such as the one shown in **Figure 4D** for animal # 1. The next columns in **Table 1** represent the lag times of the dendrogram-based group averages.

For example, for the first animal, the first zero crossing of the autocorrelation function for dendrogram-based average LFP of group 1 is around $\tau \approx 1500\Delta t$ (see **Figure 5A**), whereas the first minimum of the AMI is around $\tau \approx 2000\Delta t$ (see **Figure 5B**).

Our data were stored as single-column text files representing the LFP recordings with a sampling rate of $\Delta t = 10^{-4}$ *s*. Tisean command for estimating the lag time from autocorrelation function was *autocor dataFile.txt -p -o*, where the option $-p$ specified periodic continuation of data and $-o$ specified that the expected output will be returned to a file named *dataFile.txt.co*, which is plotted in **Figure 5A**.

Tisean command for estimating the lag time from the AMI was *mutual dataFile.txt -D10000 -o*, where the option $-D10000$ specified the range of lag times for which AMI was computed and stored in the file *dataFile.txt.mut*, which is plotted in **Figure 5B**.

5.3. Embedding Dimension

The method of false nearest neighbors (FNN) estimates the embedding dimension d_E by repeatedly increasing the embedding dimension until the orbits of the phase space flow do not intersect or overlap with each other. We used a lag time $\tau = 2200\Delta t$ and estimated the embedding dimension using FNN method with ratios f between 2 and 20 (see **Figure 6A**). As expected, for large ratios of distances, e.g., $f > 7$, the percentage of FNNs drops to almost zero for an embedding dimension $d_E = 3$.

The actual Tisean routine used was *false_nearest dataFile.txt -f2 -d2200 -o*, which calculated the percentage of FNNs with a ratio $f \geq 2$, a lag time $d = 2200\Delta t$, with the default phase space dimensions from 1 to 5 (see **Figure 6A**). **Figure 6A** clearly indicates that an embedding dimension $d_E = 3$ is sufficient.

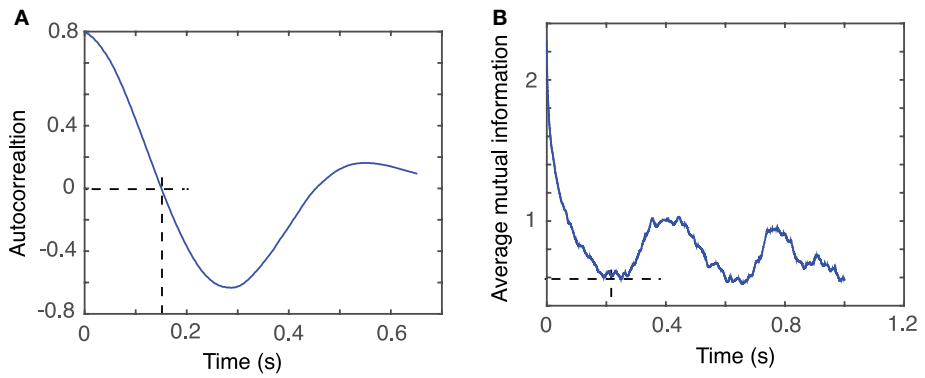


FIGURE 5 | Time lag estimation. The first zero crossing of autocorrelation function is around $\tau \approx 1500\Delta t$ **(A)** and the first minimum of the average mutual information is around $\tau \approx 2000\Delta t$ **(B)** with $\Delta t = 10^{-4}$ *s*.

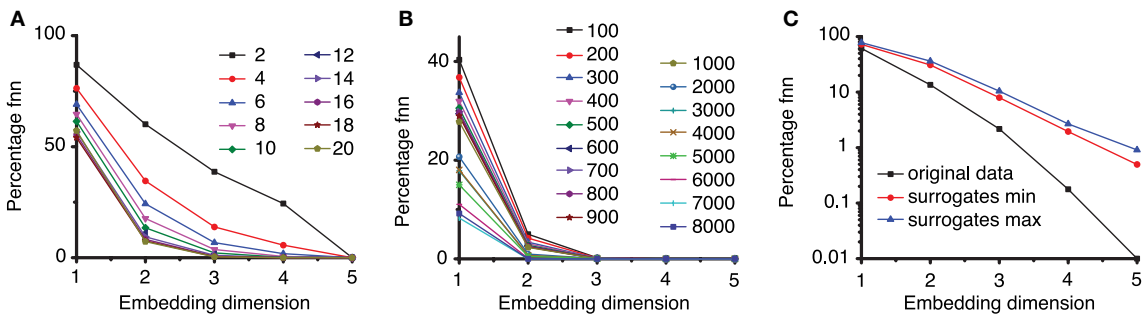


FIGURE 6 | Percentage of false nearest neighbors. (A) For a too small ratio $f < 7$ of distances between neighbor points in different embedding dimensions, the percentage of false nearest neighbors is high and only drops near zero for very large embedding dimensions. For larger ration $f > 7$ all percentages drop to almost zero false nearest neighbors for an embedding dimension of $d_E = 3$. This suggests that an optimum ratio is above $f = 7$, in agreement with results from others (Abarbanel, 1996; Konstantinou, 2002). **(B)** To avoid spurious spatial correlations due to inherent temporal correlation between too closely spaced points in a time series, the percentage of FNN was estimated with variable Theiler window (*t*). **(C)** The percentage of FNN is also a good discriminating statistics. For the third group of data from the first animal, the logarithmic plot shows that the percentage of FNN for the original data (solid squares) is always smaller than any of the 100 surrogates. Only the envelopes of the minimum (solid circles), respectively, maximum (solid triangles) values of FNN are shown.

Any estimate of dimension, especially when it is based on correlation among data points, assumes that pairs of points are drawn randomly and independently according to the scale invariant measure of the attractor. However, points occurring close in time are not independent and lead to spuriously low estimates of embedding dimension. To avoid this issue, points closer than some minimum time (called the Theiler window) can be excluded from calculations (Grassberger, 1987; Theiler, 1990). Heuristic examples of estimates of Theiler window are three times the correlation time (Heath, 2000), $(d-1)\tau$, or other *ad hoc* values based on space-time separation plots (Provenzale et al., 1992).

In our estimation of embedding dimension with FNN method, we also tested a wide range of Theiler windows from 100 to 8000 sampling times (**Figure 6B**) in order to make sure that no spurious temporal correlation among data points led us to a too low estimation of the embedding dimension. All plots of the fraction of FNNs indicated that $d_E = 3$ is still a good choice of the embedding dimension. The actual Tisean routine was *false_nearest dataFile.txt -f20 -d2200 -t100 -o*, which calculates the percentage of FNNs with a ratio greater than $f = 20$, a lag time of $d = 2200\Delta t$, a Thriller windows $-t$ of $100\Delta t$ for all embedding dimensions from 1 to 5 (see **Figure 4B**).

The attractors were reconstructed (see **Figure 7**) using the time lag τ and embedding dimension d_E as determined above. AAs seen form **Figures 7A1–A5**, the dendrogram-based preprocessing separated quite well the LFP waveforms in "similar" groups such that randomly selected LFPs from the same group remained close to each other at all times (see red and green traces in **Figures 7B1–B5**). The reconstruction of individual trials was performed with their corresponding delay (lag) times (see **Figure 4D** for the distribution of all delay times for the first animal). We also showed the reconstructed group average (blue thick trace in **Figures 7B1–B5**) not because it represents the "true" attractor, but rather as a visual cue to help us gauge if the phase space trajectories of the individual trials remained close to each other and at all times. As expected from the dendrogram-based preprocessing, the first three groups gave very similar reconstructed attractors. The shape of attractors from the first three groups could be roughly described as a continuous circular loop twisted in an "8"-shaped object (see **Figures 7B1–B3**). Since the group average (blue thick line) is less noisy than the individual trials (red and green lines) it serves as a visual aid toward identifying the shape of the attractor suggested by the individual trials. The shape of the first group's attractor (**Figures 7B1–B3**) could be viewed as an "8"-shaped loop bent around its midpoint (see also Supplementary Materials Video). However, by increasing the lag time, the "8"-shaped attractor can be "untangled" such that the two loops look more like the circles shown in **Figures 7B2,B3**. For example, in **Figure 8** we showed two examples of the same trials (red and green lines) together with their corresponding group average (thick black trace) that were reconstructed in the three dimensional phase space using different delay times. In **Figure 8A** for $\tau = 1900$ we clearly notice the twisted "8" shaped attractor that looks straight in **Figure 8B** for a delay time of $\tau = 2200$. Therefore, all attractors in **Figures 7B1–B3** are topologically identical (up to some microscale details) since any of them could be morphed into another by a (circular) phase shift. Furthermore, a close inspection of the fourth's group attractor shows that it is close to the previous three and quite different from the fifth attractor.

The detailed procedure described above was also applied to the other five data sets from different animals. The results are summarized in **Figures 9–13**. For all six animals that were retained and analyzed, the zero crossings of the autocorrelation and the minimum the AMI gave consistent lag time estimations (see **Table 1**).

We found that for all six animals the optimum delay embedding dimension was $d_E = 3$. We found topologically identical attractors in all first four LFP dendrogram-based groups for animal #1 (see **Figures 7B1–B4**), which cover 90% of the recordings. The attractor is "8"-shaped and is topologically equivalent (after appropriate phase shifting) with an "untangled" attractor (see **Figure 8**).

For animal #2, all attractors belong to the same "8"-shaped class or its topologically identical counterparts (see **Figures 9B1–B5**), although the fifth group presented a very large variability.

For animal #3, there were three topologically identical dendrogram-based LFP groups that gave an "8"-shaped attractor (see **Figures 10B1–B3**), which covered 84% of recordings.

For animal #4, all attractors were topologically identical that belonged to the "8"-shaped attractor (see **Figures 11B1–B4**), although the fourth group presented a very large variability.

For animal #5, there were again three topologically identical dendrogram-based LFP groups that belonged to the "8"-shaped attractor (see **Figures 12B1–B3**), which covered 74% of recordings.

For animal #6, there were two topologically identical dendrogram-based LFP groups that belonged to the "8"-shaped attractor (see **Figures 13B1,B2**), which covered 34% of recordings.

An important characteristic of the attractors that were not included in the above category of "8"-shaped attractors or their topological equivalents is that all of them showed relatively low amplitude oscillations of the LFP. For example, while the peak-to-peak amplitude of LFP oscillations for the four topologically equivalent attractors shown in **Figures 7B1–B4** was between -0.15 and $+0.25$ arb. units, the amplitude of the LFP for the last group was between -0.075 and 0.075 arb .unit., which is a decrease by a factor of 2.6. Similarly, for animal #3, the range of LFP for the "8"-shaped attractor or their topological equivalents (see **Figures 10B1–B3**) was between -0.4 to $+0.7$ arb. units whereas for the only dissimilar group the LFP amplitude was between -0.1 to $+0.1$, a decrease in amplitude of LFP by a factor of 5.5. For animal #5, the decrease in amplitude of LFP only by a factor of 1.5 and for animal #6 the factor was 2.5. One possible explanation could be an intermittent malfunction of the laser's trigger. The dendrogram method helped us automatically sort the data set into "similar" groups before performing a delay embedding. As a result, we decreased the computational

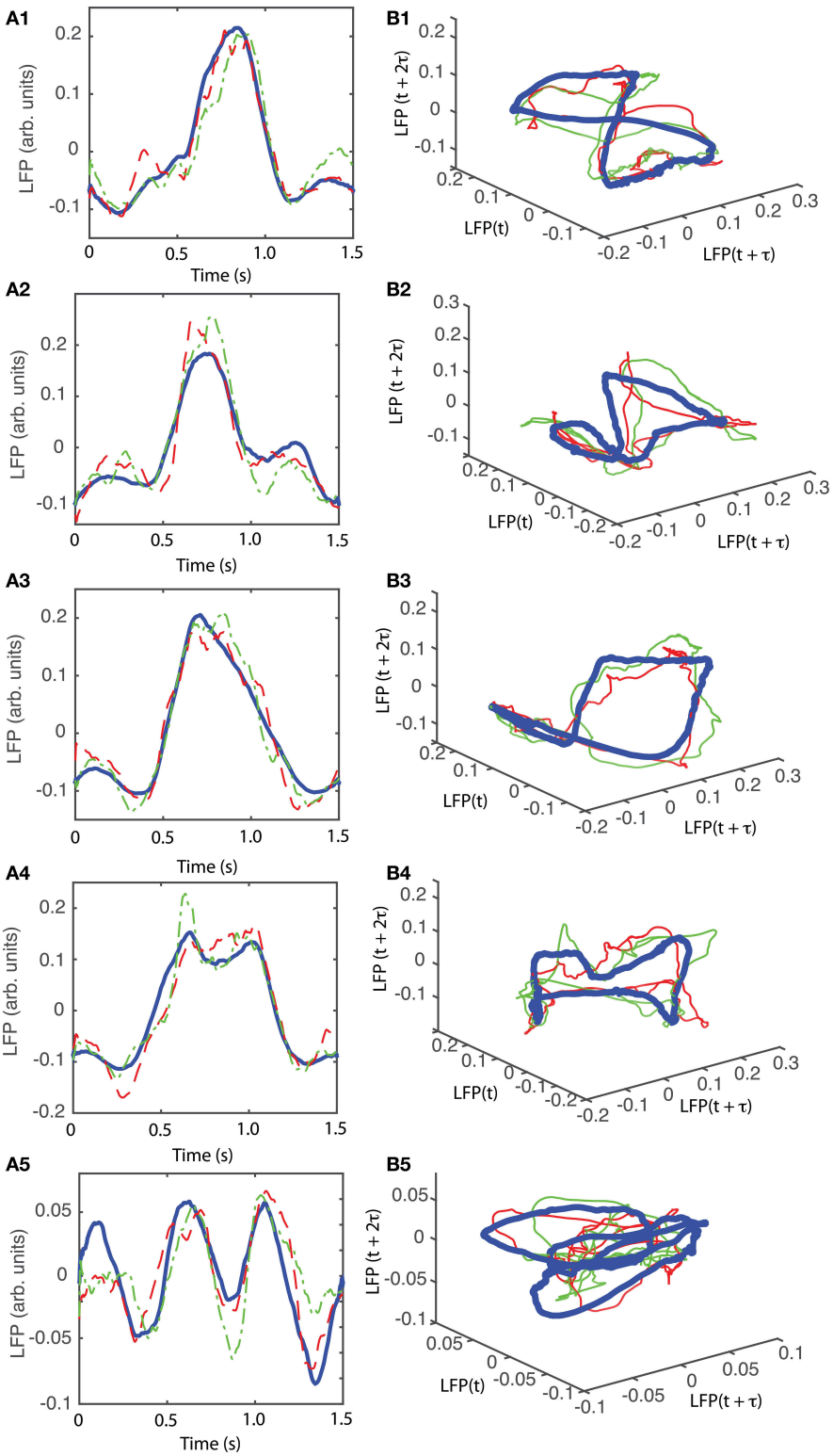


FIGURE 7 | Reconstructed 3-dimensional attractor for average activity of the local network. From each 2 s long trial only the last 1.5 s of the steady LFP recording was used. **(A1–A5)** Show the average LFP for each of the five groups of the corresponding dendrogram (blue thick line) as a visual aid to guide us gauge if the two randomly selected trials from the same group (red dashed and green dashed-dotted line) remain close to each other al all times. **(B1–B5)** Show the corresponding three dimensional reconstructed attractors. With the exception of the last group of LFP recordings, the attractors look similar after they are appropriately rotated and/or phase shifted.

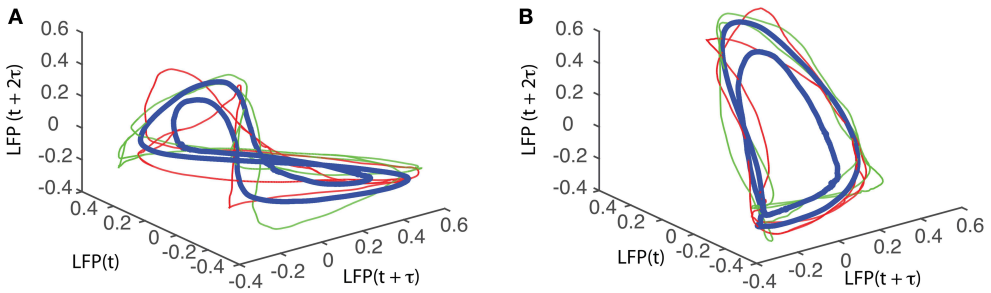


FIGURE 8 | "8"-shaped attractor and its topologically equivalents. The same data sets were reconstructed in three dimensions with different delay times to produce topologically equivalent attractors. **(A)** A typical "8" shaped attractor made of two twisted loops can be obtained for a delay time of $\tau = 1900\Delta t$. Using a delay time $\tau = 2200\Delta t$, the attractor is "untangled" and becomes quite flat **(B)**.

time by eliminating pair comparisons of all reconstructed attractors to determine which trials remain close to each other.

6. Discussion

Accurate quantification of the dynamic structure of LFPs can provide insight into the characteristics of the underlying neurophysiological processes that generated the data. In the present study, we first determined that nonlinearity is present in our LFP data using the surrogates method and two different discriminating statistics: (1) time reversal asymmetry, and (2) percentage of FNN. Time reversal asymmetry is a robust method for detecting irreversibility, which represents nonlinearity, even in the presence of a large amount of noise in the time series (Diks et al., 1995). Time reversal asymmetry statistics revealed clear differences between the original and the surrogates, with the exception of one group of data out of five for the first animal. For each of the six animals we had one group of original data for which we could not reject the null hypothesis that the time series could be produced by a linearly filtered noise at a significance level of 5% (Stam et al., 1998).

We performed also a FNN-based nonlinearity test and found that for all LFPs the percentage of FNN is always smaller for the original data trials compared to any of their surrogates. For example, any of the individual trials from the group of data for which we could not reject the null hypothesis based on time reversal asymmetry criterion had a smaller percentage of FNN than any of its 100 surrogates (see **Figure 6C**). As a result, we concluded that nonlinearity is likely present in all our data sets.

We performed two important data preprocessing that helped us reduce the computational time required for attractors identification: (1) phase shifting LFPs to correct for the phase resetting induced by light stimulus, and (2) grouping the shifted LFPs in similar patterns of activity using a dendrogram (see **Figure 4A**).

Since the light stimulus was applied every 2 s, it found the rhythmic LFP activity at different phases. As a result, it produced significantly different permanent phase shifts of the LFPs from trial to trial (see the two out-of-phase red and blue LFP recordings in **Figure 2A**). We determined the amount of phase resetting by circularly shifting the recordings (for example, compare the out-of-phase traces in **Figure 3A** against a better overlap of LFPs in **Figure 3B**). The phase resetting in neural networks is of paramount importance for large neural network synchronization. For example, in deep brain stimulation (DBS) procedures an electrical pulse is applied through an electrode to a brain region with the purpose of disrupting the synchronous activity, e.g., during epileptic seizures (Varela et al., 2001; Tass, 2003; Greenberg et al., 2010). For this purpose, stimuli are carefully designed with appropriate amplitude and duration and are precisely delivered during DBS procedures (Tass, 2003; Greenberg et al., 2010). Such procedures are based on precise measurements of phase resetting. Although we did not use electrical stimuli like in DBS, we also produced large phase resettings in background activity of mPFC. Using correlation maximization criteria, we were able to estimate quantitatively the amount of phase resetting. To our knowledge, phase shifting LFPs to maximize their pair correlation was not previously used in the context of measuring the amount of phase resetting in optogentic experiments.

Although dendrogram grouping is not absolutely necessary for attractors identification, it reduced the computational time required for data analysis. For example, for $N = 100$ trials we should have performed $N(N+1)/2 \approx 5000$ pair comparisons to find if and which reconstructed phase space trajectories remained close to each other, therefore, hinting toward a possible attractor. Instead, we only checked if the individual trials from the same group remained close to each other (see red and green traces in **Figures 7B1–B5**). By analyzing all possible pairs of trials we would have eventually reached the same conclusion, i.e., that the individual trials from group 1 (**Figure 7B1**) do not remain close to the reconstructed trajectories from group 5 (see **Figure 7B5**).

We showed that the recorded LFPs from mPFC of ChR2 expressing PV+ interneurons could be successfully embedded in a three dimensional space. For this purpose, we presented a detailed analysis of delay embedding procedure for LFPs in response to a brief 10 ms light pulse. Both the autocorrelation and the AMI gave consistently close estimations of delay, or

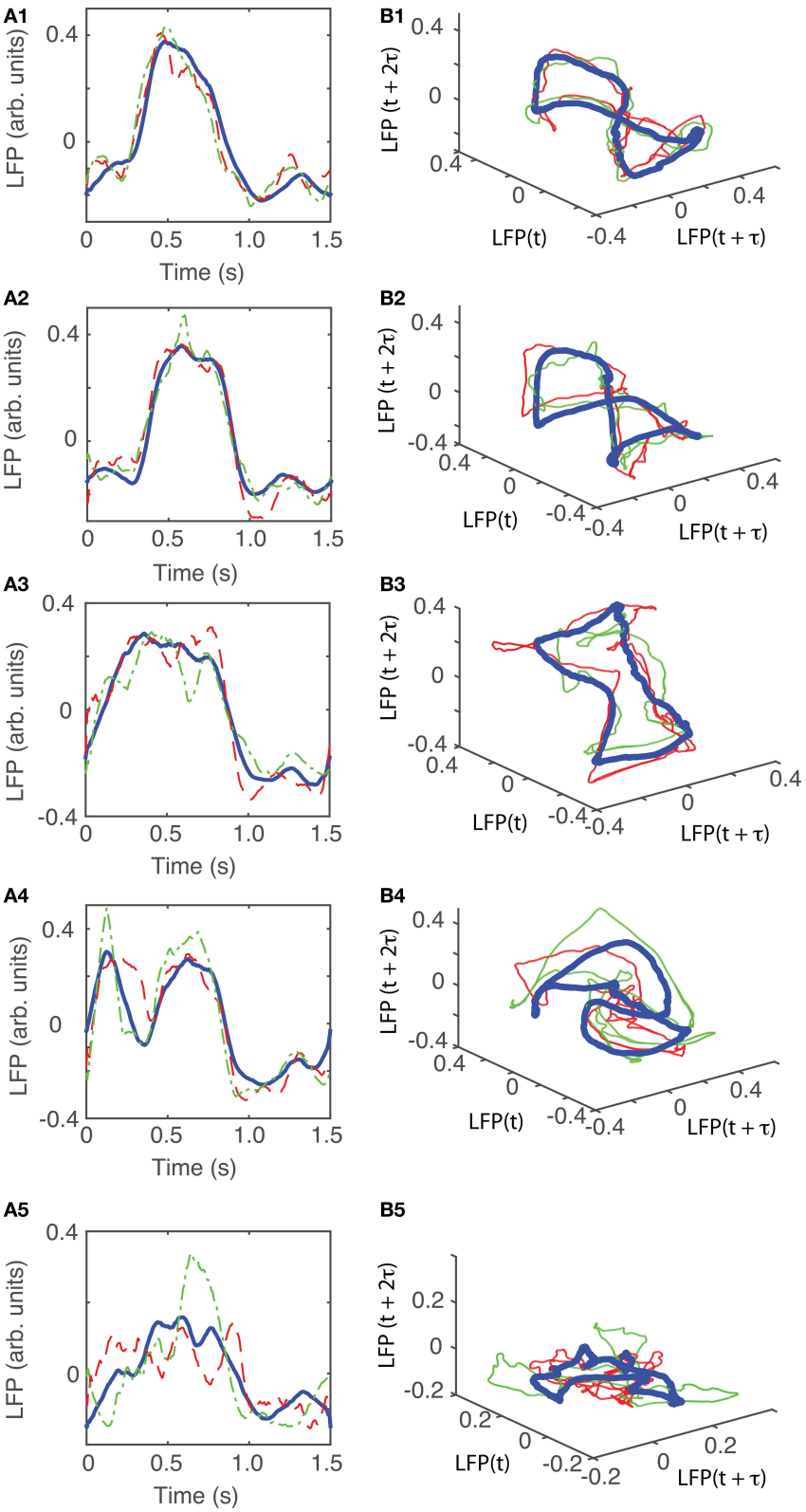


FIGURE 9 | Reconstructed 3-dimensional attractor for animal #2. (A1–A5) Show the average LFP for each of the five main groups of the corresponding dendrogram (blue thick line) and two randomly selected trials form the same group (red dashed and green dashed-dotted line). **(B1–B5)** Show the corresponding three dimensional reconstructed attractors. With the exception of the last group of LFP recordings, the attractors look similar after they are appropriately rotated and/or phase shifted.

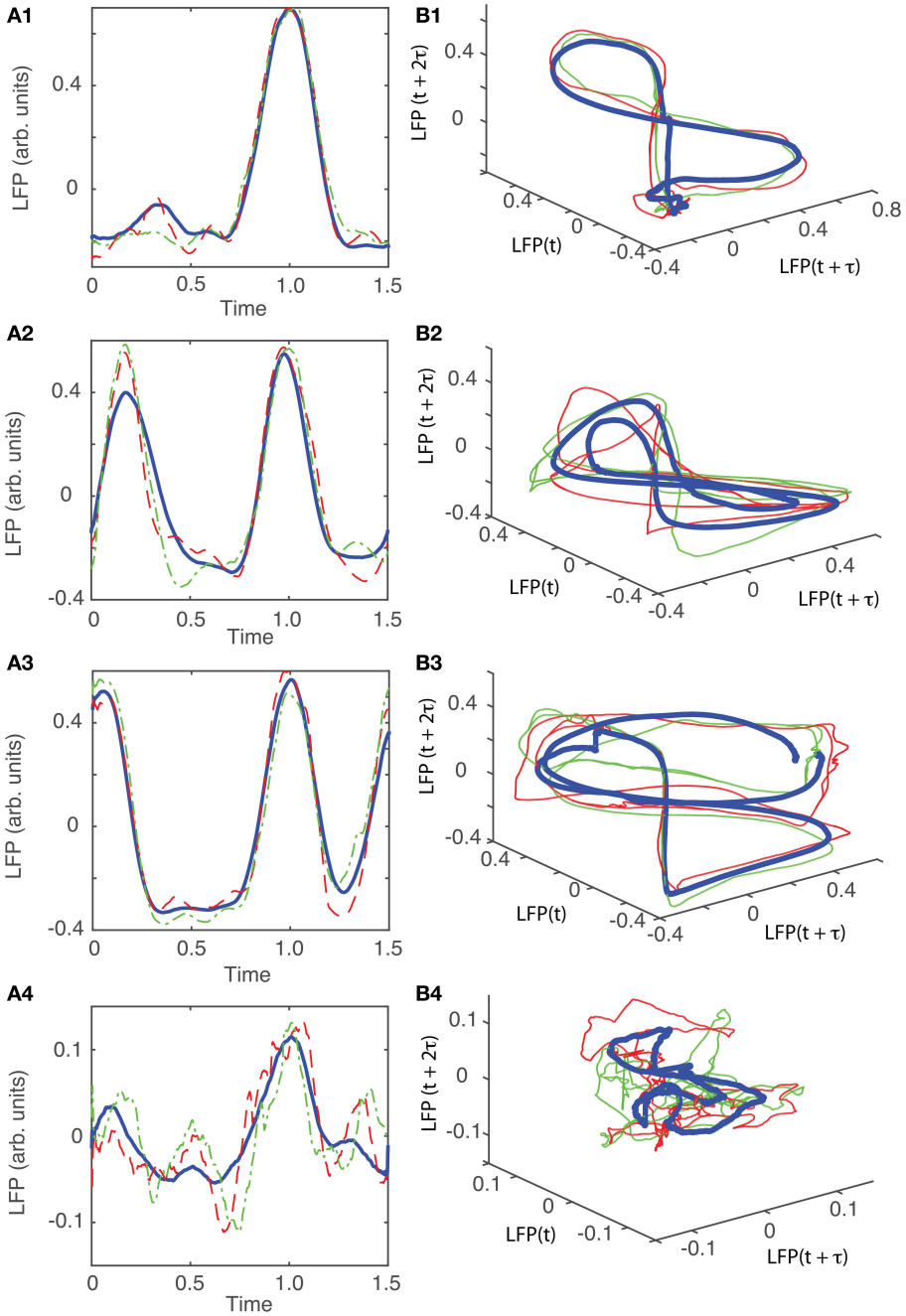


FIGURE 10 | Reconstructed 3-dimensional attractor for animal #3. (A1–A4) Show the average LFP for each of the four main groups of the corresponding dendrogram (blue thick line) and two randomly selected trials form the same group (red dashed and green dashed-dotted line). **(B1–B4)** Show the corresponding three dimensional reconstructed attractors. With the exception of the last group of LFP recordings, the attractors look similar after they are appropriately rotated and/or phase shifted.

lag, time (see **Table 1**). We found that a sufficient embedding dimension was $d_E = 3$ for all six animals. The embedding dimension estimation based on the FNN method was stable for a broad range of lag times around the optimally predicted values. We also considered a wide range of values both for the ratio of the distances between neighbors in successively larger phase spaces (parameter f in FNN routine—see Section 4.2) and different Thieler window (parameter t in FNN routine).

We found the same "8"-shaped attractor, or its topologically equivalent counterparts after appropriate phase shifting, in all six animals, which covers overed 80% of recorded data.

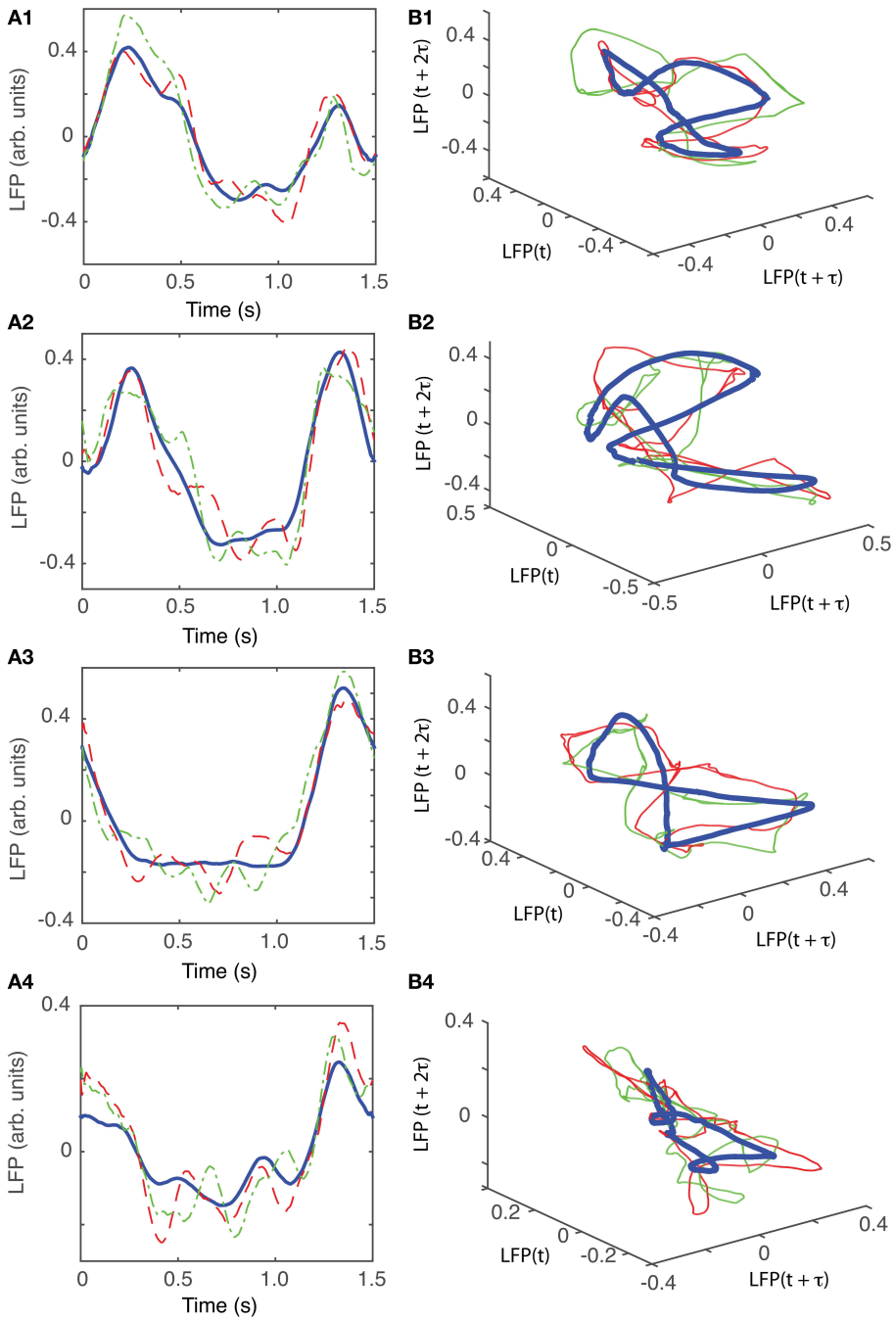


FIGURE 11 | Reconstructed 3-dimensional attractor for animal #4. (A1–A4) Show the average LFP for each of the four main groups of the corresponding dendrogram (blue thick line) and two randomly selected trials form the same group (red dashed and green dashed-dotted line). **(B1–B4)** Show the corresponding three dimensional reconstructed attractors. With the exception of the last group of LFP recordings, the attractors look similar after they are appropriately rotated and/or phase shifted.

All the other attractors were produced by low-amplitude and higher frequency oscillations of LFPs, which led to a more complex structure of the attractor. One possible reason for such a clear separation into two classes of attractors across all animals could be due to neural network bistability, i.e., depending on the phase of the light stimulus the network's activity could lead to one attractor (the "8"-shaped) or a more complex geometry. Another possible, much simpler, explanation could be that the recording quality was intermittently degraded by unknown factors, such as laser trigger malfunction, etc. Future LFP recordings are required to test such hypotheses.

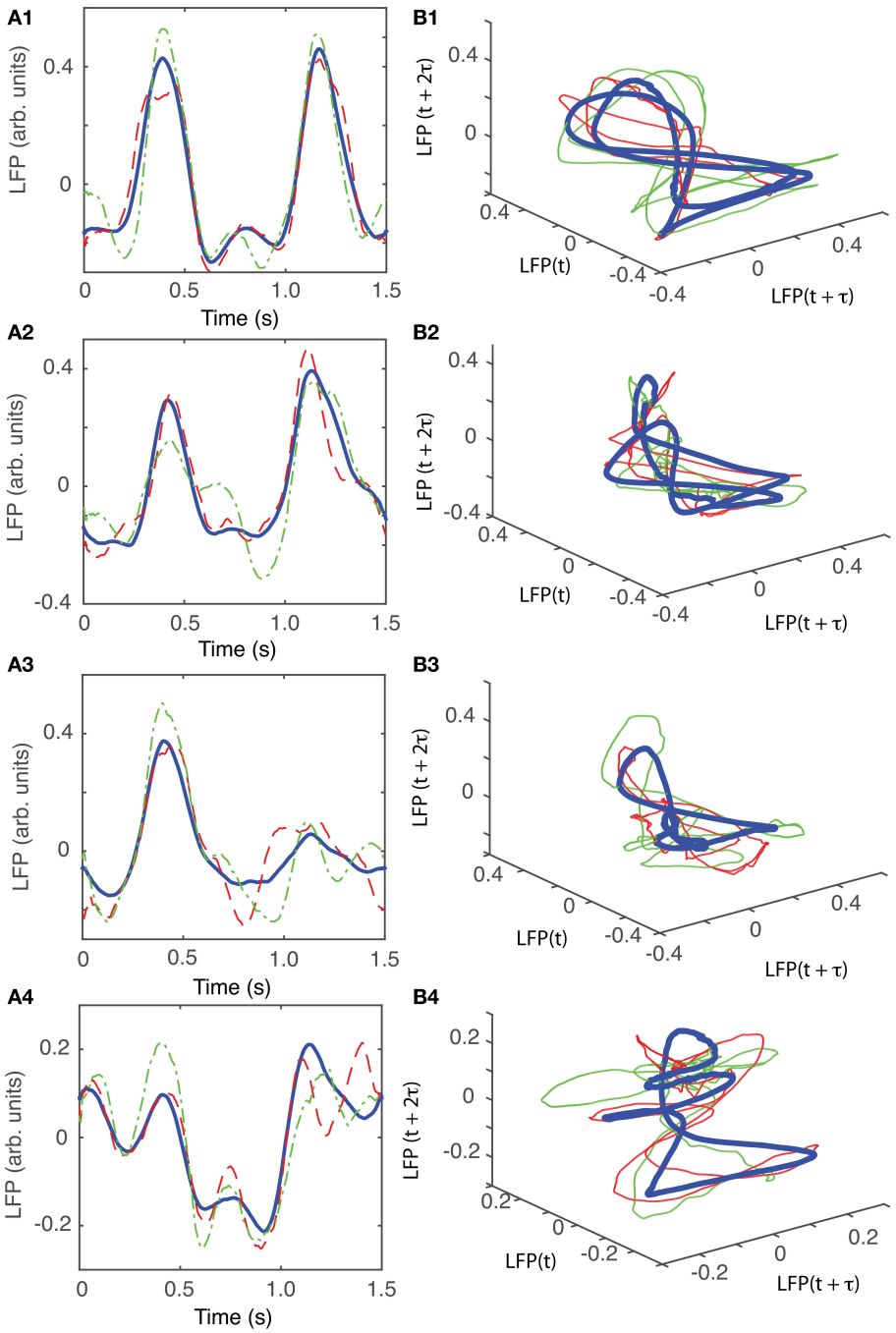


FIGURE 12 | Reconstructed 3-dimensional attractor for animal #3. (A1–A4) Show the average LFP for each of the four main groups of the corresponding dendrogram (blue thick line) and two randomly selected trials form the same group (red dashed and green dashed-dotted line). **(B1–B4)** Show the corresponding three dimensional reconstructed attractors. With the exception of the last group of LFP recordings, the attractors look similar after they are appropriately rotated and/or phase shifted.

Additionally, the low-dimensional attractor that we identified opens the possibility of fitting the experimental data to a three-dimensional model for the purpose of better understanding the dynamics of the network, e.g., through bootstrap method (Efron, 1982) .

7. Conclusions

The activity of medial prefrontal cortex of six optogenetic mice was periodically perturbed with brief laser pulses. The pair correlations between recorded LFPs were enhanced

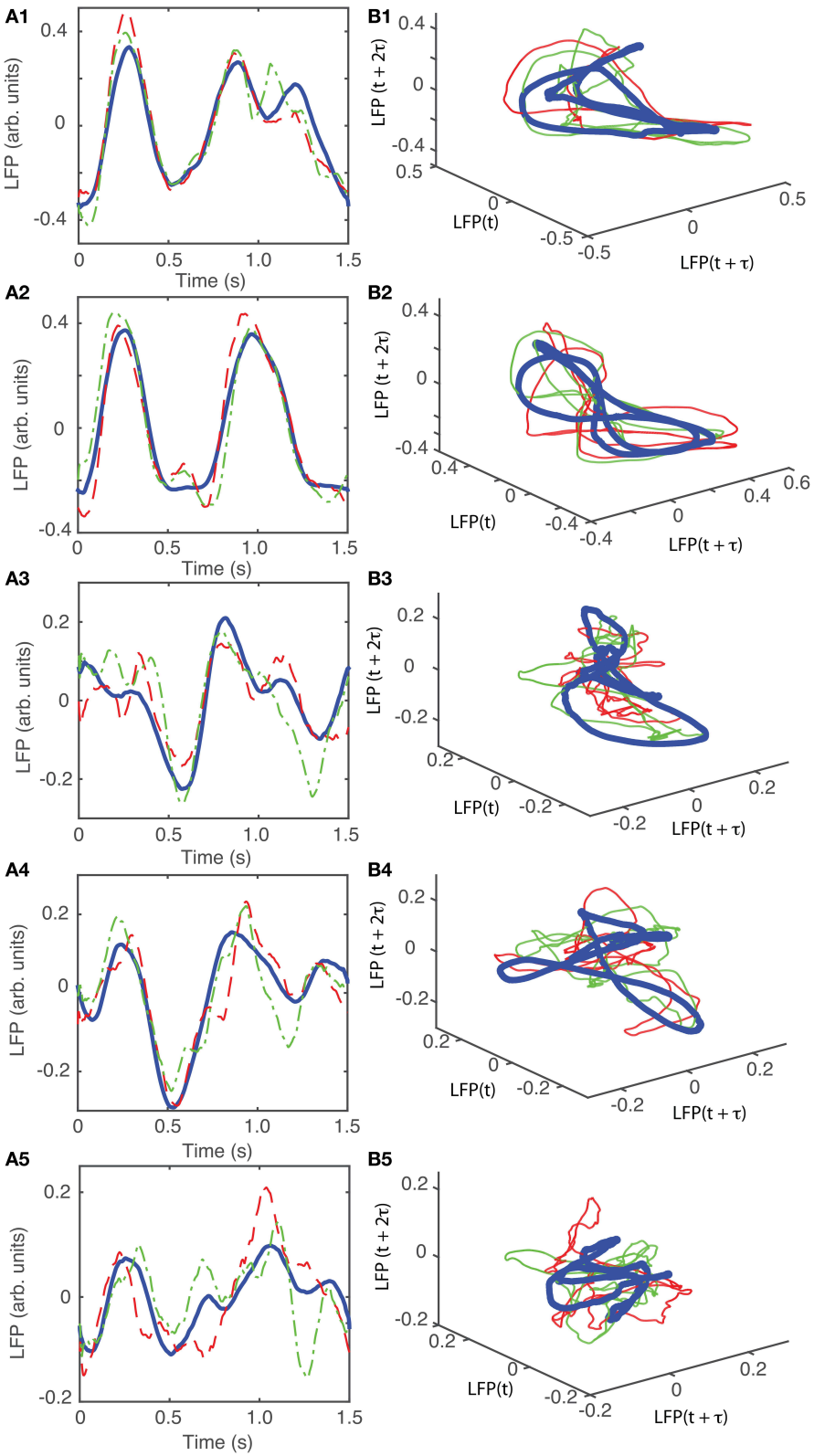


FIGURE 13 | Reconstructed 3-dimensional attractor for animal # 6. (A1–A5) Show the average LFP for each of the five main groups of the corresponding dendrogram (blue thick line) and two randomly selected trials form the same group (red dashed and green dashed-dotted line). **(B1–B5)** Show the corresponding three dimensional reconstructed attractors. With the exception of the last group of LFP recordings, the attractors look similar after they are appropriately rotated and/or phase shifted.

by appropriate phase shifting them to account for the light-induced phase resetting of network activity. The phase space dynamics was reconstructed using delay embedding method. We found that the reconstructed attractors are three dimensional and they have similar shapes across different animals.

Author Contributions

SO tested data nonlinearity, corrected data for phase resetting using crosscorrelation, computed dendrogram-based statistics, carried out numerical simulations for delay-embedding, and wrote the manuscript. PL contributed to delay-embedding numerical simulations. TT and AL performed the experiments and reviewed the manuscript.

Funding

SO acknowledges support for this research from NSF-CAREER award IOS 1054914 and MUSC bridge funding (AL).

Supplementary Material

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fncom.2015.00125

References

Abarbanel, H. (ed.). (1996). *Analysis of Observed Chaotic Data*. New York, NY: Springer.

Babloyantz, A., and Destexhe, A. (1986). Low-dimensional chaos in an instance of epilepsy. *Proc. Natl. Acad. Sci. U.S.A.* 83, 3513–3517.

Broome, B. M., Jayaraman, V., and Laurent, G. (2006). Encoding and decoding of overlapping odor sequences. *Neuron* 51, 467–482. doi: 10.1016/j.neuron.2006.07.018

Broomhead, D., and King, G. P. (1986). Extracting qualitative dynamics from experimental data. *Phys. D* 20, 217–236.

Burgess, A. P., and Ali, L. (2002). Functional connectivity of gamma eeg activity is modulated at low frequency during conscious recollection. *Int. J. Psychophysiol.* 46, 91–100. doi: 10.1016/S0167-8760(02)00108-3

Casdagli, M., Eubank, S., Farmer, J. D., and Gibson, J. (1991). State space reconstruction in the presence of noise. *Phys. D* 51, 52–98.

Diks, C., van Houwelingen, J., Takens, F., and DeGoede, J. (1995). Reversibility as a criterion for discriminating time series. *Phys. Lett. A* 201, 221–228.

Dilgen, J. E., Tompa, T., Saggu, S., Naselaris, T., and Lavin, A. (2013). Optogenetically evoked gamma oscillations are disturbed by cocaine administration. *Front. Cell. Neurosci.* 7:213. doi: 10.3389/fncel.2013.00213

Ebersole, J., and Pedley, T. (2003). *Current Practice of Clinical Electroencephalography*. LWW Medical Book Collection (Philadelphi, PA: Lippincott Williams & Wilkins).

Efron, B. (1982). *The Jackknife, the Bootstrap and Other Resampling Plans*. Philadelphi, PA: Society for Industrial and Applied Mathematics.

Fraser, A. M., and Swinney, H. L. (1986). Independent coordinates for strange attractors from mutual information. *Phys. Rev. A* 33, 1134–1140.

Fujiwara-Tsukamoto, Y., and Isomura, Y. (2008). Neural mechanism underlying generation of synchronous oscillations in hippocampal network. *Brain Nerve* 60, 755–762.

Grassberger, P., and Procaccia, I. (1983). Characterization of strange attractors. *Phys. Rev. Lett.* 50, 346–349.

Grassberger, P. (1987). Evidence for climatic attractors. *Nature* 362, 524.

Greenberg, B. D., Gabriels, L. A., Malone, D. A. Jr., Rezai, A. R., Friehs, G. M., Okun, M. S., et al. (2010). Deep brain stimulation of the ventral internal capsule/ventral striatum for obsessive-compulsive disorder: worldwide experience. *Mol. Psychiatry* 15, 64–79. doi: 10.1038/mp.2008.55

Heath, R. A. (ed.). (2000). *Nonlinear Dynamics: Techniques and Applications in Psychology*. Mahwah, NJ: Psychology Press.

Hegger, R., Kantz, H., and Schreiber, T. (1999). Practical implementation of nonlinear time series methods: the tisean package. *Chaos* 9, 413–435.

Holzfuss, J., and Mayer-Kress, G. (1986). "An approach to error-estimation in the application of dimension algorithms," in *Dimensions and Entropies in Chaotic Systems*, Vol. 32 *of Springer Series in Synergetics*, ed G. Mayer-Kress (Berlin; Heidelberg: Springer), 114–122.

Iasemidis, L. D., Shiau, D.-S., Chaovalitwongse, W., Sackellares, J. c., Pardalos, P. M., Principe, J., et al. (2003). Adaptive epileptic seizure prediction system. *IEEE Trans. Biomed. Eng.* 50, 616–627. doi: 10.1109/TBME.2003.810689

Iasemidis, L. (2003). Epileptic seizure prediction and control. *IEEE Trans. Biomed. Eng.* 50, 549–558. doi: 10.1109/TBME.2003.810705

Jerger, K. K., Netoff, T. I., Francis, J. T., Sauer, T., Pecora, L., Weinstein, S. L., et al. (2001). Early seizure detection. *J. Clin. Neurophysiol.* 18, 259–268. doi: 10.1097/00004691-200105000-00005

Jung, K.-Y., Kim, J.-M., and Kim, D. W. (2003). Nonlinear dynamic characteristics of electroencephalography in a high-dose pilocarpine-induced status epilepticus model. *Epilepsy Res.* 54, 179–188. doi: 10.1016/S0920-1211(03)00079-2

Kahana, M. J., Sekuler, R., Caplan, J. B., Kirschen, M., and Madsen, J. R. (1999). Human theta oscillations exhibit task dependence during virtual maze navigation. *Nature* 399, 781–784.

Kantz, H., and Schreiber, T. (eds.). (1997). *Non-linear Time Series Analysis*. Cambridge: Cambridge University Press.

Kaplan, J., and Yorke, J. (1979). "Chaotic behavior of multidimensional difference equations," in *Functional Differential Equations and Approximation of Fixed Points,* Vol. 730 *of Lecture Notes in Mathematics*, eds H.-O. Peitgen and H.-O. Walther (Berlin; Heidelberg: Springer), 204–227.

Kennel, M. B., Brown, R., and Abarbanel, H. D. I. (1992). Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Phys. Rev. A* 45, 3403–3411.

King, G. P., Jones, R., and Broomhead, D. (1987). Phase portraits from a time series: a singular system approach. *Nucl. Phys. B* 2, 379–390.

Kirihara, K., Rissling, A., Swerdlow, N., Braff, D., and Light, G. A. (2012). Hierarchical organization of gamma and theta oscillatory dynamics in schizophrenia. *Biol. Psychiatry* 71, 873–880. doi: 10.1016/j.biopsych.2012.01.016

Konstantinou, K. I. (2002). Deterministic non-linear source processes of volcanic tremor signals accompanying the 1996 vatnajÖkull eruption, central iceland. *Geophys. J. Int.* 148, 663–675. doi: 10.1046/j.1365-246X.2002.01608.x

Kralemann, B., Cimponeriu, L., Rosenblum, M., Pikovsky, A., and Mrowka, R. (2008). Phase dynamics of coupled oscillators reconstructed from data. *Phys. Rev. E* 77:066205. doi: 10.1103/PhysRevE.77.066205

Kugiumtzis, D. (2002). "Surrogate data test on time series," in *Modelling and Forecasting Financial Data,* Vol. 2 *of Studies in Computational Finance*, eds A. Soofi and L. Cao (Springer), 267–282.

Lisman, J., and Jensen, O. (2013). The theta-gamma neural code. *Neuron* 77, 1002–1016. doi: 10.1016/j.neuron.2013.03.007

Miltner, W. H., Braun, C., Arnold, M., Witte, H., and Taub, E. (1999). Coherence of gamma-band eeg activity as a basis for associative learning. *Nature* 397, 434–436.

Morris, C., and Lecar, H. (1981). Voltage oscillations in the barnacle giant muscle fiber. *Biophys. J.* 35, 193–213.

Oprisan, S. A., and Buhusi, C. V. (2013). How noise contributes to time-scale invariance of interval timing. *Phys. Rev. E* 87:052717. doi: 10.1103/PhysRevE.87.052717

Oprisan, S., and Buhusi, C. (2014). What is all the noise about in interval timing? *Philos. Trans. R. Soc. B Biol. Sci.* 369, 20120459. doi: 10.1098/rstb.2012.0459

Oprisan, S., and Canavier, C. (2002). The influence of limit cycle topology on the phase resetting curve. *Neural Comput.* 14, 1027–2002. doi: 10.1162/089976602753633376

Oprisan, S., and Canavier, C. (2006). Technique for eliminating nonessential components in the refinement of a model of dopamine neurons. *Neurocomputing* 69, 1030–1034. doi: 10.1016/j.neucom.2005.12.039

Oprisan, S. A, Thirumalai, V. V., and Canavier, C. (2003). Dynamics from a time series: can we extract the phase resetting curve from a time series? *Biophys. J.* 84, 2919–2928. doi: 10.1016/s0006-3495(03)70019-8

Oprisan, S. (2002). An application of the least-squares method to system parameters extraction from experimental data. *Chaos* 12, 27–32. doi: 10.1063/1.1436501

Oprisan, S. (2009). Reducing the complexity of computational models of neurons using bifurcation diagrams. *Rev. Roum. Chim.* 54, 465–475.

Oprisan, S. (2013). Local linear approximation of the jacobian matrix better captures phase resetting of neural limit cycle oscillators. *Neural Comput.* 26, 132–157. doi: 10.1162/NECO_a_00536

Osborne, A., and Provencale, A. (1989). Finite correlation dimension for stochastic systems with power-law spectra. *Phys. D* 35, 357–381.

Packard, N. H., Crutchfield, J. P., Farmer, J. D., and Shaw, R. S. (1980). Geometry from a time series. *Phys. Rev. Lett.* 45, 712–716.

Päivinen, N., Lammi, S., Pitkänen, A., Nissinen, J., Penttonen, M., and Gronförs, T. (2005). Epileptic seizure detection: a nonlinear viewpoint. *Comput. Methods Prog. Biomed.* 79, 151–159. doi: 10.1016/j.cmpb.2005.04.006

Parra, P., Gulyās, A. I., and Miles, R. (1998). How many subtypes of inhibitory cells in the hippocampus? *Neuron* 20, 983–993.

Provenzale, A., Smith, L., Vio, R., and Murante, G. (1992). Distinguishing between low-dimensional dynamics and randomness in measured time series. *Phys. D* 58, 31–49.

Roux, F., and Uhlhaas, P. (2014). Working memory and neural oscillations: alpha-gamma versus theta-gamma codes for distinct wm information? *Trends Cogn. Sci.* 18, 16–25. doi: 10.1016/j.tics.2013.10.010

Schiff, S., and Chang, T. (1992). Differentiation of linearly correlated noise from chaos in a biologic system using surrogate data. *Biol. Cybern.* 67, 387–393.

Schnitzler, A., and Gross, J. (2005). Normal and pathological oscillatory communication in the brain. *Nat. Rev. Neurosci.* 6, 285–296. doi: 10.1038/nrn1650

Schreiber, T., and Schmitz, A. (2000). Surrogate time series. *Phys. D* 142, 346–382. doi: 10.1016/S0167-2789(00)00043-9

Schuster, H. G. and Just, W. (eds.). (2005). *Deterministic Chaos: An Introduction, 4th, Revised and Enlarged Edition*. Weinheim: WILEY-VCH Verlag GmbH and Co. KGaA.

Sen, A. K., Litak, G., and Syta, A. (2007). Cutting process dynamics by nonlinear time series and wavelet analysis. *Chaos* 17, 023133. doi: 10.1063/1.2749329

Small, M. (2005). *Applied Nonlinear Time Series Analysis: Applications in Physics, Physiology and Finance*. World Scientific Series in Nonlinear Science, Series A (Toh Tuck Link: World Scientific).

Stam, C., Nicolai, J., and Keunen, R. (1998). Nonlinear dynamical analysis of periodic lateralized epileptiform discharges. *Clin. Electroencephalogr.* 292, 101–105.

Stefánsson, A., Koncar, N., and Jones, A. (1997). A note on the gamma test. *Neural Comput. Appl.* 5, 131–133.

Stopfer, M., Jayaraman, V., and Laurent, G. (2003). Intensity versus identity coding in an olfactory system. *Neuron* 39, 991–1004. doi: 10.1016/j.neuron.2003.08.011

Takens, F. (1981). "Detecting strange attractors in turbulence," in *Dynamical Systems and Turbulence, Warwick 1980,* Vol. 898 *of Lecture Notes in Mathematics*, eds D. Rand and L.-S. Young (Berlin; Heidelberg: Springer), 366–381.

Tass, P. A. (2003). A model of desynchronizing deep brain stimulation with a demand-controlled coordinated reset of neural subpopulations. *Biol. Cybern.* 89, 81–88. doi: 10.1007/s00422-003-0425-7

Theiler, J., Eubank, S., Longtin, A., Galdrikian, B., and Farmer, J. (1992). Testing for nonlinearity in time series: the method of surrogate data. *Phys. D* 58, 77–94.

Theiler, J. (1990). Estimating fractal dimension. *J. Opt. Soc. Am. A* 7, 1055–1073.

Traub, R., Jefferys, J., and Whittington, M. (1997). Simulation of gamma rhythms in networks of interneurons and pyramidal cells. *J. Comput. Neurosci.* 4, 141–150.

van der Heyden, M. J., Velis, D. N., Hoekstra, B. P., Pijn, J. P., van Emde Boas, W., van Veelen, C. W., et al. (1999). Non-linear analysis of intracranial human eeg in temporal lobe epilepsy. *Clin. Neurophysiol.* 110, 1726–1740.

Varela, F., Lachaux, J.-P., Rodriguez, E., and Martinerie, J. (2001). The brainweb: phase synchronization and large-scale integration. *Nat. Rev. Neurosci.* 2, 229–239. doi: 10.1038/35067550

Vlachos, I., and Kugiumtzis, D. (2010). Nonuniform state-space reconstruction and coupling detection. *Phys. Rev. E* 82:016207. doi: 10.1103/PhysRevE.82.016207

Yuan, G.-C., Lozier, M. S., Pratt, L. J., Jones, C. K. R. T., and Helfrich, K. R. (2004). Estimating the predictability of an oceanic time series using linear and nonlinear methods. *J. Geophys. Res.* 109, C08002. doi: 10.1029/2003JC002148

Zeng, X., Eykholt, R., and Pielke, R. A. (1991). Estimating the lyapunov-exponent spectrum from short time series of low precision. *Phys. Rev. Lett.* 66, 3229–3232.

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.

frontiers in
COMPUTATIONAL NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 06 October 2014
doi: 10.3389/fncom.2014.00125

A pooling-LiNGAM algorithm for effective connectivity analysis of fMRI data

Lele Xu[1], Tingting Fan[1], Xia Wu[1,2,3,4]*, KeWei Chen[5], Xiaojuan Guo[1], Jiacai Zhang[1] and Li Yao[1,3,4]

[1] College of Information Science and Technology, Beijing Normal University, Beijing, China
[2] State Key Laboratories of Transducer Technology, Shanghai Institute of Technical Physics, Chinese Academy of Sciences, Shanghai, China
[3] State Key Laboratory of Cognitive Neuroscience and Learning, IDG/McGovern Institute for Brain Research, Beijing Normal University, Beijing, China
[4] Center for Collaboration and Innovation in Brain and Learning Sciences, Beijing Normal University, Beijing, China
[5] Department of Mathematics and Statistics, Banner Good Samaritan PET Center, Banner Alzheimer's Institute, Arizona State University, Phoenix, AZ, USA

Edited by:
Tobias Alecio Mattei, Ohio State University, USA

Reviewed by:
Antonio Politi, Consiglio Nazionale delle Ricerche, Italy
Le Wang, Boston University, USA

***Correspondence:**
Xia Wu, College of Information Science and Technology, Beijing Normal University, No. 19 Xin Jie Kou Wai Da Jie, Beijing, 100875, China
e-mail: wuxia@bnu.edu.cn

The Independent Component Analysis (ICA)—linear non-Gaussian acyclic model (LiNGAM), an algorithm that can be used to estimate the causal relationship among non-Gaussian distributed data, has the potential value to detect the effective connectivity of human brain areas. Under the assumptions that (a): the data generating process is linear, (b) there are no unobserved confounders, and (c) data have non-Gaussian distributions, LiNGAM can be used to discover the complete causal structure of data. Previous studies reveal that the algorithm could perform well when the data points being analyzed is relatively long. However, there are too few data points in most neuroimaging recordings, especially functional magnetic resonance imaging (fMRI), to allow the algorithm to converge. Smith's study speculates a method by pooling data points across subjects may be useful to address this issue (Smith et al., 2011). Thus, this study focus on validating Smith's proposal of pooling data points across subjects for the use of LiNGAM, and this method is named as pooling-LiNGAM (pLiNGAM). Using both simulated and real fMRI data, our current study demonstrates the feasibility and efficiency of the pLiNGAM on the effective connectivity estimation.

Keywords: effective connectivity, causal structure, group analysis, functional magnetic resonance imaging (fMRI), linear non-Gaussian acyclic model (LiNGAM), pooling-LiNGAM (pLiNGAM)

INTRODUCTION

Functional connectivity and effective connectivity analyses have been widely used in the neuroimaging communities (Friston, 1994; Biswal et al., 1995; Greicius et al., 2003). Functional connectivity reflects the temporal correlations between spatially remote brain regions (Friston et al., 1993), and effective connectivity evaluates the influence that one brain region exerts on others (Friston, 1994). With the ability to describe the directionality of information transferred within a brain network, effective connectivity has become a hot topic in cognitive neuroscience research.

A variety of analysis methods have been developed for estimating effective connectivity, such as the Structural Equation Modeling (McLntosh and Gonzalez Lima, 1994), Dynamic Causal Modeling (Friston et al., 2003), Granger Causality Mapping (Goebel et al., 2003), and Bayesian Network (Zheng and Rajapakse, 2006). In a number of functional magnetic resonance imaging (fMRI) effective connectivity studies, the Gaussian assumption is usually made (Geiger and Heckerman, 1994; Bollen, 1998), however, most of fMRI data possess non-Gaussion distributions. Structural Equation Modeling and Dynamic Causal Modeling are model-driven methods and may be not suitable for resting-state fMRI data (Heckerman, 2008) or for situations where the prior knowledge is insufficient. Bayesian Network is a data-driven method but requires the data to be Gaussian-distributed (Shachter and Kenley, 1989; Baker et al., 1994; Wu and Lewin, 1994). Granger Causality Mapping uses a vector autoregressive model to estimate the effective connectivity among regions. It is also data-driven and only requires the data to be wide-sense stationary and has a zero mean (Goebel et al., 2003). However, Granger Causality Mapping is sensitive to noise and down sampling, thus it may generate spurious causality under some circumstances (Geiger and Heckerman, 1994; Chen et al., 2006; Shimizu et al., 2006).

A new method named linear non-Gaussian acyclic model (LiNGAM) algorithm was proposed by Shimizu et al. (2006) and suggested to be a promising tool to estimate the causal relationship among non-Gaussian distributed data. The fundamental difference of LiNGAM from most classical effective connectivity methods is the assumption of non-Gaussian distributions. The LiNGAM algorithm utilizes higher-order distributional statistics [Independent Component Analysis (ICA)] to estimate causal relations (Shimizu et al., 2006). This algorithm is data-driven and uses the following assumptions: (a) the data generating process is linear, (b) no unobserved confounders are present, and (c) disturbance variables follow non-Gaussian distributions. With a linear, non-Gaussian setting, LiNGAM can estimate the full causal model without undetermined parameters (Shimizu and Kano, 2008), whereas methods with Gaussian data need more information to work, such as the causal ordering of variables (Shimizu et al., 2006).

The LiNGAM algorithm could perform more stably in simulated data with more data points, e.g., the number of data points ≥1000 (Smith et al., 2011). However, the number of data points is fairly small (usually no more than 300) in most fMRI experiments. One viable strategy to address this issue is to pooling data points across subjects, in this way, a larger number of data points could be submitted to the LiNGAM algorithm. In this study, this method is called as pooling-LiNGAM (pLiNGAM), and the pooling subject can be termed as the virtual subject (V-subject).

The pooling of data points from multiple subjects actually belongs to group analysis method. There are mainly three categories of group analysis techniques, including the "virtual-typical-subject" (VTS) method, the "individual-structure" (IS) method, and "common-structure" (CS) method. The VTS method assumes that every subject within a group performs the same function and has the same connectivity network, and it does not consider inter-subject variability (Li et al., 2008). The IS method learns a network for each subject separately and then performs group analysis on the individually learned networks (Goncalves et al., 2001; Li et al., 2007). It considers inter-subject variability but may not integrate group data tightly enough (Li et al., 2008). The CS method imposes the same network structure on each subject, while allowing different parameters across subjects (Mechelli et al., 2002; Kim et al., 2007). It considers the group similarity at the structural level and inter-subject variability at the parameter level (Li et al., 2008). Each technique has its own advantages. Specifically, the VTS approach fits the data when inter-subject variability is assumed minimal, for example healthy subjects; the IS approach fits the data with large inter-subject variability, such as patients with large ranged clinical scores; while the CS approach otherwise (Li et al., 2008). The pLiNGAM used in this paper belongs to the VTS technique, thus our current study only considered the case where the inter-subject variability is low, such as the healthy subjects group.

In this paper, we aimed to demonstrate the feasibility of pLiNGAM on the estimation of effective connectivity by pooling data points across subjects. First, in order to examine the validity of pLiNGAM, the simulated fMRI data that is described in Smith's study (Smith et al., 2011) was adopted. Then, to verify the practicability of pLiNGAM algorithm, the real fMRI data was further used.

MATERIALS AND METHODS

METHODS

In this section, the original LiNGAM theory and the proposed pLiNGAM theory will be introduced.

LiNGAM theory

The LiNGAM algorithm has the following properties:

(a) Suppose x_i ($i \in \{1,\ldots,m\}$, x_i stands for the observed variables) can be arranged in their causal order $k(i)$. For example, as in the Gaussian Bayesian theory, there are two observed variables x and y, if x is the parent node of y, then the causal order of x and y satisfy the relation of $k(x) > k(y)$. The generating process of variables x_i is recursive (Shimizu and Kano, 2008) and can be represented graphically by a directed acyclic graph (Pearl, 2000; Spirtes et al., 2000).

(b) Each variable x_i is a linear function of the preceding/parent variables, a "disturbance" term e_i, and an optional constant term c_i, that is

$$x_i = \sum_{k(j) < k(i)} b_{ij} x_j + e_i + c_i \tag{1}$$

where b_{ij} is the weight coefficient, $k(i)$ is the causal order for each variable.

(c) The disturbances e_i are non-Gaussian distributions, non-zero variances, and independent of each other.
After subtracting the mean from each variable x_i and rewriting the equation in a matrix form, the following equation can be obtained:

$$x = Bx + e \tag{2}$$

where $\boldsymbol{x}$ is data vector containing the component x_i, $\boldsymbol{B}$ is the weight coefficients matrix and can be permuted to a strict lower triangular matrix if the causal ordering of variables is known (strict lower triangular matrix is defined as the lower triangular matrix with all zeros on the diagonal) and $\boldsymbol{e}$ is a disturbance term. Then, we can have:

$$x = Ae \tag{3}$$

where $\boldsymbol{A} = (\boldsymbol{I} - \boldsymbol{B})^{-1}$. Matrix $\boldsymbol{A}$ can be permuted to lower triangular (all diagonal elements are non-zero). For Equation (3), the independence and non-Gaussianity of $\boldsymbol{e}$ define the special ICA model.

ICA is commonly used to discover hidden sources from a set of observed data when the sources are non-Gaussian and maximally independent. In this algorithm, FastICA (Hyvärinen and Oja, 1997) is chosen to estimate the sources $\boldsymbol{e}$ and the weight coefficients matrix $\boldsymbol{B}$. However, there are two essential indeterminacies that ICA cannot solve: the order of independent components and the scaling of independent component amplitudes (Comon, 1994). In LiNGAM algorithm, the first indeterminacy can be solved by reordering the components following the rule that matrix $\boldsymbol{B}$ is a strict lower triangular matrix. If the results cannot be reordered to lower triangular, approaches have been produced to set the upper triangular elements to zero by changing the matrix as little as possible (Goebel et al., 2003). The second indeterminacy is usually handled by fixing the weights of their corresponding observed variables to unity. To assess the significance of the estimated connectivity for the LiNGAM algorithm, three statistical tests are usually performed to prune the edges of the estimated network: (a) Wald test, testing the significance of b_{ij}; (b) chi-square test, examining an overall fit of the model assumptions; and c) difference chi-square test, comparing nested models (Shimizu et al., 2006).

pooling-LiNGAM (pLiNGAM) theory

To avoid the fatigue of subjects and ensure the quality of the data, researchers often conduct relatively short fMRI experiments.

The length of time for data acquisition from these experiments is usually limited, such as 480 s (8 min), thus may result in the unstable results of LiNGAM algorithm. To address this issue, the pLiNGAM algorithm of pooling data over multiple subjects is proposed (Smith et al., 2011).

In this method, long enough fMRI data points are obtained for an artificial subject, referred to as the "V-subject," by pooling several single subjects. As a V-subject is constructed from more than one single subject, it is preferred to assume that the inter-subject variability can be ignored. Here we provide formulated forms of extended LiNGAM, which is pLiNGAM. Suppose there are n subjects, then each variable $\boldsymbol{x} = (x_{i1}, x_{i2}, \ldots, x_{in})$ $(i \in \{1, \ldots, m\})$ is a linear function of the preceding/parent variables and a "disturbance" term $\boldsymbol{e} = (e_{i1}, e_{i2}, \ldots, e_{in})$ and an optional constant term $\boldsymbol{c} = (c_{i1}, c_{i2}, \ldots, c_{in})$, that is

$$(x_{i1}, x_{i2}, \ldots, x_{in}) = \sum_{k(j)<k(i)} b'_{ij}(x_{j1}, x_{j2}, \ldots, x_{jn}) + (e_{i1}, e_{i2}, \ldots, e_{in}) + (c_{i1}, c_{i2}, \ldots, c_{in}) \quad (4)$$

where b'^{ij} is the weight coefficient, $k(i)$ belongs to the causal order and $\boldsymbol{e} = (e_{i1}, e_{i2}, \ldots, e_{in})$ is non-Gaussian distributions, non-zero variances and independent of each other.

Then the mean is subtracted from each variable $x = (x_{i1}, x_{i2}, \ldots, x_{in})$, the equation can be rewritten in a matrix form as:

$$\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ \cdots & \cdots & \cdots & \cdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1m} \\ \cdots & \cdots & \cdots & \cdots \\ b_{m1} & b_{m2} & \cdots & b_{mm} \end{bmatrix} \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ \cdots & \cdots & \cdots & \cdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} + \begin{bmatrix} e_{11} & e_{12} & \cdots & e_{1n} \\ \cdots & \cdots & \cdots & \cdots \\ e_{m1} & e_{m2} & \cdots & e_{mn} \end{bmatrix} \quad (5)$$

If we abbreviate the matrixes, (5) can be expressed as:

$$x' = B'x' + e' \quad (6)$$

where $\boldsymbol{x}'$ denotes the variable matrix, $\boldsymbol{B}'$ is the weight coefficients matrix and can be permuted to a strict lower triangular matrix according to the causal ordering of variables. Then we can get the form of Equation (6) the same as Equation (2).

Based on the Equation (6), we can also get Equation (7) that defines the special ICA model as follows:

$$x' = A'e' \quad (7)$$

where $A' = (I - B')^{-1}$.

The specific steps of pLiNGAM based on V-subjects consist of the following steps:

(1) Generate V-subjects. First, randomly select m $(1 \leq m \leq n)$ subjects (the length of a single subject is L_s) from the total n subjects. Then, the m subjects' data are pooled into one V-subject with a randomly order. The length of each V-subject is therefore $L_m = m^* L_s$. **Figure 1** illustrates the procedure.

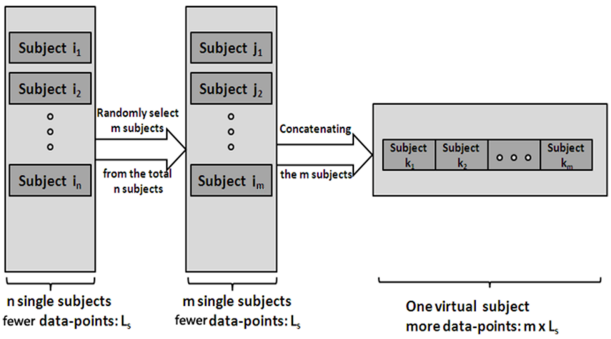


FIGURE 1 | The procedure of generating the V-subjects. Subject $i_1 \ldots i_n$ stands for the total n single subjects, which have few data-points. Subject $j_1 \ldots j_m$ stands for the m subjects selected from the total n subjects. Then the V-subject is the pooling data of the m selected subjects in a random order.

(2) Apply LiNGAM algorithm to the V-subjects. Default parameters of the ICA-LiNGAM algorithm are used, except for the "*skew*" instead of the "*tanh*" nonlinearity because the "*skew*" nonlinearity presents better results (Smith et al., 2011).

The error of the pLiNGAM algorithm is measured by the false positive ratio (FPR), false negative ratio (FNR), false direction ratio (FDR) and the sum of FPR, FNR, and FDR. FPR stands for the ratio of the number of falsely added edges to the whole possible existing edges, FNR denotes the ratio of the number of falsely missed edges to the whole possible existing edges, and FDR is the ratio of the number of edges that are wrongly identified in the direction to the whole possible existing edges. Furthermore, the sum of FPR, FNR and FDR is calculated to represent the total error of pLiNGAM.

SIMULATED fMRI DATA

The simulated data are from Smith et al. in their 2011 publication (Smith et al., 2011), which have been widely used in fMRI studies (Cole et al., 2010; Smith et al., 2011). The simulations are generated using the Dynamic Causal Modeling fMRI forward model (Friston et al., 2003), in which the Dynamic Causal Modeling uses a nonlinear balloon model (Buxton et al., 1998) for the vascular dynamics. These data can provide 28 simulations, and we select the No. 7 simulation set which has 5000 data points in this paper because it has more than enough data points for the purpose of our study. The No. 7 simulation set contains 5 nodes with 250 min of data at a repetition time of 3 s. The total number of data points is 5000 (scans) for each of the 50 simulated subjects. The coefficients matrix used to generate these 50 subjects data have the same structure with slightly different coefficients.

REAL fMRI DATA

Participants

12 healthy right-handed young students, including 5 males and 7 females (mean age: 21 years) participate in our study. This study is supported by the Beijing Normal University Imaging Center. All subjects have provided written informed consent.

Data acquisition

Images are acquired using a Siemens Trio 3-Tesla scanner (Siemens, Erlangen, Germany) in the National Key Laboratory for Cognitive Neuroscience and Learning, Beijing Normal University. Participants are instructed to remain motionless, close their eyes but stay awake during the entire scanning procedure which lasts for 8 min. All of the functional data are acquired using an echo-planar imaging sequence with the following parameters: 33 axial slices, $TR = 2000$ ms, $TE = 30$ ms, acquisition voxel size, $3.13 \times 3.13 \times 3.60\,mm^3$, in-plane resolution $= 64 \times 64$ and matrix $= 64 \times 64$, 240 volumes.

Data analyses

Data preprocessing. The first five volumes of the total 240 volumes in the functional fMRI data are removed to make the signal more stable. Image preprocessing including slice timing, realignment, normalization, and smoothing (FWHM = 8 mm) are conducted using the SPM8 software (http://www.fil.ion.ucl.ac.uk/spm).

Default mode network (DMN) and regions of interest (ROIs). Group ICA is performed to the preprocessed data using the fMRI toolbox (http://mialab.mrn.org/software/#gica) to determine the default mode network (DMN). In recent years, ICA has been widely used to identify the low-frequency neural network during resting-state or cognitively undemanding fMRI scans (Calhoun et al., 2001; Greicius and Menon, 2004; van de Ven et al., 2004). The Group ICA includes two rounds of principal component analysis, ICA separation and back-reconstruction. In ICA separation, the Extended Infomax algorithm is used (Lee et al., 1999). To select the independent component that best matches the DMN, a DMN template is developed based on a dataset of regions reported by Greicius et al. (Greicius and Menon, 2004). Subsequently, the DMN at the single subject level is acquired, and one sample t-test ($p < 0.05$, false discovery rate corrected) is performed (**Figure 2**). **Figure 2** shows the regions with significant connectivity at the resting state including the medial prefrontal cortex (mPFC), posterior cingulate cortex (PCC), left/right inferior parietal cortex (lIPC/rIPC), left/right lateral and inferior temporal cortex (lITC/rITC), and left/right (para) hippocampus (lHC/rHC). Then, these eight core DMN regions are selected as nodes (ROIs) for the LiNGAM analysis. The coordinates of the eight maximally activated voxels in the core DMN ROIs are given in **Table 1**, and the ROIs are generated with a sphere with 6 mm-radius centered at the voxel with the maxima local T-value. Then, the data points of each ROI are extracted with the software rest (http://restfmri.net/forum/index.php).

pLiNGAM on the real fMRI data. Before applying the pLiNGAM on the real fMRI data, the distribution of the V-subject obtained from the real fMRI data is examined by the One-Sample Kolmogorov–Smirnov Test. If the distribution is non-Gaussian, then the LiNGAM will be used on the V-subject to estimate the effective connectivity network among the eight core DMN ROIs.

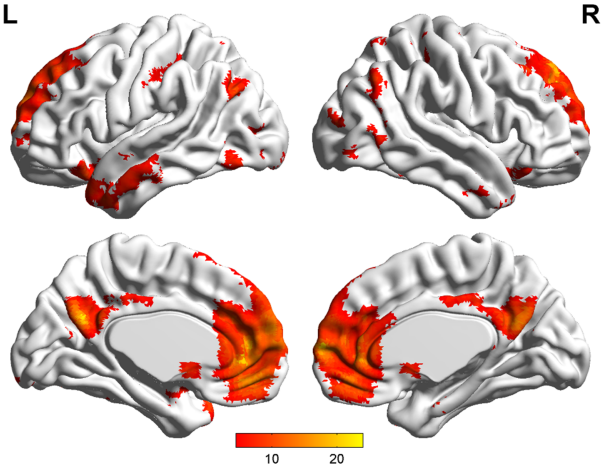


FIGURE 2 | DMN identified by group ICA ($p < 0.05$, false discovery rate corrected).

Table 1 | The coordinates of all the ROIs for real fMRI data ($p < 0.05$, false discovery rate corrected).

ROI	BA	MNI coordinate			*T*-value
		x	*y*	*z*	
PCC	23/31	0	−57	20	20.31
*m*PFC	10	−2	62	8	19.22
*l*IPC	39	−43	−67	33	9.99
*r*IPC	39	45	−60	29	7.32
*r*HC	28/35	25	−14	−23	6.47
*l*ITC	20/21	−59	−15	−16	5.50
*r*ITC	20/21	59	−12	−20	5.19
*l*HC	28/35	−22	−15	−22	4.50

BA, Brodmann's area; mPFC, medial prefrontal cortex; PCC, posterior cingulate cortex; lIPC/rIPC, left/right inferior parietal cortex; lITC/rITC, left/right lateral and inferior temporal cortex; lHC/rHC, left/right (para) hippocampus.

RESULTS

SIMULATED VALIDATION

To verify the feasibility of pLiNGAM on the estimation of effective connectivity of fMRI data, some simulation validation are performed, including the desired number of data points that is needed to make the results of LiNGAM stable, the feasibility of the pooling of data points across multiple subjects, the effectiveness of V-subjects in pLiNGAM and the influence of pooling order on pLiNGAM.

Desired number of data points of LiNGAM

The simulated data is used to investigate the desirable number of data points that can make the LiNGAM algorithm stable. Part of the total data points (5000 data points) of each single subject is applied to the LiNGAM. Part of data points in each subject are selected at the beginning of the total data points and the length of the points ranges from 200 to 5000. To avoid the influence of differences between subjects, the LiNGAM algorithm is applied

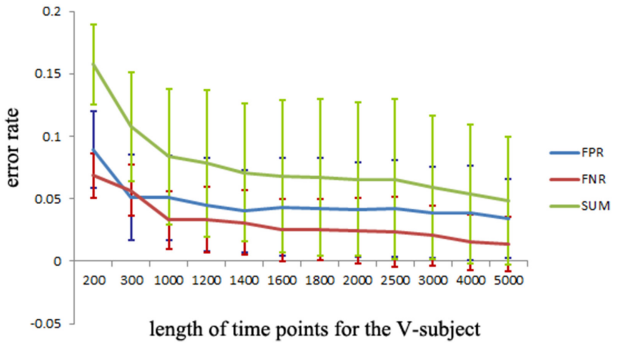


FIGURE 3 | FPR, FNR, FDR and the sum of FPR, FNR, and FDR (SUM) of the simulated subjects with different length of data points for LiNGAM algorithm (FDR is 0, so it is not shown in the figure).

to 50 subjects and the FPR, FNR, and FDR are calculated by averaging the fifty results. The average FPR, FNR, and FDR and the sum of FPR, FNR and FDR are shown in **Figure 3** (FDR is 0, so it is not shown in the figure). Three statistical tests: Wald test, chi-square test, and difference chi-square test ($p = 0.05$) are performed to prune the edges of the estimated network. **Figure 3** illustrates that both FPR and FNR are consistently decreasing as the number of data points increases. The sum of FPR and FNR reduces to approximate 7% when the length of data points arrives 5000. Because of the limitation of the number of total data points, this algorithm is not tested with longer data points.

Feasibility of subject pooling

To confirm pooling over subjects' data is a feasible method, the following two validations are performed.

(a) First, test if the pooling step could keep the distribution of the data non-Gaussian. Use the Kolmogorov–Smirnov Test to examine the distribution of the data. For the simulated data, the distribution of each single subject and several V-subjects is tested. The V-subjects were constructed as shown in **Figure 1**. Each of these V-subjects is pooled with several (range from 1 to 25) single subjects (each with 200 data points), then 25 V-subjects with the length of data points ranging from 200 to 5000 can be constructed. The results of the Kolmogorov–Smirnov Test in **Table 2** show that all the single subjects and the V-subjects are significant non-Gaussian distribution. Furthermore, we note that the main difference of the distribution of the single subjects or V-subjects from Gaussian is the "peakedness," then one classical measurement of the "peakedness" for non-Gaussian distribution named Kurtosis Test is adopted (Hyvärinen and Oja, 2000). The results show that the data has different kurtosis value from 3, e.g., 3.41, 3.98, 4.99 (the kurtosis value of Gaussian distribution is 3), further indicating the deviation of the data from Gaussian distribution. All these results indicate that the V-subjects are feasible to the LiNGAM algorithm.

(b) Second, test if the pooling step could improve the accuracy of the estimated model, in other words, test whether the result of pooling of subjects is better than that of single subject. Three groups of data are modeled: single-subject_2000 (G1), V-subject_2000 (G2), and single-subject_200 (G3). More specifically, the single-subject_2000 group consists of 10 subjects and each single subject has 2000 data points. The V-subject_2000 group is a V-subject with 2000 data points, which are pooled from 10 single subjects with 200 data points each. The single-subject_200 group consists of 10 single subjects and each single subject has 200 data points. The 10 subjects used in this paper are randomly selected from the total 50 subjects and the pooling order is random.

Table 2 | The *p*-value [mean (STD)] of One-Sample Kolmogorov–Smirnov Test of 5 ROIs for the simulated fMRI data.

Subjects \ ROIs No.	1	2	3	4	5
Single subject	8.68E-90 (6.077E-89)	8.13E-105 (5.69E-104)	2.23E-118 (1.53E-117)	1.83E-104 (1.28E-103)	4.11E-112 (2.87E-111)
V-subject	1.16E-153 (8.13E-153)	2.82E-148 (1.97E-147)	6.38E-179 (0)	4.82E-166 (0)	1.79E-183 (0)

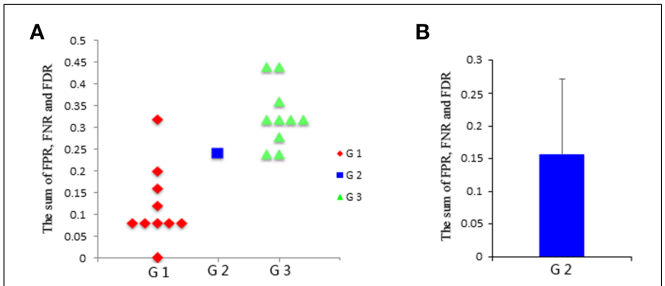


FIGURE 4 | (A) The sum of FPR, FNR, and FDR of three groups: G1: single-subject_2000, G2: V-subject_2000, and G3: single-subject_200; **(B)** the sum [mean (SD)] of FPR, FNR, and FDR of the G2 group across 50 different V-subject_2000s.

Then, the FPR, FNR, and FDR of these three groups are calculated, and the sum of FPR, FNR, and FDR for the three groups is shown in **Figure 4A**. The results clearly show that the G1 group has a smaller sum of FPR, FNR, and FDR compared to the other two groups, and the G2 group has a smaller sum of FPR, FNR, and FDR than the G3 group. Furthermore, one sample t-test is performed on G3 and G1 respectively to verify whether the mean of G3 or G1 is significantly different from G2. The results are encouraging ($T = -4.291$, $p = 0.002$ for G1; $T = 3.973$, $p = 0.003$ for G3). These statistical results denote that the G1 group shows better results than both the G2 group and G3 group, and G2 group shows better results than G3 group, which indicating that subject pooling is feasible for the LiNGAM algorithm, and pLiNGAM can offer better results when data points were few for the single subjects. Furthermore, to test if the error rate of the G2 group is stable across different subsets of 10 single subjects, 50 V-subject_2000 are constructed by randomly selecting 10 single subjects. The sum of FPR, FNR, and FDR of these V-subject_2000 are then calculated, and the results show that the

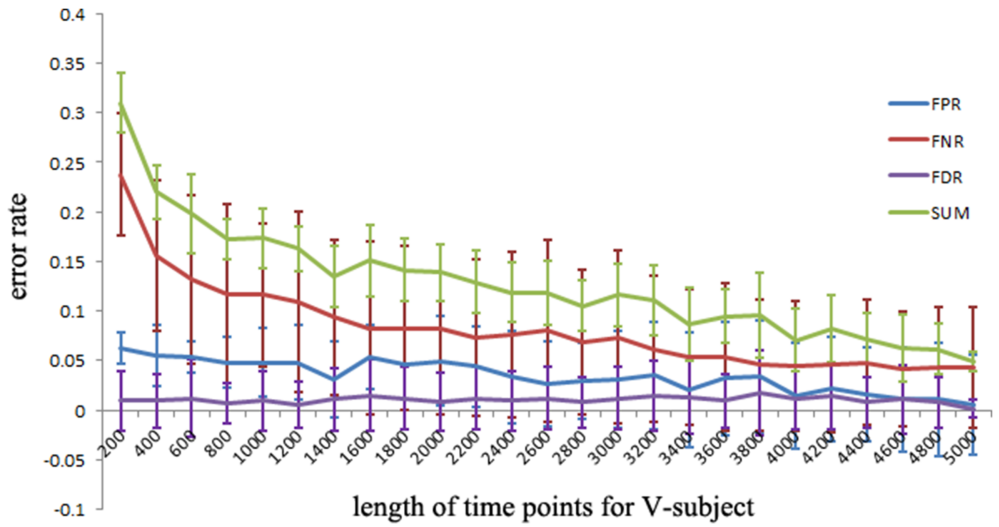


FIGURE 5 | The FPR, FNR, FDR and the sum of FPR, FNR, and FDR (SUM) of the simulated V-subjects for pLiNGAM algorithm. The length of data points of V-subjects ranges from 200 to 5000.

error rate of the G2 group is stable across different selections of the 10 subjects (**Figure 4B**).

pLiNGAM with V-subjects

To explore the FPR, FNR, and FDR estimated using the pLiNGAM with the V-subjects, the V-subjects are constructed according to the schematic shown in **Figure 1**. Each of the V-subjects is pooled with several (range from 1 to 25) single subjects (200 data points). For example, in each single subject, 200 data points are selected at the beginning of the total data points, then the 6 single subjects with data points of 200 are combined to form one V-subject with data points of 1200. The length of data points of each V-subject ranges from 200 to 5000. To ensure the reliability of the results, 50 V-subjects are constructed for each length of data points. **Figure 5** demonstrates that when data points are more than 2000, the sum of FPR, FNR, and FDR reaches 15%, which is better than most other effective connectivity methods (Cole et al., 2010; Smith et al., 2011).

Influence of the pooling order

To determine whether the order of pooling subjects has any effect on the estimated network, the following test is conducted. 10 single subjects are randomly selected from the total 50 subjects. Among 3628800 possible orders, 3000 orders are randomly selected to examine this effect. For each of the 3000 pooling orders, a V-subject is generated. Then, the pLiNGAM algorithm is applied to these V-subjects and the FPR, FNR, and FDR are calculated. Our results show that the estimated network has no relation with the order of pooling, which is consistent with the fact that the major advantage of concatenation of data points across subjects in ICA is ordering the components in different subjects in the same way (Calhoun et al., 2001).

REAL fMRI VALIDATION

The distribution of the V-subject from the real fMRI data follows non-Gaussian according to the One-Sample Kolmogorov–Smirnov Test (shown in **Table 3**). Thus, the pLiNGAM is applicable for the real fMRI data.

Table 3 | The result of One-Sample Kolmogorov–Smirnov Test of 8 ROIs for real fMRI data.

Parameters \ ROIs no.	1	2	3	4	5	6	7	8
Kolmogorov–Smirnov Z	8.09	10.84	3.64	4.46	5.03	6.64	7.03	9.89
Sig. (2-tailed)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

"Sig" represents the asymmetry significance. When the two-tailed asymptotic significance of each ROI is less than 0.05, the test distribution is not normal.

In this section, the stability of causal network is tested on the real fMRI data using pLiNGAM, and the results of effective connectivity for the real fMRI data are also displayed.

The stability of effective connectivity on real fMRI data

pLiNGAM is tested with different subsets of subjects from the real fMRI data to validation the robustness and stability of the result. Several subjects, $n = 3$ for example, are randomly selected from all the subjects (a total of 12 subjects) to construct the V-subject for 100 times (3, 4, 5, 6, 7, 8, and 9 subjects are tested respectively to ensure the procedure of random selection be repeated for 100 times, while 1, 2, 10, 11, 12 subjects can't be randomly selected for 100 times and are not used for testing). For each number of subjects, the causal network is analyzed for 100 times, and the common structure of the 100 causal networks is then considered as a baseline to calculate the FPR, FNR, and FDR of each causal network. Then the average of the sum of the FPR, FNR, and FDR is taken as the variability of the results. As it is shown in **Figure 7**, the variability of different subsets of the subjects is not high (about 0.26 for different number of subjects). This variability is comparable with the results of many algorithms that are mentioned in Smith et al., such as Granger, Bayes net and so on (Smith et al., 2011). Furthermore, the variability of different subsets of the subjects is stable along with the increased number of

subjects (slightly decrease). These results indicate the stability and robustness of the causal networks that obtained by pLiNGAM.

The results of effective connectivity for real fMRI data

Figure 6 shows the effective connectivity model of DMN during the resting state investigated by the pLiNGAM algorithm (using all the 12 subjects). From **Figure 6**, we can conclude the following connections: *m*PFC→*r*HC/*r*IPC/*l*ITC/PCC/*r*ITC/*l*IPC/*l*HC, *r*IPC →PCC/*r*HC/*l*HC/*l*ITC, *r*ITC→*r*IPC/*l*IPC/PCC/*l*ITC/*l*HC/*r*HC, *l*ITC→PCC/*l*HC/*r*HC, *l*IPC→PCC/*r*IPC/*r*HC/*l*HC/*l*ITC ($p < 0.05$, Wald statistics). Seven direct connections are detected between *m*PFC, *r*HC, *r*ITC, *l*HC, *l*IPC, PCC, and the other ROIs. Interestingly, all links associated with *m*PFC are out-going connections, and all links associated with *r*HC are in-going connections. Furthermore, six of the total seven links associated with *r*ITC are out-going connections, and six of the total seven links associated with *l*HC are in-going connections. In addition, five of the total seven links associated with *l*IPC are out-going connections, and five of the total seven links associated with PCC are in-going connections.

DISCUSSION

This study employs the pLiNGAM algorithm to explore the effective connectivity of fMRI data with the V-subject. The results demonstrate that the pLiNGAM is feasible for both simulated and real fMRI data.

The pLiNGAM algorithm has several advantages in estimating the effective connectivity of brain areas. First, the simulated fMRI data demonstrate that pLiNGAM produces a more robust effective connectivity model with the V-subject than the original single subject. With a small number of data points, however, the computational stability of pLiNGAM cannot be guaranteed because in ICA estimation, the weight matrix ***B*** often converges on different values when there are not enough data points (Goebel et al., 2003). Second, this algorithm is based on the assumptions of non-Gaussianity of disturbance variables, linearity and an acyclic model, which allow the identification of the full causal model. Previous methods (Pearl, 2000; Shimizu and Kano, 2008) based on the assumption of Gaussianity require additional information (such as the causal order of variables) to obtain a full causal model (Shimizu et al., 2006). Third, a V-subject composed of more than one subject can provide more valuable information compared to a single subject. Fourth, the sum of FPR, FNR and FDR for the V-subjects can fall to 15% (**Figure 5**), which is smaller than most of other approaches (45%) (Cole et al., 2010; Smith et al., 2011).

Our results of the simulated data show that the sum of FPR, FNR, and FDR can just reduce to approximate 7% but not 0% when there are sufficient number of data points (shown in **Figure 5**), indicating that we can't obtain a perfect network of the simulated data even if the data points are long enough. This situation is explainable. A sampling step was done in the procedure of generating the simulated data (Smith et al., 2011), thus may result in the loss of information about the data. Furthermore, some noises are also added into the simulated data (Smith et al., 2011). All these process may cause the imperfect performance of pLiNGAM even when the data points are long enough.

The subject pooling has been verified to be a reasonable method through the simulated fMRI data. Then this method is applied to the real fMRI data, and the results show that the causal network is reliable and stable across different subsets of subjects, which further indicated the feasible application of pLiNGAM in the situation with low inter-subject variability. Furthermore, most of the links associated with the PCC are in-going connections, demonstrating that the PCC acts as a confluent node. Similar conclusions have been acquired in the previous studies (Li et al., 2012; Yan et al., 2013). In addition, the links associated with *m*PFC show

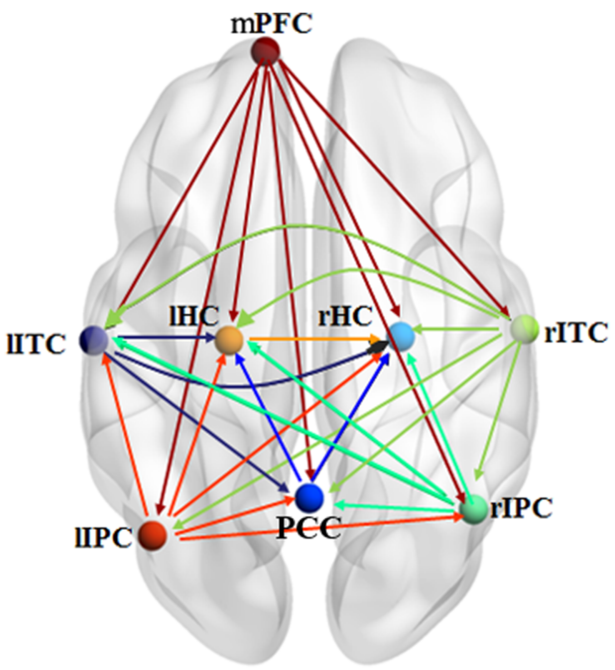


FIGURE 6 | Effective connectivity model of DMN during the resting state explored by pLiNGAM. The different line colors indicate connections originating from different nodes. The effective connectivity has been corrected using Wald statistics, chi-square test and difference chi-square test with $p < 0.05$ as the significant level.

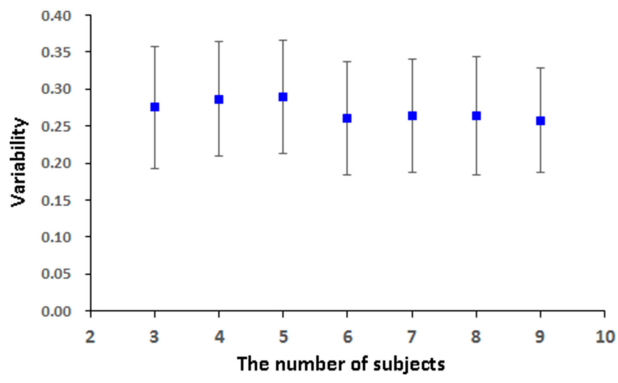


FIGURE 7 | The variability [mean (STD)] of the 100 V-subjects constructed from different number of subjects (3, 4, 5, 6, 7, 8, and 9 respectively) for the real fMRI data with pLiNGAM algorithm.

good consistency because all links are out-going connections. Li et al.'s (2012) study also supports this result.

The variability in **Figure 7** for the real fMRI data is not significantly decreasing (slightly decreasing) as the number of subjects increases, which is different from the results of the V-subject in **Figure 5**. This may because that the variability in the real fMRI data is more stable than that of the simulated data, thus having reached the flat part toward the tail like that in **Figure 5**. To a certain extent, the variability is stable (slightly decrease) along with the increased number of subjects for the real fMRI data, which indicates the stability and robustness of the causal networks that obtained by pLiNGAM. In any way, further detailed explorations are needed to delve into this problem in our future study.

While having many merits, the pLiNGAM method still has several limitations. First, it only performs well when the inter-subject variability is low. pLiNGAM is one form of the "VTS" technique (Li et al., 2008), which assumes that every subject within a group performs the same function and has the same connectivity network. Other group analysis method based on LiNGAM, such as the algorithm proposed in Shimizu (2012), assumes that each subject shares a causal ordering but different connection strengths, which is similar with the "CS" approach (Li et al., 2008). So this algorithm in Shimizu (2012) may perform worse than pLiNGAM when the inter-subject variability is low (e.g., the healthy subject group), while better than pLiNGAM when inter-subject variability is a little larger (e.g., patient group). Therefore, more efforts are needed to improve pLiNGAM in order to be applicable for more general situations. Second, the V-subjects have more data points, thus may result in longer calculation time. In addition, the calculation time also depends on group sizes and the number of ROIs (Hyvärinen and Oja, 1997). Third, the assumption of an acyclic model may be a limitation to the fMRI data. This assumption implies that information can only be transmitted from one ROI to another, but not transmitted back. However, feedback is an important feature for biological systems, such as cortico-subcortical loops (Lynch and Tian, 2006). In any way, further exploration is needed to improve the pLiNGAM algorithm.

ACKNOWLEDGMENTS

This work was supported by the Key Program of National Natural Science Foundation of China (91320201), the Funds for International Cooperation and Exchange of the National Natural Science Foundation of China (61210001), the Excellent Young Scientist Program of China (61222113), and Program for New Century Excellent Talents in University (NCET-12-0056).

REFERENCES

Baker, J. R., Weisskoff, R. M., Stern, C. E., Kennedy, D. N., Jiang, A., Kwong, K. K., et al. (1994). "Statistical assessment of functional MRI signal change," in *Proceedings of the 2nd Annual Meeting of the Society of Magnetic Resonance*, 626.

Biswal, B., Zerrin Yetkin, F., Haughton, V. M., and Hyde, J. S. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar mri. *Magn. Reson. Med.* 34, 537–541. doi: 10.1002/mrm.1910340409

Bollen, K. A. (1998). *Structural Equation Models*. John Wiley & Sons, Ltd.

Buxton, R. B., Wong, E. C., and Frank, L. R. (1998). Dynamics of blood flow and oxygenation changes during brain activation: the balloon model. *Magn. Reson. Med.* 39, 855–864. doi: 10.1002/mrm.1910390602

Calhoun, V. D., Adali, T., Pearlson, G. D., and Pekar, J. J. (2001). A method for making group inferences from functional MRI data using independent component analysis. *Hum. Brain Mapp.* 14, 140–151. doi: 10.1002/hbm.1048

Chen, Y., Bressler, S. L., Knuth, K. H., Truccolo, W. A., and Ding, M. (2006). Stochastic modeling of neurobiological time series: power, coherence, Granger causality, and separation of evoked responses from ongoing activity. *Chaos* 16, 26113. doi: 10.1063/1.2208455

Cole, D. M., Smith, S. M., and Beckmann, C. F. (2010). Advances and pitfalls in the analysis and interpretation of resting-state FMRI data. *Front. Syst. Neurosci.* 4:8. doi: 10.3389/fnsys.2010.00008

Comon, P. (1994). Independent component analysis, a new concept? *Signal Process* 36, 287–314. doi: 10.1016/0165-1684(94)90029-9

Friston, K. J. (1994). Functional and effective connectivity in neuroimaging: a synthesis. *Hum. Brain Mapp.* 2, 56–78. doi: 10.1002/hbm.460020107

Friston, K. J., Frith, C. D., Liddle, P. F., and Frackowiak, R. (1993). Functional connectivity: the principal-component analysis of large (PET) data sets. *J Cerebr Blood F Met* 13, 5. doi: 10.1038/jcbfm.1993.4

Friston, K. J., Harrison, L., and Penny, W. (2003). Dynamic causal modelling. *Neuroimage* 19, 1273–1302. doi: 10.1016/S1053-8119(03)00202-7

Geiger, D., and Heckerman, D. (1994). "Learning gaussian networks," in *Proceedings of the 10th International Conference on Uncertainty in Artificial Intelligence* (Morgan Kaufmann Publishers Inc.), 235–243.

Goebel, R., Roebroeck, A., Kim, D., and Formisano, E. (2003). Investigating directed cortical interactions in time-resolved fMRI data using vector autoregressive modeling and Granger causality mapping. *Magn. Reson. Imaging* 21, 1251–1261. doi: 10.1016/j.mri.2003.08.026

Goncalves, M. S., Hall, D. A., Johnsrude, I. S., and Haggard, M. P. (2001). Can meaningful effective connectivities be obtained between auditory cortical regions? *Neuroimage* 14, 1353–1360. doi: 10.1006/nimg.2001.0954

Greicius, M. D., Krasnow, B., Reiss, A. L., and Menon, V. (2003). Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. *Proc. Natl. Acad. Sci. U.S.A.* 100, 253–258. doi: 10.1073/pnas.0135058100

Greicius, M. D., and Menon, V. (2004). Default-mode activity during a passive sensory task: uncoupled from deactivation but impacting activation. *J. Cogn. Neurosci.* 16, 1484–1492. doi: 10.1162/0898929042568532

Heckerman, D. (2008). "A tutorial on learning with Bayesian networks," in *Innovations in Bayesian Networks* (Berlin; Heidelberg: Springer), 33–82.

Hyvärinen, A., and Oja, E. (1997). A fast fixed-point algorithm for independent component analysis. *Neural Comput.* 9, 1483–1492. doi: 10.1162/neco.1997.9.7.1483

Hyvärinen, A., and Oja, E. (2000). Independent component analysis: algorithms and applications. *Neural Netw.* 13, 411–430. doi: 10.1016/S0893-6080(00)00026-5

Kim, J., Zhu, W., Chang, L., Bentler, P. M., and Ernst, T. (2007). Unified structural equation modeling approach for the analysis of multisubject, multivariate functional MRI data. *Hum. Brain Mapp.* 28, 85–93. doi: 10.1002/hbm.20259

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

Li, J., Li, R., Chen, K., Yao, L., and Wu, X. (2012). Temporal and instantaneous connectivity of default mode network estimated using Gaussian Bayesian network frameworks. *Neurosci. Lett.* 513, 62–66. doi: 10.1016/j.neulet.2012.02.008

Li, J., Wang, Z. J., and McKeown, M. J. (2007). "A multi-subject, dynamic Bayesian networks (dbns) framework for brain effective connectivity," in *Acoustics, Speech and Signal Processing, IEEE International Conference*, 429–432.

Li, J., Wang, Z. J., Palmer, S. J., and McKeown, M. J. (2008). Dynamic Bayesian network modeling of fMRI: a comparison of group-analysis methods. *Neuroimage* 41, 398–407. doi: 10.1016/j.neuroimage.2008.01.068

Lynch, J. C., and Tian, J. (2006). Cortico-cortical networks and cortico-subcortical loops for the higher control of eye movements. *Prog. Brain Res.* 151, 461–501. doi: 10.1016/S0079-6123(05)51015-X

McLntosh, A. R., and Gonzalez Lima, F. (1994). Structural equation modeling and its application to network analysis in functional brain imaging. *Hum. Brain Mapp.* 2, 2–22. doi: 10.1002/hbm.460020104

Mechelli, A., Penny, W. D., Price, C. J., Gitelman, D. R., and Friston, K. J. (2002). Effective connectivity and intersubject variability: using a multisubject network to test differences and commonalities. *Neuroimage* 17, 1459–1469. doi: 10.1006/nimg.2002.1231

Pearl, J. (2000). *Causality: Models, Reasoning and Inference.* Cambridge: MIT press.

Shachter, R. D., and Kenley, C. R. (1989). Gaussian influence diagrams. *Manage. Sci.* 35, 527–550. doi: 10.1287/mnsc.35.5.527

Shimizu, S. (2012). Joint estimation of linear non-Gaussian acyclic models. *Neurocomputing* 81, 104–107. doi: 10.1016/j.neucom.2011.11.005

Shimizu, S., Hoyer, P. O., Hyvärinen, A., and Kerminen, A. (2006). A linear non-Gaussian acyclic model for causal discovery. *J. Mach. Learn. Res.* 7, 2003–2030.

Shimizu, S., and Kano, Y. (2008). Use of non-normality in structural equation modeling: application to direction of causation. *J. Stat. Plan. Infer.* 138, 3483–3491. doi: 10.1016/j.jspi.2006.01.017

Smith, S. M., Miller, K. L., Salimi-Khorshidi, G., Webster, M., Beckmann, C. F., Nichols, T. E., et al. (2011). Network modelling methods for FMRI. *Neuroimage* 54, 875–891. doi: 10.1016/j.neuroimage.2010.08.063

Spirtes, P., Glymour, C. N., and Scheines, R. (2000). *Causation, Prediction, and Search.* MIT press.

van de Ven, V. G., Formisano, E., Prvulovic, D., Roeder, C. H., and Linden, D. E. (2004). Functional connectivity as revealed by spatial independent component analysis of fMRI measurements during rest. *Hum. Brain Mapp.* 22, 165–178. doi: 10.1002/hbm.20022

Wu, D., and Lewin, J. S. (1994). "Evaluation of non-parametric statistical measures and data clustering for functional MR data analysis," in *Proceedings of the SMR 2nd Annual Meeting* (San Francisco, CA), 629.

Yan, H., Zhang, Y., Chen, H., Wang, Y., and Liu, Y. (2013). Altered effective connectivity of the default mode network in resting-state amnestic type mild cognitive impairment. *J. Int. Neuropsychol. Soc.* 19, 400–409. doi: 10.1017/S1355617712001580

Zheng, X., and Rajapakse, J. C. (2006). Learning functional structure from fMR images. *Neuroimage* 31, 1601–1613. doi: 10.1016/j.neuroimage.2006.01.031

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.

Received: 18 July 2014; paper pending published: 29 July 2014; accepted: 17 September 2014; published online: 06 October 2014.

Citation: Xu L, Fan T, Wu X, Chen K, Guo X, Zhang J and Yao L (2014) A pooling-LiNGAM algorithm for effective connectivity analysis of fMRI data. Front. Comput. Neurosci. ***8****:125. doi: 10.3389/fncom.2014.00125*

This article was submitted to the journal Frontiers in Computational Neuroscience.

frontiers in
COMPUTATIONAL NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 18 February 2015
doi: 10.3389/fncom.2015.00016

EEG entropy measures in anesthesia

Zhenhu Liang[1], Yinghua Wang[2,3], Xue Sun[1], Duan Li[4], Logan J. Voss[5], Jamie W. Sleigh[5], Satoshi Hagihira[6] and Xiaoli Li[2,3]*

[1] *Institute of Electrical Engineering, Yanshan University, Qinhuangdao, China*
[2] *State Key Laboratory of Cognitive Neuroscience and Learning and IDG/McGovern Institute for Brain Research, Beijing Normal University, Beijing, China*
[3] *Center for Collaboration and Innovation in Brain and Learning Sciences, Beijing Normal University, Beijing, China*
[4] *Institute of Information Science and Engineering, Yanshan University, Qinhuangdao, China*
[5] *Department of Anesthesia, Waikato Hospital, Hamilton, New Zealand*
[6] *Department of Anesthesiology, Osaka University Graduate School of Medicine, Osaka, Japan*

Edited by:
Tobias Alecio Mattei, Ohio State University, USA

Reviewed by:
Raoul Rashid Nigmatullin, Kazan Federal University, Russia
Fengyu Cong, Dalian University of Technology, China

***Correspondence:**
Xiaoli Li, State Key Laboratory of Cognitive Neuroscience and Learning and IDG/McGovern Institute for Brain Research; Center for Collaboration and Innovation in Brain and Learning Sciences, Beijing Normal University, Beijing 100875, China
e-mail: xiaoli@bnu.edu.cn

Highlights:

- Twelve entropy indices were systematically compared in monitoring depth of anesthesia and detecting burst suppression.
- Renyi permutation entropy performed best in tracking EEG changes associated with different anesthesia states.
- Approximate Entropy and Sample Entropy performed best in detecting burst suppression.

Objective: Entropy algorithms have been widely used in analyzing EEG signals during anesthesia. However, a systematic comparison of these entropy algorithms in assessing anesthesia drugs' effect is lacking. In this study, we compare the capability of 12 entropy indices for monitoring depth of anesthesia (DoA) and detecting the burst suppression pattern (BSP), in anesthesia induced by GABAergic agents.

Methods: Twelve indices were investigated, namely Response Entropy (RE) and State entropy (SE), three wavelet entropy (WE) measures [Shannon WE (SWE), Tsallis WE (TWE), and Renyi WE (RWE)], Hilbert-Huang spectral entropy (HHSE), approximate entropy (ApEn), sample entropy (SampEn), Fuzzy entropy, and three permutation entropy (PE) measures [Shannon PE (SPE), Tsallis PE (TPE) and Renyi PE (RPE)]. Two EEG data sets from sevoflurane-induced and isoflurane-induced anesthesia respectively were selected to assess the capability of each entropy index in DoA monitoring and BSP detection. To validate the effectiveness of these entropy algorithms, pharmacokinetic/pharmacodynamic (PK/PD) modeling and prediction probability (P_k) analysis were applied. The multifractal detrended fluctuation analysis (MDFA) as a non-entropy measure was compared.

Results: All the entropy and MDFA indices could track the changes in EEG pattern during different anesthesia states. Three PE measures outperformed the other entropy indices, with less baseline variability, higher coefficient of determination (R^2) and prediction probability, and RPE performed best; ApEn and SampEn discriminated BSP best. Additionally, these entropy measures showed an advantage in computation efficiency compared with MDFA.

Conclusion: Each entropy index has its advantages and disadvantages in estimating DoA. Overall, it is suggested that the RPE index was a superior measure. Investigating the advantages and disadvantages of these entropy indices could help improve current clinical indices for monitoring DoA.

Keywords: EEG, anesthesia, entropy, pharmacokinetic/pharmacodynamic modeling, depth of anesthesia monitoring

INTRODUCTION

In the operating room, general anesthesia is important to guarantee successful surgery and ensure patients' safety and comfort. For anesthesia, the reliable monitoring of anesthetic drug effects on the brain is a clinical concern for anesthesiologists (Monk et al., 2005). The central nervous system (CNS) is the main target of anesthetic drugs. Originated in CNS, the electroencephalogram (EEG) reflects the neural activities of brain, and has been widely used as a surrogate parameter to quantify the anesthetic drug effect (Rampil, 1998; Bruhn et al., 2006; Jameson and Sloan,

2006). However, only limited information can be obtained from the EEG signals purely by waveform observation. With the development of signal processing, various methods have been applied to analyze, identify or detect mental disorders and consciousness mechanisms from EEG signals (Okogbaa et al., 1994; Natarajan et al., 2004; Abásolo et al., 2006), as well as evaluating the effects of anesthesia.

In recent decades, numerous attempts have been made to develop an index for describing anesthetic drug effects on the brain, including zero crossing frequency, spectral edge, wavelet analysis, high-order spectral analysis etc. These studies laid the foundation of commercial EEG-based monitors of depth of anesthesia (DoA), such as BIS (Aspect Medical Systems, Newton, MA) (Bruhn et al., 2006; Ellerkmann et al., 2010) and M-entropy (GE Healthcare, Helsinki, Finland) (Viertiö-Oja et al., 2004; Bruhn et al., 2006). Many of these methods are derived from linear theories. However, various studies have shown that the EEG is a non-stationary signal that exhibits non-linear or chaotic behaviors (Elbert et al., 1994; Pritchard et al., 1995; Zhang et al., 2001; Natarajan et al., 2004). This prompted many researchers to adopt non-linear analysis methods in anesthesia study, for example largest Lyapunov exponent (Fell et al., 1996), Hurst exponent (Alvarez-Ramirez et al., 2008), fractal analysis (Klonowski et al., 2006; Gifani et al., 2007; Liang et al., 2012), detrended fluctuation analysis (DFA) (Jospin et al., 2007; Nguyen-Ky et al., 2010b), recurrence analysis (Huang et al., 2006), and non-linear entropies (Bruhn et al., 2001; Li et al., 2008a). In particular, non-linear entropy methods describing the complexity of EEG signals, have received considerable attention.

The word "entropy" was first proposed as a thermodynamic principle by Clausius (1867). It describes the distribution probability of molecules of gaseous or fluid systems. In 1949, Claude E. Shannon introduced entropy into information theory to describe the distribution of signal components (Shannon and Weaver, 1949). So far, numerous entropy algorithms have been proposed and used to quantify DoA, covering Spectral entropy [which includes Response Entropy (RE) and State entropy (SE)] (Viertiö-Oja et al., 2004; Klockars et al., 2012), Approximate entropy (ApEn) (Bruhn et al., 2000), Sample entropy (SampEn) (Richman and Moorman, 2000), Fuzzy entropy (FuzzyEn) (Chen et al., 2007), Shannon Permutation entropy (SPE) (Li et al., 2008a, 2012), Shannon Wavelet entropy (SWE) (Särkelä et al., 2007), and Hilbert-Huang spectral entropy (HHSE) (Li et al., 2008b).

Spectral Entropy is the method applied in the commercial M-Entropy Module (Viertiö-Oja et al., 2004). It consists of two parameters: Response Entropy (RE) and State Entropy (SE). SE primarily includes the spectrum of the EEG signal from 0.8 to 32 Hz, and RE includes electromyogram activity from 0.8 to 47 Hz (Viertiö-Oja et al., 2004). Shannon Wavelet entropy (SWE) is the Shannon entropy in the wavelet domain, which indicates signal variation at each frequency scale (Rosso et al., 2001). And the Hilbert–Huang spectral entropy (HHSE) is the Shannon entropy based on the Hilbert–Huang transform proposed by Huang et al. (1998). HHSE has been successfully applied to the anesthetic EEG signals (Li et al., 2008b).

The above methods are based on the frequency spectrum. Whereas many entropy methods are based on the time series and phase space analysis. ApEn is an algorithm derived from the Kolmogorov-Sinai entropy (Pincus, 1991). It quantifies the predictability of subsequent amplitude values of a signal. A previous investigation showed that ApEn correlates well with the concentration of desflurane (Bruhn et al., 2000). However, ApEn lacks relative consistency and is highly dependent on data length, SampEn was proposed to overcome ApEn's limitation by removing self-matching and relieving its bias (Richman and Moorman, 2000). SampEn has been used for analyzing EEG signals (Montirosso et al., 2010; Yoo et al., 2012). FuzzyEn was proposed by Chen et al. (2007). It is based on the fuzzy membership functions to define the vectors' similarity, using the soft and continuous boundaries of fuzzy functions to ensure the continuity and the validity of FuzzyEn's definition (Chen et al., 2009). SPE was introduced by Bandt and Pompe (2002). It is a complexity measure based on symbolic dynamics (Bandt and Pompe, 2002). Because of its simple concept and fast computation, SPE has been widely used in EEG signal analysis (Cao et al., 2004; Li et al., 2007, 2008a). Furthermore, its derivatives, multi-scale permutation entropy (Li et al., 2010) and composite permutation entropy index (Olofsen et al., 2008) have been successfully applied to analyze EEG signals during anesthesia.

However, "No one knows what entropy really is, so in a debate you will always have the advantage." This statement is true for EEG analysis today (Ferenets et al., 2006). Each entropy index has its own advantages and disadvantages, but how does their performance compare when evaluating the effect of anesthesia on brain activity? To this end, some researchers have compared the performance of different entropy methods for anesthesia monitoring (Sleigh et al., 2001, 2005; Bein, 2006). Unfortunately, these articles analyzed no more than three entropies. To our knowledge, a systematic comparison of the performance of them in assessing anesthesia drug effect is lacking. In this study, we aim to compare the capability of several commonly used entropy indices for monitoring DoA.

We noticed that definitions of all the above entropies are based on Shannon information theory, which belongs to a short-range or extensive concept. However, the physical systems especially the biomedical systems are often characterized by either long-range interactions, long-term memories, or multifractality (Zunino et al., 2008). To describe these characters, two generalized forms of entropy were proposed: Renyi entropy (Renyi, 1970) and Tsallis entropy (q-entropy) (Tsallis et al., 1998). For example Tsallis entropy has a parameter q for non-extensity. If $q > 1$, the entropy is more sensitive to events that occur often, whereas if $0 < q < 1$ it is more sensitive to the events that occur seldom (Maszczyk and Duch, 2008). In the limit $q \to 1$, it coincides with Shannon entropy. These generalized entropies can provide additional informational about the importance of specific events, such as outliers or rare events. The two classes of entropies and their combinations with current signal processing methods have been already applied in EEG analysis (Bezerianos et al., 2003; Tong et al., 2003; Inuso et al., 2007) and often been proved advantageous than the Shannon version (Zunino et al., 2008; Arefian et al., 2009). To make the research more instructive, we believe it useful to investigate these non-extensive entropy measures along with those extensive Shannon entropies in DoA monitoring. In

this study, we involved the Tsallis wavelet entropy (TWE) and Renyi wavelet entropy (RWE) proposed by Rosso et al. (2003, 2006), as well as the Tsallis permutation entropy (TPE) proposed by Zunino et al. (2008) and a new Renyi permutation entropy (RPE).

For illustrative purpose, we divide the entropies into two families:

(1) Entropies in the time-frequency domain: RE, SE, SWE, TWE, RWE, and HHSE;
(2) Entropies in the time domain: ApEn, SampEn, FuzzyEn, SPE, TPE, and RPE.

In this work, their performance for monitoring DoA were compared. Using data sets obtained during sevoflurane and isoflurane anesthesia, we quantified for each index the responsiveness to loss of consciousness, computation complexity and the ability to detect BSP. Pharmacokinetic/pharmacodynamic (PK/PD) modeling and prediction probability statistics were applied to evaluate the efficiency of each index for tracking anesthetic concentration. Additionally, in order to prove the efficiency of the entropy approaches, two non-linear dynamic methods: DFA (Jospin et al., 2007) and multifractal DFA (MDFA) (Kantelhardt et al., 2002) are compared.

ENTROPY INDICES

The computation of each entropy index is briefly described as follows.

SPECTRAL ENTROPY (RE AND SE)

Spectral Entropy quantifies the probability density function (PDF) of the signal power spectrum in the frequency domain. The detail of the Spectral Entropy algorithm can be seen in Inouye et al. (1991) and Rezek and Roberts (1998). Spectral Entropy consists of the RE and the SE. RE is computed over a frequency range from 0.8 to 47 Hz while SE is computed over the frequency range from 0.8 to 32 Hz. The normalization step for RE and SE are defined as follows:

$$RE = \frac{H_{sp_{0.8-47}}}{\log\left(N_{0.8-47}\right)} \tag{1}$$

$$SE = \frac{H_{sp_{0.8-32}}}{\log\left(N_{0.8-47}\right)} \tag{2}$$

where $H_{sp_{0.8-47}}$ and $H_{sp_{0.8-32}}$ means the sum of spectral power between 0.8 and 47 Hz, and 0.8 to 32 Hz, respectively. And $N_{0.8-47}$ equals the total number of frequency components in the range 0.8–47 Hz. Spectral Entropy describes the degree of skewness in the frequency distribution. For example, in the normalized case, the Spectral Entropy of a pure sine wave with a single spectral peak is 0, while that of white noise is 1.

WAVELET ENTROPY (SWE, TWE, AND RWE)

WE differentiates specific brain states under spontaneous or stimulus-related conditions and recognizes the time localizations of a dynamic process. To calculate Wavelet Entropy, wavelet energy E_j of a signal is determined at each scale j as follows:

$$E_j = \sum_{k=1}^{L_j} d(k)^2 \tag{3}$$

where k and L_j are the summation index and the number of coefficients at each scale j with in a given epoch, respectively. The total energy over all scales is obtained by:

$$E_{total} = \sum_j E_j = \sum_j \sum_{k=1}^{L_j} d_j(k)^2 \tag{4}$$

Then wavelet energy is divided by total energy to obtain the relative wavelet energy at each scale j:

$$p_j = \frac{E_j}{E_{total}} = \frac{E_j}{\sum_j E_j} = \frac{\sum_{k=1}^{L_j} d(k)^2}{\sum_j \sum_{k=1}^{L_j} d_j(k)^2} \tag{5}$$

SWE is calculated from Shannon entropy of p_j distribution between scales as follows:

$$S^{(s)} = -\sum_j p_j \log p_j \tag{6}$$

The detail of the algorithm used in this study can be seen in Särkelä et al. (2007).

And the TWE is defined as,

$$S_q^{(T)} = \frac{1}{q-1} \sum_j \left[p_j - \left(p_j\right)^q\right] \tag{7}$$

where q is a non-extensity parameter.

Based on the definition of Renyi entropy (Renyi, 1970), the RWE is defined as Rosso et al. (2006):

$$S_a^{(R)} = \frac{1}{1-a} log\left[\sum_j \left(p_j\right)^a\right] \tag{8}$$

For $S_q^{(S)}$, the normalized SWE is

$$SWE = S^{(s)} / \log N_J \tag{9}$$

where N_J is the number of wavelet resolution levels.

And $S_q^{(T)}$ is normalized by dividing $\left[1 - N_J^{1-q}\right]/(q-1)$, defined by Rosso et al. (2003):

$$TWE = \frac{S_q^{(T)}}{\left[1 - N_J^{1-q}\right]/(q-1)} \tag{10}$$

Further, the normalized $S_a^{(R)}$ is defined as Maszczyk and Duch (2008):

$$RWE = \frac{S_a^{(R)}}{\log N_J} \tag{11}$$

The values of three WE measures depend on the wavelet basis function, the number of decomposed layers (n) and the data

length (N). Furthermore, the TWE and RWE are related to the parameters q and a respectively. Among these parameters, the wavelet basis function is most important. Because of the lack of a fixed criterion, it is very difficult to select an appropriate wavelet basis function in practical applications and many studies choose it based on experiments. The details of the selection process in this study can be found in Supplement Material 1.

HILBERT-HUANG SPECTRAL ENTROPY (HHSE)

HHSE is based on the Hilbert-Huang transform, which applies the Shannon entropy concept to the Hilbert-Huang spectrum. The detail of the algorithm is seen in Li et al. (2008b). For a given non-stationary signal $x(t)$, the EMD method decomposes the signal into a series of intrinsic mode functions (IMFs), $C_n\,(1, 2, \ldots, M)$, where M is the number of IMFs. The signal $x(t)$ can be written by:

$$x\,(t) = \sum\nolimits_{i=1}^{n-1} imf(t)_i + r_n\,(t) \tag{12}$$

Apply the Hilbert transform to the IMF components,

$$Z\,(t) = imf\,(t) + iH\left[imf(t)\right] = a\,(t)\,e^{i\int \omega(t)dt} \tag{13}$$

in which $a\,(t) = \sqrt{imf^2\,(t) + H^2\left[imf(t)\right]}$, $\omega\,(t) = \frac{d}{dt}\left[\arctan\left(H\left[imf(t)\right]/imf(t)\right)\right]$, where $\omega\,(t)$ and $a(t)$ are the instantaneous frequency and amplitude, respectively, of the IMFs.

The Hilbert-Huang marginal spectrum is defined by:

$$h\,(\omega) = \int H\,(\omega, t)\,dt \tag{14}$$

To simplify the representation, the Hilbert-Huang spectrum is denoted as a function of frequency (f) instead of angular frequency (ω). The marginal spectrum is normalized by:

$$\hat{h}\left(f\right) = h(f)/\sum h(f) \tag{15}$$

Next, the Shannon entropy concept is applied to the Hilbert-Huang spectrum, and Hilbert-Huang spectral entropy is obtained by:

$$HHSE = -\sum\nolimits_f \hat{h}\left(f\right) log\left(\hat{h}\left(f\right)\right) \tag{16}$$

The HHSE values are mainly affected by the frequency resolution and data length (N). For accurate computation, the frequency resolution is chosen as 0.1 Hz. N directly influences the EMD. In general, the boundary effect may be induced if N is too large or too small, which can contaminate the data and distort the power spectrum. The selection of N in this study is given in Supplement Material 1.

APPROXIMATE ENTROPY (ApEn)

ApEn is derived from Kolmogorov entropy. It was introduced by Pincus (1991). It can be used to analyze a finite length signal and describe its unpredictability or randomness. Its computation involves embedding the signal into the phase space and estimating the rate of increment in the number of phase space patterns within a predefined value r, when the embedding dimension of phase space increases from m to $m + 1$.

For a time series $x\,(i)$, $1 \le i \le N$ of finite length N, reconstitute the $N - m + 1$ vectors $X_m(i)$ following the form:

$$X_m\,(i) = \{x\,(i)\,, x\,(i+1)\,, \ldots, x(i+m-1)\}\,, \quad i = 1, 2, \ldots, N-m+1 \tag{17}$$

where m is the embedding dimension.

Let $C_i^m(r)$ be the probability that any vector $X_m(j)$ is within distance r of $X_m\,(i)$, defined as:

$$C_i^m\,(r) = \frac{1}{N-m+1}\sum\nolimits_{j=1}^{N-m+1}\Theta\left(d_{ij}^m - r\right); \quad i, j = 1, 2, \ldots, N-m+1 \tag{18}$$

where d is the distance between the vectors $X_m(i)$ and $X_m\left(j\right)$, defined as:

$$d_{ij}^m = d\left[X_i^m, X_j^m\right] = \max\left(\left|x\,(i+k) - x(j-k)\right|\right), \quad k = 0, 1, \ldots, m \tag{19}$$

and Θ is the Heaviside function.

After that, define a parameter $\Phi^m(r)$:

$$\Phi^m\,(r) = (N-m+1)^{-1}\sum\nolimits_{i=1}^{N-m+1}\ln C_i^m\,(r) \tag{20}$$

Next, when the dimension changes to $m + 1$, the above process is repeated.

$$\Phi^{m+1}\,(r) = (N-m)^{-1}\sum\nolimits_{i=1}^{N-m}\ln C_i^{m+1}\,(r) \tag{21}$$

Finally, the approximate entropy is defined by:

$$ApEn\,(m, r, N) = \Phi^m\,(r) - \Phi^{m+1}(r) \tag{22}$$

The detailed algorithm is seen in Bruhn et al. (2000). The ApEn index is influenced by data length (N), tolerance (r) and embedding dimension (m). According to Pincus (1991) and Bruhn et al. (2000), N is recommended to be 1000, r 0.1∼0.25 of the standard deviation of the signal and m 2∼3. The selection of these parameters is described in Supplement Material 1.

SAMPLE ENTROPY (SampEn)

The SampEn proposed by Richman and Moorman (2000) is based on ApEn but differs from it in three ways to remove bias:

(1) SampEn eliminates self-matches;
(2) To avoid $\ln 0$ caused by removing self-matches, SampEn computes the additional operation of the total number of template well-matches prior to the logarithmic operation.

(3) In order to have an equal number of patterns for both embedding dimension m and $m+1$, the time series reconstitution in SampEn have $N-m$ rows instead of $N-m+1$ in ApEn in embedding dimension m.

The first step of calculating SampEn is the same as ApEn. When the embedding dimension is m, the total number of template matches is:

$$B^m(r) = (N-m)^{-1} \sum_{i=1}^{N-m} C_i^m(r) \tag{23}$$

Similarly, when the embedding dimension is $m+1$, the total number of template matches is:

$$A^m(r) = (N-m)^{-1} \sum_{i=1}^{N-m} C_i^{m+1}(r) \tag{24}$$

Finally, the SampEn of the time series is estimated by:

$$\text{SampEn (r, m, N)} = -\ln \frac{\text{A}^{\text{m}}(\text{r})}{\text{B}^{\text{m}}(\text{r})} \tag{25}$$

SampEn is based on ApEn, so its parameter selection procedure is similar to that of ApEn (see Supplement Material 1).

FUZZY ENTROPY (FuzzyEn)

Zadeh introduced the concept of "fuzzy set" (Zadeh, 1965). Fuzzy set provides a mechanism for measuring the degree to which a pattern belongs to a given class, by introducing the concept of "membership degree" having a fuzzy function $u_c(x)$. The nearer the value $u_c(x)$ is to unity, the higher the membership grade of x in the set C will be. Inspired by this, Chen et al. (2007) developed the FuzzyEn based on SampEn. FuzzyEn uses the fuzzy membership function $u(d_{ij}^m, r)$ to obtain the similarity between X_i^m and X_j^m instead of the Heaviside function.

FuzzyEn is based on SampEn, so its parameter selection is similar to that of SampEn (see Supplement Material 1).

PERMUTATION ENTROPY (SPE, TPE, AND RPE)

There are three types of PE measures involved in this study. PE is an ordinal analysis method, in which a given time series is divided into a series of ordinal patterns for describing the order relations between the present and a fixed number of equidistant past values (Bandt, 2005). The advantage of this method is its simplicity, robustness and low computational complexity (Li et al., 2007).

For an N-point normalized time series $\{x(i) : 1 \le i \le N\}$, firstly the time series is reconstructed:

$$X_i = \{x(i), x(i+\tau), \ldots, x(i+(m-1)\tau)\},$$
$$i = 1, 2, \ldots, N-(m-1)\tau \tag{26}$$

where τ is the time delay, m is the embedding dimension.

Then, rearrange X_i in an increasing order:

$$\{x(i+(j_1-1)\tau) \le x(i+(j_2-1)\tau)) \le \cdots \le x(i+(j_m-1)\tau\} \tag{27}$$

There are $m!$ permutations for m dimensions. Each vector X_i can be mapped to one of the $m!$ permutations.

Next, the probability of the jth permutation occurring p_j can be defined as:

$$p_j = \frac{n_j}{\sum_{j=1}^{m!} n_j} \tag{28}$$

where n_j is the number of times the jth permutation occurs.

Based on the probability of the jth permutation p_j, we define SPE, TPE and RPE as follows.

SPE is just the Shannon entropy associated with the probability distribution p_j:

$$S_1^{(s)} = -\sum_{j=1}^{m!} p_j \log p_j \tag{29}$$

And the normalized SPE is:

$$SPE_n = \frac{S_1^{(S)}}{S_{1,\max}^{(s)}} = \frac{\sum_{j=1}^{m!} p_j \log p_j}{\log(m!)} \tag{30}$$

Based on the definition of Tsallis entropy, Zunino et al., proposed the normalized TPE and defined it as Zunino et al. (2008):

$$TPE = \frac{\sum_{j=1}^{m!} \left(p_j - p_j^q\right)}{1-(m!)^{1-q}} \tag{31}$$

Furthermore, the normalized RPE measure based on the Renyi entropy and permutation probability distribution p_j is:

$$RPE_n = \frac{\log \sum_{j=1}^{m!} P_j^a}{(1-a)\ln m!} \tag{32}$$

In Li et al. (2008a, 2010, 2012), SPE was used to evaluate the effect of sevoflurane and isoflurane anesthesia on the brain. In this study, the parameters of $m=6$ and $\tau=1$ are selected for sevoflurane anesthesia as proposed in Li et al. (2008a). The SPE's parameters for isoflurane anesthesia are the same as those proposed by Li et al. (2012). TPE and RPE are first used in DoA measure, therefore selection of the appropriate parameters of TPE and RPE should be based on the experiments. The details of the selection process is shown in Supplement Material 1.

MATERIALS AND STATISTICAL METHODS

SUBJECTS AND EEG RECORDINGS

EEG data set during sevoflurane-induced anesthesia

In this study, the first data set we used was from a previous study (McKay et al., 2006), in which 19 patients aged 18–63 years were recruited from Waikato Hospital, Hamilton, New Zealand. The subjects were scheduled for elective gynecologic, general, or orthopedic surgery. All patients fasted for at least 6 h before anesthesia and received no premedication. Patients were American Society of Anesthesiologists physical status I or II and signed written informed consent following approval by the Waikato Hospital ethics committee.

Before application of Ag/AgCl electrodes, the skin was carefully cleaned with an alcohol swab to ensure electrode-skin impedance of less than 7.5 kΩ. A composite electrode, the Entropy™ Sensor, composed of a self-adhering flexible band holding three electrodes were used to record the EEG signals between the forehead and temple (active = FpZ, earth = Fp1, and reference = F8). RE and SE were measured every 5 s with a plug-in M-Entropy S/5 Module (Datex-Ohmeda). The sevoflurane concentration was measured at the mouth at 100/s (McKay et al., 2006). All data were recorded and stored on a laptop computer. Off-line analysis was performed using the MATLAB (version 8, MathWorks Inc.) software.

EEG data set during isoflurane-induced anesthesia

The second data set contains 29 patients (9 men and 20 women, age 33–77 year) receiving elective abdominal surgery during combined isoflurane general anesthesia and epidural anesthesia (Hagihira et al., 2002). These patients had no neurologic or psychiatric disorders and didn't receive medication with any drugs known to influence anesthesia. The data recordings were approved by the Osaka Prefectural Habikino Hospital and all patients gave written informed consent.

Each patient was injected intramuscularly with 0.5 mg atropine before entering the operating room. Initially, an epidural catheter was placed at the appropriate spinal location. Then, after confirming the effect of epidural analgesia, 3 mg/kg thiopental was used to induce anesthesia. Anesthesia was subsequently maintained with isoflurane, oxygen, and nitrogen after tracheal intubation. Vecuronium was given as required. Lidocaine 1% (80–110 mg/h; initial dose, 90–100 mg) was administered epidurally. Patients received controlled ventilation to maintain adequate oxygenation and normocapnia. To keep mean blood pressure at 60 mmHg, dopamines were administered as required at a dose of 2–5 μg/(kg·min).

Before induction of anesthesia, five EEG electrodes (A1, A2, FP1, FP2, and FPz) were attached to the patients according to the International 10–20 System. FPz was used as the ground electrode. The EEG signal used was recorded from a unipolar lead (FP1-A1) through a 514 X-2 EEG telemetry system (GE Marquette, Tokyo, Japan) with sample frequency of 512 Hz (another Fp2-A2 channel was not analyzed). Isoflurane was initially increased to 1.5% and then stepped down to 0.7%. The end-tidal concentration of isoflurane was purposely maintained at set levels (1.5, 1.3, 1.1, 0.9, and 0.7%) for 30 min at each level. The EEG recordings at 0.3 and 0.5% isoflurane were collected immediately after the operation. The concentration of isoflurane was continuously monitored and recorded by Canomac (Datex, Helsinki, Finland). The BSP was evident in six of the 29 EEG recordings.

The two data sets used can be obtained by asking the authors of corresponding original papers.

EEG PREPROCESSING

All the EEG recordings were preprocessed by following the steps outlined in Li et al. (2010) before further analysis. Firstly, data points whose amplitude values exceeded a threshold determined by mean and standard deviation (SD) statistics were removed as outliers. Then, the filter function filter.m was used to remove the frequency components higher than 60 Hz. This FIR filter ensures that phase information is not distorted. Thirdly the stationary wavelet transform was used to reduce electro-oculogram (EOG) artifact. Finally, an inverse filter was used to detect and remove EMG and other high-amplitude transient artifacts.

PHARMACOKINETIC/PHARMACODYNAMIC MODELING

To derive the relationship between effect-site anesthetic drug concentration and the measured EEG index, PK/PD modeling was used. These methods have been successfully used to evaluate the proposed EEG indices (Li et al., 2008a; Olofsen et al., 2008). It describes the relationship between drug dose and its effect through two successive physiological processes (McKay et al., 2006). The pharmacokinetic (PK) side of the model describes the changes in blood concentration of the drug over time, while the pharmacodynamic (PD) aspect shows the relation between the concentration of drug at its effect site and its measured effect. The simplest effect site model is a first order model, defined as:

$$dC_{\text{eff}}/dt = k_{\text{eo}}(C_{\text{et}} - C_{\text{eff}}) \tag{33}$$

where C_{eff} denotes the effect-site concentration, k_{eo} is the first-order rate constant for efflux from the effect compartment and C_{et} is the end-tidal concentration.

In addition, a non-linear inhibitory sigmoid $E_{\max}$ model was used to describe the relationship between the estimated C_{eff} and the measured EEG indices.

$$\text{Effect} = E_{\max} - (E_{\max} - E_{\min}) \times \frac{C_{\text{eff}}^{\gamma}}{EC_{50}^{\gamma} + C_{\text{eff}}^{\gamma}} \tag{34}$$

where Effect is the processed EEG measure, $E_{\max}$ and $E_{\min}$ respectively are the maximum and minimum Effect for each individual, EC_{50}^{γ} is the drug concentration that causes 50% of the maximum Effect and γ is the slope of the concentration–response relationship.

The coefficient of determination R^2 is calculated by:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} \left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{n} \left(y_i - \bar{y}\right)^2} \tag{35}$$

where y_i is the measured Effect for a given time and $\hat{y}_i$ is corresponding modeled Effect.

C_{eff} is estimated by iteratively running the above model with a series of k_{eo} values, with the optimal k_{eo} yielding the greatest R^2 for each patient.

MDFA EXPONENT

Kantelhardt et al., proposed the MDFA method to describe the non-stationary time series, which is based on a generalization DFA method (Kantelhardt et al., 2002). Nguyen-Ky et al., used the moving-average DFA method to monitoring the DoA and the results showed that DFA could accurately estimate a patient's hypnotic state (Nguyen-Ky et al., 2010a).

For a time series $x(t)$ of length N, the main computation procedure of MDFA consists of three steps.

Step 1. Construct the profile as the equation showed below,

$$y(j) = \sum_{i=}^{j} [x(i) - \langle x \rangle] \tag{36}$$

where $\langle x \rangle$ represents the average value of $x(t)$.

Step 2. Divide the new profile $\{y(j)\}$ into $N_s = N/s$ non-overlapping segments of equal length s. Since the record length N may not be a multiple of the considered time scale s, a short part at the end of the profile will remain in most cases. In order not to disregard this part of record, the same procedure is repeated starting from the other end of the profile $\{y(j)\}$. Thus, $2N_s$ segments are obtained altogether.

Step 3. Calculate the local trend for each segment by a least-square fit of the data and calculate the variance $F^2(s, v)$. Thus, the qth order fluctuation function is calculated as follows:

$$F_q(s) = \left\{ \frac{1}{2N_s} \sum_{v=1}^{2N} \left[F^2(s, v)^{q/2} \right] \right\}^{1/q} \tag{37}$$

If $q = 0$, then

$$F_0(s) = \exp\left\{ \frac{1}{4N_s} \sum_{v=1}^{2N_s} \ln\left[F^2(s, v) \right] \right\} \tag{38}$$

It is obvious that when $q = 2$, we have the standard DFA procedure.

MFDFA characterizes the evolution of $F_q(s)$ is a function of the segment length s. Modeling fluctuations that present a power-law behavior between $F_q(s)$ and s, $F_q(s) \propto s^{h(q)}$, where the $h(q)$ is generalized Hurst exponent.

For the multifractal time series, the scaling behavior is sensitive with the parameter q. For positive q, $h(q)$ describes the scaling behavior of the segments with large fluctuations. On the contrary, for negative q, $h(q)$ is sensitive to small fluctuations. For more detail of the MDFA method, see in Kantelhardt et al. (2002).

In this study, we only considered the influence of q with the MDFA measure. The selection of parameter is described in Supplement Material 1.

STATISTICAL ANALYSIS

To further evaluate the correlation between the measured EEG index and underlying anesthetic drug effect, prediction probability (P_k) statistics were applied, as described in Smith et al. (1996). Given two random data points with different C_{eff}, P_k describes the probability that the measured EEG index correctly predicts the C_{eff} of the two points. Its definition is:

$$P_k = \frac{P_c + P_{tx}/2}{P_c + P_d + P_{tx}} \tag{39}$$

where P_c, P_d and P_{tx} separate the probability that two data points drawn at random, independently and with replacement from the population are a concordance, a discordance or an x-only tie. A value of 1 means that the EEG index is perfectly concordant with C_{eff}; whereas a value of 0.5 means the EEG index is obtained by chance. When the monotonic relation between the drug concentration and the EEG index is negative, the resultant P_k value is replaced by $1 - P_k$.

In addition, The Kolmogorov–Smirnov test was used to determine whether the data sets were normally distributed. To assess the index stability during the awake state and the sensitivity to the induction process, the relative coefficient of variation (CV) (Li et al., 2008a) was used. Kruskal-Wallis test was used to determine the significant difference of the index values between awake, induction, anesthesia and recovery states.

RESULTS

First we used these entropy measures on EEG data from sevoflurane anesthesia. **Figure 1A** shows a preprocessed EEG recording and the derivative from the EEG signal during the whole sevoflurane induction process, from awake to induction, then to deep anesthesia, and finally to recovery. With deepening anesthesia, the mean amplitude of the EEG gradually increased and then the amplitude decreased in the state of recovery. The concurrent end-tidal sevoflurane concentration is represented by the black line given in **Figure 1B**. It can be regarded as the drug concentration in blood, derived from the recorded sevoflurane concentration at the mouth (represented by gray line). The changes in RE, SE, SWE, TWE, RWE, HHSE, ApEn, SampEn, FuzzyEn, SPE, TPE, RPE, and MDFA corresponding to the EEG recording are successively given in **Figures 1C–K**. As can be seen, all the entropy indices generally followed the changes in EEG pattern as the drug concentration increased and decreased. And MDFA had the opposite trend with entropy indices.

Then we analyzed the EEG recording during isoflurane anesthesia using the same entropy algorithms and MDFA methods. **Figures 2A,B** are the EEG recording and isoflurane end-tidal concentration respectively. It can be seen that the drug concentration increased and then decreased. **Figures 2C–K** shows the same entropy and MDFA indices as **Figures 1C–K**, and demonstrate equivalent trends, in line with changes in drug concentration.

Loss of consciousness (LOC) is the most important clinical time point during anesthesia. We investigated the ability of these entropies in tracking LOC. **Figure 3** demonstrates the changes in each index around LOC, from LOC−30 s to LOC+30 s for all subjects during sevoflurane anesthesia. For these plots, index values were normalized to between 0 and 1. It can be seen in **Figures 3A–N** that MDFA(−8) decreased most rapidly, followed by SWE. Thus, the MDFA with $q = -8$ appeared to be the most sensitive to LOC. To verify this, we calculated the absolute slope values (mean ± SD) of the linear-fitted polynomials vs. time for these indices, as shown in **Figure 3O**. As can be seen, the absolute slope value for MDFA(−8) (0.44 ± 0.22) is largest, followed by SWE (0.43 ± 0.23).

To further compare the ability of the indices to distinguish different anesthesia states, the sevoflurane anesthesia procedure was divided into four states, i.e., awake, induction, deep anesthesia, and recovery. For each index, a box plot is given in **Figure 4**. The data was not normally distributed, so the statistics of the 19 patients undergoing sevoflurane anesthesia were expressed as median (min—max), as shown in **Table 1**. All the entropy indices monotonically decreased as anesthesia deepened, then increased

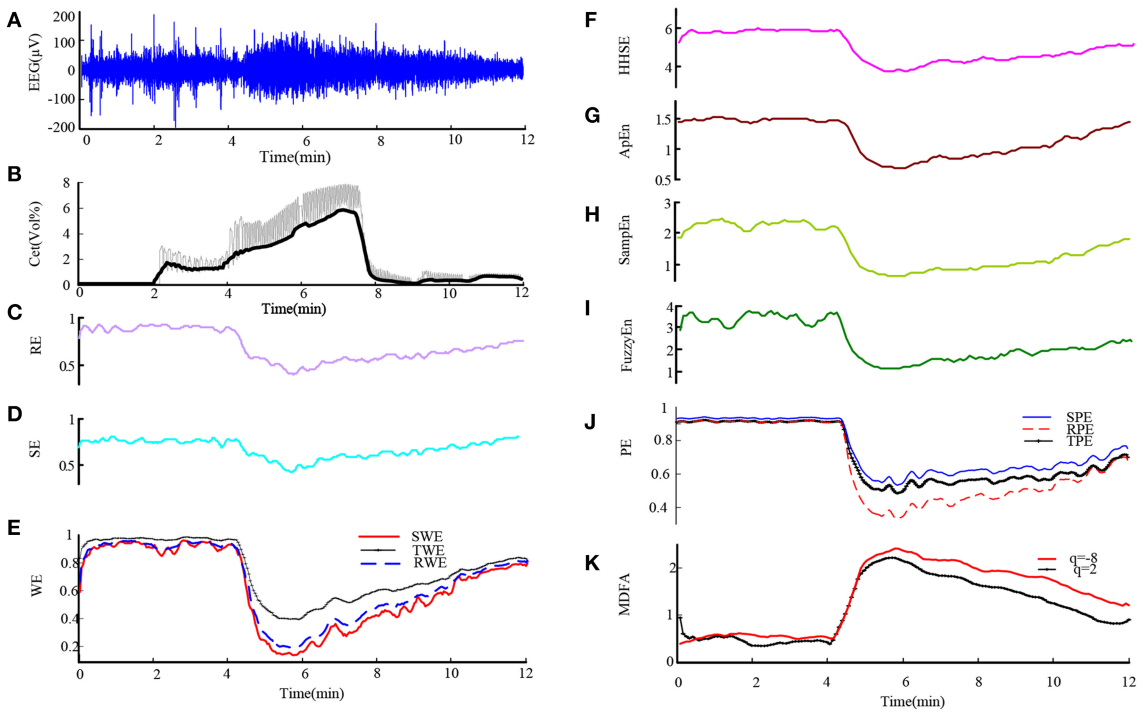


FIGURE 1 | An EEG recording from a patient undergoing sevoflurane anesthesia and corresponding entropy indices vs. time. (A) Preprocessed EEG recording. **(B)** Sevoflurane concentration recorded at the mouth (gray line) and the derived end-tidal sevoflurane concentration (black line). **(C–J)** The time course of the studied EEG derivative. The indices are calculated over a window of 10 s with an overlap of 75%. **(K)** The time course of MDFA at $q = 2$ [MDFA(2)] and $q = -8$ [MDFA(−8)]. The window and overlap selection are similar with entropy measures.

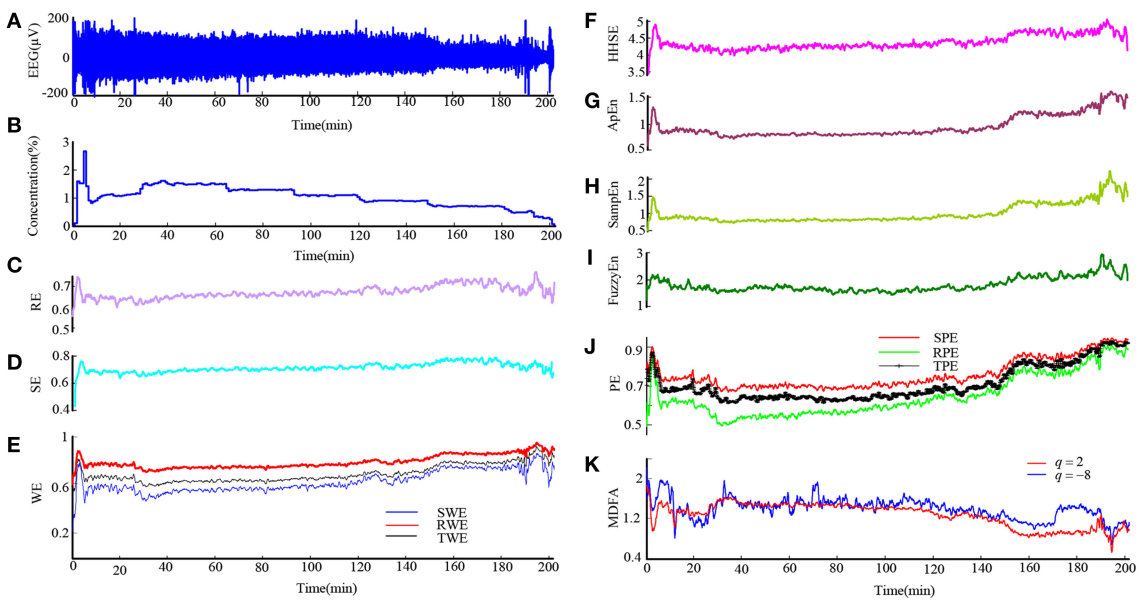


FIGURE 2 | An EEG recording from a patient in isoflurane anesthesia and calculated indices. (A) Preprocessed EEG recording, re-sampled at 128 Hz. **(B)** Recording of the isoflurane end-tidal concentration. **(C–J)** Time course of entropy indices, with a time interval of 10 s and 5 s overlap. **(K)** Time course of MDFA measures with a time interval of 10 and 5 s overlap.

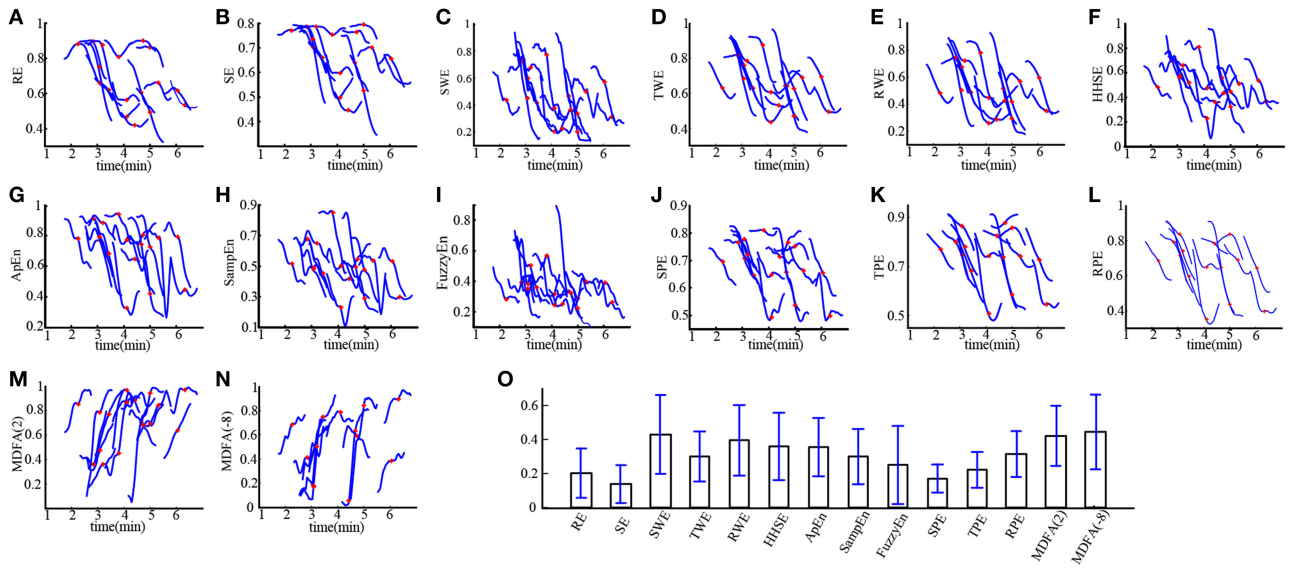


FIGURE 3 | Entropy and MDFA analysis around the time of LOC for subject undergoing sevoflurane anesthesia ($n = 19$). (A–N) The normalized indices around LOC (from LOC − 30 s to LOC + 30 s) for all subjects. The red plus sign denotes the point of LOC. **(O)** Statistical analysis of the absolute slope of the linear-fitted polynomials vs. time for studied indices. Bar height indicates the mean value, and the lower and upper line are the 95% confidence interval of each index.

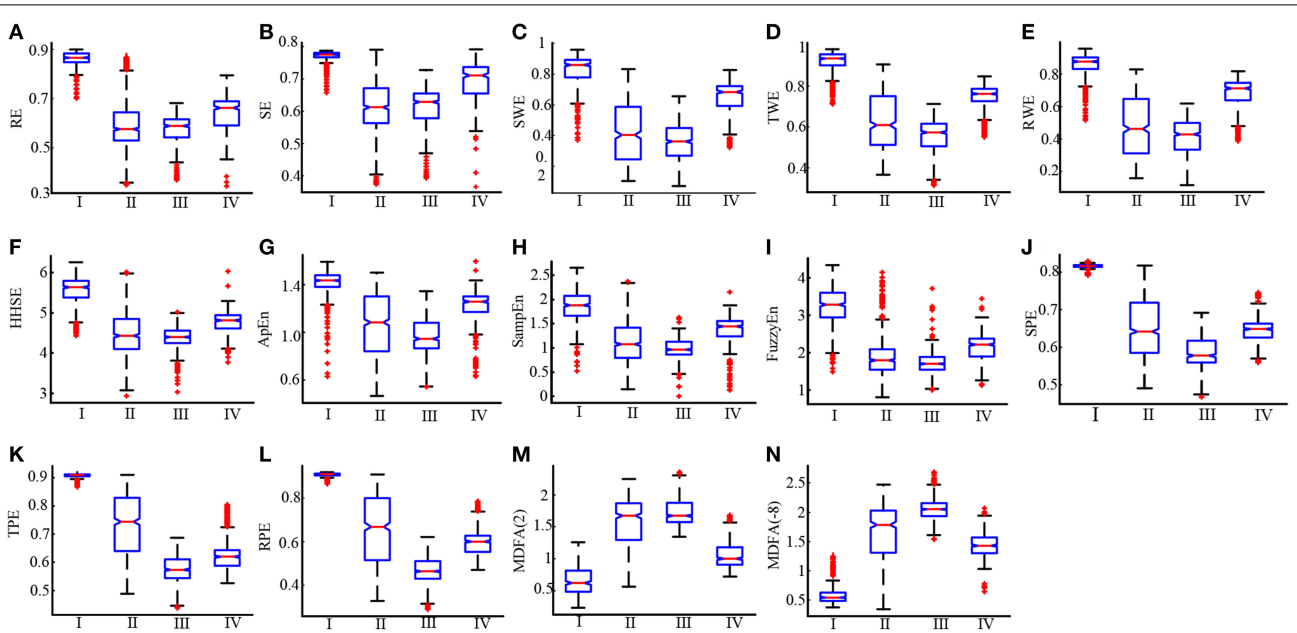


FIGURE 4 | Box plots of RE, SE, SWE, TWE, RWE, HHEn, ApEn, SampEn, FuzzyEn, SPE, TPE, RPE, MDFA(2) and MDFA(-8) (A–N) at awake (I), induction (II), deep anesthesia (III) and recovery (IV) states.

during recovery. The MDFA indices have an opposite trend with the entropy measures. These are consistent with the results in **Figure 1**. The overlap of three types of PE (SPE, TPE, and RPE) values between the awake and deep anesthesia states were smaller than the other indices. This means the PE has a better ability to separate these states and a greater robustness for individual differences.

To estimate the baseline variability and the sensitivity to the induction process of each index, the CV value of all the indices for the sevoflurane data set are computed and the results are given in **Table 2**. During the awake state, the CV value of SampEn was 0.095, which was the highest; The CV value of TPE was 0.003, significantly lower than MDFA(2) (0.240) and MDFA(−8) (0.125) and the other indices. The CV values of SPE and RPE were lower

Table 1 | The statistics of the studied indices at different anesthetic states [median (min-max)].

	Awake	Induction	Deep anesthesia	RoC
RE	0.87 (0.65–0.90)	0.58 (0.35–0.89)	0.59 (0.37–0.68)	0.66 (0.34–0.79)
SE	0.77 (0.65–0.79)	0.61 (0.37–0.79)	0.63 (0.39–0.73)	0.71 (0.37–0.79)
SWE	0.86 (0.37–0.96)	0.40 (0.10–0.83)	0.36 (0.07–0.66)	0.68 (0.32–0.83)
TWE	0.93 (0.71–0.98)	0.61 (0.37–0.91)	0.57 (0.32–0.71)	0.76 (0.55–0.85)
RWE	0.88 (0.52–0.96)	0.46 (0.16–0.83)	0.43 (0.12–0.62)	0.71 (0.39–0.82)
HHSE	5.63 (4.43–6.26)	4.43 (2.93–6.01)	4.40 (3.02–5.02)	4.81 (3.76–6.03)
ApEn	1.44 (0.63–1.59)	0.95 (0.54–1.35)	1.08 (0.47–1.50)	1.26 (0.63–1.60)
SampEn	1.88 (0.52–2.65)	1.08 (0.15–2.37)	0.97 (0.01–1.63)	1.44 (0.13–2.16)
FuzzyEn	3.28 (1.49–4.33)	1.80 (0.81–4.14)	1.70 (1.01–3.72)	2.22 (1.13–3.44)
SPE	0.81 (0.79–0.83)	0.64 (0.49–0.82)	0.58 (0.46–0.82)	0.65 (0.56–0.75)
TPE	0.91 (0.87–0.92)	0.74 (0.49–0.91)	0.57 (0.44–0.69)	0.62 (0.53–0.80)
RPE	0.91 (0.87–0.92)	0.67 (0.33–0.91)	0.46 (0.29–0.62)	0.60 (0.47–0.79)
MDFA (2)	0.62 (0.23–1.26)	1.67 (0.56–2.25)	1.67 (1.35–2.36)	1.00 (0.72–1.68)
MDFA (−8)	0.54 (0.38–1.32)	1.79 (0.35–2.47)	2.05 (1.54–2.68)	1.43 (0.84–2.06)

RE, response entropy in the M-entropy module; SE, state entropy; SWE, Shannon wavelet entropy; TWE, Tsallis wavelet entropy; RWE, Renyi wavelet entropy; HHSE, Hilbert-Huang spectral entropy; ApEn, approximate entropy; SampEn, sample entropy; FuzzyEn, fuzzy entropy; SPE, Shannon permutation entropy; TPE, Tsallis permutation entropy; RPE, Renyi permutation entropy; MDFA(2), Multifractal detrended fluctuation analysis with $q = 2$; MDFA(-8), Multifractal detrended fluctuation analysis with $q = -8$.

Table 2 | The CV of the studied indices at different anesthetic states.

	Awake	Induction	Deep	RoC
RE	0.025	0.149	0.047	0.052
SE	0.016	0.122	0.047	0.050
SWE	0.080	0.338	0.177	0.077
TWE	0.024	0.161	0.063	0.038
RWE	0.043	0.276	0.127	0.057
HHSE	0.029	0.089	0.027	0.024
ApEn	0.040	0.193	0.064	0.043
SampEn	0.095	0.259	0.087	0.094
FuzzyEn	0.089	0.193	0.088	0.073
SPE	0.006	0.115	0.028	0.025
TPE	0.003	0.138	0.030	0.028
RPE	0.004	0.219	0.043	0.041
MDFA(2)	0.240	0.176	0.046	0.100
MDFA(-8)	0.125	0.256	0.047	0.097

than other indices as well. The lower CV value of PE illustrates that PE measures were less sensitive to noise, while MDFA methods were least robust against noise. During induction, the CV of SWE (0.338) was the highest. This demonstrates that SWE had a faster response speed compared to the other indices.

In order to verify the performance of all the indices for monitoring DoA and detecting the burst suppression state, we analyzed the isoflurane anesthesia data set, in which some subjects entered into the burst suppression state during deep anesthesia. The results are given in histogram form and shown in **Figure 5**. All the indices except SE and MDFA decreased with increasing isoflurane concentration. During burst suppression, only ApEn and SampEn continued to decrease. This means that the ApEn and SampEn algorithms could be used to evaluate DoA including detection of the burst suppression state, without the need for Supplementary Methods. The tabulated results for each index at the different isoflurane concentrations and BSP are presented in **Table 3**. The CV of the indices show that PE (0.033) outperformed the others in awake state (0% concentration) (see **Table 4**). And the CV of two MDFA measures were relative higher in awake state. It indicate that MDFA algorithms were no better than some entropy measures in anti-noise performance.

To further compare the performance of the studied indices, PK/PD modeling was performed to describe the relationship between the index values and the estimated sevoflurane and isoflurane effect-site concentration. **Tables 5, 6** give these parameters for isoflurane and sevoflurane anesthesia respectively, in which the maximum coefficient of determination (R^2) gives the correlation between the index values and the anesthetic effect site concentration. **Figures 6A,B** show the R^2 values of the indices for the two data sets. **Figure 6A** shows the R^2 values for sevoflurane. It can be seen that R^2 for TPE (0.95, 95% confidence interval 0.92–0.98) was significantly higher than the other entropy indices. **Figure 6B** shows R^2 values for isoflurane. Again, R^2 for SPE (0.81) was higher than the other entropy indices. Although R^2 of MDFA with $q = 8$ was relative higher in sevoflurane anesthesia, the value in isoflurane anesthesia was lower. The statistical analysis also shows that for the same entropy algorithm, the mean R^2 value for sevoflurane was significantly higher than for isoflurane.

To assess the performance of the indices to correctly predict drug effect-site concentrations, we evaluated the prediction probability P_k of all the indices from the PK/PD modeling for all the subjects, as shown in **Figures 7A,B**. And the statistical results are shown in **Table 7**. Overall, most P_k values of indices for sevoflurane were higher than for isoflurane. For sevoflurane, P_k of RPE and MDFA were equal (0.87, 95% confidence interval is 0.83–0.90 and 0.83–0.92 respectively), slightly higher than RWE (0.85) and TWE 0.81 (95% confidence interval 0.79–0.84). Also, P_k of

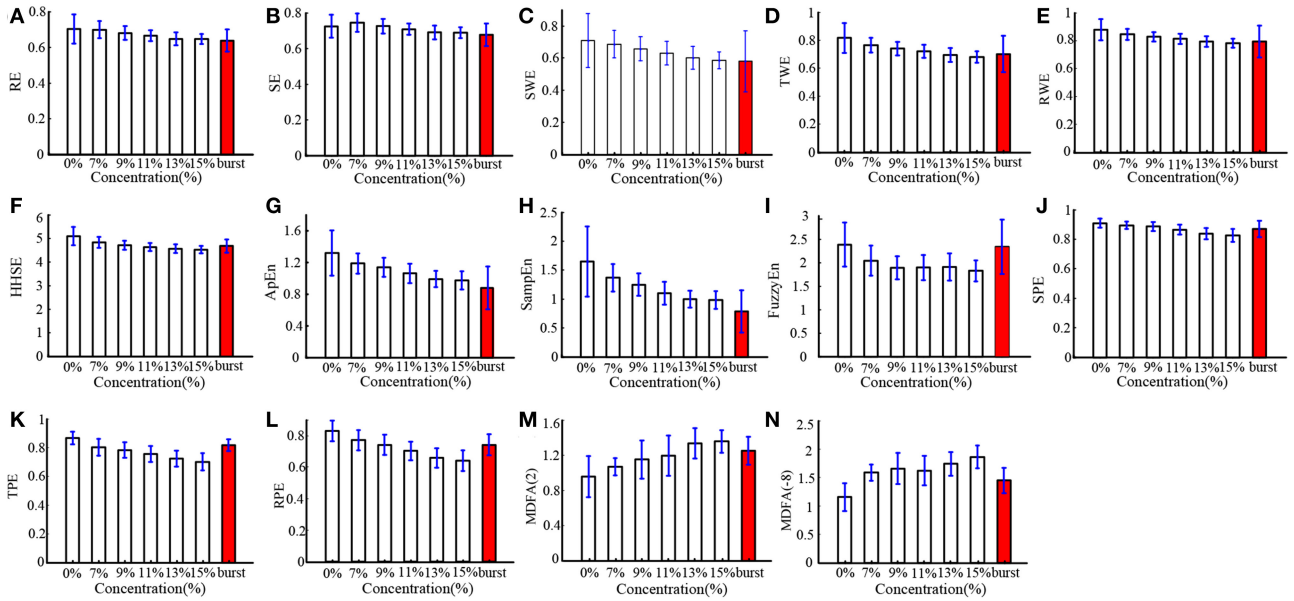


FIGURE 5 | Histograms of entropy (A–L) and MDFA (M,N) indices for patients induced with different isoflurane concentrations, including 0, 7, 9, 11, 13, 15% and the concentration at which burst suppression occurred. The burst suppression state is highlighted by the red bar.

Table 3 | The statistics of the studied indices at different isoflurane concentrations [median (min-max)].

	Concentrations and BSP						
	0%	**7%**	**9%**	**11%**	**13%**	**15%**	**BSP**
RE	0.70 (0.42–0.91)	0.70 (0.46–0.80)	0.68 (0.49–0.79)	0.67 (0.50–0.77)	0.65 (0.43–0.73)	0.65 (0.55–0.72)	0.65 (0.37–0.82)
SE	0.73 (0.45–0.83)	0.75 (0.49–0.85)	0.73 (0.52–0.85)	0.71 (0.54–0.83)	0.70 (0.46–0.78)	0.69 (0.59–0.77)	0.69 (0.39–0.83)
SWE	0.73 (0.03–0.95)	0.70 (0.02–0.87)	0.67 (0.30–0.87)	0.63 (0.34–0.81)	0.61 (0.30–0.80)	0.60 (0.41–0.74)	0.62 (0–0.94)
TWE	0.82 (0.24–0.97)	0.76 (0.22–0.91)	0.74 (0.56–0.88)	0.71 (0.14–0.85)	0.70 (0.52–0.86)	0.67 (0.55–0.80)	0.72 (0.12–0.97)
RWE	0.89 (0.36–0.98)	0.85 (0.33–0.95)	0.83 (0.69–0.93)	0.81 (0.24–0.91)	0.80 (0.66–0.91)	0.78 (0.68–0.87)	0.82 (0.19–0.98)
HHSE	5.06 (3.53–5.95)	4.82 (3.57–5.48)	4.71 (3.93–5.33)	4.64 (3.62–5.24)	4.58 (3.66–5.07)	4.53 (4.02–4.95)	4.70 (3.38–5.33)
ApEn	1.45 (0.07–1.60)	1.17 (0.06–1.55)	1.14 (0.82–1.48)	1.06 (0.01–1.42)	0.98 (0.63–1.34)	0.95 (0.73–1.29)	0.90 (0.07–1.51)
SampEn	1.75 (0.03–2.58)	1.31 (0.02–2.18)	1.22 (0.78–1.90)	1.10 (0.01–1.78)	0.99 (0.40–1.49)	0.95 (0.38–1.42)	0.78 (0.02–1.88)
FuzzyEn	2.37 (0.56–3.93)	2.00 (0.33–3.29)	1.86 (1.23–2.89)	1.86 (0.61–3.04)	1.87 (1.13–3.17)	1.81 (1.29–2.65)	2.45 (0.32–3.47)
SPE	0.92 (0.66–0.94)	0.90 (0.39–0.94)	0.89 (0.76–0.94)	0.87 (0.41–0.94)	0.84 (0.69–0.92)	0.82 (0.69–0.92)	0.88 (0.47–0.92)
TPE	0.88 (0.73–0.92)	0.79 (0.65–0.92)	0.78 (0.61–0.91)	0.76 (0.59–0.89)	0.72 (0.59–0.88)	0.69 (0.58–0.85)	0.82 (0.67–0.89)
RPE	0.85 (0.59–0.91)	0.76 (0.60–0.90)	0.74 (0.55–0.90)	0.70 (0.36–0.87)	0.66 (0.47–0.85)	0.63 (0.48–0.81)	0.75 (0.55–0.86)
MDFA (2)	0.96 (0.41–1.61)	1.07 (0.81–1.42)	1.23 (0.56–1.56)	1.20 (0.69–1.66)	1.31 (0.92–1.81)	1.37 (1.01–1.74)	1.27 (0.77–1.95)
MDFA (-8)	1.21 (0.55–2.13)	1.58 (1.19–2.22)	1.69 (1.04–2.32)	1.62 (0.98–2.36)	1.71 (1.09–2.36)	1.88 (1.32–2.59)	1.42 (0.32–2.89)

RPE was higher than that of TPE and SPE. Similarly, P_k of RWE was highest in three WE methods. It means that Renyi entropy had a better performance in predicting drug effect-site concentrations comparing with Shannon entropy and Tsallis entropy. The differences between RPE and the other indices were statistically significant (all $p < 0.05$, paired t-test), except for MDFA(-8). And the difference between RPE and TPE, SPE were statistically significant ($p = 0.03$ and 0.01 respectively, paired t-test), which means that RPE had a stronger ability to track the sevoflurane effect-site concentration during anesthesia. In order to get a more intuitive comparison, the best curve fits of all indices against the effect-site concentration are demonstrated for both sevoflurane (**Figure 8**) and isoflurane (**Figure 9**).

To compare the timeliness performance of each index in tracking DoA, we recorded the computing time of each index for the same subject. 20 EEG recordings from the two data sets were selected. The calculate epoch length (N) of each algorithm is equal to 10 s, and the overlap select 5.0 s. The computing time for 1 min of EEG data compared for each index is given in **Table 8**. The fastest index was WE (0.025 ± 0.001 s). The RE/SE and PE computation times were 0.096 ± 0.008 s and 0.545 ± 0.016 s respectively. The MDFA (16.338 ± 0.280 s) was the slowest.

The desktop computer used for this test had the following configuration: Intel Core i3 CPU, 4 cores at 2.93 GHz, with 2 GB of RAM, running Windows XP professional operating system.

DISCUSSION AND CONCLUSION

In this study, we investigated the performance of 12 entropy algorithms to assess the effect of GABAergic anesthetic agents on EEG activity, including RE, SE, SWE, TWE, RWE, HHSE, ApEn, SampEn, FuzzyEn, SPE, TPE, and RPE. Two data sets including sevoflurane and isoflurane anesthesia were employed as the test samples for evaluating the entropy algorithms. We compared their performance in estimating the DoA and detecting the burst suppression pattern. PK/PD modeling and prediction probability statistics were applied to assess their effectiveness. In addition, we compared the MDFA measure with all entropy indices to test the efficiency of entropy approach.

The twelve entropy measures could be divided into two classes: time-domain-based and time-frequency-domain-based analyses. On one hand, ApEn, SampEn, FuzzyEn, and PE are time domain analysis methods. All these entropy algorithms are based on nonlinear theories, and the first three are phase space analytical methods (Chen et al., 2009). PE is based on ordinal pattern analysis of the time series (Bandt, 2005). Considering that the EEG has non-linear characteristics, these four methods have their advantages. For example, FuzzyEn and PE are less sensitive to the signal quality and calculation length (Pincus, 1991; Li et al., 2008a). Relative to ApEn and SampEn, FuzzyEn can resolve more detail in the time series and has more accurate definition in theory (Chen et al., 2009). On the other hand, RE, SE, WE, and HHSE indices are based on the time-frequency domain. The start point of RE and SE is the spectral entropy, which has the particular advantage that the contributions to entropy from any particular frequency range are explicitly separated. In order to achieve optimal response time, RE and SE adopt a variable time window for each particular frequency-called time-frequency balanced spectral entropy (Viertiö-Oja et al., 2004). Compared to the variable time windows of RE and SE, the window function of WE is variable in both time and frequency domains. The HHSE algorithm is based on the EMD and Hilbert transform (Li et al., 2008b). The advantage of this method is that it can estimate the instantaneous amplitude and phase/frequency. Also it can break down a complicated signal without a basis function (such as sine or wavelet functions) into several oscillatory modes that are embedded in this complicated signal. The marginal spectrum gives a more accurate and nearly continuous distribution of EEG energy, which is completely different from the Fourier spectrum (Li et al., 2008b).

Table 4 | The CV of indices for different isoflurane concentrations.

	Concentrations and BSP						
	0%	**7%**	**9%**	**11%**	**13%**	**15%**	**BSP**
RE	0.118	0.070	0.057	0.046	0.056	0.045	0.097
SE	0.089	0.070	0.057	0.046	0.055	0.044	0.093
SWE	0.237	0.125	0.114	0.111	0.118	0.090	0.328
TWE	0.130	0.070	0.064	0.065	0.071	0.060	0.187
RWE	0.087	0.047	0.042	0.045	0.048	0.040	0.143
HHSE	0.077	0.048	0.041	0.037	0.040	0.035	0.060
ApEn	0.216	0.108	0.106	0.114	0.103	0.119	0.308
SampEn	0.368	0.172	0.156	0.178	0.147	0.154	0.466
FuzzyEn	0.196	0.156	0.131	0.141	0.152	0.122	0.249
SPE	0.033	0.028	0.033	0.038	0.046	0.053	0.064
TPE	0.052	0.073	0.069	0.074	0.078	0.085	0.050
RPE	0.079	0.083	0.086	0.086	0.095	0.101	0.090
MDFA(2)	0.24	0.08	0.19	0.19	0.13	0.09	0.13
MDFA(−8)	0.21	0.09	0.17	0.16	0.12	0.11	0.15

Table 5 | The PK/PD modeling parameters for sevoflurane.

	$t_{1/2}k_{eo}$(min)	γ	E_{max}	E_{min}	EC_{50}	R^2
RE	0.04 ± 0.03	8.25 ± 7.62	0.46 ± 0.09	0.13 ± 0.06	1.19 ± 0.60	0.80 ± 0.14
SE	0.06 ± 0.06	5.22 ± 2.32	0.35 ± 0.09	0.14 ± 0.05	1.71 ± 0.93	0.72 ± 0.16
SWE	0.07 ± 0.02	4.01 ± 3.12	1.01 ± 0.16	0.15 ± 0.07	1.42 ± 0.51	0.79 ± 0.12
TWE	0.03 ± 0.01	3.81 ± 1.86	0.50 ± 0.10	0.05 ± 0.16	1.54 ± 0.63	0.86 ± 0.06
RWE	0.04 ± 0.02	5.95 ± 3.98	0.58 ± 0.10	0.12 ± 0.07	1.68 ± 0.60	0.85 ± 0.06
HHSE	0.05 ± 0.02	4.15 ± 3.43	1.99 ± 0.41	0.62 ± 0.34	1.56 ± 1.15	0.80 ± 0.06
ApEn	0.05 ± 0.02	8.22 ± 6.62	0.82 ± 0.17	0.22 ± 0.11	1.84 ± 0.52	0.78 ± 0.11
SampEn	0.05 ± 0.02	5.68 ± 4.45	1.46 ± 0.38	0.40 ± 0.22	1.64 ± 0.62	0.75 ± 0.12
FuzzyEn	0.06 ± 0.04	2.75 ± 1.54	2.14 ± 0.40	0.58 ± 0.32	1.05 ± 0.38	0.69 ± 0.17
SPE	0.70 ± 0.32	4.65 ± 1.57	0.32 ± 0.05	0.08 ± 0.03	1.30 ± 0.33	0.94 ± 0.04
TPE	0.18 ± 0.01	6.98 ± 3.19	0.39 ± 0.04	0.02 ± 0.12	1.33 ± 0.37	0.96 ± 0.02
RPE	0.02 ± 0.01	4.67 ± 3.25	0.50 ± 0.14	0.10 ± 0.16	1.40 ± 0.48	0.95 ± 0.03
MDFA(2)	0.07 ± 0.03	4.92 ± 3.10	0.27 ± 0.15	1.37 ± 0.32	1.52 ± 0.49	0.88 ± 0.06
MDFA(-8)	0.05 ± 0.02	4.54 ± 2.57	0.03 ± 0.27	1.67 ± 0.14	1.33 ± 0.40	0.94 ± 0.03

$t_{1/2}k_{eo}$, blood effect-site equilibration constant; γ, slope parameter of the concentration-response relation; E_{max}, EEG parameter value corresponding to the maximum drug effect; E_{min}, EEG parameter value corresponding to the minimum drug effect; EC_{50}, concentration that causes 50% of the maximum effect; R^2, maximum coefficients of determination.

Table 6 | Parameters of PK/PD models for isoflurane.

	$t_{1/2}k_{eo}$(min)	γ	E_{max}	E_{min}	EC_{50}	R^2
RE	0.04 ± 0.04	28.88 ± 61.28	0.20 ± 0.04	2.91 ± 0.81	0.91 ± 0.20	0.64 ± 0.07
SE	0.05 ± 0.05	33.32 ± 70.92	0.21 ± 0.04	−1.27 ± 0.50	0.74 ± 0.19	0.65 ± 0.08
SWE	0.05 ± 0.07	19.44 ± 47.62	0.40 ± 0.09	0.14 ± 0.19	1.01 ± 0.20	0.72 ± 0.09
TWE	0.03 ± 0.03	4.80 ± 7.32	0.32 ± 0.11	0.07 ± 0.19	1.00 ± 0.31	0.74 ± 0.09
RWE	0.02 ± 0.01	3.87 ± 6.82	0.23 ± 0.05	0.05 ± 0.15	0.98 ± 0.33	0.75 ± 0.09
HHSE	0.02 ± 0.01	16.70 ± 27.10	1.29 ± 0.58	−5.03 ± 14.83	5.00 ± 10.90	0.72 ± 0.08
ApEn	0.06 ± 0.06	6.46 ± 6.48	0.74 ± 0.27	0.25 ± 0.32	0.75 ± 0.21	0.69 ± 0.17
SampEn	0.03 ± 0.02	5.32 ± 6.73	12.95 ± 13.50	6.79 ± 0.81	0.87 ± 0.28	0.72 ± 0.10
FuzzyEn	0.02 ± 0.01	7.82 ± 15.16	9.21 ± 32.21	0.52 ± 0.42	0.72 ± 0.37	0.61 ± 0.14
SPE	0.06 ± 0.2	3.32 ± 7.35	0.13 ± 0.12	−0.01 ± 0.21	1.30 ± 1.41	0.81 ± 0.07
RPE	0.02 ± 0.01	1.94 ± 5.51	0.42 ± 0.44	0.04 ± 0.34	0.77 ± 0.22	0.78 ± 0.09
TPE	0.01 ± 0.01	5.55 ± 6.64	0.90 ± 2.37	0.08 ± 0.09	0.68 ± 0.24	0.76 ± 0.07
MDFA(2)	0.01 ± 0.02	4.54 ± 10.73	0.17 ± 0.24	0.33 ± 0.45	0.41 ± 0.50	0.78 ± 0.09
MDFA(−8)	0.02 ± 0.01	11.54 ± 20.60	0.02 ± 1.52	1.07 ± 0.51	0.68 ± 0.23	0.69 ± 0.11

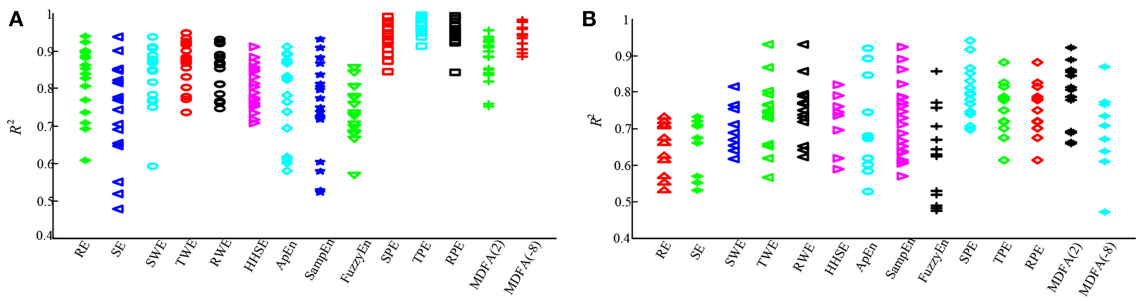


FIGURE 6 | Statistical analysis of the sevoflurane and isoflurane anesthesia datasets for each of the entropy and MDFA indices. (A) Maximum coefficient of determination values for sevoflurane anesthesia ($n = 19$). For comparison, the R^2 values for each index are expressed by a different sign and color. **(B)** The R^2 value of the same entropy indices for isoflurane anesthesia ($n = 20$).

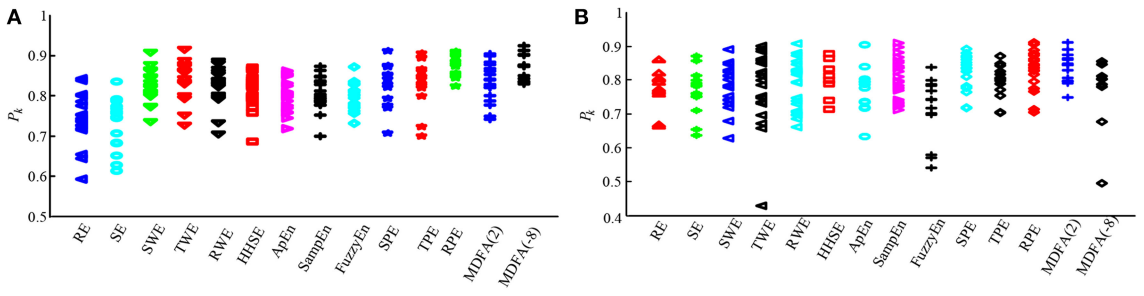


FIGURE 7 | Statistical analysis of prediction probability (P_k) values for sevoflurane and isoflurane anesthesia. (A) The P_k values for each entropy and MDFA index under sevoflurane anesthesia ($n = 19$). **(B)** The P_k values for each index during isoflurane anesthesia ($n = 20$).

Although each entropy algorithm has theoretical advantages with respect to the characterization of EEG recordings during GABAergic anesthesia, we still need to assess the practical performance from several perspectives. In qualitative terms, all the indices are effective at tracking the changes of drug concentration through the EEG analysis. As demonstrated in the presented figures and tables, all the entropies decreased with deepening anesthesia. However, there are quantitative differences between indices for different anesthesia states. This is because the principles underlying each algorithm are entirely different. Entropies based on the time domain, ApEn for example, measure the predictability of future amplitude values of the electroencephalogram based on the knowledge of one or two previous amplitude values. With increasing GABAergic anesthetic drug

concentration, the EEG signals become more regular, which leads to a reduction in the ApEn value. Entropies based on the time-frequency domain, such as RE and SE, also decrease with increasing DoA because the EEG shifts to a simpler frequency pattern as the anesthetic dose increases (Rampil, 1998).

In all 12 entropy measures, the TWE, RWE, TPE, and RPE are based on the Tsallis entropy and Renyi entropy theory respectively. Tsallis entropy and Renyi entropy theory are considered generalized concept of entropy compared to Shannon entropy. Similar to Renyi entropy, the Tsallis entropy uses the non-extensive parameter q to measure the information of specific events. The results showed that TPE and RPE were better than SPE in assessing the effect of anesthesia. Similar results can also be seen in TWE, RWE, and SWE. There are no studies using TPE or RPE in DoA monitoring before. The excellent performance indicates their potential usefulness in anesthesia analysis.

Table 7 | The P_k statistics for sevoflurane and isoflurane anesthesia for each entropy and MDFA index.

Entropy index	P_k sevoflurane	P_k isoflurane
RE	0.74 ± 0.06	0.78 ± 0.06
SE	0.73 ± 0.06	0.77 ± 0.07
SWE	0.83 ± 0.04	0.78 ± 0.07
TWE	0.84 ± 0.05	0.77 ± 0.10
RWE	0.85 ± 0.05	0.78 ± 0.07
HHSE	0.81 ± 0.04	0.80 ± 0.06
ApEn	0.80 ± 0.04	0.77 ± 0.07
SampEn	0.81 ± 0.03	0.81 ± 0.06
FuzzyEn	0.80 ± 0.03	0.71 ± 0.09
SPE	0.83 ± 0.05	0.82 ± 0.05
TPE	0.83 ± 0.06	0.80 ± 0.05
RPE	0.87 ± 0.03	0.83 ± 0.06
MDFA(2)	0.83 ± 0.05	0.83 ± 0.04
MDFA(−8)	0.87 ± 0.03	0.76 ± 0.11

Furthermore, the coefficient of determination and prediction probability statistics were used to assess the correlation of each index with the anesthetic drug effect site concentration. Three PE measures had a higher P_k and R^2 compared with the other indices. Also, MDFA at $q = 2$ had a relative higher P_k and R^2 in all indices. Comparing anesthetic drugs, the R^2 values for sevoflurane anesthesia were higher than for isoflurane anesthesia, while the P_k values were similar (see **Figures 5**, **6** and **Table 3**). This means that the entropy measures were better able to track sevoflurane than isoflurane effect site concentration.

Four additional measures were considered for evaluation of each entropy index. First, the CV was used to evaluate the sensitivity of each index to artifacts during the awake state (Li et al., 2008b, 2010). The results showed that PE outperformed the other indices on this level. In all entropy measures, SWE had the highest CV during anesthesia induction, indicating that this index was superior at discriminating between the awake and anesthetized states. Secondly, the performance for estimating the point of LOC was considered. Although all the entropy measures could distinguish between awake and anesthetized states (see **Figure 4**), the speed of transition (slope) between the two states was fastest

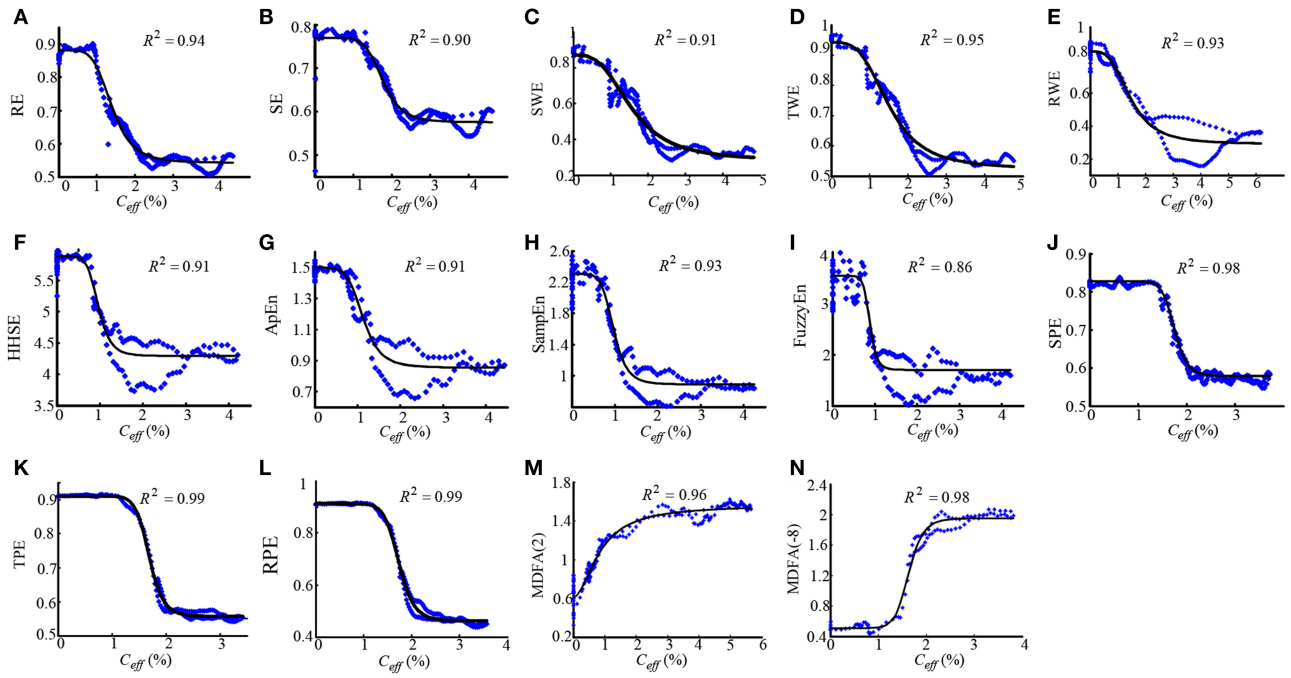


FIGURE 8 | Dose-response curves between the RE(A), SE(B), SWE(C), TWE(D), RWE(E), HHSE(F), ApEn(G), SampEn(H), FuzzyEn(I), SPE(J), TPE(K), RPE (L), MDFA(2) (M), MDFA(-8) (N) and the sevoflurane C_{eff} for the best fit, with the greatest value of R^2 show above the figures. The dots denote the measured EEG indices values. The solid lines denote the PK/PD modeled EEG index values.

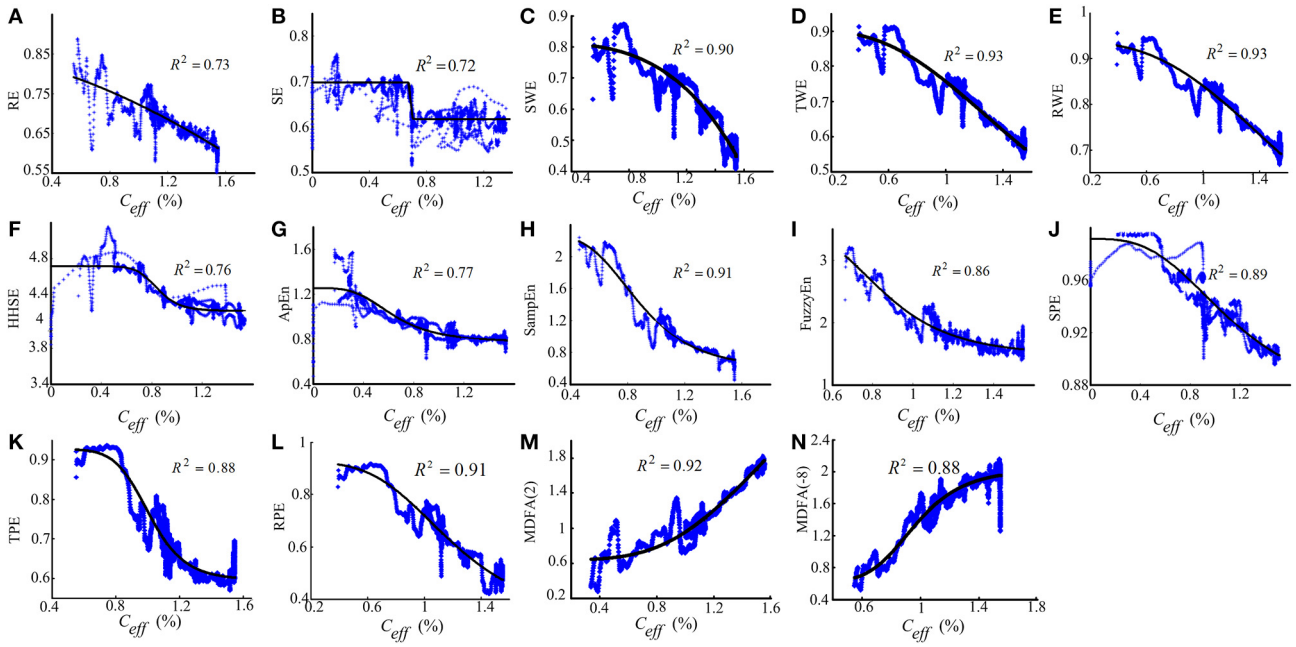


FIGURE 9 | The similar description as Figure 8 with the dose-response curves between entropy indices and isoflurane effect-site concentration.

Table 8 | The computing time for different entropy and MDFA indices for 1 min data length.

Entropy index	Calculation time(s)
RE/SE	0.096 ± 0.008
SWE/RWE/TWE	0.025 ± 0.001
HHSE	14.718 ± 1.563
ApEn	2.490 ± 0.098
SampEn	2.541 ± 0.073
FuzzyEn	4.785 ± 0.119
SPE/RPE/TPE	0.545 ± 0.016
MDFA	16.338 ± 0.280

for SWE, while SE had the slowest transition. Thirdly, the performance for discriminating different drug concentrations was considered, especially the ability to distinguish the burst suppression state. The mean ± SD value of the indices showed that all the entropy measures can distinguish different drug concentrations, while only ApEn and SampEn have the ability to distinguish burst suppression from the other states. This means that, if using PE as a DoA index, an additional method for detecting the burst suppression pattern would need to be incorporated, such as Nonlinear Energy Operator (NLEO) (Särkelä et al., 2002). The results are in accordance with the findings during desflurane anesthesia for ApEn (Bruhn et al., 2000) and sevoflurane anesthesia for PE and HHSE (Li et al., 2008b, 2010). Finally, the computing time was used to assess algorithm complexity. The results showed that the WE index is the fastest algorithm of all the entropy indices tested. HHSE was the slowest: its computing time for the same data length was about 580 times longer that for WE. In order to improve the computational efficiency, the parallelized method based on the graphics processing unit has been proposed (Chen et al., 2010).

The efficiency of these entropy measures were compared with other two non-linear dynamic measures, the MDFA with $q = 2$ and -8, where MDFA with $q = 2$ is a standard DFA measure. The results and statistics show that MDFA were better in some aspects compared to some of entropy measures, such as sharper slope in LOC, higher P_k and R^2 for sevoflurane (almost equal to RPE) measure. However, there are several shortcomings in MDFA measures. First, CVs of MDFA in awake state were higher compared to those of entropy indices. Second, MDFA could not distinguish the burst suppression state from other states. Most importantly, the computing time of MDFA is the longest in all algorithms, even longer than HHSE, which means that MDFA algorithms are not suitable for real time DoA monitoring. Therefore, entropy approaches are capable for monitoring the EEG changes in anesthesia, and are often advantageous in computation efficiency.

Although this study covers a number of entropy methods and two types of anesthesia, the research has its limitations. For instance, errors caused by individual variability, e.g., age, physical wellness, intraoperative tolerance are hard to control because of the difficulty in data collection in clinical practice. Besides, Interactions between EEG activities and drug concentrations could be studied using finer-grained paradigm, for instance by increasing the drug concentration in a stepwise pattern. Additionally, optimal parameters for each entropy measure may not have been achieved and need further investigation.

This study doesn't provide an absolute measure of "depth" of clinical anesthesia, nor of consciousness for the prevention of intra-operative recall; but rather focuses on understanding the inner workings of each entropy index, and explores whether these indices correlate with GABAergic drug effect. Having a good

understanding of the strengths and weaknesses of each measure is necessary before possibly applying them within a clinical context.

In conclusion, each entropy measure has its advantages, and several indices show promise as a simple open-source method for quantifying the brain effects of GABAergic drugs. In particular, the PE indices perform better than other entropy indices as an EEG derivative in several aspects, especially for RPE measure. However, further work is required to accurately quantify the burst suppression pattern. Also, to be useful as a clinical measure, each algorithm still needs additional parameter and computation efficiency optimizations.

ACKNOWLEDGMENTS

This research was supported by National Natural Science Foundation of China (No. 61304247, 61203210 and 61271142), China Postdoctoral Science Foundation (2014M551051) and Applied basic research project in Hebei province (No. 12966120D).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://www.frontiersin.org/journal/10.3389/fncom.2015.00016/abstract

REFERENCES

Abásolo, D., Hornero, R., Espino, P., Alvarez, D., and Poza, J. (2006). Entropy analysis of the EEG background activity in Alzheimer's disease patients. *Physiol. Meas.* 27, 241. doi: 10.1088/0967-3334/27/3/003

Alvarez-Ramirez, J., Echeverria, J. C., and Rodriguez, E. (2008). Performance of a high-dimensional R/S method for Hurst exponent estimation. *Phys. A Statist. Mech. Appl.* 387, 6452–6462. doi: 10.1016/j.physa.2008.08.014

Arefian, N. M., Zali, A. R., Seddighi, A. S., Fathi, M., Teymourian, H., Dabir, S., et al. (2009). Clinical analysis of eeg parameters in prediction of the depth of anesthesia in different stages: a comparative study. *Tanaffos* 8, 46–53.

Bandt, C. (2005). Ordinal time series analysis. *Ecol. Modell.* 182, 229–238. doi: 10.1016/j.ecolmodel.2004.04.003

Bandt, C., and Pompe, B. (2002). Permutation entropy: a natural complexity measure for time series. *Phys. Rev. Lett.* 88, 174102. doi: 10.1103/PhysRevLett.88.174102

Bein, B. (2006). Entropy. *Best Pract. Res. Clin. Anaesthesiol.* 20, 101–109. doi: 10.1016/j.bpa.2005.07.009

Bezerianos, A., Tong, S., and Thakor, N. (2003). Time-dependent entropy estimation of EEG rhythm changes following brain ischemia. *Ann. Biomed. Eng.* 31, 221–232. doi: 10.1114/1.1541013

Bruhn, J., Lehmann, L. E., Röpcke, H., Bouillon, T. W., and Hoeft, A. (2001). Shannon entropy applied to the measurement of the electroencephalographic effects of desflurane. *Anesthesiology* 95, 30–35. doi: 10.1097/00000542-200107000-00010

Bruhn, J., Myles, P., Sneyd, R., and Struys, M. (2006). Depth of anaesthesia monitoring: what's available, what's validated and what's next? *Br. J. Anaesth.* 97, 85–94. doi: 10.1093/bja/ael120

Bruhn, J., Röpcke, H., and Hoeft, A. (2000). Approximate entropy as an electroencephalographic measure of anesthetic drug effect during desflurane anesthesia. *Anesthesiology* 92, 715–726. doi: 10.1097/00000542-200003000-00016

Cao, Y., Tung, W., Gao, J., Protopopescu, V., and Hively, L. (2004). Detecting dynamical changes in time series using the permutation entropy. *Phys. Rev. Ser. E* 70, 46217–46217. doi: 10.1103/PhysRevE.70.046217

Chen, D., Li, D., Xiong, M., Bao, H., and Li, X. (2010). GPGPU-aided ensemble empirical-mode decomposition for EEG analysis during anesthesia. *Inform. Technol. Biomed. IEEE Trans.* 14, 1417–1427. doi: 10.1109/TITB.2010.2072963

Chen, W., Wang, Z., Xie, H., and Yu, W. (2007). Characterization of surface EMG signal based on fuzzy entropy. *Neural Syst. Rehabil. Eng. IEEE Trans.* 15, 266–272. doi: 10.1109/TNSRE.2007.897025

Chen, W., Zhuang, J., Yu, W., and Wang, Z. (2009). Measuring complexity using FuzzyEn, ApEn, and SampEn. *Med. Eng. Phys.* 31, 61–68. doi: 10.1016/j.medengphy.2008.04.005

Clausius, R. (1867). *The Mechanical Theory of Heat: with its Applications to the Steam-Engine and to the Physical Properties of Bodies.* London: John van Voorst.

Elbert, T., Ray, W. J., Kowalik, Z. J., Skinner, J. E., Graf, K. E., and Birbaumer, N. (1994). Chaos and physiology: deterministic chaos in excitable cell assemblies. *Physiol. Rev.* 74, 1–48.

Ellerkmann, R. K., Soehle, M., Riese, G., Zinserling, J., Wirz, S., Hoeft, A., et al. (2010). The Entropy Module and Bispectral Index as guidance for propofol-remifentanil anaesthesia in combination with regional anaesthesia compared with a standard clinical practice group. *Anaesth. Intensive Care* 38, 159–166.

Fell, J., Röschke, J., Mann, K., and Schäffner, C. (1996). Discrimination of sleep stages: a comparison between spectral and nonlinear EEG measures. *Electroencephalogr. Clin. Neurophysiol.* 98, 401–410. doi: 10.1016/0013-4694(96)95636-9

Ferenets, R., Lipping, T., Anier, A., Jantti, V., Melto, S., and Hovilehto, S. (2006). Comparison of entropy and complexity measures for the assessment of depth of sedation. *Biomed. Eng. IEEE Trans.* 53, 1067–1077. doi: 10.1109/TBME.2006.873543

Gifani, P., Rabiee, H., Hashemi, M., Taslimi, P., and Ghanbari, M. (2007). Optimal fractal-scaling analysis of human EEG dynamic for depth of anesthesia quantification. *J. Franklin Inst.* 344, 212–229. doi: 10.1016/j.jfranklin.2006.08.004

Hagihira, S., Takashina, M., Mori, T., Mashimo, T., and Yoshiya, I. (2002). Changes of electroencephalographic bicoherence during isoflurane anesthesia combined with epidural anesthesia. *Anesthesiology* 97, 1409–1415. doi: 10.1097/00000542-200212000-00012

Huang, L., Wang, W., and Singare, S. (2006). "Recurrence quantification analysis of EEG predicts responses to incision during anesthesia," in *Neural Information Processing*, eds I. King, J. Wang, L.-W. Chan, and D. Wang (Hong Kong: Springer), 58–65.

Huang, N. E., Shen, Z., Long, S. R., Wu, M. L. C., Shih, H. H., Zheng, Q. N., et al. (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R. Soc. Math. Phys. Eng. Sci.* 454, 903–995. doi: 10.1098/rspa.1998.0193

Inouye, T., Shinosaki, K., Sakamoto, H., Toi, S., Ukai, S., Iyama, A., et al. (1991). Quantification of EEG irregularity by use of the entropy of the power spectrum. *Electroencephalogr. Clin. Neurophysiol.* 79, 204–210. doi: 10.1016/0013-4694(91)90138-T

Inuso, G., La Foresta, F., Mammone, N., and Morabito, F. C. (2007). "Brain activity investigation by EEG processing: wavelet analysis, kurtosis and Renyi's entropy for artifact detection," in *Information Acquisition, 2007. ICIA'07. International Conference on: IEEE* (Jeju), 195–200.

Jameson, L. C., and Sloan, T. B. (2006). Using EEG to monitor anesthesia drug effects during surgery. *J. Clin. Monit. Comput.* 20, 445–472. doi: 10.1007/s10877-006-9044-x

Jospin, M., Caminal, P., Jensen, E. W., Litvan, H., Vallverdú, M., Struys, M. M., et al. (2007). Detrended fluctuation analysis of EEG as a measure of depth of anesthesia. *Biomed. Eng. IEEE Trans.* 54, 840–846. doi: 10.1109/TBME.2007.893453

Kantelhardt, J. W., Zschiegner, S. A., Koscielny-Bunde, E., Havlin, S., Bunde, A., and Stanley, H. E. (2002). Multifractal detrended fluctuation analysis of nonstationary time series. *Phys. A Statist. Mech. Appl.* 316, 87–114. doi: 10.1016/S0378-4371(02)01383-3

Klockars, J. G., Hiller, A., Munte, S., Van Gils, M. J., and Taivainen, T. (2012). Spectral entropy as a measure of hypnosis and hypnotic drug effect of total intravenous anesthesia in children during slow induction and maintenance. *Anesthesiology* 116, 340–351. doi: 10.1097/ALN.0b013e3182410b5e

Klonowski, W., Olejarczyk, E., Stepien, R., Jalowiecki, P., and Rudner, R. (2006). Monitoring the depth of anaesthesia using fractal complexity method. *Complex. Mundi. Emerg. Pattern. Nat.* 333–342. doi: 10.1142/9789812774217_0031

Li, D., Liang, Z., Wang, Y., Hagihira, S., Sleigh, J. W., and Li, X. (2012). Parameter selection in permutation entropy for an electroencephalographic measure of isoflurane anesthetic drug effect. *J. Clin. Monit. Comput.* 27, 113–123. doi: 10.1007/s10877-012-9419-0

Li, D., Li, X., Liang, Z., Voss, L. J., and Sleigh, J. W. (2010). Multiscale permutation entropy analysis of EEG recordings during sevoflurane anesthesia. *J. Neural Eng.* 7:046010. doi: 10.1088/1741-2560/7/4/046010

Li, X., Cui, S., and Voss, L. J. (2008a). Using permutation entropy to measure the electroencephalographic effects of sevoflurane. *Anesthesiology* 109, 448. doi: 10.1097/ALN.0b013e318182a91b

Li, X., Li, D., Liang, Z., Voss, L. J., and Sleigh, J. W. (2008b). Analysis of depth of anesthesia with Hilbert–Huang spectral entropy. *Clin. Neurophysiol.* 119, 2465–2475. doi: 10.1016/j.clinph.2008.08.006

Li, X., Ouyang, G., and Richards, D. A. (2007). Predictability analysis of absence seizures with permutation entropy. *Epilepsy Res.* 77, 70. doi: 10.1016/j.eplepsyres.2007.08.002

Liang, Z., Li, D., Ouyang, G., Wang, Y., Voss, L. J., Sleigh, J. W., et al. (2012). Multiscale rescaled range analysis of EEG recordings in sevoflurane anesthesia. *Clin. Neurophysiol.* 123, 681–688. doi: 10.1016/j.clinph.2011.08.027

Maszczyk, T., and Duch, W. (2008). "Comparison of Shannon, Renyi and Tsallis entropy used in decision trees," in *Artificial Intelligence and Soft Computing–ICAISC 2008*, eds L. Rutkowski, R. Tadeusiewicz, L. A. Zadeh, and J. M. Zurada (Zakopane: Springer), 643–651.

McKay, I. D. H., Voss, L. J., Sleigh, J. W., Barnard, J. P., and Johannsen, E. K. (2006). Pharmacokinetic-pharmacodynamic modeling the hypnotic effect of sevoflurane using the spectral entropy of the electroencephalogram. *Anesth. Analg.* 102, 91–97. doi: 10.1213/01.ane.0000184825.65124.24

Monk, T. G., Saini, V., Weldon, B. C., and Sigl, J. C. (2005). Anesthetic management and one-year mortality after noncardiac surgery. *Anesth. Analg.* 100, 4. doi: 10.1213/01.ANE.0000147519.82841.5E

Montirosso, R., Riccardi, B., Molteni, E., Borgatti, R., and Reni, G. (2010). Infant's emotional variability associated to interactive stressful situation: a novel analysis approach with Sample Entropy and Lempel–Ziv Complexity. *Infant Behav. Dev.* 33, 346–356. doi: 10.1016/j.infbeh.2010.04.007

Natarajan, K., Acharya, R., Alias, F., Tiboleng, T., and Puthusserypady, S. K. (2004). Nonlinear analysis of EEG signals at different mental states. *Biomed. Eng. Online* 3:7. doi: 10.1186/1475-925X-3-7

Nguyen-Ky, T., Wen, P., and Li, Y. (2010a). An improved detrended moving-average method for monitoring the depth of anesthesia. *Biomed. Eng. IEEE Trans.* 57, 2369–2378. doi: 10.1109/TBME.2010.2053929

Nguyen-Ky, T., Wen, P., and Li, Y. (2010b). Improving the accuracy of depth of anaesthesia using modified detrended fluctuation analysis method. *Biomed. Signal Process. Control* 5, 59–65. doi: 10.1016/j.bspc.2009.03.001

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. doi: 10.1016/0003-6870(94)90054-X

Olofsen, E., Sleigh, J., and Dahan, A. (2008). Permutation entropy of the electroencephalogram: a measure of anaesthetic drug effect. *Br. J. Anaesth.* 101, 810–821. doi: 10.1093/bja/aen290

Pincus, S. M. (1991). Approximate entropy as a measure of system complexity. *Proc. Natl. Acad. Sci. U.S.A.* 88, 2297–2301. doi: 10.1073/pnas.88.6.2297

Pritchard, W. S., Duke, D. W., and Krieble, K. K. (1995). Dimensional analysis of resting human EEG II: surrogate-data testing indicates nonlinearity but not low−dimensional chaos. *Psychophysiology* 32, 486–491. doi: 10.1111/j.1469-8986.1995.tb02100.x

Rampil, I. J. (1998). A primer for EEG signal processing in anesthesia. *Anesthesiology* 89, 980–1002. doi: 10.1097/00000542-199810000-00023

Renyi, A. (1970). *Probability Theory*. Amsterdam: North-Holland.

Rezek, I., and Roberts, S. J. (1998). Stochastic complexity measures for physiological signal analysis. *Biomed. Eng. IEEE Trans.* 45, 1186–1191. doi: 10.1109/10.709563

Richman, J. S., and Moorman, J. R. (2000). Physiological time-series analysis using approximate entropy and sample entropy. *Am. J. Physiol. Heart Circ. Physiol.* 278, H2039–H2049.

Rosso, O. A., Blanco, S., Yordanova, J., Kolev, V., Figliola, A., Schurmann, M., et al. (2001). Wavelet entropy: a new tool for analysis of short duration brain electrical signals. *J. Neurosci. Methods* 105, 65–76. doi: 10.1016/S0165-0270(00)00356-3

Rosso, O., Martin, M., Figliola, A., Keller, K., and Plastino, A. (2006). EEG analysis using wavelet-based information tools. *J. Neurosci. Methods* 153, 163–182. doi: 10.1016/j.jneumeth.2005.10.009

Rosso, O., Martin, M., and Plastino, A. (2003). Brain electrical activity analysis using wavelet-based informational tools (II): tsallis non-extensivity and complexity measures. *Phys. A Statist. Mech. Appl.* 320, 497–511. doi: 10.1016/S0378-4371(02)01529-7

Särkelä, M., Mustola, S., Seppänen, T., Koskinen, M., Lepola, P., Suominen, K., et al. (2002). Automatic analysis and monitoring of burst suppression in anesthesia. *J. Clin. Monit. Comput.* 17, 125–134. doi: 10.1023/A:1016393904439

Särkelä, M. O. K., Ermes, M. J., Van Gils, M. J., Yli-Hankala, A. M., Jäntti, V. H., and Vakkuri, A. P. (2007). Quantification of epileptiform electroencephalographic activity during sevoflurane mask induction. *Anesthesiology* 107, 928–938. doi: 10.1097/01.anes.0000291444.68894.ee

Shannon, C. E., and Weaver, W. (1949). *The Mathematical Theory of Communication.* Urbana: University of Illinois Press.

Sleigh, J., Voss, L., and Barnard, J. (2005). What are electroencephalogram entropies really measuring? *Int. Cong. Ser.* 1283, 231–234. doi: 10.1016/j.ics.2005.06.048

Sleigh, J. W., Olofsen, E., Dahan, A., de Goede, J., and Steyn-Ross, A. (2001). "Entropies of the EEG: the effects of general anaesthesia," in *Paper Presented at the 5th International Conference on Memory, Awareness and Consciousness* (New York, NY).

Smith, W. D., Dutton, R. C., and Smith, T. N. (1996). Measuring the performance of anesthetic depth indicators. *Anesthesiology* 84, 38–51. doi: 10.1097/00000542-199601000-00005

Tong, S., Bezerianos, A., Malhotra, A., Zhu, Y., and Thakor, N. (2003). Parameterized entropy analysis of EEG following hypoxic–ischemic brain injury. *Phys. Lett. A* 314, 354–361. doi: 10.1016/S0375-9601(03)00949-6

Tsallis, C., Mendes, R., and Plastino, A. R. (1998). The role of constraints within generalized nonextensive statistics. *Phys. A Statist. Mech. Appl.* 261, 534–554. doi: 10.1016/S0378-4371(98)00437-3

Viertiö-Oja, H., Maja, V., Särkelä, M., Talja, P., Tenkanen, N., Tolvanen-Laakso, H., et al. (2004). Description of the Entropy™ algorithm as applied in the Datex-Ohmeda S/5™ Entropy Module. *Acta Anaesthesiol. Scand.* 48, 154–161. doi: 10.1111/j.0001-5172.2004.00322.x

Yoo, C. S., Jung, D. C., Ahn, Y. M., Kim, Y. S., Kim, S. G., Yoon, H., et al. (2012). Automatic detection of seizure termination during electroconvulsive therapy using sample entropy of the electroencephalogram. *Psychiatry Res.* 195, 76–82. doi: 10.1016/j.psychres.2011.06.020

Zadeh, L. A. (1965). Fuzzy sets. *Inform. Control* 8, 338–353. doi: 10.1016/S0019-9958(65)90241-X

Zhang, X. S., Roy, R. J., and Jensen, E. W. (2001). EEG complexity as a measure of depth of anesthesia for patients. *Biomed. Eng. IEEE Trans.* 48, 1424–1433. doi: 10.1109/10.966601

Zunino, L., Pérez, D., Kowalski, A., Martín, M., Garavaglia, M., Plastino, A., et al. (2008). Fractional Brownian motion, fractional Gaussian noise, and Tsallis permutation entropy. *Phys. A Statist. Mech. Appl.* 387, 6057–6068. doi: 10.1016/j.physa.2008.07.004

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.

Received: 15 August 2014; accepted: 28 January 2015; published online: 18 February 2015.

Citation: Liang Z, Wang Y, Sun X, Li D, Voss LJ, Sleigh JW, Hagihira S and Li X (2015) EEG entropy measures in anesthesia. Front. Comput. Neurosci. ***9****:16. doi: 10.3389/fncom.2015.00016*

This article was submitted to the journal Frontiers in Computational Neuroscience.

frontiers in Computational Neuroscience

ORIGINAL RESEARCH
published: 02 June 2015
doi: 10.3389/fncom.2015.00066

Detection of subjects and brain regions related to Alzheimer's disease using 3D MRI scans based on eigenbrain and machine learning

Yudong Zhang[1]*, *Zhengchao Dong*[2], *Preetha Phillips*[3], *Shuihua Wang*[1,4], *Genlin Ji*[1,5], *Jiquan Yang*[5] and *Ti-Fei Yuan*[6]*

[1] School of Computer Science and Technology, Nanjing Normal University, Nanjing, China, [2] Division of Translational Imaging and MRI Unit, New York State Psychiatric Institute, Columbia University, New York, NY, USA, [3] School of Natural Sciences and Mathematics, Shepherd University, Shepherdstown, WV, USA, [4] School of Electronic Science and Engineering, Nanjing University, Nanjing, China, [5] Jiangsu Key Laboratory of 3D Printing Equipment and Manufacturing, Nanjing, China, [6] School of Psychology, Nanjing Normal University, Nanjing, China

OPEN ACCESS

Edited by:
Tobias Alecio Mattei, Brain and Spine Center - InvisionHealth - Kenmore Mercy Hospital, USA

Reviewed by:
Fahad Sultan, University Tübingen, Germany
Petia D. Koprinkova-Hristova, Bulgarian Academy of Sciences, Bulgaria

***Correspondence:**
Yudong Zhang, School of Computer Science and Technology, Nanjing Normal University, 1 Wenyuan, Nanjing, Jiangsu 210023, China
zhangyudong@njnu.edu.cn;
Ti-Fei Yuan, School of Psychology, Nanjing Normal University, 22 Ninghai Rd., Nanjing, Jiangsu 210008, China
ytf0707@126.com

Received: 12 February 2015
Accepted: 17 May 2015
Published: 02 June 2015

Citation:
Zhang Y, Dong Z, Phillips P, Wang S, Ji G, Yang J and Yuan T-F (2015) Detection of subjects and brain regions related to Alzheimer's disease using 3D MRI scans based on eigenbrain and machine learning. Front. Comput. Neurosci. 9:66. doi: 10.3389/fncom.2015.00066

Purpose: Early diagnosis or detection of Alzheimer's disease (AD) from the normal elder control (NC) is very important. However, the computer-aided diagnosis (CAD) was not widely used, and the classification performance did not reach the standard of practical use. We proposed a novel CAD system for MR brain images based on eigenbrains and machine learning with two goals: accurate detection of both AD subjects and AD-related brain regions.

Method: First, we used maximum inter-class variance (ICV) to select key slices from 3D volumetric data. Second, we generated an eigenbrain set for each subject. Third, the most important eigenbrain (MIE) was obtained by Welch's *t*-test (WTT). Finally, kernel support-vector-machines with different kernels that were trained by particle swarm optimization, were used to make an accurate prediction of AD subjects. Coefficients of MIE with values higher than 0.98 quantile were highlighted to obtain the discriminant regions that distinguish AD from NC.

Results: The experiments showed that the proposed method can predict AD subjects with a competitive performance with existing methods, especially the accuracy of the polynomial kernel (92.36 ± 0.94) was better than the linear kernel of 91.47 ± 1.02 and the radial basis function (RBF) kernel of 86.71 ± 1.93. The proposed eigenbrain-based CAD system detected 30 AD-related brain regions (Anterior Cingulate, Caudate Nucleus, Cerebellum, Cingulate Gyrus, Claustrum, Inferior Frontal Gyrus, Inferior Parietal Lobule, Insula, Lateral Ventricle, Lentiform Nucleus, Lingual Gyrus, Medial Frontal Gyrus, Middle Frontal Gyrus, Middle Occipital Gyrus, Middle Temporal Gyrus, Paracentral Lobule, Parahippocampal Gyrus, Postcentral Gyrus, Posterial Cingulate, Precentral Gyrus, Precuneus, Subcallosal Gyrus, Sub-Gyral, Superior Frontal Gyrus, Superior Parietal Lobule, Superior Temporal Gyrus, Supramarginal Gyrus, Thalamus, Transverse Temporal Gyrus, and Uncus). The results were coherent with existing literatures.

Conclusion: The eigenbrain method was effective in AD subject prediction and discriminant brain-region detection in MRI scanning.

Keywords: Alzheimer's disease, Welch's *t*-test, magnetic resonance imaging, machine learning, machine vision, eigenbrain, support vector machine, particle swarm optimization

Introduction

Alzheimer's disease (AD) is not a normal part of aging. It is a type of dementia that causes problems with memory, thinking, and behavior. Symptoms usually develop slowly and worsen over time. Symptoms may become severe enough to interfere with daily life, and lead to death (Hahn et al., 2013). There is no cure for this disease. In 2006, 26.6 million people worldwide suffered from this disease. AD is predicted to affect 1 in 85 people globally by 2050, and at least 43% of prevalent cases need high level of care (Brookmeyer et al., 2007). As the world is evolving into an aging society, the burdens and impacts caused by AD on families and the society has also increased significantly. In the US, healthcare on people with AD currently costs roughly $100 billion per year and is predicted to cost $1 trillion per year by 2050 (Miller et al., 2012).

Early and accurate detection of AD is beneficial for the management of the disease (Han et al., 2011). Presently, a multitude of neurologists and medical researchers have been dedicating considerable time and energy toward this goal, and promising results have been continually springing up (Xinyun et al., 2011). Magnetic resonance imaging (MRI) is an imaging technique that produces high quality images of the anatomical structures of the human body, especially in the brain, and provides rich information for clinical diagnosis and biomedical research (Shamonin et al., 2014). The diagnostic values of MRI are greatly enhanced by the automated and accurate classification of the MR images (Goh et al., 2014; Zhang et al., 2015a,b). It already plays an important role in detecting AD subjects from normal elder controls (NC) (Angelini et al., 2012; Smal et al., 2012; Nambakhsh et al., 2013; Hamy et al., 2014; Jeurissen et al., 2014).

In earlier cases, most diagnosis work was done to measure manually or semi-manually a priori region of interest (ROI) of magnetic resonance (MR) images, based on the fact that AD patients suffer more cerebral atrophy compared to NCs (Kubota et al., 2006; Anagnostopoulos et al., 2013). Most of these ROI-based analyses focused on the shrinkage of hippocampus and cortex, and enlarged ventricles (Pennanen et al., 2004). Somehow, the ROI-based methods suffer from some limitations. First, the methods focus on the ROIs need prior knowledge. Second, the accuracy of early detection depends heavily on the experiences of the examiners. Third, the mutual information among the voxels is difficult to operate (Xinyun et al., 2011; Lee et al., 2013). Finally, there is no evidence that other regions (except hippocampus and entorhinal cortex) did not provide any information related to AD. Also, the auto-segmentation of ROI is not feasible in practice, and examiners tend to segment the brain manually.

On the other hand, multivariate approaches that consider all the voxels in a scan as one observation offer an alternative method to ROI-based methods. The advantages of multivariate approaches are that they are data driven, which means that the analyses are fully based on the data without any prior knowledge and that the interactions among voxels and error effects are assessed statistically. However, multivariate approaches suffer from either the curse of dimension problem or the small sample size problems or the lack of the capability, to make statistical inferences about regionally specific changes (Álvarez et al., 2009b).

The Eigenbrain was an excellent multivariate approach that solves both the curse of dimensionality and the problems in small sample size. It was proposed by Alvarez et al. (2009a) and Lopez et al. (2009), and was applied on Single Photon Emission Computed Tomography (SPECT) images. In their research, the eigenbrain approach was shown to efficiently reduce the feature space from $\sim 5 \times 10^5$ to only $\sim 10^2$, and therefore, was able to achieve excellent classification accuracy. In this study, we make a tentative test of applying eigenbrains in MRI scans for AD detection.

Support vector machine (SVM) has been arguably regarded as one of the most excellent classification methods in machine learning (Zhang and Wu, 2012a). Original SVMs are linear classifiers, and do not perform well on nonlinear data. Hence, we introduced in the kernel SVMs (KSVMs), which extends original linear SVMs to nonlinear SVM classifiers by applying the kernel function to replace the dot product form in the original SVMs (Gomes et al., 2012). Compared with the original plain SVM, the KSVMs allows one to fit the maximum-margin hyperplane in a transformed feature space (Garcia et al., 2010). The transformation may be nonlinear and the transformed space is high dimensional; thus although the classifier is a hyperplane in the high-dimensional feature space, it may be nonlinear in the original input space (Hable, 2012).

The aim of our study was to develop a novel classification system based on eigenbrain and machine learning, in order to grow a computer-aided diagnosis (CAD) system for the early detection of AD subjects and AD-related brain regions. Our goal was not to replace clinicians, but to provide an assisting tool. The rest of the paper was organized as follows: the next section reviewed relates literatures from two aspects: the extracted features and the classification methods. Section The Proposed Method describes the methodology of the proposed CAD. Section Experiments and Results contains the experiments and results. Section Discussion analyzes the reason behind the experiment results. Finally, Section Conclusion and Future Research is devoted to conclusion and future research. For ease in reading, the acronyms and

their meanings of this study are listed in Table 12 in the appendix.

The **contributions** of the paper fell within the following five aspects: (i) We generalized the Eigenbrain to MR images, and proved its effectiveness; (ii) We proposed a hybrid eigenbrain-based CAD system that can not only detect AD from NC, but also detect brain regions that related to AD. (iii) We proved the proposed method had classification accuracy comparable to state-of-the-art methods, and the detected brain regions were in line with 16 existing literatures. (iv) We used inter-class variance (ICV) and Welch's *t*-test (WTT) to reduce redundant data; (v) We found POL kernel is better than linear and RBF kernel for this study.

Literature Review

In common convention, the automatic classification consisted of two stages: feature extraction and classifier construction. We reviewed over ten literatures, and analyzed themthrough the two stages.

Features of MR Images

Scholars have proposed numerous methods to extract various features[1]. Chaplot et al. (2006) used the approximation coefficients obtained by discrete wavelet transform (DWT). Maitra and Chatterjee (2006) employed the Slantlet transform, which is an improved version of DWT. Their feature vector of each image was created by considering the magnitudes of Slantlet transform outputs corresponding to six spatial positions that were chosen according to a specific logic. El-Dahshan et al. (2010) extracted the approximation and detail coefficients of 3-level DWT. Plant et al. (2010) used brain region cluster (BRC). They suggested to use information gain (IG) to rate the interestingness of a voxel, and applied clustering algorithm to identify groups of adjacent voxels with a high discriminatory power. Zhang et al. (2011) exclusively used the approximation coefficients of 3-level decomposition, and used PCA to reduce the features. Ramasamy and Anandhakumar (2011) used fast Fourier transform (FFT) as features. Saritha et al. (2013) proposed a novel feature of wavelet-entropy, and employed spider-web plots to further reduce features. Zhang et al. (2013) employed digital wavelet transform to extract features then used principal component analysis (PCA) to reduce the feature space. Savio and Grana (2013) proposed to use deformation-based morphometry (DBM) techniques, and proposed five features as Jacobian map, modulated GM (MGM), trace of Jacobian matrix (TJM), magnitude of the displacement field, and Geodesic Anisotropy (GEODAN). In addition, they suggested the use of Pearson's correlation (PEC), Bhattacharyya distance (BD), and WTT to measure the significance of voxel site. Das et al. (2013) suggested to use Ripplet transform, followed by PCA to reduce features. Kalbkhani et al. (2013) modeled the detail coefficients of 2-level DWT by generalizing autoregressive conditional heteroscedasticity (GARCH) statistical model, and the parameters of GARCH model were considered as the primary feature vector. Zhang et al. (2014) used an undersampling (US) technique on the volumetric image, followed by singular value decomposition (SVD) to select features. El-Dahshan et al. (2014) proposed to add a preprocessing technique that used pulse-coupled neural network (PCNN) for image segmentation. Zhou et al. (2015) used wavelet-entropy as the feature space. Zhang et al. (2015a) used discrete wavelet packet transform (DWPT), and harnessed Tsallis entropy to obtain features from DWPT coefficients. Yang et al. (2015) selected wavelet-energy as the features.

From the literature used, the DWT based features were proven to be efficient. In this study, we suggested using a novel feature of eigenbrain, which was used for SPECT images but was never been used in MR images.

Classification Model in MRI

There are numerous classification models, but only a few of them are suitable for MR images. Chaplot et al. (2006) employed the self-organizing map (SOM) neural network and SVM. Maitra and Chatterjee (2006) used the common artificial neural network (ANN). El-Dahshan et al. (2010) used ANN and K-nearest neighbor (KNN) classifiers. Plant et al. (2010) used SVM, Bayes statistics, and voting feature intervals (VFI) to derive the quantitative index of pattern matching. Zhang et al. (2011) suggested to use ANN. The weights of ANN were trained by scaled-conjugate-gradient method. Ramasamy and Anandhakumar (2011) proposed to use Expectation and Maximization Gaussian Mixture Model algorithm (EM-GMM). Saritha et al. (2013) used the probabilistic neural network (PNN). Zhang et al. (2013) constructed a kernel SVM with RBF kernel, using particle swarm optimization (PSO) to optimize the parameters *C* and sigma. Savio and Grana (2013) chose SVM, and used grid search for tuning parameters. Das et al. (2013) used least-square SVM, and their 5×5 CV showed high classification accuracy. Kalbkhani et al. (2013) tested the KNN and SVM models. Zhang et al. (2014) proposed to combine KSVM and decision tree, and their method was dubbed KSVM-DT. El-Dahshan et al. (2014) used feed forward back-propagation neural network (FFBPNN). Zhou et al. (2015) used a Naive Bayes classifier (NBC) as classification method. Zhang et al. (2015a) used a generalized eigenvalue proximal SVM (GEPSVM) with RBF kernel. Yang et al. (2015) used SVM as the classifier, and employed biogeography-based optimization (BBO) to train the classifier.

After reviewing the latest literatures that were related to classifiers, we found that SVMs had significant advantages of high accuracy, elegant mathematical tractability, and direct geometric interpretation, compared with other classification methods (Collins and Pape, 2011). In addition, it did not need a large number of training samples to avoid overfitting (Li et al., 2010). Kernel technique further enhanced the performance of SVM. Therefore, KSVM was harnessed in this study.

The Proposed Method

Preprocessing on Volumetric Data

For each individual, all available 3 or 4 volumetric 3D MR brain images were motion-corrected, and coregistered to form

[1]Some abbreviations are modified to avoid conflict within this paper.

an averaged 3D image. Then, those 3D images were spatially normalized to the Talairach coordinate space and brain-masked. CDR was interpreted as the target (label). It is a numeric scale quantifying the severity of symptoms of dementia (Williams et al., 2013). The patient's cognitive and functional performances were assessed in six areas: memory, orientation, judgment and problem solving, community affairs, home and hobbies, and personal care. In this study, we chose two types of CDR, i.e., the subjects with CDR of 0 were considered as NC and subjects with CDR of 1 were considered as AD (Marcus et al., 2007).

Calculating eigenbrains on the entire brain was difficult. Instead, we proposed a simplified method that selected several key slices that capture structures indicative of AD from NC. The procedure was as follows: we established the ICV v as

$$v(k) = \left\| \mu_{\mathrm{AD}}(Slice = k) - \mu_{\mathrm{NC}}(Slice = k) \right\|^2 \quad (1)$$

where k was the index of key slice, μ_{AD} and μ_{NC} represented the mean of gray-level values of the kth slice of AD subjects and NC subjects, respectively, $||.||^2$ represented the l_2-norm. Then, we selected the key-slices of ICV larger than 50% of maximum ICV, with 10× undersampling factor (i.e., every 10 slices).

In addition, the slice direction can be chosen as either axial, sagittal, or coronal. Usually coronal direction will give a clearer view than the other two directions. **Figure 1** showed that the coronal slice had an advantage over other directions in that it can cover three of the most important tissues within one slice. Those tissues were seen as indicative of AD. These tissues are the cerebral cortex, the ventricle, and the hippocampus. If we used axial or sagittal slice, then we may need to record two or even more slices to cover those tissues. Therefore, we chose the coronal direction for key slice selection, with the aim of recording only one slice.

Eigenbrain

AD has different physical structures from NC. Revisit **Figure 1** which indicated the AD subjects had severe atrophy of the cerebral cortex (region i), severely enlarged ventricles (region ii), and extreme shrinkage of hippocampus (region iii). Therefore, eigenbrain tried to capture those different characteristic changes of anatomical structures between AD and NC.

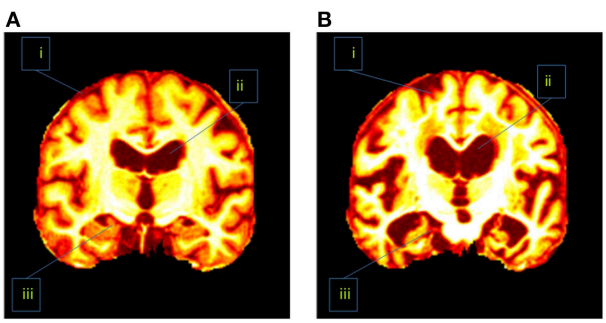


FIGURE 1 | Difference between (A) a healthy brain and (B) an AD brain. The labeled three regions are (i) cerebral cortex (ii) ventricle, and (iii) hippocampus.

Eigenbrain is carried out by PCA, which is a statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components (PC). For 2D images the PCs are extended naturally to the 2D eigenbrains.

Suppose $\boldsymbol{X}$ is a given data matrix with size of $N \times A$, where N represents the number of samples and A number of attributes (For a 256 × 256 image, we need to vectorize it to a 1 × 65536 vector, hence $A = 65536$). First, we normalized the dataset matrix $\boldsymbol{X}$, so that each sample in the normalized matrix $\boldsymbol{Z}$ was mean-centered and unit-variance scaled, by subtracting its mean value and dividing the difference by its standard deviation.

$$Z \leftarrow \frac{X - \mu(X)}{\sigma(X)} \quad (2)$$

Next, we estimated the covariance matrix $\boldsymbol{C}$ with size of $A \times A$ by

$$C \leftarrow \frac{1}{N-1} Z^T Z \quad (3)$$

Here we used $N - 1$ instead of N in order to produce an unbiased estimator of the variance (See Bessel's correction (Russell and Cohn, 2012) for details).

Third, we perform the eigendecomposition of $\boldsymbol{C}$:

$$C = U \wedge U^{-1} \quad (4)$$

where $\boldsymbol{U}$ is an $A \times (N-1)$ matrix, whose columns are the eigenvectors of covariance matrix $\boldsymbol{C}$, matrix $\boldsymbol{\Lambda}$ is an $(N-1) \times (N-1)$ diagonal matrix whose diagonal elements are eigenvalues of $\boldsymbol{C}$, each corresponding to an eigenvector of A. It is common to sort the eigenvalue matrix $\boldsymbol{\Lambda}$ and eigenvector matrix $\boldsymbol{U}$ in order of decreasing eigenvalue $u_1 > u_2 > \ldots > u_N$. To view the ith eigenbrain $u(i)$, the ith column of $\boldsymbol{U}$ was reshaped to an image. Suppose the ith column of $\boldsymbol{U}$ contains 65536 elements, then the reshaped image was 256 × 256.

$$u(i) = \mathrm{reshape}(U(:, i)) \quad (5)$$

Note that in our situation ($N \sim 10^2$ and $A \sim 10^4$, where $\sim$ denotes the same order of magnitude), the computation burdens of eigendecomposition of equation (4) are enormous. It can be accelerated by replacing $\boldsymbol{C}$ in equation (3) with $\boldsymbol{C}'$, since $N << A$.

$$C' \leftarrow \frac{1}{N-1} Z Z^T \quad (6)$$

The size of $\boldsymbol{C}'$ is $N \times N$, which can significantly reduce the computation burden. Using Matlab, the eigenbrain can be done by a simple "PCA" command without considering these issues. The flowchart of calculating eigenbrain is shown in **Figure 2**.

The eigenvalues represent the distribution of energy of the source data among each of the eigenbrains, where the eigenbrains form a basis for original data.

To further select an eigenbrain that is the most statistically significant, we employ the two-sample location test. Saritha et al.

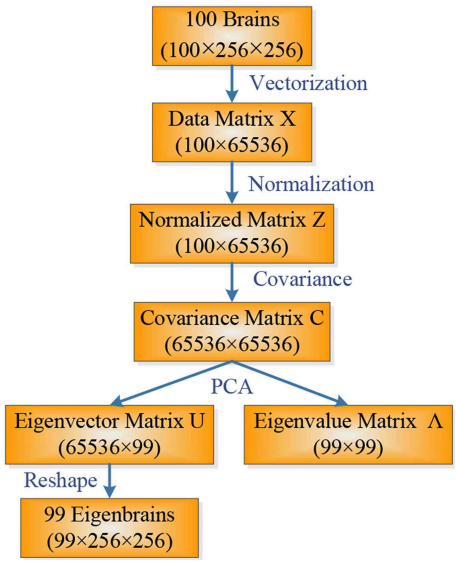


FIGURE 2 | Flowchart of calculating eigenbrain.

(2013) selected the Student's *t*-test which assumes both the means and variances of the two data are equal. The assumption of equal variances was not necessary and can be dropped; while the assumption of equal means is essential to select significantly important eigenbrains. Therefore, we used WTT that is an adaption of the Student's *t*-test and checks nothing except the two populations that have equal means.

The null hypothesis is that the eigenvalues of AD and NC have equal means, without assuming they have equal variances. The alternative hypothesis is they have unequal means. WTT was carried out at the 95% confidence interval. The eigenvalues of the selected most important eigenbrain (MIE) were used as input features for following classification.

Region Detection

We proposed a visual interpretation method of Eigenbrain to detect regions that can distinguish AD and NC, which is not reported in literatures of Alvarez et al. (2009a) and Lopez et al. (2009). The interpretation in a four-stage process is listed in **Table 1**.

Classifier

SVM was used as the classifier. In addition, sequential minimal optimization (SMO) is chosen to train SVM due to simple and fast speed (Zhang and Wu, 2012b). Traditional linear SVMs cannot separate intricately distributed data. In order to generalize SVMs to create nonlinear hyperplane, the kernel trick is applied. The KSVMs allows us to fit the maximum-margin hyperplane in a transformed feature space (Liu et al., 2014). The transformation may be nonlinear and the transformed space is a higher dimensional space. Though the classifier is a hyperplane in the higher-dimensional feature space, it may be nonlinear in the original input space.

TABLE 1 | Four-stage region detection method.

Region detection
Step 1 We selected the most important eigenbrain (MIE).
Step 2 We performed an absolution operation on MIE, since there are both positive and negative elements in the MIE matrix.
Step 3 We highlighted those voxels with the values higher than 0.98 quantile, i.e., 98th percentile.
Step 4 We outputted the anatomical label information of selected voxels using Talairach Daemon software, the output of which contained five levels: hemisphere, lobe, gyrus, tissue, and cell.

TABLE 2 | Assessment of classification performance.

Measure	Definition
Accuracy	$(TP + TN) / (TP + TN + FP + FN)$
Sensitivity (Recall)	$TP / (TP + FN)$
Specificity	$TN / (TN + FP)$
Precision	$TP / (TP + FP)$

TABLE 3 | Pseudocode of proposed method.

Step 1 Input 3D MRI data and corresponding CDR labels.

Step 2 Select key slices by ICV larger than 50% of maximum, with 10× undersampling factor.

Step 3 Generate eigenbrain set for each key slice.

Step 4 Select the MIE by WTT with 95% confidence interval.

Step 5 (**Output 1**): Submit eigenvalues of MIE to the classifier, and report its performance based on 50 × 10 CV.

Step 6 (**Output 2**): Report the discriminant regions by the absolute coefficient values higher than 0.98 quantile.

The radial basis function (RBF) kernel is one of the most widely used kernels with the form as Zhang and Wu (2012b).

$$\kappa(x_m, x_n) = \exp\left(-\frac{\|x_m - x_n\|}{2\sigma^2}\right) \tag{7}$$

where κ is the kernel function, σ the scaling factor, and x_m and x_n are vectors in the input space.

Another commonly used kernel is polynomial (POL) kernel defined as

$$\kappa(x_m, x_n) = \left(x_m^T x_n + c\right)^d \tag{8}$$

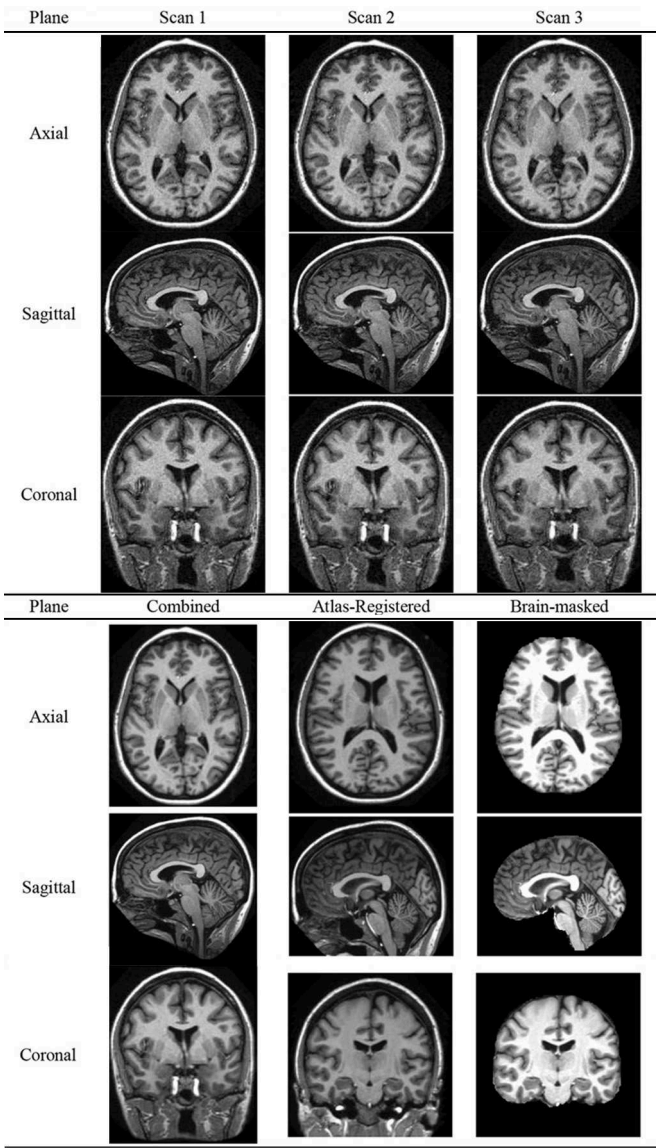
where d is the degree of polynomial, and c a soft margin constant trading off the influence of higher-order vs. lower-order terms in the polynomial.

Based on the two kernels, we tested RBF-KSVM and POL-KSVM for our models. To obtain the best parameter of kernels (the scaling factor σ of RBF, or the degree d and soft margin constant c of POL), PSO was employed since it has been used successfully to tune parameters of KSVM in various problems (Aich and Banerjee, 2014; Khazaee and Zadeh, 2014; Xue et al., 2014).

TABLE 4 | Subject demographics status.

	NC	AD
Number of subjects	98	28
Male/Female	26/72	9/19
Age	75.91 ± 8.98	77.75 ± 6.99
Education	3.26 ± 1.31	2.57 ± 1.31
SES	2.51 ± 1.09	2.87 ± 1.29
CDR	0	1
MMSE	28.95 ± 1.20	21.67 ± 3.75

TABLE 5 | Preprocessing of a specified subject.



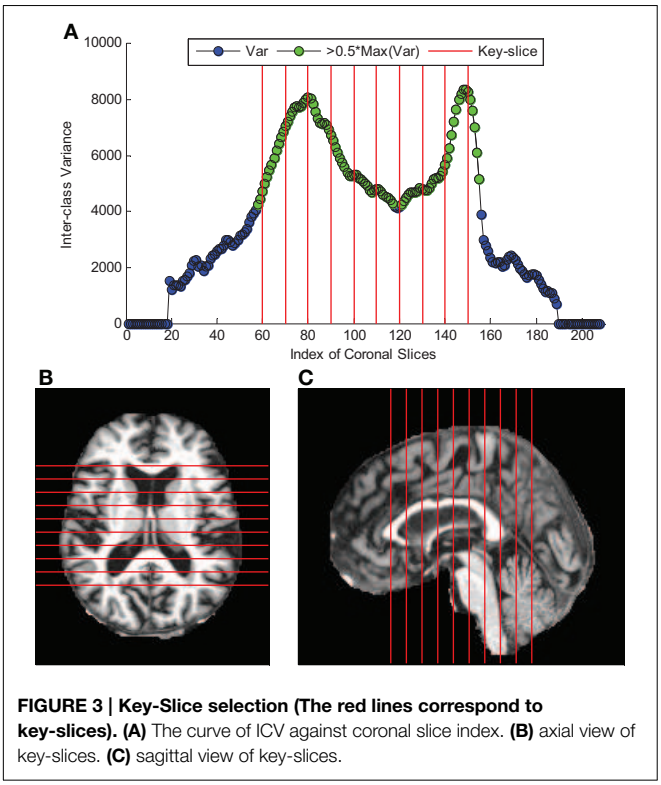


FIGURE 3 | Key-Slice selection (The red lines correspond to key-slices). (A) The curve of ICV against coronal slice index. **(B)** axial view of key-slices. **(C)** sagittal view of key-slices.

K-fold CV was employed, and K was assigned with a value of 10 considering the best compromise between computational cost and reliable estimates, i.e., the dataset is randomly divided into 10 mutually exclusively subsets of approximately equal size, in which $10 - 1 = 9$ subsets were used as training set and the last subset was used as the validation set. The procedure that was mentioned above was repeated 10 times, so each subset was used once for validation. The K results from the K folds were combined together, to yield a single estimation of the whole dataset.

The *K*-fold CV repeated 50 times, i.e., we carried out a 50 × 10-fold CV. For each time, we used four measures: accuracy, sensitivity, specificity, and precision (**Table 2**), to assess the performance. Here TP, FP, TN, and FN represented the instance number of true positive, false positive, true negative, and false negative, respectively. We considered a correctly identified AD case as a true positive, following the common convention. Summarizing the 50 repetitions, we reported the final measures of both the mean and standard deviation (SD) of the four measures.

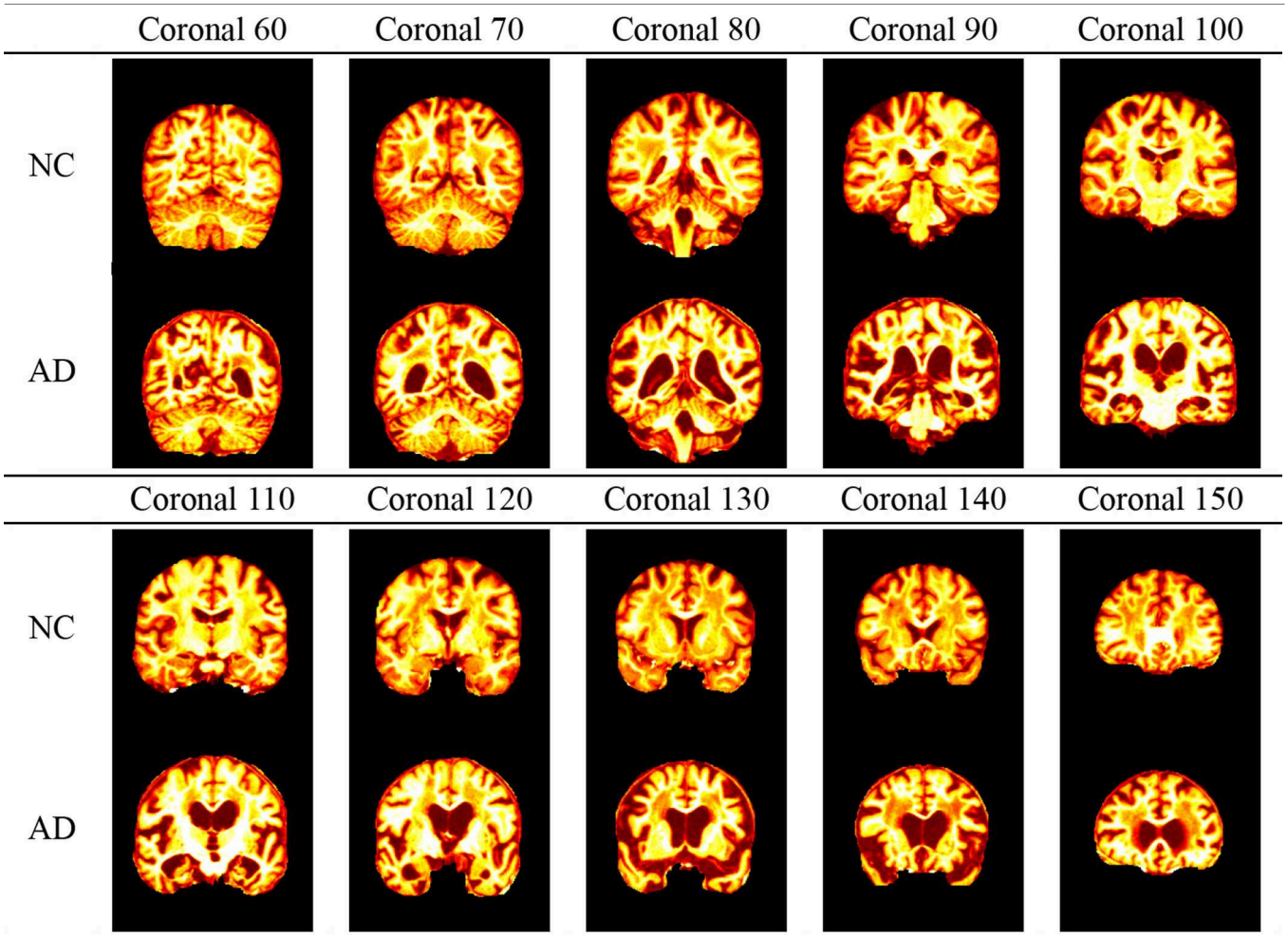
Implementation

The purpose of the proposed method is two-fold: (i) to find discriminant voxels that distinguish AD from NC; and (ii) to develop a CAD system and report its performance. The pseudocode is listed in **Table 3**.

Experiments and Results

The programs were in-house developed using Matlab 2014a, and ran on IBM laptop with 3 GHz Intel i3 dual-processor and 8 GB RAM. Readers could repeat our results on any machine where Matlab is available.

TABLE 6 | Difference between NC and AD on key-slices.



Data Source

We downloaded the dataset from Open Access Series of Imaging Studies (OASIS) (Ardekani et al., 2013, 2014). We chose the cross-sectional dataset corresponding to MRI scans of individuals at a single time point (Bin Tufail et al., 2012). The OASIS dataset consists of 416 subjects aged 18–96, who are all right-handed. We excluded subjects under 60 years old and those with missing records and then picked 126 subjects (98 NCs and 28 ADs) from the rest of the subjects. The demographic statuses of the included subjects were summarized in **Table 4**. Here SES, CDR, and MMSE represent socioeconomic status, clinical dementia rating, and mini-mental state examination, respectively.

Preprocessing

Table 5 shows an example of the combination of 3 individual scans of a subject. The resolution is $1 \times 1 \times 1.25$ mm. The preprocessing performed motion-correction on the 3D MR images, registered them to form a combined image in the native acquisition space, and resampled to $1 \times 1 \times 1$ mm. Afterwards, the combined image was spatially normalized to the Talairach coordinate space, and brain-extracted (**Table 5**).

Key-slice Selection by ICV

The curve of ICV against slice index was shown in **Figure 3A**. We selected 10 coronal slices (60, 70, 80, 90, 100, 110, 120, 130, 140, and 150). Their corresponding ICVs were all higher than 50% of the maximum. **Figures 3B,C** showed the axial and sagittal view of the 10 key-slices. **Table 6** showed the comparison between NC and AD in the selected 10 key-slices.

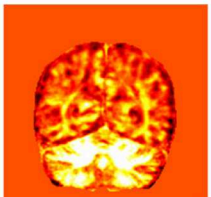
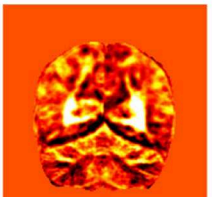
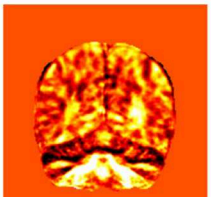
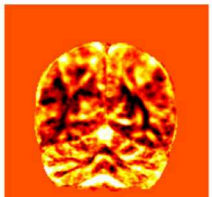
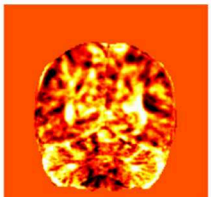
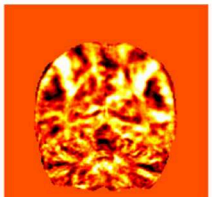
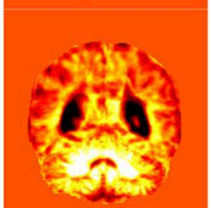
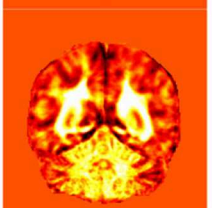
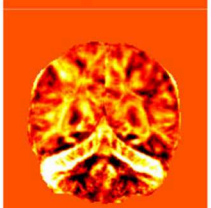
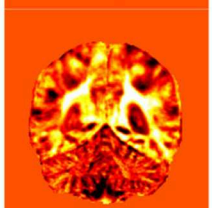
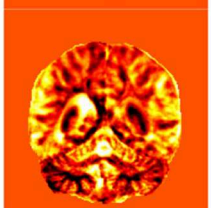
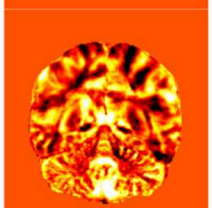
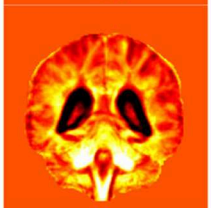
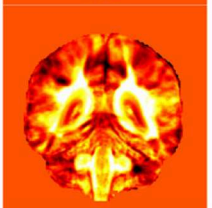
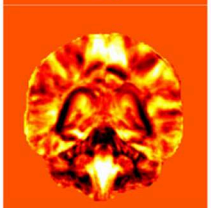
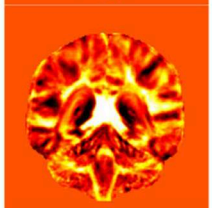
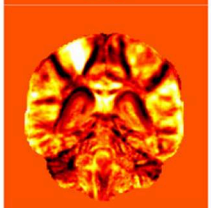
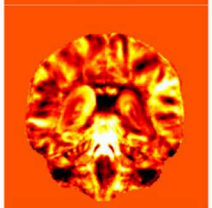
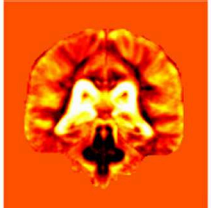
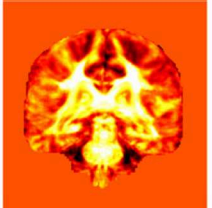
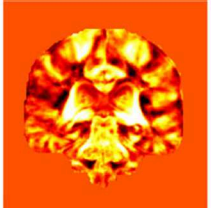
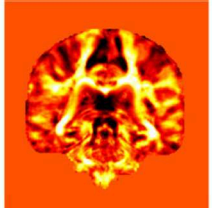
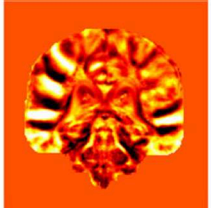
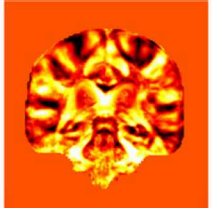
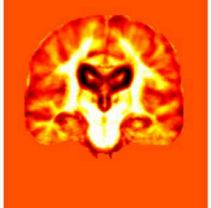
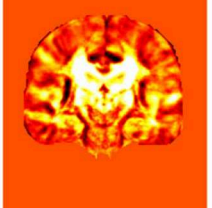
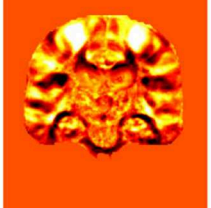
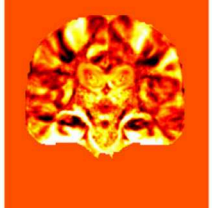
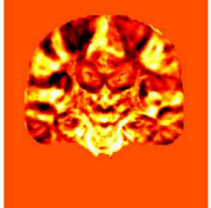
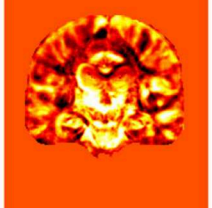
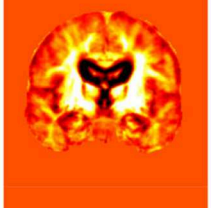
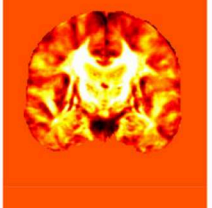
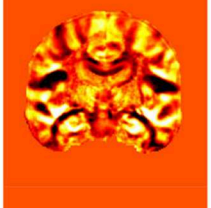
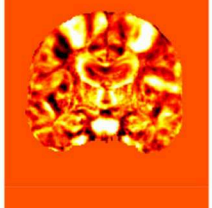
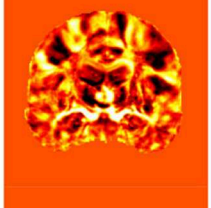
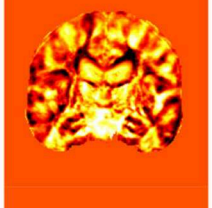
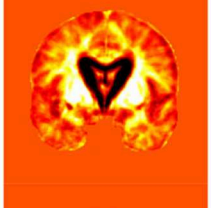
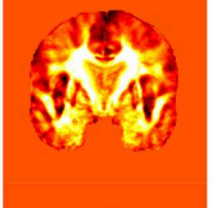
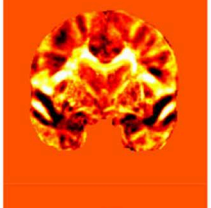
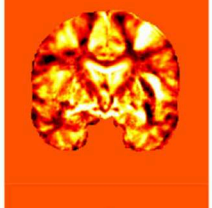
Eigenbrains

Table 7 showed the eigenbrain results obtained by running PCA on the slices of all subjects. For each slice, we had a set of 125 eigenbrains in total. Due to the page limit, we selected and listed the first 6 eigenbrains. The eigenbrains were sorted in the order of decreasing eigenvalues. In general, the eigenbrains in the previous columns were more important than in latter columns.

Most Important Eigenbrain

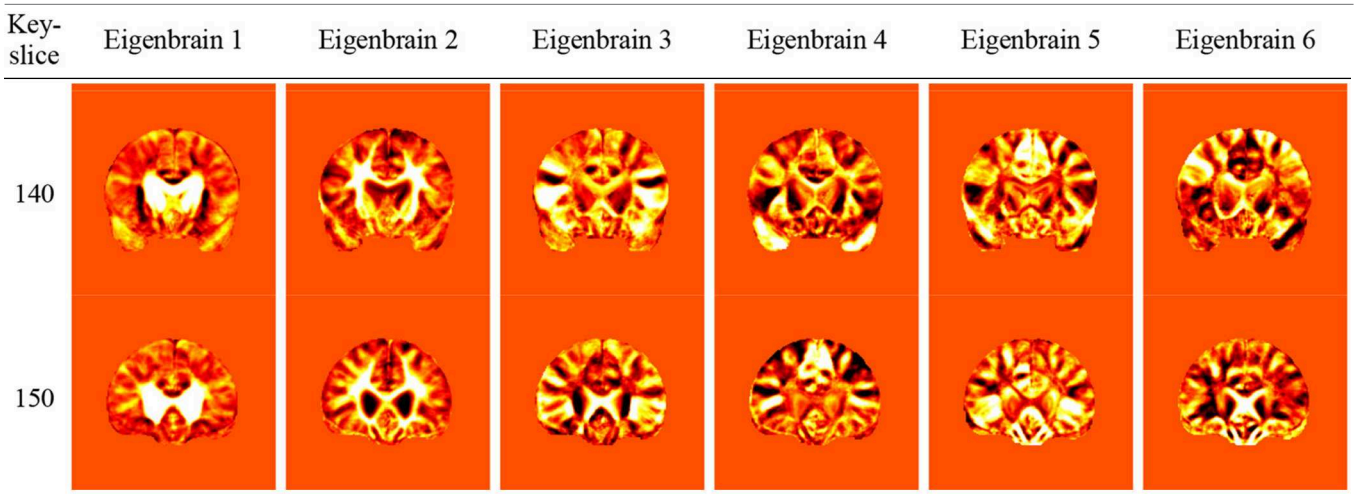
WTT was conducted to give quantified proof of why the first eigenbrain was MIE. We performed WTT for the first six eigenbrains of all key-slices between eigenvalues to characterize those that were AD and those that were NC. The results were

TABLE 7 | Eigenbrain results.

Key-slice	Eigenbrain 1	Eigenbrain 2	Eigenbrain 3	Eigenbrain 4	Eigenbrain 5	Eigenbrain 6
60						
70						
80						
90						
100						
110						
120						
130						

(Continued)

TABLE 7 | Continued



shown in **Table 8**, and p-values less than 0.05 were marked in bold. Only the first eigenvalues of all slices were less than 0.05; therefore, the first eigenbrain was indeed the MIE, and we assigned the eigenvalues of MIE of all 10 key-slices (namely, $10 \times 1 = 10$ features) of each subject to classification.

Classification Comparison

The two classes in order were AD and NC, following common convention. Here we designed three tasks. The first did not use the kernel technique, i.e., the basic linear SVM; the second used RBF-KSVM; and the third used POL-KSVM. The kernel parameters and error penalty were optimized by PSO method. The classification results were listed in **Table 9**, in addition with the results of state-of-the-art methods.

Region Detection

We carried out the region detection procedure from MIE as Section Region Detection described. **Table 10** showed the result, in which the green points represented the discriminant voxels.

Here we reported the discriminative regions interpreted by eigenbrain in **Table 11**, where BA represented Brodmann area.

Discussion

It is clearly observed in **Table 6** that the selected coronal slices are significant in detecting AD from NC. In particular, the AD subjects show the cerebrospinal fluid (CSF) in the areas occupied by brain matter in the NC subjects. We conclude that 10× is reasonable because of following three reasons: (1) The 10× key-slice undersampling (i.e., select only one slice from 10 consecutive slices) yields a coarser brain while still capturing most tissues in the brain (Compare **Table 6** with **Figure 1**). (2) It is very hard to define a fitness (optimization) function to find the optimal undersampling factor. (3) The classification system has a good accuracy in distinguishing AD from NC, and it detects correct AD-related brain regions (See **Tables 9**, **11**). As there are spatial redundancy for neighboring coronal slices, the undersampling could reduce this redundancy to a rather small degree.

Overall, the eigenbrains in **Table 7** capture both similarities and differences of structural features between AD and NC. The first eigenbrain capture the significant feature of AD from NC, and the second and following eigenbrains capture general brain structure. Revisiting the hippocampus part in the first eigenbrain of all key-slices, it is easily perceived that the body lateral ventricles area of AD are highlighted, which is indeed a distinct attribute between AD and NC. Our experiment extends the eigenbrain on SPECT images by Alvarez et al. (2009a) and Lopez et al. (2009) and shows that eigenbrain is also suitable for MRI scans.

The p-values in **Table 8** show that the first eigenvalue λ_1 are all less than 0.05 for all key-slices. It indicates that mean values of λ_1 of AD and NC are significantly different. Hence, the most dominating eigenvalue characterizing AD and NC is the one corresponding to the first eigenbrain. For other eigenvalues, merely 1 of 10 p-values is less than 0.05, which indicates that those eigenbrains are not dominating features indicative of AD from NC. Therefore, the first eigenvalue is MIE and was selected.

Classification results in **Table 9** compare the proposed three classifiers with state-of-the-art methods, in which Zhang's results (Table 7 in Zhang et al., 2014) are calculated through a single K-fold CV experiment. Plant's results (Task 1 in Table 3 Plant et al., 2010) offer the means together with 95% confidence intervals. Savio's results (Table 5 Savio and Grana, 2013) give the means with SD. For the proposed methods, it is **unexpected** that the POL-KSVM produces better classification accuracy of 92.36 ± 0.94 than linear SVM of 91.47 ± 1.02 and RBF-KSVM of 86.71 ± 1.93, because RBF was reported as the most widely used kernel. Our results are better than or comparable to other approaches to AD prediction from MR brain images of NC, e.g., US + SVD-PCA + SVM-DT of 90% (Zhang et al., 2014), BRC + IG + SVM of 90% (Plant et al., 2010), BRC + IG + Bayes of 92% (Plant et al., 2010), MGM + PEC + SVM of 92.07% (Savio and Grana, 2013), GEODAN + BD + SVM of 92.09% (Savio and

TABLE 8 | WTT of the first six eigenvalues of 10 key-slices.

	λ_1			λ_2			λ_3		
Slice	**NC**	**AD**	**p**	**NC**	**AD**	**p**	**NC**	**AD**	**p**
60	−3.36 ± 20.01	11.75 ± 27.91	**0.01**	2.82 ± 18.77	−9.87 ± 27.93	**0.03**	0.11 ± 18.95	−0.39 ± 21.44	0.91
70	−6.84 ± 25.60	23.92 ± 28.33	**0.00**	0.43 ± 21.20	−1.50 ± 36.97	0.79	1.84 ± 19.88	−6.44 ± 22.86	0.09
80	−7.48 ± 29.05	26.18 ± 27.04	**0.00**	−0.65 ± 22.00	2.26 ± 33.36	0.67	−0.25 ± 21.84	0.87 ± 25.08	0.83
90	6.79 ± 32.04	−23.75 ± 24.86	**0.00**	0.42 ± 21.94	−1.46 ± 32.98	0.78	−1.88 ± 20.16	6.57 ± 21.48	0.07
100	−6.93 ± 34.25	24.27 ± 30.89	**0.00**	2.51 ± 23.05	−8.79 ± 31.63	0.09	0.63 ± 20.16	−2.22 ± 23.74	0.57
110	−6.95 ± 31.89	24.31 ± 24.10	**0.00**	0.48 ± 25.03	−1.67 ± 32.93	0.75	1.95 ± 18.17	−6.81 ± 29.05	0.14
120	−5.93 ± 31.60	20.74 ± 23.14	**0.00**	−0.33 ± 24.02	1.14 ± 31.84	0.82	−1.07 ± 16.73	3.74 ± 25.61	0.35
130	5.02 ± 28.13	−17.56 ± 28.09	**0.00**	−1.40 ± 21.70	4.90 ± 27.75	0.27	−0.59 ± 17.75	2.06 ± 19.20	0.52
140	4.27 ± 25.02	−14.94 ± 22.06	**0.00**	−1.34 ± 18.13	4.70 ± 27.10	0.27	3.12 ± 17.91	−10.93 ± 14.69	**0.00**
150	5.51 ± 18.50	−19.30 ± 30.21	**0.00**	−2.22 ± 18.08	7.78 ± 24.66	0.05	1.42 ± 16.56	−4.97 ± 13.98	0.05
	λ_4			λ_5			λ_6		
Slice	**NC**	**AD**	**p**	**NC**	**AD**	**p**	**NC**	**AD**	**p**
60	−1.27 ± 15.47	4.43 ± 25.32	0.27	1.51 ± 14.13	−5.29 ± 23.59	0.16	−1.29 ± 13.10	4.50 ± 23.71	0.22
70	1.99 ± 17.76	−6.95 ± 22.50	0.06	−0.03 ± 16.69	0.09 ± 23.25	0.98	−0.96 ± 16.08	3.35 ± 20.79	0.32
80	1.46 ± 21.14	−5.12 ± 18.85	0.12	−0.72 ± 17.80	2.52 ± 24.31	0.51	−1.34 ± 17.47	4.68 ± 21.78	0.19
90	0.31 ± 19.66	−1.09 ± 23.73	0.78	−0.54 ± 18.05	1.89 ± 24.49	0.63	−1.80 ± 16.79	6.29 ± 23.33	0.10
100	−1.56 ± 18.77	5.47 ± 21.18	0.12	0.84 ± 16.32	−2.95 ± 25.35	0.46	−0.53 ± 15.58	1.85 ± 24.87	0.63
110	−0.31 ± 19.32	1.07 ± 17.30	0.72	0.54 ± 16.78	−1.87 ± 22.19	0.60	−1.09 ± 16.07	3.83 ± 20.43	0.25
120	−0.32 ± 16.83	1.13 ± 21.16	0.74	−2.21 ± 18.00	7.74 ± 10.70	**0.00**	−1.31 ± 14.81	4.57 ± 21.45	0.18
130	1.61 ± 17.00	−5.62 ± 18.51	0.07	1.39 ± 14.21	−4.86 ± 23.47	0.19	2.01 ± 15.42	−7.04 ± 17.25	**0.02**
140	2.11 ± 16.81	−7.39 ± 16.29	**0.01**	0.44 ± 15.37	−1.56 ± 17.70	0.59	1.21 ± 14.37	−4.24 ± 17.85	0.15
150	1.17 ± 13.52	−4.11 ± 18.51	0.17	0.27 ± 14.35	−0.94 ± 13.89	0.69	0.17 ± 13.52	−0.58 ± 15.14	0.82

P-values less than 0.05 are in bold.

TABLE 9 | Comparison of classification results.

	Accuracy	**Sensitivity**	**Specificity**	**Precision**
EXISTING METHODS				
US + SVD-PCA + SVM-DT (Zhang et al., 2014)	90	94	71	N/A
BRC + IG + SVM (Plant et al., 2010)	90.00 [77.41, 96.26]	96.88 [82.01, 99.84]	77.78 [51.92, 92.63]	N/A
BRC + IG + Bayes (Plant et al., 2010)	92.00 [79.89, 97.41]	93.75 [77.78, 98.27]	88.89 [63.93, 98.05]	N/A
BRC + IG + VFI (Plant et al., 2010)	78.00 [63.67, 88.01]	65.63 [46.78, 80.83]	100.00 [78.12, 100]	N/A
MGM + PEC + SVM (Savio and Grana, 2013)	92.07 ± 1.12	86.67 ± 4.71	N/A	95.83 ± 5.89
GEODAN + BD + SVM (Savio and Grana, 2013)	92.09 ± 2.60	80.00 ± 4.00	N/A	88.09 ± 5.33
TJM + WTT + SVM (Savio and Grana, 2013)	92.83 ± 0.91	86.33 ± 3.73	N/A	85.62 ± 0.85
PROPOSED METHODS				
ICV + Eigenbrain + WTT + SVM	91.47 ± 1.02	90.17 ± 1.66	91.84 ± 1.09	93.21 ± 2.43
ICV + Eigenbrain + WTT + RBF-KSVM	86.71 ± 1.93	85.71 ± 1.91	86.99 ± 2.30	66.12 ± 4.16
ICV + Eigenbrain + WTT + POL-KSVM	92.36 ± 0.94	83.48 ± 3.27	94.90 ± 1.09	82.28 ± 2.78

Grana, 2013), and TJM + WTT + SVM of 92.83% (Savio and Grana, 2013). There were many other methods (Gray et al., 2012; Arbizu et al., 2013; Chaves et al., 2013; Dukart et al., 2013; Cohen and Klunk, 2014) proposed for detecting AD from NC, however, they treated images from other modalities (such as SPECT and PET). Therefore, it is not appropriate to compare the proposed methods with them. We will test our methods on SPECT and PET images in the future.

Table 11 shows that eigenbrains interpret the discriminant voxels involving the following regions reported in existing literatures: Anterior Cingulate (BA-24, BA-32) (Schultz et al., 2014), Caudate Nucleus (Head, body, and tail) (Möller et al., 2015), Cerebellum (Colloby et al., 2014), Cingulate Gyrus (BA-23, BA-24, BA-31) (Yu et al., 2014), Claustrum (De Reuck et al., 2014), Inferior Frontal Gyrus (BA-47) (Eliasova et al., 2014), Inferior Parietal Lobule (BA-40) (Wang et al., 2015), Insula

TABLE 10 | Discriminant voxels.

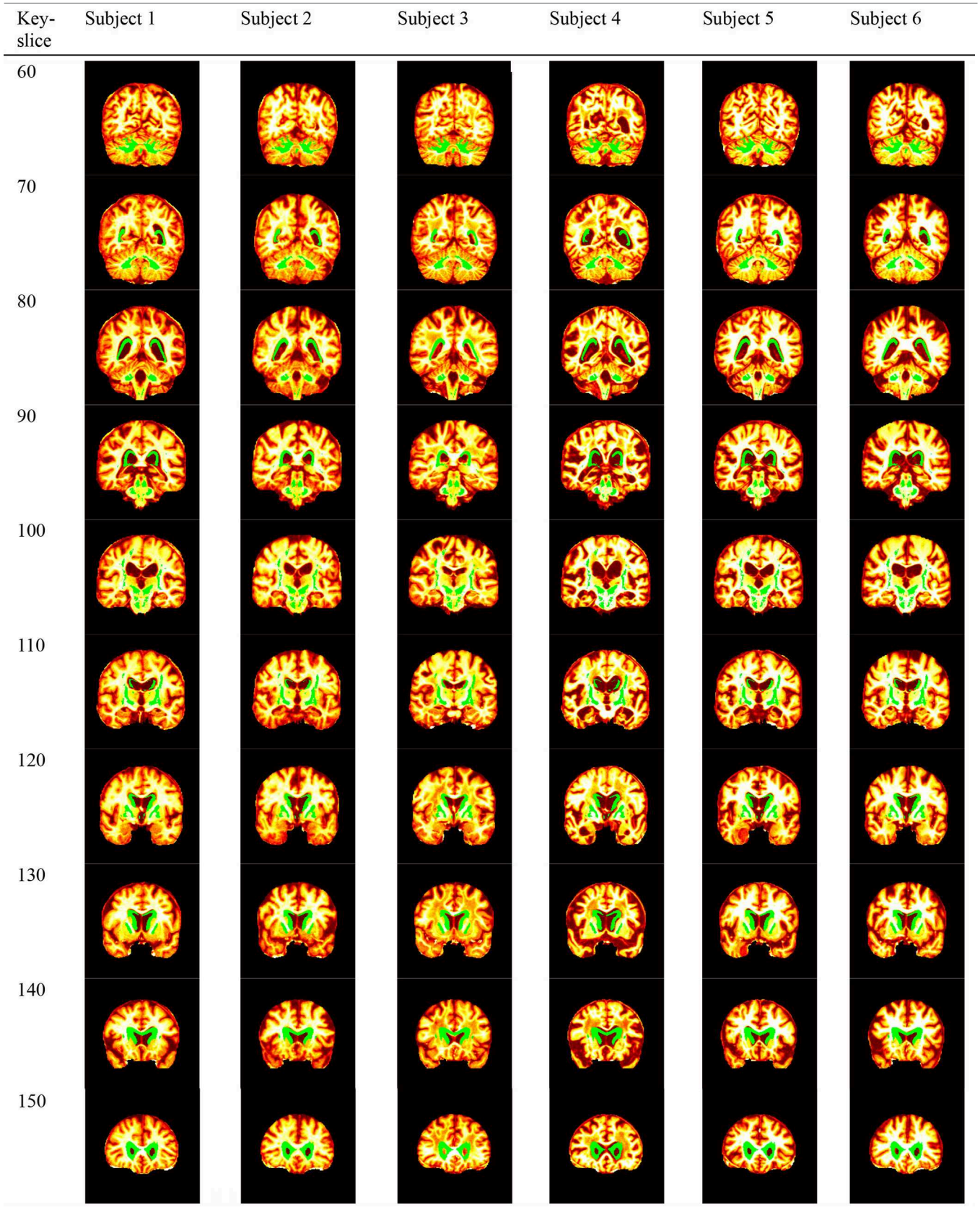


TABLE 11 | Regions found by Eigenbrain.

Regions	# of voxels	Reported by
Anterior cingulate (BA-24, BA-32)	35	Schultz et al., 2014
Caudate nucleus (Head, body, and tail)	407	Möller et al., 2015
Cerebellum	65	Colloby et al., 2014
Cingulate gyrus (BA-23, BA-24, BA-31)	343	Yu et al., 2014
Claustrum	14	De Reuck et al., 2014
Inferior frontal gyrus (BA-47)	71	Eliasova et al., 2014
Inferior parietal lobule (BA-40)	29	Wang et al., 2015
Insula (BA-13)	23	He et al., 2015
Lateral ventricle	410	Voevodskaya et al., 2014
Lentiform nucleus	569	Möller et al., 2015
Lingual gyrus	71	Lehmann et al., 2013
Medial frontal gyrus (BA-10, BA-11, BA-25, BA-6)	416	Kang et al., 2013
Middle frontal gyrus (BA-11)	52	Schultz et al., 2014
Middle occipital gyrus	22	Lehmann et al., 2013
Middle temporal gyrus	50	Aubry et al., 2015
Paracentral lobule (BA-3, BA-4, BA-5, BA-6, BA-7)	210	Kang et al., 2013
Parahippocampal gyrus (Amygdala, BA-28, BA-35, Hippocampus)	276	Eskildsen et al., 2015
Postcentral gyrus (BA-5)	10	Kang et al., 2013
Posterior cingulate	27	Shinohara et al., 2014
Precentral gyrus (BA-4)	11	Kang et al., 2013
Precuneus (BA-7, BA-31)	557	Kang et al., 2013
Subcallosal gyrus (BA-25, BA-34, BA-47)	82	Paakki et al., 2010
Sub-Gyral (BA-40, Corpus Callosum, Hippocampus)	589	Streitburger et al., 2012
Superior frontal gyrus	70	Chen et al., 2014
Superior parietal lobule	269	Quiroz et al., 2013
Superior temporal gyrus (BA-38)	12	Paakki et al., 2010
Supramarginal gyrus	14	Quiroz et al., 2013
Thalamus (Medial Geniculum Body, Pulvinar, Ventral Lateral Nucleus)	35	He et al., 2015
Transverse Temporal Gyrus (BA-41)	26	Kim et al., 2012
Uncus (BA-28)	25	Bangen et al., 2014

(BA-13) (He et al., 2015), Lateral Ventricle (Voevodskaya et al., 2014), Lentiform Nucleus (Möller et al., 2015), Lingual gyrus (Lehmann et al., 2013), Medial Frontal Gyrus (BA-10, BA-11, BA-25, BA-6) (Kang et al., 2013), Middle Frontal Gyrus (BA-11) (Schultz et al., 2014), Middle Occipital Gyrus (Lehmann et al., 2013), Middle Temporal Gyrus (Aubry et al., 2015), Paracentral Lobule (BA-3, BA-4, BA-5, BA-6, BA-7) (Kang et al., 2013), Parahippocampal Gyrus (Amygdala, BA-28, BA-35, Hippocampus) (Eskildsen et al., 2015), Postcentral Gyrus (BA-5) (Kang et al., 2013), Posterior Cingulate (Shinohara et al., 2014), Precentral Gyrus (BA-4) (Kang et al., 2013), Precuneus (BA-7, BA-31) (Kang et al., 2013), Subcallosal Gyrus (BA-25, BA-34, BA-47) (Paakki et al., 2010), Sub-Gyral (BA-40, Corpus Callosum, Hippocampus) (Streitburger et al., 2012), Superior Frontal Gyrus (Chen et al., 2014), Superior Parietal Lobule (Quiroz et al., 2013), Superior Temporal Gyrus (BA-38) (Paakki et al., 2010), Supramarginal Gyrus (Quiroz et al., 2013), Thalamus (Medial Geniculum Body, Pulvinar, Ventral Lateral Nucleus) (He et al., 2015), Transverse Temporal Gyrus (BA-41) (Kim et al., 2012), and Uncus (BA-28) (Bangen et al., 2014).

Nevertheless, some regions reported to be associated with AD are not interpreted by Eigenbrain, such as subthalamic nucleus (De Reuck et al., 2014). The reason may lie in three aspects. First, the quantile of our method is assigned with a value of 0.98, which is considered high. Reducing the quantile value may include more regions. Second, some literature used other advanced imaging modalities, such as MRSI and fMRI for metabolism detection and function analysis. Third, the key-slice selection procedure may miss important regions.

From another point of view, **Table 11** demonstrates the power of the eigenbrain. Our study uses only one feature (eigenbrain) on 10 key-slices of a simple 3D structural MR image, nevertheless, our findings cover 30 related regions reported by over twenty literatures, which used various feature extraction methods and advanced imaging technologies.

The **contributions** of the paper fall within the following five aspects: (i) We generalize the Eigenbrain to MR images, and prove its effectiveness; (ii) We propose a hybrid eigenbrain-based CAD system that can not only detect AD from NC, but also detect brain regions that related to AD. (iii) We prove the proposed method has a classification accuracy comparable to state-of-the-art methods, and the detected brain regions are in line with 16 existing literatures. (iv) We use ICV and WTT to reduce redundant data; (v) we find POL kernel is better than linear and RBF kernel for this study.

In conclusion, the advantages of eigenbrain are three-fold: (i) it reaches very high classification accuracy, which was better than or competitive with state-of-the-art methods (Plant et al., 2010; Savio and Grana, 2013; Zhang et al., 2014); (ii) it can directly find discriminant voxels/regions within the whole brain; (iii) it can be combined with other features, in order to increase the classification performance. On the other hand, the disadvantages of eigenbrain also exist: (i) it is essentially two-dimensional, which does not reduce the redundancy along the slice direction; (ii) it needs preprocessing of spatial registration, which costs large amount of computation resources.

To the policy-makers, this study suggests the eigenbrain technique can achieve comparable results to traditional methods. It may offer a ray of hope for AD diagnosis with unconventional means with the combination of eigenbrain and machine learning. This preclinical study suggests that hospitals and medical laboratories enroll more computer scientists and engineers, with the aim of developing efficient AD diagnosis and region detection systems.

Conclusion and Future Research

We presented an automated and accurate classification method that was based on eigenbrains and machine learning, in order to detect AD subjects and AD-related brain regions using 3D MR

images. The results showed the proposed POL-KSVM method achieved 92.36% accuracy, which was competitive with state-of-the-art methods.

In the future, we will focus our research in the following aspects: (i) We shall generalize the eigenbrain to three dimensional, so the procedure of key-slice selection can be removed; (ii) We shall test other kernels for SVM, and try to replace KSVM with other advanced pattern recognition tools. (iii) Eigenbrain can be used in combination with DWT-based features and others, and an increase in classification accuracy is expected.

Acknowledgments

This work was supported by NSFC (610011024, 61273243, 51407095), Program of Natural Science Research of Jiangsu Higher Education Institutions (13KJB460011, 14KJB520021), Jiangsu Key Laboratory of 3D Printing Equipment and Manufacturing (BM2013006), Key Supporting Science and Technology Program (Industry) of Jiangsu Province (BE2012201, BE2014009-3, BE2013012-2), Special Funds for Scientific and Technological Achievement Transformation Project in Jiangsu Province (BA2013058), and Nanjing Normal University Research Foundation for Talented Scholars (2013119XGQ0061, 2014119XGQ0080). The authors acknowledge their gratitude to the OASIS dataset that came from NIH grants P50AG05681, P01 AG03991, R01 AG021910, P50 MH071616, U24 RR021382 and R01 MH56584.

Supplementary Material

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fncom.2015.00066/abstract

References

Aich, U., and Banerjee, S. (2014). Modeling of EDM responses by support vector machine regression with parameters selected by particle swarm optimization. *Appl. Math. Model.* 38, 2800–2818. doi: 10.1016/j.apm.2013.10.073

Alvarez, I., Gorriz, J. M., Ramirez, J., Salas-Gonzalez, D., Lopez, M., Puntonet, C. G., et al. (2009a). Alzheimer's diagnosis using eigenbrains and support vector machines. *Electron. Lett.* 45, 342–343. doi: 10.1049/el.2009.3415

Álvarez, I., Górriz, J. M., Ramírez, J., Salas-Gonzalez, D., López, M., Segovia, F., et al. (2009b). "Alzheimer's diagnosis using eigenbrains and support vector machines," in *Bio-Inspired Systems: Computational and Ambient Intelligence*, Vol. 5517, eds J. Cabestany, F. Sandoval, A. Prieto, and J. Corchado (Berlin: Springer), 973–980.

Anagnostopoulos, C. N., Giannoukos, I., Spenger, C., Simmons, A., Mecocci, P., Soininen, H., et al. (2013). "Classification models for Alzheimer's disease Detection," in *Engineering Applications of Neural Networks*, Vol. 384(Pt II), eds L. Iliadis, H. Papadopoulos, and C. Jayne (Berlin; Heidelberg: Springer), 193–202. doi: 10.1007/978-3-642-41016-1_21

Angelini, E. D., Delon, J., Bah, A. B., Capelle, L., and Mandonnet, E. (2012). Differential MRI analysis for quantification of low grade glioma growth. *Med. Image Anal.* 16, 114–126. doi: 10.1016/j.media.2011.05.014

Arbizu, J., Prieto, E., Martinez-Lage, P., Marti-Climent, J. M., Garcia-Granero, M., Lamet, I., et al. (2013). Automated analysis of FDG PET as a tool for single-subject probabilistic prediction and detection of Alzheimer's disease dementia. *Eur. J. Nucl. Med. Mol. Imaging* 40, 1394–1405. doi: 10.1007/s00259-013-2458-z

Ardekani, B. A., Bachman, A. H., Figarsky, K., and Sidtis, J. J. (2014). Corpus callosum shape changes in early Alzheimer's disease: an MRI study using the OASIS brain database. *Brain Struct. Funct.* 219, 343–352. doi: 10.1007/s00429-013-0503-0

Ardekani, B. A., Figarsky, K., and Sidtis, J. J. (2013). Sexual dimorphism in the human corpus callosum: an MRI study using the OASIS brain database. *Cereb. Cortex* 23, 2514–2520. doi: 10.1093/cercor/bhs253

Aubry, S., Shin, W., Crary, J. F., Lefort, R., Qureshi, Y. H., Lefebvre, C., et al. (2015). Assembly and interrogation of Alzheimer's disease genetic networks reveal novel regulators of progression. *PLoS ONE* 10:25. doi: 10.1371/journal.pone.0120352

Bangen, K. J., Nation, D. A., Clark, L. R., Harmell, A. L., Wierenga, C. E., Dev, S. I., et al. (2014). Interactive effects of vascular risk burden and advanced age on cerebral blood flow. *Front. Aging Neurosci.* 6:159. doi: 10.3389/fnagi.2014.00159

Bin Tufail, A., Abidi, A., Siddiqui, A. M., and Younis, M. S. (2012). "Multiclass classification of initial stages of Alzheimer's disease using structural MRI phase images," in *Proceedings of the IEEE International Conference in Control System, Computing and Engineering (ICCSCE)* (Penang: IEEE), 317–321. doi: 10.1109/ICCSCE.2012.6487163

Brookmeyer, R., Johnson, E., Ziegler-Graham, K., and Arrighi, H. M. (2007). Forecasting the global burden of Alzheimer's disease. *Alzheimers Dement.* 3, 186–191. doi: 10.1016/j.jalz.2007.04.381

Chaplot, S., Patnaik, L. M., and Jagannathan, N. R. (2006). Classification of magnetic resonance brain images using wavelets as input to support vector machine and neural network. *Biomed. Signal Process. Control* 1, 86–92. doi: 10.1016/j.bspc.2006.05.002

Chaves, R., Ramirez, J., Gorriz, J. M., and Alzheimer's Dis, N. (2013). Integrating discretization and association rule-based classification for Alzheimer's disease diagnosis. *Expert Syst. Appl.* 40, 1571–1578. doi: 10.1016/j.eswa.2012.09.003

Chen, Y., Liu, Z., Zhang, J., Xu, K., Zhang, S., Wei, D., et al. (2014). Altered brain activation patterns under different working memory loads in patients with Type 2 diabetes. *Diabetes Care* 37, 3157–3163. doi: 10.2337/dc14-1683

Cohen, A. D., and Klunk, W. E. (2014). Early detection of Alzheimer's disease using PiB and FDG PET. *Neurobiol. Dis.* 72, 117–122. doi: 10.1016/j.nbd.2014.05.001

Collins, M. P., and Pape, S. E. (2011). The potential of support vector machine as the diagnostic tool for schizophrenia: a systematic literature review of neuroimaging studies. *Eur. Psychiatry* 26, 117–122. doi: 10.1016/S0924-9338(11)73068-1

Colloby, S. J., O'Brien, J. T., and Taylor, J. P. (2014). Patterns of cerebellar volume loss in dementia with Lewy bodies and Alzheimer's disease: A VBM-DARTEL study. *Psychiatry Res.* 223, 187–191. doi: 10.1016/j.pscychresns.2014.06.006

Das, S., Chowdhury, M., and Kundu, M. K. (2013). Brain MR image classification using multiscale geometric analysis of ripplet. *Prog. Electromagn. Res.* 137, 1–17. doi: 10.2528/PIER13010105

De Reuck, J. L., Deramecourt, V., Auger, F., Durieux, N., Cordonnier, C., Devos, D., et al. (2014). Iron deposits in post-mortem brains of patients with neurodegenerative and cerebrovascular diseases: a semi-quantitative 7.0 T magnetic resonance imaging study. *Eur. J. Neurol.* 21, 1026–1031. doi: 10.1111/ene.12432

Dukart, J., Mueller, K., Barthel, H., Villringer, A., Sabri, O., Schroeter, M. L., et al. (2013). Meta-analysis based SVM classification enables accurate detection of Alzheimer's disease across different clinical centers using FDG-PET and MRI. *Psychiatry Res.* 212, 230–236. doi: 10.1016/j.pscychresns.2012.04.007

El-Dahshan, E. S. A., Hosny, T., and Salem, A. B. M. (2010). Hybrid intelligent techniques for MRI brain images classification. *Digit. Signal Process.* 20, 433–441. doi: 10.1016/j.dsp.2009.07.002

El-Dahshan, E. S. A., Mohsen, H. M., Revett, K., and Salem, A. B. M. (2014). Computer-aided diagnosis of human brain tumor through MRI: a survey and a new algorithm. *Expert Syst. Appl.* 41, 5526–5545. doi: 10.1016/j.eswa.2014.01.021

Eliasova, I., Anderkova, L., Marecek, R., and Rektorova, I. (2014). Non-invasive brain stimulation of the right inferior frontal gyrus may improve attention in early Alzheimer's disease: a pilot study. *J. Neurol. Sci.* 346, 318–322. doi: 10.1016/j.jns.2014.08.036

Eskildsen, S. F., Coupé, P., Fonov, V. S., Pruessner, J. C., and Collins, D. L. (2015). Structural imaging biomarkers of Alzheimer's disease: predicting disease progression. *Neurobiol. Aging* 36(Suppl. 1), S23–S31. doi: 10.1016/j.neurobiolaging.2014.04.034

Garcia, S., Fernandez, A., Luengo, J., and Herrera, F. (2010). Advanced nonparametric tests for multiple comparisons in the design of experiments in computational intelligence and data mining: experimental analysis of power. *Inf. Sci.* 180, 2044–2064. doi: 10.1016/j.ins.2009.12.010

Goh, S., Dong, Z., Zhang, Y., DiMauro, S., and Peterson, B. S. (2014). Mitochondrial dysfunction as a neurobiological subtype of autism spectrum disorder: evidence from brain imaging. *JAMA Psychiatry* 71, 665–671. doi: 10.1001/jamapsychiatry.2014.179

Gomes, T. A. F., Prudêncio, R. B. C., Soares, C., Rossi, A. L. D., and Carvalho, A. (2012). Combining meta-learning and search techniques to select parameters for support vector machines. *Neurocomputing* 75, 3–13. doi: 10.1016/j.neucom.2011.07.005

Gray, K. R., Wolz, R., Heckemann, R. A., Aljabar, P., Hammers, A., Rueckert, D., et al. (2012). Multi-region analysis of longitudinal FDG-PET for the classification of Alzheimer's disease. *Neuroimage* 60, 221–229. doi: 10.1016/j.neuroimage.2011.12.071

Hable, R. (2012). Asymptotic normality of support vector machine variants and other regularized kernel methods. *J. Multivar. Anal.* 106, 92–117. doi: 10.1016/j.jmva.2011.11.004

Hahn, K., Myers, N., Prigarin, S., Rodenacker, K., Kurz, A., Förstl, H., et al. (2013). Selectively and progressively disrupted structural connectivity of functional brain networks in Alzheimer's disease—Revealed by a novel framework to analyze edge distributions of networks detecting disruptions with strong statistical evidence. *Neuroimage* 81, 96–109. doi: 10.1016/j.neuroimage.2013.05.011

Hamy, V., Dikaios, N., Punwani, S., Melbourne, A., Latifoltojar, A., Makanyanga, J., et al. (2014). Respiratory motion correction in dynamic MRI using robust data decomposition registration – Application to DCE-MRI. *Med. Image Anal.* 18, 301–313. doi: 10.1016/j.media.2013.10.016

Han, J. W., Kim, T. H., Lee, S. B., Park, J. H., Lee, J. J., Huh, Y., et al. (2011). 327 Diagnostic Stability of Mild Cognitive Impairment Subtype. *Asian J. Psychiatry* 4(Suppl. 1), S65–S66. doi: 10.1016/s1876-2018(11)60250-5

He, W., Liu, D., Radua, J., Li, G., Han, B., and Sun, Z. (2015). Meta-analytic comparison between PIB-PET and FDG-PET results in Alzheimer's disease and MCI. *Cell Biochem. Biophys.* 71, 17–26. doi: 10.1007/s12013-014-0138-7

Jeurissen, B., Leemans, A., and Sijbers, J. (2014). Automated correction of improperly rotated diffusion gradient orientations in diffusion weighted MRI. *Med. Image Anal.* 18, 953–962. doi: 10.1016/j.media.2014.05.012

Kalbkhani, H., Shayesteh, M. G., and Zali-Vargahan, B. (2013). Robust algorithm for brain magnetic resonance image (MRI) classification based on GARCH variances series. *Biomed. Signal Process. Control* 8, 909–919. doi: 10.1016/j.bspc.2013.09.001

Kang, K., Yoon, U., Lee, J. M., and Lee, H. W. (2013). Idiopathic normal-pressure hydrocephalus, cortical thinning, and the cerebrospinal fluid tap test. *J. Neurol. Sci.* 334, 55–62. doi: 10.1016/j.jns.2013.07.014

Khazaee, A., and Zadeh, A. E. (2014). ECG beat classification using particle swarm optimization and support vector machine. *Front. Comput. Sci.* 8, 217–231. doi: 10.1007/s11704-014-2398-1

Kim, J. S., Lee, S. H., Park, G., Kim, S., Bae, S. M., Kim, D. W., et al. (2012). Clinical implications of quantitative electroencephalography and current source density in patients with Alzheimer's disease. *Brain Topogr.* 25, 461–474. doi: 10.1007/s10548-012-0234-1

Kubota, T., Ushijima, Y., and Nishimura, T. (2006). A region-of-interest (ROI) template for three-dimensional stereotactic surface projection (3D-SSP) images: initial application to analysis of Alzheimer disease and mild cognitive impairment. *Int. Congr. Ser.* 1290, 128–134. doi: 10.1016/j.ics.2005.11.104

Lee, W., Park, B., and Han, K. (2013). Classification of diffusion tensor images for the early detection of Alzheimer's disease. *Comput. Biol. Med.* 43, 1313–1320. doi: 10.1016/j.compbiomed.2013.07.004

Lehmann, M., Ghosh, P. M., Madison, C., Laforce, R., Corbetta-Rastelli, C., Weiner, M. W., et al. (2013). Diverging patterns of amyloid deposition and hypometabolism in clinical variants of probable Alzheimer's disease. *Brain* 136, 844–858. doi: 10.1093/brain/aws327

Li, D., Yang, W., and Wang, S. (2010). Classification of foreign fibers in cotton lint using machine vision and multi-class support vector machine. *Comput. Electron. Agric.* 74, 274–279. doi: 10.1016/j.compag.2010.09.002

Liu, F. Y., Zhou, L. P., Shen, C. H., and Yin, J. P. (2014). Multiple kernel learning in the primal for multimodal Alzheimer's disease classification. *IEEE J. Biomed. Health Inform.* 18, 984–990. doi: 10.1109/JBHI.2013.2285378

Lopez, M., Ramirez, J., Gorriz, J. M., Alvarez, I., Salas-Gonzalez, D., Segovia, F., et al. (2009). "Automatic system for Alzheimer's disease diagnosis using eigenbrains and bayesian classification rules," *Bio-Inspired Systems: Computational and Ambient Intelligence*, Vol. 5517, eds J. Cabestany, A. Prieto, F. Sandoval, and J. M. Corchado (Berlin: Springer-Verlag Berlin), 949–956.

Maitra, M., and Chatterjee, A. (2006). A Slantlet transform based intelligent system for magnetic resonance brain image classification. *Biomed. Signal Process. Control* 1, 299–306. doi: 10.1016/j.bspc.2006.12.001

Marcus, D. S., Wang, T. H., Parker, J., Csernansky, J. G., Morris, J. C., and Buckner, R. L. (2007). Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. *J. Cogn. Neurosci.* 19, 1498–1507. doi: 10.1162/jocn.2007.19.9.1498

Miller, V., Erlien, S., and Piersol, J. (2012). *Identifying Dementia in MRI Scans using Machine Learning*. Stanford, CA: Stanford University.

Möller, C., Dieleman, N., van der Flier, W. M., Versteeg, A., Pijnenburg, Y., Scheltens, P., et al. (2015). More atrophy of deep gray matter structures in frontotemporal dementia compared to Alzheimer's disease. *J. Alzheimers Dis.* 44, 635–647. doi: 10.3233/JAD-141230

Nambakhsh, C. M. S., Yuan, J., Punithakumar, K., Goela, A., Rajchl, M., Peters, T. M., et al. (2013). Left ventricle segmentation in MRI via convex relaxed distribution matching. *Med. Image Anal.* 17, 1010–1024. doi: 10.1016/j.media.2013.05.002

Paakki, J. J., Rahko, J., Long, X., Moilanen, I., Tervonen, O., Nikkinen, J., et al. (2010). Alterations in regional homogeneity of resting-state brain activity in autism spectrum disorders. *Brain Res.* 1321, 169–179. doi: 10.1016/j.brainres.2009.12.081

Pennanen, C., Kivipelto, M., Tuomainen, S., Hartikainen, P., Hänninen, T., Laakso, M. P., et al. (2004). Hippocampus and entorhinal cortex in mild cognitive impairment and early AD. *Neurobiol. Aging* 25, 303–310. doi: 10.1016/S0197-4580(03)00084-8

Plant, C., Teipel, S. J., Oswald, A., Böhm, C., Meindl, T., Mourao-Miranda, J., et al. (2010). Automated detection of brain atrophy patterns based on MRI for the prediction of Alzheimer's disease. *Neuroimage* 50, 162–174. doi: 10.1016/j.neuroimage.2009.11.046

Quiroz, Y. T., Stern, C. E., Reiman, E. M., Brickhouse, M., Ruiz, A., Sperling, R. A., et al. (2013). Cortical atrophy in presymptomatic Alzheimer's disease presenilin 1 mutation carriers. *J. Neurol. Neurosurg. Psychiatry* 84, 556–561. doi: 10.1136/jnnp-2012-303299

Ramasamy, R., and Anandhakumar, P. (2011). "Brain tissue classification of MR images using fast fourier transform based expectation- maximization gaussian mixture model," in *Advances in Computing and Information Technology,* Vol. 198, D. C. Wyld, M. Wozniak, N. Chaki, N. Meghanathan, and D. Nagamalai (Berlin, Springer-Verlag Berlin), 387–398.

Russell, J., and Cohn, R. (2012). *Bessel's Correction*.

Saritha, M., Joseph, K. P., and Mathew, A. T. (2013). Classification of MRI brain images using combined wavelet entropy based spider web plots and probabilistic neural network. *Pattern Recognit. Lett.* 34, 2151–2156. doi: 10.1016/j.patrec.2013.08.017

Savio, A., and Grana, M. (2013). Deformation based feature selection for computer aided diagnosis of Alzheimer's Disease. *Expert Syst. Appl.* 40, 1619–1628. doi: 10.1016/j.eswa.2012.09.009

Schultz, S. A., Larson, J., Oh, J., Koscik, R., Dowling, M. N., Gallagher, C. L., et al. (2014). Participation in cognitively-stimulating activities is associated with brain structure and cognitive function in preclinical Alzheimer's disease. *Brain Imaging Behav.* doi: 10.1007/s11682-014-9329-5. [Epub ahead of print].

Shamonin, D. P., Bron, E. E., Lelieveldt, B. P. F., Smits, M., Klein, S., and Staring, M. (2014). Fast Parallel Image Registration on CPU and GPU for Diagnostic Classification of Alzheimer's Disease. *Front. Neuroinform.* 7:50. doi: 10.3389/fninf.2013.00050

Shinohara, M., Fujioka, S., Murray, M. E., Wojtas, A., Baker, M., Rovelet-Lecrux, A., et al. (2014). Regional distribution of synaptic markers and APP correlate with distinct clinicopathological features in sporadic and familial Alzheimer's disease. *Brain* 137, 1533–1549. doi: 10.1093/brain/awu046

Smal, I., Carranza-Herrezuelo, N., Klein, S., Wielopolski, P., Moelker, A., Springeling, T., et al. (2012). Reversible jump MCMC methods for fully automatic motion analysis in tagged MRI. *Med. Image Anal.* 16, 301–324. doi: 10.1016/j.media.2011.08.006

Streitburger, D. P., Möller, H. E., Tittgemeyer, M., Hund-Georgiadis, M., Schroeter, M. L., and Mueller, K. (2012). Investigating structural brain changes of dehydration using voxel-based morphometry. *PLoS ONE* 7:e44195. doi: 10.1371/journal.pone.0044195

Voevodskaya, O., Simmons, A., Nordenskjold, R., Kullberg, J., Ahlstrom, H., Lind, L., et al. (2014). The effects of intracranial volume adjustment approaches on multiple regional MRI volumes in healthy aging and Alzheimer's disease. *Front. Aging Neurosci.* 6:264. doi: 10.3389/fnagi.2014.00264

Wang, Z., Xia, M., Dai, Z., Liang, X., Song, H., He, Y., et al. (2015). Differentially disrupted functional connectivity of the subregions of the inferior parietal lobule in Alzheimer's disease. *Brain Struct. Funct.* 220, 745–762. doi: 10.1007/s00429-013-0681-9

Williams, M. M., Storandt, M., Roe, C. M., and Morris, J. C. (2013). Progression of Alzheimer's disease as measured by clinical dementia rating sum of boxes scores. *Alzheimers Dement.* 9(1, Suppl.), S39-S44. doi: 10.1016/j.jalz.2012.01.005

Xinyun, C., Wenlu, Y., and Xudong, H. (2011). "ICA-based classification of MCI vs HC. Natural Computation (ICNC)," *Seventh International Conference*, Vol. 3 (Shanghai: IEEE), 1658–1662. doi: 10.1109/ICNC.2011.6022275

Xue, Z. H., Du, P. J., and Su, H. J. (2014). Harmonic analysis for hyperspectral image classification integrated with PSO optimized SVM. *J. Select. Topics Appl. Earth Obs. Remote Sens IEEE* 7, 2131–2146. doi: 10.1109/JSTARS.2014.2307091

Yang, G., Zhang, Y., Yang, J., Ji, G., Dong, Z., Wang, S., et al. (2015). Automated classification of brain images using wavelet-energy and biogeography-based optimization. *Multimed. Tools Appl.* 1–17. doi: 10.1007/s11042-015-2649-7

Yu, Q., Peng, Y., Mishra, V., Ouyang, A., Li, H., Zhang, H., et al. (2014). Microstructure, length, and connection of limbic tracts in normal human brain development. *Front. Aging Neurosci.* 6:228. doi: 10.3389/fnagi.2014.00228

Zhang, Y., Dong, Z., Wang, S., Ji, G., and Yang, J. (2015a). Preclinical Diagnosis of Magnetic Resonance (MR) Brain Images via Discrete Wavelet Packet Transform with Tsallis Entropy and Generalized Eigenvalue Proximal Support Vector Machine (GEPSVM). *Entropy* 17, 1795–1813. doi: 10.3390/e170a41795

Zhang, Y., Dong, Z., Wu, L., and Wang, S. (2011). A hybrid method for MRI brain image classification. *Expert Syst. Appl.* 38, 10049–10053. doi: 10.1016/j.eswa.2011.02.012

Zhang, Y., Wang, S., and Dong, Z. (2014). Classification of Alzheimer disease based on structural magnetic resonance imaging by kernel support vector machine decision tree. *Prog. Electromagn. Res.* 144, 171–184. doi: 10.2528/PIER13121310

Zhang, Y., Wang, S., Ji, G., and Dong, Z. (2013). An MR brain images classifier system via particle swarm optimization and kernel support vector machine. *Scientific World Journal* 2013:130134. doi: 10.1155/2013/130134

Zhang, Y., Wang, S., Ji, G., and Dong, Z. (2015b). Exponential wavelet iterative shrinkage thresholding algorithm with random shift for compressed sensing magnetic resonance imaging. *IEEJ Trans. Electr. Electron. Eng.* 10, 116–117. doi: 10.1002/tee.22059

Zhang, Y., and Wu, L. (2012a). Classification of fruits using computer vision and a multiclass support vector machine. *Sensors* 12, 12489–12505. doi: 10.3390/s120912489

Zhang, Y., and Wu, L. (2012b). An MR brain images classifier via principal component analysis and kernel support vector machine. *Prog. Electromagn. Res.* 130, 369–388. doi: 10.2528/PIER12061410

Zhou, X., Wang, S., Xu, W., Ji, G., Phillips, P., Sun, P., et al. (2015). "Detection of pathological brain in MRI scanning based on wavelet-entropy and naive bayes classifier," in *Bioinformatics and Biomedical Engineering,* Vol. 9043, eds F. Ortuño and I. Rojas (Granada: Springer International Publishing), 201–209. doi: 10.1007/978-3-319-16483-0_20

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.

frontiers in Computational Neuroscience

ORIGINAL RESEARCH
published: 19 May 2015
doi: 10.3389/fncom.2015.00054

On the distinguishability of HRF models in fMRI

Paulo N. Rosa[1]*, *Patricia Figueiredo*[2] and *Carlos J. Silvestre*[3]

[1] Flight Systems Business Unit, Aerospace, Defense & Systems Department, Deimos Engenharia, Lda., Lisboa, Portugal, [2] Institute for Systems and Robotics and Department of Bioengineering, Instituto Superior Técnico, Universidade de Lisboa, Portugal, [3] Department of Electrical and Computer Engineering, Faculty of Science and Technology, University of Macau, Taipa, Macau, China

Modeling the Hemodynamic Response Function (HRF) is a critical step in fMRI studies of brain activity, and it is often desirable to estimate HRF parameters with physiological interpretability. A biophysically informed model of the HRF can be described by a non-linear time-invariant dynamic system. However, the identification of this dynamic system may leave much uncertainty on the exact values of the parameters. Moreover, the high noise levels in the data may hinder the model estimation task. In this context, the estimation of the HRF may be seen as a problem of model falsification or invalidation, where we are interested in distinguishing among a set of eligible models of dynamic systems. Here, we propose a systematic tool to determine the distinguishability among a set of physiologically plausible HRF models. The concept of absolutely input-distinguishable systems is introduced and applied to a biophysically informed HRF model, by exploiting the structure of the underlying non-linear dynamic system. A strategy to model uncertainty in the input time-delay and magnitude is developed and its impact on the distinguishability of two physiologically plausible HRF models is assessed, in terms of the maximum noise amplitude above which it is not possible to guarantee the falsification of one model in relation to another. Finally, a methodology is proposed for the choice of the input sequence, or experimental paradigm, that maximizes the distinguishability of the HRF models under investigation. The proposed approach may be used to evaluate the performance of HRF model estimation techniques from fMRI data.

Keywords: HRF, fMRI, BOLD fMRI, distinguishability, model selection, experimental paradigm

OPEN ACCESS

Edited by:
Tobias Alecio Mattei, Brain & Spine Center - InvisionHealth - Kenmore Mercy Hospital, USA

Reviewed by:
Nelson Jesús Trujillo-Barreto, University of Manchester, UK
Daniele Marinazzo, University of Ghent, Belgium

***Correspondence:**
Paulo N. Rosa, Flight Systems Business Unit, Aerospace, Defense & Systems Department, Deimos Engenharia, Av. D. Joao II, Lt 1.17.01, 10th floor, 1998-023 Lisboa, Portugal
paulo.rosa@deimos.com.pt

Received: *22 July 2014*
Accepted: *24 April 2015*
Published: *19 May 2015*

Citation:
Rosa PN, Figueiredo P and Silvestre CJ (2015) On the distinguishability of HRF models in fMRI. Front. Comput. Neurosci. 9:54. doi: 10.3389/fncom.2015.00054

Introduction

The hemodynamic response function (HRF) describes the local changes in cerebral blood flow, volume, and oxygenation associated with neuronal activity, and it is extensively used to model Blood Oxygen Level Dependent (BOLD) signals measured using functional Magnetic Resonance Imaging (fMRI) (Logothetis and Wandell, 2004). In general, fMRI experiments are used to map networks of brain activity that are associated with a specific stimulus or task, or that are functionally correlated during rest. Mapping of stimulus/task-related BOLD changes is most frequently achieved by fitting a general linear model (GLM) to the data, consisting on the stimulus/task time course convolved with a pre-specified HRF model (Friston et al., 1994), assuming a linear time invariant system (Boynton et al., 1996). Although the exact mechanisms underlying the HRF are not yet completely known, the consistency

of its observed shape allowed for canonical (parameterized) HRF models to be derived (Friston et al., 1998). In particular, double-gamma HRF models are commonly employed in fMRI analysis. Nevertheless, extensive HRF variability has been reported across brain regions (Handwerker et al., 2004), scanning sessions (Aguirre et al., 1998), tasks (Cohen and Ugurbil, 2002), physiological modulations (Liu et al., 2004), subjects (Handwerker et al., 2004), and populations (D'Esposito et al., 2003), which may hinder or confound the measurement of BOLD changes associated with brain activity, limiting the interpretability of fMRI studies.

Common approaches attempting to take into account HRF variability allow for greater flexibility in the HRF shape and dynamics by describing it through a set of basis functions in a GLM framework. They include using the partial derivatives with respect to time and dispersion of a canonical HRF (Friston et al., 1998), finite impulse response (FIR) basis sets (Glover, 1999), and specially designed basis functions (Woolrich et al., 2004). An approach that also takes into account the spatial localization of the HRF was very recently proposed in Vincent et al. (2014). While a small number of basis functions cannot accurately model the whole range of HRF shapes and delays, at the other extreme, deconvolution of the BOLD response is a very noisy process. Critically, these approaches do not provide a biophysical foundation for the HRF model, hence limiting the physiological interpretability of the associated parameters. Moreover, they do not explain empirically observed non-linearities in the BOLD responses (Birn et al., 2001).

Biophysically informed non-linear models of the HRF have been proposed, based on the combination of the Balloon model, describing the dynamic changes in deoxyhemoglobin content as a function of blood oxygenation and blood volume (Buxton et al., 1998), with a model of the blood flow dynamics during brain activation, where neuronal activity is approximated by the stimulus/task input scaled by a factor called neural efficiency (Friston et al., 2000). In the original work that proposed this model, the associated parameters were estimated by using a Volterra kernel expansion to characterize the system dynamics (Friston et al., 2000). Later, a Bayesian estimation framework was introduced, allowing for the use of a priori distributions of the parameter values and the production of the respective posterior probability distributions given the data by using Expectation-Maximization methods (Friston, 2002). This HRF model and respective estimation procedure have further been incorporated in Dynamic Causal Models (DCM) developed to study effective connectivity among networks of brain regions from fMRI data (Friston et al., 2003). More recently, the methods of dynamic expectation maximization, variational filtering, and generalized filtering have also been proposed for model inversion (estimation) in this context (Friston et al., 2008).

Several extensions of the Balloon model have since been considered (Buxton et al., 2004), as well as a metabolic/hemodynamic model that takes the metabolic dynamics into account in order to incorporate the separate roles played by excitatory and inhibitory neuronal activities in the generation of the BOLD signal (Sotero and Trujillo-Barreto, 2007). A few alternative approaches for the estimation of these HRF models and related extensions have also been proposed (Riera et al., 2004). In Riera et al. (2004), a fully stochastic model was presented in order to include physiological noise in the hemodynamic states, in addition to the measurement noise in the observations. A local linearization filter was used for estimating the hemodynamic states as well as the model parameters. In Sotero et al. (2009), a similar approach was used for estimating the metabolic/hemodynamic model proposed by the same group. In contrast to these linearization-based approaches, Johnston et al. (2008) used particle filters so as to truly accommodate the model non-linearities. More recently, Havlicek et al. (2011) proposed non-linear cubature Kalman filtering as a means to invert models of coupled dynamical systems, which furnishes posterior estimates of both the hidden states and the parameters of the system, including any unknown exogenous input.

In fMRI experiments, the system input is given by the stimulus/task time course, which is generally designed as a series of events alternating with baseline periods at specified inter-stimulus intervals (ISIs). A number of studies have addressed the problem of systematically assessing the quality of fMRI experimental designs, both in terms of the ability to detect stimulus/task-related BOLD activation (detection power) and the ability to estimate the HRF model (estimation efficiency) in a given amount of imaging time (Dale, 1999; Liu et al., 2001). Different methodologies have been proposed to determine the optimal design of fMRI experiments for maximal estimation efficiency (Buracas and Boynton, 2002; Wager and Nichols, 2003; Maus et al., 2012), and a few studies have compared different HRF models and the associated estimation efficiency, focusing on specific parameters of interest such as the response latency and duration (Lindquist and Wager, 2007; Lindquist et al., 2009). Importantly, the authors were concerned with the physiological plausibility of the estimated HRF parameters and with their independence, such that differences in one parameter are not confounded with differences in another parameter. However, these studies were based on parameterized HRF models with no direct biophysical groundings, which severely limited the desired physiological interpretability. To our knowledge, no study has so far investigated the effect of experimental design on the estimation of biophysically informed models of the HRF.

When the HRF model is expressed as a dynamic system, the identifiability of this system must be established in order to guarantee that the HRF models inferred from the input/output data are physiologically plausible. It has been shown that the sensitivity of the HRF system input/output behavior to the model parameters is in general small, which means that, when many parameters are estimated together, their values can be varied over a large range with only small changes in the system output (Deneux and Faugeras, 2006). In these cases, the problem of model estimation may be treated as a model falsification (or invalidation) problem, in which we are interested in distinguishing among a set of eligible dynamic systems (Silvestre et al., 2010a). The simplest model falsification problem one can think of is that of stating whether or not a given model is compatible with the current observed input/output data. However, it is important to notice that a model can never be validated in practice. Indeed, the model being compatible with the input/output data up to time t does not imply that it should be compatible at time $t + \delta$ where $\delta > 0$. Therefore, one can

only say that a given model is not falsified (or invalidated) by the current input/output data. On the other hand, a model is obviously invalidated or falsified once it is not compatible with the observations. Hence, we usually refer to model falsification rather than model validation, since the latter is not achievable in practice. The related problem of model (in)distinguishability arises in a wide range of decision architectures, especially in those that are used in noisy and/or uncertain environments, where more than a single eligible model is compatible with the observed input/output dataset. The distinguishability of two models is in general affected by the input signals, particularly by the uncertainty on the input time-delay and on its magnitude. In fact, model invalidation requires a kind of persistence of the excitation condition in the exogenous inputs, so that the magnitude of the system output signal is large enough when compared to the noise level of the data acquisition process—see (Grewal and Glover, 1976; Walter et al., 1984) and references therein.

In this paper, we extend the results in Silvestre et al. (2010b), by first introducing the concept of absolutely input-distinguishable systems and showing that, for systems with forced responses, the distinguishability between two models can be significantly affected by the shape and magnitude of the external input signals. Moreover, several types of uncertainty, such as unknown input time-delays and uncertain magnitudes of the input signal, can also be adverse to model invalidation. We then exploit the concept of absolutely input-distinguishable systems, in order to optimize the estimation efficiency of fMRI experimental designs through the maximization of the distinguishability among a set of physiologically plausible HRF models. It is stressed that one of the main motivations for the work described herein is the development of a technique that helps define an optimal sequence of stimuli, so that the differences between the models in the set of plausible HRFs become apparent. Hence, the methodology proposed in this paper provides a first step to the so-called experimental paradigm design, while also shedding light on the intrinsic limitations of HRF parameter estimation based on fMRI.

Methods

The Balloon Model proposed by Buxton et al. (1998), and further analyzed and complemented with the flow dynamics by Friston et al. (2000), consists of a non-linear differential equation that describes the dynamics of normalized values of the blood flow b_f, with s being the vasodilatatory and activity dependent signal that increases the flow b_f, the veins deoxyhemoglobin content q, and the blood venous volume v, which are considered 1 at rest. This non-linear dynamic system can be described by

$$\left.\begin{array}{ll} \dot{s} = \varepsilon u - k_s s - k_f(b_f - 1) & \triangleq F_1 \\ \dot{b}_f = s & \triangleq F_2 \\ \dot{v} = \frac{1}{\tau}\left(b_f - v^{\frac{1}{\alpha}}\right) & \triangleq F_3 \\ \dot{q} = \frac{1}{\tau}\left(b_f \frac{1-(1-E_o)^{1/b_f}}{E_o} - v^{\left(\frac{1}{\alpha}-1\right)} q\right) & \triangleq F_4 \end{array}\right\} \triangleq F(x, \theta, u)$$

$$y = V_o\left[k_1\left(1-q\right) + k_2\left(1-\frac{q}{v}\right) + k_3\left(1-v\right)\right] = f(x, \theta, u) \qquad (1)$$

where $x = [x_1, x_2, x_3, x_4]^{\mathrm{T}} = [s, b_f, v, q]^{\mathrm{T}}$, E_o is the resting net oxygen extraction fraction by capillary bed, ε is the efficacy with which neuronal activity causes an increase in signal, $1/k_s$ and $1/k_f$ are time constants, τ is the mean transit time, and α is a stiffness exponent that specifies the flow-volume relationship of the venous balloon. The output of this model, $y(t)$, is the BOLD signal and represents a complex response controlled by different parameters, that range from the blood oxygenation, to the cerebral blood flow, and cerebral blood volume, and reflects the regional increase in metabolism due to enhancing of the neural activity. In the output equation, V_o is the resting blood volume fraction, and k_1, k_2, and k_3 are constants.

The response of the system described by Equation (1), with the parameters in **Table 1** and with initial state $x^{\mathrm{T}}(0) = [0\ 1\ 1\ 1]$, to a rectangular input signal, is depicted in **Figure 1**, for different integration periods.

The linear approximation of the model of the system leads to pronouncedly different responses, when compared to the non-linear system. An alternative to this, as described in sequel, is to consider a so-called bilinear model, which accurately mimics the non-linear behavior for sufficiently small integration periods.

Linearization and Discretization of the Model

The model described by Equation (1) is highly non-linear and parameter-dependent, thus barely allowing any systematic analysis of the associated expected behavior. Hence, to make the problem tractable from a mathematical point of view, the (bi)linearization of the HRF is considered in this paper. This approach allows the use of a widely spread framework for analysis, namely that of the linear time-varying systems. **Figure 1** shows that a close match of the HRF can be obtained by using a bilinear approximation (linear on the state, if the input is fixed, and linear on the input, if the state is fixed). Therefore, in this subsection, a (bi)linearization is derived that approximates the non-linear model locally and that is able to describe the state of the system at a given time, $x(kT_s)$, as a function of the state several sampling periods before, $x\left((k-N)T_s\right)$.

In particular, linearizing Equation (1) around $x(\cdot) = x^*$ and $u(\cdot) = 0$, i.e., writing the associated Taylor expansion and truncating it at the linear term, one obtains (omitting the time-dependence of the variables, for the sake of readability):

$$\dot{x} \approx F(x^*, \theta, 0) + \left.\frac{\partial F(x, \theta, u)}{\partial x}\right|_{x^*,\theta,0}(x - x^*) + \sum_i u_i \left(\left.\frac{\partial^2 F(x, \theta, u)}{\partial x \partial u_i}\right|_{x^*,\theta,0}(x - x^*) + \left.\frac{\partial F(x, \theta, u)}{\partial u_i}\right|_{x^*,\theta,0}\right),$$

where

$$\frac{\partial F}{\partial x} = \begin{bmatrix} -k_s & -k_f & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & \frac{1}{\tau} & -\frac{x_3^{\left(\frac{1}{\alpha}-1\right)}}{\alpha\tau} & 0 \\ 0 & \frac{\partial F_4}{\partial x_2} & \frac{\partial F_4}{\partial x_3} & -\frac{x_3^{\left(\frac{1}{\alpha}-1\right)}}{\tau} \end{bmatrix}, \qquad (2)$$

TABLE 1 | Parameters for the non-linear model described by Equation (1).

Parameter	ε	$k_s[s^{-1}]$	$k_f[s^{-1}]$	τ[s]	α	E_o	V_o	k_1	k_2	k_3
Value	0.065	0.550	0.410	1.280	0.880	0.920	4.88	$7E_0$	2.0	$2E_0 - 0.2$

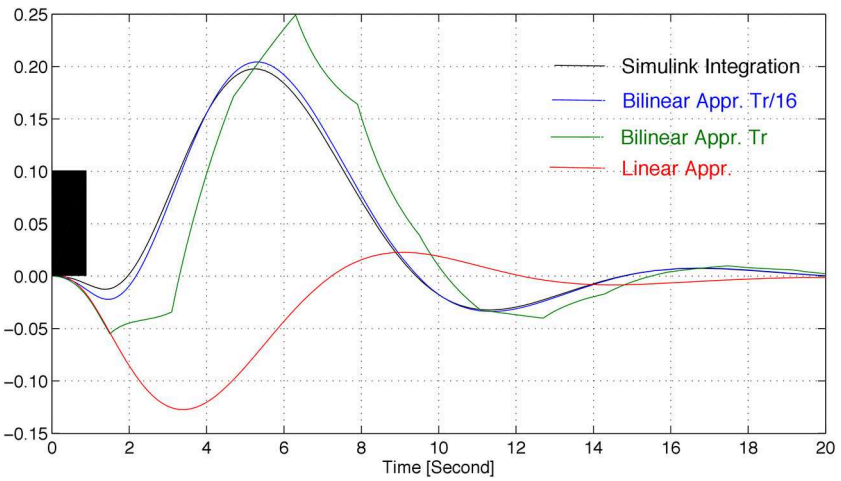


FIGURE 1 | Approximations of the response of model Equation (1) to a rectangular input signal (in black), for the parameters of Table 1, where *TR* denotes the repetition time.

$$\frac{\partial F_4}{x_2} = \frac{1}{\tau}\left[\frac{1-(1-E_o)^{\frac{1}{x_2}}}{E_o} + \frac{\log(1-E_o)(1-E_o)^{\frac{1}{x_2}}}{E_o x_2}\right],$$
$$\frac{\partial F_4}{x_3} = \frac{1}{\tau}\left(1-\frac{1}{\alpha}\right)x_3^{\left(\frac{1}{\alpha}-2\right)}x_4.$$

and with output equation described by

$$\frac{\partial y}{\partial x} = \begin{bmatrix} 0, & 0, & -k_3 V_o + k_2 V_o q v^{-2}, & -k_1 V_o - k_2 V_o v^{-1} \end{bmatrix}.$$

Moreover, given that F_1 depends linearly upon u, we have that $\frac{\partial^2 F}{\partial x \partial u_i} = 0$.

Using the transformation proposed in Friston et al. (2000), one finally obtains the following dynamics:

$$\dot{\tilde{x}} = A\tilde{x} + \sum_i u_i E_i \tilde{x}, \tag{3}$$

where $\tilde{x} = \begin{bmatrix} 1 & x \end{bmatrix}^{\mathrm{T}}$,

$$A \triangleq \begin{bmatrix} 0 & 0 \\ \left(F(x^*,\theta,u) - \frac{\partial F(x^*,\theta,u)}{\partial x}x^*\right) & \frac{\partial F(x^*,\theta,0)}{\partial x} \end{bmatrix},$$
$$E_i \triangleq \begin{bmatrix} 0 & 0 \\ \frac{\partial F(x^*,\theta,0)}{\partial u_i} & 0 \end{bmatrix},$$

and $\frac{\partial F(x^*,\theta,0)}{\partial u_i} = \begin{bmatrix} \varepsilon & 0 & 0 & 0 \end{bmatrix}^{\mathrm{T}}$.

Uncertain Dynamic Model Description

It should be noticed that the dynamics in Equation (3) are bilinear in the state and input variables. This non-linear term hinders the distinguishability analysis proposed in Rosa and Silvestre (2011) and, thus, a more suitable description is derived in herein.

For the sake of simplicity, we start by redefining $x(t) \triangleq \tilde{x}(t)$ and $x^*(t) \triangleq [1 \ \ (x^*(t))^{\mathrm{T}}]^{\mathrm{T}}$. It was previously shown that the continuous-time dynamic model of the HRF, for a single input, can be approximated by

$$\begin{cases} \dot{x}(t) = (A(t) + B_o(t)u(t) + \Delta(t)B_1(t)u(t))x(t), \\ \qquad\qquad\qquad\qquad\qquad\qquad x(0) = x^*(0), \\ y(t) = h(x(t)), \end{cases} \tag{4}$$

with $t \geq 0$, and where $\Delta : \mathrm{R}^+ \to \mathrm{R}$ was also included to represent an input uncertainty subject to $|\Delta(t)| \leq 1$ for all $t \geq 0$, and where $B_o = E_1$. This input uncertainty can be seen as a surrogate for uncertainty in the stimulation signal. The initial state is denoted by $x(0) \in \mathrm{R}^n$, and n is the number of states of the system. Moreover, we assume that

$$B_1(t) = \eta B_o(t),$$

with known $\eta \in \mathrm{R}$. We also define $B(t) = B_o(t) + \Delta(t)B_1(t)$.

To proceed with the derivation of a discrete-time description of the HRF model in Equation (4), for a given sampling period, T_s, the following assumptions are posed:

Assumption 1: The input signal, $u(\cdot)$, is constant during sampling periods, i.e., $u(t) = u(kT_s)$, for all $t \in [kT_s, \ (k+1)T_s[$.

Assumption 2: The input uncertainty, $\Delta(\cdot)$, is constant during sampling periods, i.e., $\Delta(t) = \Delta(kT_s)$, for all $t \in [kT_s, \ (k+1)T_s[$.

Assumption 3: The maps $A(\cdot)$, $B_o(\cdot)$, and $B_1(\cdot)$, are constant during sampling periods, i.e., $A(t) = A(kT_s)$, $B_o(t) = B_o(kT_s)$, and $B_1(t) = B_1(kT_s)$, for all $t \in [kT_s,\ (k+1)T_s[$.

Under these assumptions, the system in Equation (4) can be rewritten as

$$\begin{cases} \dot{x}(t) &= \tilde{A}\left(k, \Delta(k)\right) x(t), \quad x(0) = x^*(0), \\ y(t) &= g(\tilde{x}(t)), \end{cases} \tag{5}$$

for $\tilde{x}(t) \in [kT_s, (k+1)T_s]$, and where

$$\tilde{A}\left(k, \Delta(k)\right) = A_o(k) + \Delta(kT_s)A_1(k),$$

with

$$A_o(k) = A(kT_s) + B_o(kT_s)u(kT_s),$$

and

$$A_1(k) = B_1(kT_s)u(kT_s).$$

In the sequel, we will abbreviate $x(k) = x(kT_s)$, for the sake of simplicity. We are now in conditions of stating the following proposition:

Proposition 1: Define

$$I^* = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{bmatrix},$$

and

$$\phi(k) = V(k)\Lambda^*(k)V^{-1}(k)e^{A(kT_s)T_s} - V(k)\Lambda^*(k)V^{-1}(k) - I^*,$$

where $V(k)\Lambda(k)V^{-1}(k) = A(kT_s)T_s$ is the spectral decomposition of $A(kT_s)T_s$ with $\Lambda(k)$ diagonal and $\Lambda_{11}(k) = 0$, and

$$\Lambda^*_{ij}(k) = \begin{cases} \frac{1}{\Lambda_{ij}(k)}, & \text{if i = j and } \Lambda_{ij} \neq 0, \\ 0, & \text{otherwise.} \end{cases}$$

Furthermore, let

$$\begin{aligned} G_o(k) &= e^{A(k)} + B_o(k)u(k) + \phi(k)B_o(k)u(k) \text{ and} \\ G_1(k) &= B_1(k)u(k) + \phi(k)B_1(k)u(k). \end{aligned}$$

Then, the system in Equation (5) is described by

$$\begin{cases} x(k+1) &= G\left(k, \Delta(k)\right) x(k), \quad x(0) = x^*(0), \\ y(k) &= h(x(k)), \end{cases} \tag{6}$$

where

$$G\left(k, \Delta(k)\right) = G_o(k) + \Delta(k)G_1(k),$$

and for $x(k) = x(kT_s)$. *Proof:* See Appendix A in Supplementary Material.

Notice that Equation (6), with $G\left(k, \Delta(k)\right) = G_o(k) + \Delta(k)G_1(k)$, associated with the linearization of the output map, g, is a full description of the HRF dynamics by means of a linear model with known matrices, $G_o(k)$ and $G_1(k)$, and an uncertain parameter, $\Delta(k)$. This description, however, is bilinear in the state, $x(k)$, and model uncertainty, $\Delta(k)$. This bilinear relationship is tainted once we describe the state $x(k+1)$ as a function of $x(k-1)$. Nevertheless, notice that

$\left(G_o(k+1) + \Delta G_1(k+1)\right)\left(G_o(k) + \Delta G_1(k)\right) = G_o(k+1)G_o(k) + \Delta\left(G_1(k+1)G_o(k) + G_o(k+1)G_1(k)\right)$, since $G_1(k+1)G_1(k) = 0$ and where, for the time being, we considered that Δ is constant (but unknown), i.e., $\Delta(k) = \Delta$ for all k. To see this, notice that

$$\begin{aligned} G_1(k+1)G_1(k) &= (B_1(k+1) + \phi(k+1)B_1(k+1))(B_1(k) \\ &\quad +\phi(k)B_1(k)) \\ &= \underbrace{B_1(k+1)B_1(k)}_{=0} + B_1(k+1)\phi(k)B_1(k) \\ &\quad +\phi(k+1)\underbrace{B_1(k+1)B_1(k)}_{=0} + \\ &\quad +\phi(k+1)B_1(k+1)\phi(k)B_1(k), \end{aligned}$$

and that $B_1(k+1)\phi(k)B_1(k) = 0$, due to the fact that the first row of ϕ is zero, and that all but the first column of B_1 are also zero.

By proceeding in a similar manner, we conclude that

$$\begin{aligned} &\left(G_o(k+m) + \Delta G_1(k+m)\right) \cdots \left(G_o(k) + \Delta G_1(k)\right) \\ &\qquad = \Psi_o(k+m) + \Delta\Psi_1(k+m), \end{aligned}$$

where

$$\Psi_o(k+m) = G_o(k+m) \cdots G_o(k),$$

and

$$\begin{cases} \Psi_1(k) &= G_1(k), \\ \Psi_1(k+m) &= G_o(k+m)\Psi_1(k+m-1) \\ &\quad +G_1(k+m)\Psi_0(k+m-1). \end{cases}$$

Hence, the state $x(k+m+1)$ can be written as

$$x(k+m+1) = \left(\Psi_o(k+m) + \Psi_1(k+m)\Delta\right) x\left(k\right)$$

Furthermore, the non-linear output Equation of (1) can be linearized as

$$y(x) = y(x^*) + \left.\frac{\partial y}{\partial x}\right|_{x^*} (x - x^*), \tag{8}$$

which, in turn, can alternatively written as:

$$z = y(x) - y(x^*) + \left.\frac{\partial y}{\partial x}\right|_{x^*} x^* = \left.\frac{\partial y}{\partial x}\right|_{x^*} x. \tag{9}$$

where $z(t)$ can be seen as the measurement for the linear time-varying system obtained by the linearization of Equation (1).

Absolutely Distinguishable Systems

The problem of indistinguishability typically arises from large amplitudes of the measurement noise, small intensity of the input excitation signals, model uncertainty, and uncertain initial conditions. In particular, if the Signal-to-Noise Ratio (SNR) of the measurements is not sufficiently large, one may be able to explain the observed variables by using more than a single dynamic model, from the set of *eligible* models. A similar conclusion applies if the intensity of the input signal is not sufficient to excite the dynamics of the system.

This section will therefore propose a methodology to systematically derive conditions that guarantee the distinguishability of a set of dynamic models, regardless of the noise sequences and initial states.

Systems with Uncertain Initial State

We start by analyzing the case where the dynamics of the system are known, although the initial state is uncertain and the measured variables are corrupted by bounded noise. Using Equation (8), we have that

$$y(k) = \underbrace{y(x^*(k)) - C(k)x^*(k)}_{\bar{y}(k)} + C(k)x(k) + n(k), \qquad (10)$$

where

$$C(k) = \left.\frac{\partial y}{\partial x}\right|_{x^*(k)},$$

and where $n(k)$ is the measurement noise. Consider that a given input sequence, $u(0), \cdots, u(N)$, feeds the inputs of systems S_A and S_B, respectively described by

$$S_A : \begin{cases} x_A(k+1) &= G_A(k)x_A(k), \\ y_A(k) &= \bar{y}_A(k) + C_A(k)x_A(k) + n_A(k), \end{cases}$$

$$S_B : \begin{cases} x_B(k+1) &= G_B(k)x_B(k), \\ y_B(k) &= \bar{y}_B(k) + C_B(k)x_B(k) + n_B(k), \end{cases}$$

where y_A and y_B are defined as in Equation (10), and $|n_A(k)| \leq \frac{\bar{n}}{2}$, $|n_B(k)| \leq \frac{\bar{n}}{2}$. Moreover, we assume that $x_A(0) \in X_o$ and $x_B(0) \in X_o$, where $X_o \in \mathrm{R}^n$ is a convex polytope. Let $\phi_i = [n_i^{\mathrm{T}}, u_i^{\mathrm{T}}]^{\mathrm{T}}$, denote the measurement noise, $n_i \in W \subseteq \mathrm{R}^{n_n}$, and input signals, $u_i \in U \subseteq \mathrm{R}^{n_u}$, at time instant i.

Definition 1: Systems S_A and S_B are said *absolutely* (X_0, U, W)*-input distinguishable in* N *sampling times* if, for any non-zero

$$\big(x_A(0), x_B(0), \phi_1, \phi_2, \cdots, \phi_N\big) \in X_o \times X_o \times \overbrace{\Phi \times \cdots \times \Phi}^{N\text{times}},$$

where $\phi_i \in W \times U =: \Phi \subseteq \mathrm{R}^{n_u+n_d}$ for $i = 0, 1, \cdots, N$, there exists $k \in \{0, 1, \cdots, N\}$ such that

$$y_A(k) \neq y_B(k).$$

Moreover, two systems are said *absolutely* (X_o, U, W)*-input distinguishable* if there exists $N \geq 0$ such that they are absolutely (X_o, U, W)-input distinguishable in N sampling times.

Let $U = (u(0), u(1), \cdots, u(N))$ and

$$W = \left\{ \big(n(0), n(1), \cdots, n(N)\big) : \underset{0 \leq k \leq N}{\forall} |n(k)| \leq \frac{\bar{n}}{2} \right\}.$$

The following proposition can be used to state whether a pair of systems is distinguishable or not.

Proposition 2: Systems S_A and S_B are absolutely (X_o, U, W)-input distinguishable in N sampling times if and only if a solution to the following linear problem does not exist:

$$\begin{bmatrix} C_A(0) & -C_B(0) \\ -C_A(0) & C_B(0) \\ C_A(1)G_A(0) & -C_B(0)G_B(0) \\ -C_A(1)G_A(0) & C_B(0)G_B(0) \\ \vdots & \vdots \\ C_A(N)G_A(N-1)\cdots G_A(0) & -C_B(0)G_B(N-1)\cdots G_B(0) \\ -C_A(N)G_A(N-1)\cdots G_A(0) & C_B(0)G_B(N-1)\cdots G_B(0) \\ M_o & 0 \\ 0 & M_o \end{bmatrix} \begin{bmatrix} x_A(0) \\ x_B(0) \end{bmatrix} \leq \begin{bmatrix} \bar{n} - \bar{y}_A(0) + \bar{y}_B(0) \\ \bar{n} + \bar{y}_A(0) - \bar{y}_B(0) \\ \bar{n} - \bar{y}_A(1) + \bar{y}_B(1) \\ \bar{n} + \bar{y}_A(1) - \bar{y}_B(1) \\ \vdots \\ \bar{n} - \bar{y}_A(N) + \bar{y}_B(N) \\ \bar{n} + \bar{y}_A(N) - \bar{y}_B(N) \\ m_o \\ m_o \end{bmatrix}, \qquad (11)$$

where X_o is defined so that $x \in X_o \Leftrightarrow M_o x \leq m_o$, which can be written as $X_o = \mathrm{Set}(M_o, m_o)$.

Proof: See Appendix B in Supplementary Material.

Systems with Uncertain Model

We now consider the case where the system dynamics are uncertain and described by

$$S_A : \begin{cases} x_A(k+1) &= \big(G_o^A(k) + \Delta_A G_1^A(k)\big)\, x_A(k), \\ y_A(k) &= \bar{y}_A(k) + C_A(k)x_A(k) + n_A(k), \end{cases}$$

$$S_B : \begin{cases} x_B(k+1) &= \big(G_o^B(k) + \Delta_B G_1^B(k)\big)\, x_B(k), \\ y_B(k) &= \bar{y}_B(k) + C_B(k)x_B(k) + n_B(k), \end{cases}$$

where y_A and y_B are defined as in Equation (10), and $|n_A(k)| \leq \frac{\bar{n}}{2}$, $|n_B(k)| \leq \frac{\bar{n}}{2}$. We also assume that $|\Delta_A| \leq 1$ and $|\Delta_B| \leq 1$. Moreover, for this case we assume that X_o is a singleton, thus removing the uncertainty in the initial state. In this case, S_A and S_B denote families of systems, due to the uncertainties Δ_A and Δ_B. Therefore, the introduction of the following definition is required.

Definition 2: The families of systems S_A and S_B are said absolutely (X_o, U, W)-input distinguishable in N sampling times if, for any pair of realizations $(S_1, S_2) \in S_A \times S_B$, the systems S_1 and S_2 are absolutely (X_o, U, W)-input distinguishable in N sampling times.

Hence, we are now in condition to state the following proposition:

Proposition 3: The families of systems S_A and S_B are absolutely (X_o, U, W)-input distinguishable in N sampling times if and only if there does not exist a solution to the following linear problem:

$$\Theta_N \begin{bmatrix} \Delta_A \\ \Delta_B \end{bmatrix} \leq \theta_N, \tag{12}$$

where

$$\Theta_N = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ C_A(1)\Psi_1^A(0)x_A(0) & -C_B(1)\Psi_1^B(0)x_B(0) \\ -C_A(1)\Psi_1^A(0)x_A(0) & C_B(1)\Psi_1^B(0)x_B(0) \\ \vdots & \vdots \\ C_A(N)\Psi_1^A(N-1)x_A(0) & -C_B(N)\Psi_1^B(N-1)x_B(0) \\ -C_A(N)\Psi_1^A(N-1)x_A(0) & C_B(N)\Psi_1^B(N-1)x_B(0) \end{bmatrix}$$

and

$$\theta_N = \begin{bmatrix} \bar{n} - \bar{y}_A(0) + \bar{y}_B(0) - C_A(0)x_A(0) + C_B(0)x_B(0) \\ \bar{n} + \bar{y}_A(0) - \bar{y}_B(0) + C_A(0)x_A(0) - C_B(0)x_B(0) \\ \bar{n} - \bar{y}_A(1) + \bar{y}_B(1) - C_A(1)\Psi_o^A(0)x_A(0) + \Psi_o^B(0)x_B(0) \\ \bar{n} + \bar{y}_A(1) - \bar{y}_B(1) + C_A(1)\Psi_o^A(0)x_A(0) - \Psi_o^B(0)x_B(0) \\ \vdots \\ \bar{n} - \bar{y}_A(N) + \bar{y}_B(N) - C_A(N)\Psi_o^A(N-1)x_A(0) + \Psi_o^B(N-1)x_B(0) \\ \bar{n} + \bar{y}_A(N) - \bar{y}_B(N) + C_A(N)\Psi_o^A(N-1)x_A(0) - \Psi_o^B(N-1)x_B(0) \end{bmatrix}$$

Proof: See Appendix C in Supplementary Material.

Figure 2A depicts the impulse and step responses of the HRF model with the parameters of **Table 1**, with an uncertainty of 10% in the input signal. It should be noticed that this type of uncertainty mainly affects the amplitude of the responses of the system. Thus, the rise- and fall-times are not significantly influenced by small variations on the amplitude of the input signal.

Systems with Uncertain Input Time-Delays

In this subsection, a strategy to model uncertain input time-delays is developed. The approach presented in the sequel amounts for rewriting these uncertain input time-delays as model uncertainty.

Consider that the input signal, at sampling time k, is given by

$$u(k) = \tilde{u}(k - k_d),$$

where k_d is an integer (the uncertain delay) satisfying $|k_d| \leq \bar{k}_d$, with known $\bar{k}_d$. The value of $\tilde{u}(k)$, for each $k \geq 0$, is also assumed known and bounded. Thus, we have, for each $k \geq 0$,

$$\underline{u}(k) \leq u(k) \leq \bar{u}(k), \tag{13}$$

where $\bar{u}(k) = \max_{|m| \leq \bar{k}_d} \tilde{u}(k-m)$ and $\underline{u}(k) = \min_{|m| \leq \bar{k}_d} \tilde{u}(k-m)$. Therefore, Equation (13) can be rewritten as

$$u(k) = u_o(k) + \Delta_u(k)u_1(k),$$

where $|\Delta_u(k)| \leq 1$, $u_o(k) = \frac{\bar{u}(k)+\underline{u}(k)}{2}$, and $u_1(k) = \frac{\bar{u}(k)-\underline{u}(k)}{2}$.

Hence, unknown but bounded time-delays on the input can be treated as uncertainty on the B matrix. The impulse and step responses of the HRF model with the parameters of **Table 1**, with an uncertain input time-delay, k_d, bounded by $|k_d| \leq 3$, are depicted in **Figure 2B**. As seen in the figure, the uncertainty in the input time-delay enlarges the uncertainty in the rise- and fall-times of the output.

Systems with Uncertain Model and Input Time-Delays

For the sake of completeness, in this subsection we analyze the effects of simultaneous uncertainty on the model and on the input time-delays. The results for this scenario are depicted in **Figure 2C**. As expected, the uncertainty on the model chiefly affects the amplitude of the responses, while the uncertainty on the input time-delay changes the corresponding rise- and fall-times.

Systems with Uncertain Model and Uncertain Initial State

We now consider the case where both the system dynamics and the initial state are uncertain. The problem is set to that of concluding whether the following two families of systems are distinguishable:

$$S_A : \begin{cases} x_A(k+1) &= (G_o^A(k) + \Delta_A(k)G_1^A(k))x_A(k), \\ y_A(k) &= \bar{y}_A(k) + C_A(k)x_A(k) + n_A(k), \end{cases}$$

$$S_B : \begin{cases} x_B(k+1) &= (G_o^B(k) + \Delta_B(k)G_1^B(k))x_B(k), \\ y_B(k) &= \bar{y}_B(k) + C_B(k)x_B(k) + n_B(k), \end{cases}$$

where y_A and y_B are defined as in Equation (10), and $|n_A(k)| \leq \frac{\bar{n}}{2}$, $|n_B(k)| \leq \frac{\bar{n}}{2}$. We also assume that $|\Delta_A(k)| \leq 1$ and $|\Delta_B(k) \leq 1$. Moreover, for this case we assume that X_o is a convex polytope.

Proposition 4: Let $e_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \end{bmatrix}^{\mathrm{T}}$. The families of systems S_A and S_B are absolutely (X_o, U, W)-input distinguishable in N sampling times if and only if there does not exist a solution to the following linear problem:

$$\underset{k\in\{0,1,\cdots,N\}}{\forall} : \begin{cases} x_A(0), x_B(0) \in X_o, \\ C_A(k)x_A(k) - C_B(k)x_B(k) \leq \bar{n} - \bar{y}_A(k) + \bar{y}_B(k), \\ -C_A(k)x_A(k) + C_B(k)x_B(k) \leq \bar{n} + \bar{y}_A(k) - \bar{y}_B(k), \\ x_A(k+1) - G_o^A(k)x_A(k) - G_1^A(k)e_1 z_A(k) = 0, \\ x_B(k+1) - G_o^B(k)x_B(k) - G_1^B(k)e_1 z_B(k) = 0, \\ -e_1^{\mathrm{T}}x_A(k) \leq z_A(k) \leq e_1^{\mathrm{T}}x_A(k), \\ -e_1^{\mathrm{T}}x_B(k) \leq z_B(k) \leq e_1^{\mathrm{T}}x_B(k), \end{cases}$$

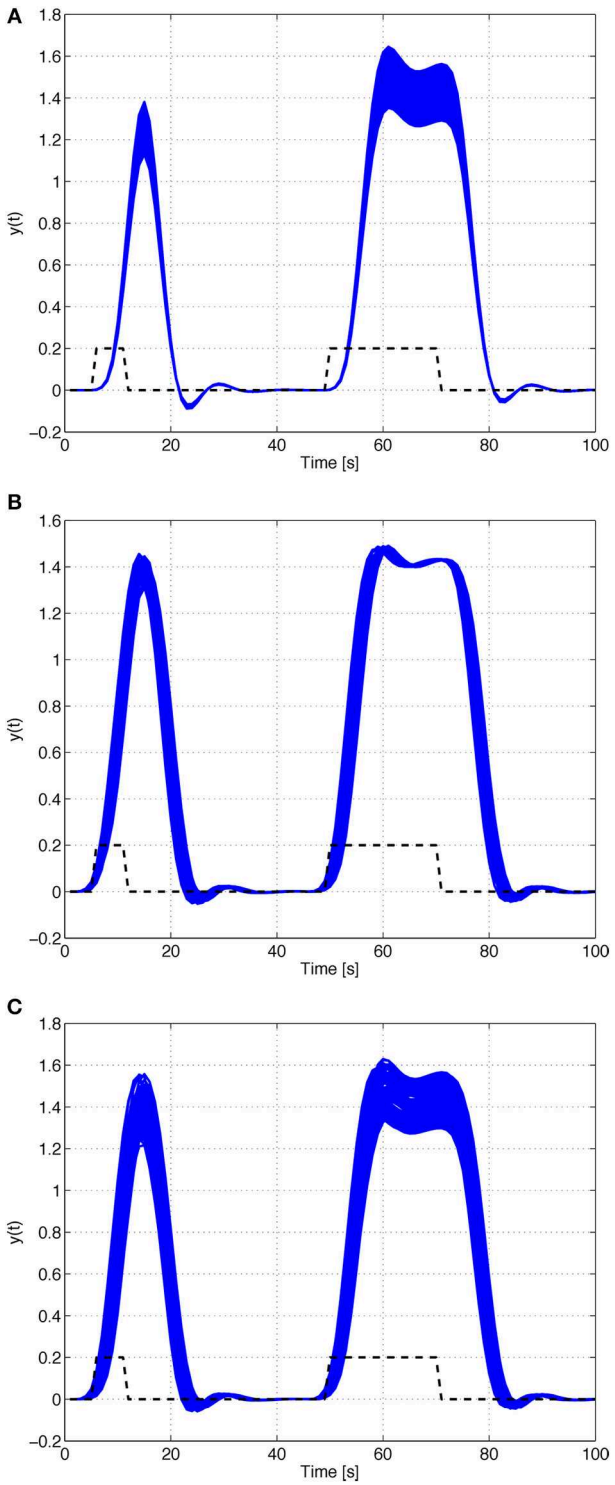


FIGURE 2 | Impulse and step responses of the HRF model with the parameters of Table 1. (A) 10% uncertainty on the input signal. **(B)** Uncertain input time-delay. **(C)** Uncertain model and input time-delay.

where the unknown variables are $x_A(0), \cdots x_A(N)$, $x_B(0), \cdots x_B(N)$, $z_A(0), \cdots z_A(N-1)$, and $z_B(0), \cdots z_B(N-1)$.

Proof: See Appendix D in Supplementary Material.

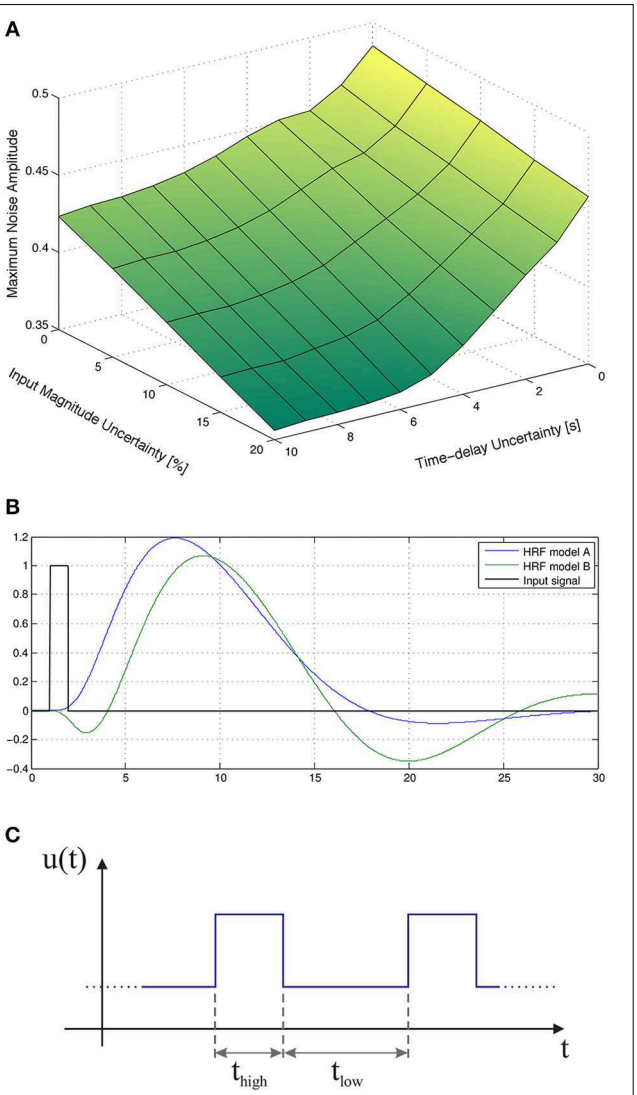


FIGURE 3 | Maximum amplitude of the measurement noise that guarantees the absolute distinguishability of two particular families of models. (A) As a function of the uncertainty on the input signal and on the corresponding time-delay. **(B)** As a function of the uncertainty on the magnitude and time-delay of the input signal, for a deterministic input signal. **(C)** As a function of the uncertainty on the magnitude and time-delay of the input signal, for a stochastic input signal with mean t_{high} and t_{low} of 12 s obtained from a uniform distribution of width 12 s.

Figure 3A depicts the maximum amplitude of the measurements noise that guarantees the absolute distinguishability of two particular families of HRF models, as a function of the uncertainty on the input signal and on the corresponding time-delay. As expected, the maximum level of sensor noise such that the two families of models are absolutely distinguishable, decreases with both types of uncertainty.

Pre-Processing of fMRI Time Series

We stress that the assumption that the additive noise in the measured signal is bounded is not conservative in practice, since outliers and other *unboundedness* behaviors can, in general, be

tackled during pre-processing, i.e., before performing the main analysis of the signals. This can be done, in particular, by low-pass filtering the signal, so that high-frequency noise is significantly attenuated.

Additionally, the following pre-processing steps are commonly applied to fMRI time series data before submitting them to statistical analysis (Jezzard et al., 2001): (i) normalization of the whole 4D fMRI dataset by scaling each volume by a single (common) scaling factor, so that subsequent analyses are valid; (ii) motion correction by alignment of all fMRI volumes to a reference volume in the time series, usually performed by applying rigid-body transformations, in order to reduce the effect of subject head motion during the experiment; and (iii) high-pass temporal filtering, usually using a local fit of a straight line (Gaussian-weighted within the line to give a smooth response), in order to remove low-frequency artifacts such as signal drifts or physiological fluctuations.

Results

In this section, we study the influence of the choice of the input signal on the distinguishability of a set of HRF models. A methodology to optimize the fMRI experimental design that takes advantage of this knowledge is also presented.

Throughout the remainder of this paper, we are going to refer to the families of HRF models *A* and *B*, described by the dynamics in Equation (1), with the physiologically plausible parameters presented in **Table 2**. Model family *B* displays a pronounced undershoot and the presence of an initial dip, in stark contrast to model family *A*.

The response of the nominal HRF models, for the parameter configurations of **Table 2**, with initial state $x^{\mathrm{T}}(0) = \begin{bmatrix} 0 & 1 & 1 & 1 \end{bmatrix}$, to a rectangular input signal of duration 1 s and unit magnitude, is depicted in **Figure 4**.

In general, the input signal is composed of a series of rectangular pulses (events) of duration t_{high} alternating with baseline periods of duration t_{low}, with a total duration of 200 s (see **Figure 5**).

In order to illustrate the characteristic behavior of HRF model family *A*, their responses to rectangular input signals of duration 5 and 20 s and unit magnitude, with an uncertain input time-delay, k_d, bounded by $|k_d| \leq 3$ s, and input uncertainty of 10%, are depicted in **Figure 6**. The uncertainty on the input time-delay enlarges the uncertainty in the rise- and fall-times of the output, while the uncertainty in the input mainly affects the amplitude of the responses of the system.

Figure 3B depicts the maximum amplitude of the measurements noise that guarantees the absolute distinguishability of the families of models *A* and *B*, for an input signal with $t_{\mathrm{low}} = 12$ s and $t_{\mathrm{high}} = 12$ s, as a function of the uncertainty on the magnitude of the input signal and on the corresponding time-delay. As expected, the maximum level of measurement noise such that the families of models *A* and *B* are absolutely distinguishable decreases with both types of uncertainty.

Furthermore, we considered a stochastic input signal, composed of a series of rectangular pulses with mean duration of $E(t_{\mathrm{high}}) = 12$ s, and mean baseline period of $E(t_{\mathrm{low}}) = 12$ s drawn from a uniform distribution of width 12 s. According to the results in the literature (see, for instance, Josephs et al., 1997; Miezin et al., 2000), we observe that, by performing random small variations on t_{high} and t_{low}, alternative trajectories of the non-linear model Equation (1) are exploited, which in turn improves the identifiability of the models, as depicted in **Figure 3C**.

We now analyze the effect of different experimental designs on the distinguishability of the families of models at hand. At this point, our goal is to find the combination of values of t_{low} and t_{high} such that the absolute distinguishability of two or more

TABLE 2 | Parameters for the families of non-linear models.

HRF	k_s [s^{-1}]	k_f [s^{-1}]	τ [s]	α	E_o
A	0.400	0.100	2.080	0.320	0.340
B	0.220	0.110	2.180	0.320	0.985

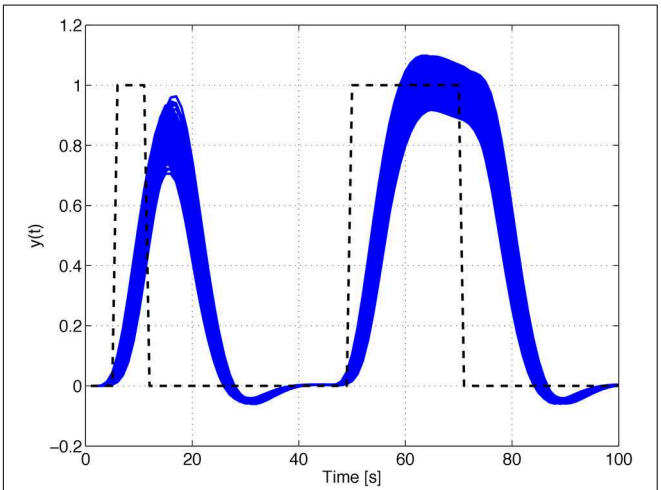


FIGURE 4 | Time responses of the nominal models of families *A* and *B*.

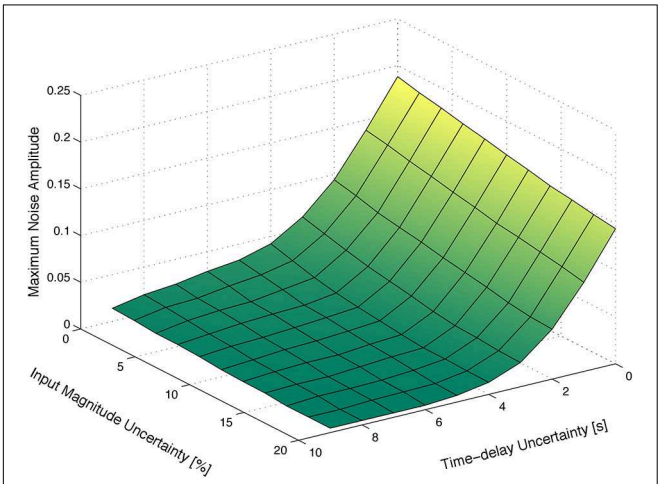


FIGURE 5 | Input signal adjustable parameters.

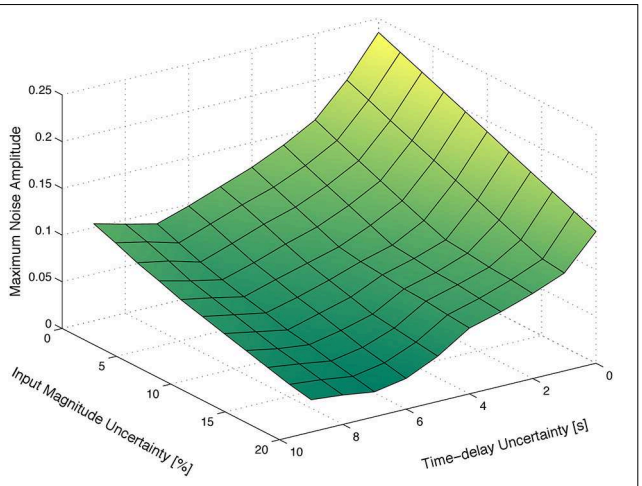


FIGURE 6 | Rectangular input responses of family *A* with uncertain input magnitude and input time-delay.

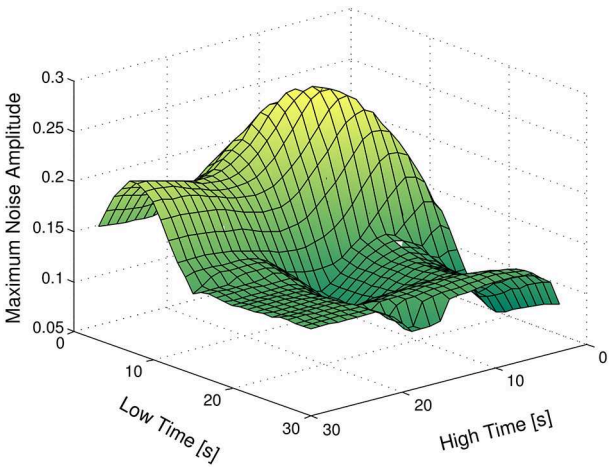


FIGURE 7 | Maximum amplitude of the measurement noise guaranteeing the absolute distinguishability of the families *A* and *B*, as a function of t_{low} and t_{high}.

families of models is guaranteed for the highest upper bound on the amplitude of the measurement noise. We denote this optimal combination of values by (t^*_{low}, t^*_{high}). The advantage of using an input signal with parameters (t^*_{low}, t^*_{high}) obviously stems from the fact that we can allow for the highest amplitude on the measurement noise, while guaranteeing the distinguishability of the families.

Figure 7 depicts the results obtained, considering no time-delay or magnitude uncertainty. As expected, input signals with very small values of t_{high} and large values of t_{low} do not have the power required to significantly stimulate the system. On the other hand, input signals with very small values of t_{high} and t_{low} are faster than the dynamics of the system, and hence do not produce remarkable changes in the output of the plant. As a final remark, the optimum value for t_{low} and t_{high} is 10 s, i.e., $t^*_{low} = 10$ s and $t^*_{high} = 10$ s.

Discussion

We have addressed the problem of the distinguishability of HRF models in the analysis of fMRI data of brain activation, based on the biophysically informed description of the HRF as a non-linear time-invariant input-state-output dynamic system. We first introduced the concept of absolutely input-distinguishable systems and then showed that the distinguishability between two HRF models, and hence system identification, is significantly affected by the external input (stimulus/task) signals. In particular, the uncertainty in the input time-delays and its magnitude may adversely affect model identification, by reducing the maximum noise level below which model distinguishability is guaranteed. We then applied the concept of absolutely input-distinguishable systems to the development of a methodology for the assessment of the HRF estimation efficiency of fMRI experimental designs, through the maximization of the distinguishability level among a set of physiologically plausible HRF models.

The main contribution of this paper is therefore 2-fold. On the one hand, we show that the distinguishability of two HRF models depends on the level of the measurement noise as well as on the characteristics of the input signal. On the other hand, we develop a methodology to optimize fMRI experimental designs for HRF estimation, which maximizes the allowable noise amplitude that does not impair the distinguishability of a set of a priori admissible dynamic systems.

In this paper, it is assumed that the system inputs can be selected or, at least, measured. This assumption is verified in a straightforward manner when external inputs are present, such as sensory stimuli or cognitive tasks. Although no explicit external inputs exist in resting-state fMRI acquisitions, it has been observed that discrete neuronal events do occur (Deco and Jirsa, 2012). Most interestingly, it has been recently suggested that such events can be identified as peaks of relatively large BOLD signal amplitude (Tagliazucchi et al., 2011), and resting-state fMRI data can then be seen as "spontaneous event-related" data (Wu et al., 2013).

Significance of HRF Estimation

The importance of estimating the HRF in fMRI experiments is based on the extensively observed variability of its shape and dynamics across brain regions, conditions, subjects, and populations, with critical consequences in the analysis of fMRI data. In fact, one direct consequence of HRF variability is that the deviation of the real HRF from the pre-specified HRF leads to a poorer model of the observed BOLD signal and hence reduces the sensitivity to detect BOLD changes (Handwerker et al., 2004). Another consequence is the potential detection of a group effect due to a systematic HRF difference, which would then be incorrectly interpreted as a neuronal effect. Moreover, when attempting to infer causality within brain networks from BOLD data, differences in HRF latency across brain regions can potentially confound the directionality of information flow (David et al., 2008; Smith et al., 2011; Murta

et al., 2012; Jorge et al., 2014). On the other hand, HRF variability may be an object of interest on its own, potentially reflecting physiological changes associated with the effects of drugs, aging or pathology, for example (Iadecola, 2004). Additionally, there is a growing interest in studying, not only the amplitude of BOLD activation, but also its dynamics, namely its latency and duration, which are reflected in the HRF (Bellgowan et al., 2003). In these cases, it would be desirable to estimate the actual HRF model underlying the BOLD signal measured in each voxel, experiment, subject or population, or otherwise account for its variability.

Despite the acknowledged need for modeling the HRF underlying fMRI BOLD data, and although different approaches have been continuously proposed in the literature for this purpose, our ability to understand HRF variability remains poor (Handwerker et al., 2012). Critically, most studies have focused on parameterized HRF models in a linear framework, while the estimation of physiologically plausible non-linear HRF models with direct biophysical interpretability has been very limited. In particular, no previous study has investigated the optimal fMRI experimental design for the estimation of such biophysical HRF models. We believe that our work therefore makes an important contribution for understanding how a biophysically informed model of the HRF may be inferred from fMRI data, as a function of experimental design and measurement noise.

Biophysically Informed HRF Modeling

Using a biophysically informed model of the HRF not only allows for a physiologically plausible interpretation of the results, but it also more accurately explains empirical BOLD data, particularly concerning commonly observed non-linearities. Importantly, in contrast to parameterized HRF models, biophysical models described by dynamic systems can account for the detailed dynamics of BOLD responses through a reduced number of parameters, while constraining it to be physiologically plausible. For example, the post-stimulus undershoot and the initial dip are two features of observed BOLD responses that naturally emerge from this dynamic system under slightly different combinations of a limited number of parameters. Although using such dynamic systems represents an additional computational effort compared to the more straightforward linear methods, this may nevertheless become the chosen approach in studies where a detailed characterization of the BOLD temporal dynamics is desirable. In particular, the combination of EEG with fMRI may greatly benefit from such approaches (Riera et al., 2005). On the other hand, important complementary information may be gained for HRF model estimation by combining BOLD recordings with the acquisition of blood flow data using Arterial Spin Labeling (ASL) or near-infrared spectroscopy (NIRS) (Huppert et al., 2006). Despite the potential advantages of such a biophysically informed dynamic system approach to fMRI data analysis, only a few studies have been dedicated to the associated problem of system identification/model estimation (Friston, 2002; Riera et al., 2004). Our study therefore makes a significant contribution to this limited body of literature, by introducing the concept of input-distinguishability of HRF models in order to inform model selection in this context.

Optimization of the Experimental Design

Previous studies systematically assessing the quality of fMRI experimental designs have again been focused on parameterized HRF models within a linear framework (Dale, 1999; Liu et al., 2001). They found that optimal estimation efficiency is obtained at the cost of reduced detection power by employing randomized rapid event-related designs. In fact, it was shown that, if the ISI is properly jittered or randomized from trial to trial, the efficiency improves monotonically with decreasing mean ISI (Dale, 1999). In general, it is found that a trade-off exists between detection power and estimation efficiency, with block designs being optimal for the former while event-related designs are optimal for the latter (Liu et al., 2001). Nevertheless, a recent report established the feasibility and test-retest reliability of estimating HRF parameters from block design fMRI data (Shan et al., 2014). In our work, we have used a randomized design by introducing uncertainty in the ISI, and we showed that smaller uncertainty leads to better distinguishability for the same noise level. Our results are therefore consistent with the literature.

Limitations

The framework adopted in this work resorts to deterministic concepts and, therefore, certain assumptions are posed on the signals acting on the system, in particular in terms of maximum amplitudes. Stochastic approaches are more flexible in that sense, but require the knowledge regarding the statistical properties of those signals, which may not be trivial to obtain, or which may be violated in practice. Therefore, a compromise between these two alternative frameworks—deterministic and stochastic—for the distinguishability of HRF models is still a subject of further research.

Conclusion

In summary, in this paper we proposed a novel approach to assess distinguishability among a set of physiologically plausible biophysically informed HRF models, and to design fMRI experiments for optimal estimation efficiency of such HRF models, with potentially great impact in further understanding HRF variability and its physiological meaning.

Acknowledgments

We acknowledge financial support by the Portuguese Science Foundation through Projects PTDC/SAU-ENB/112294/2009, PTDC/BBB-IMG/2137/2012 and FCT [UID/EEA/50009/2013], and project MYRG117(Y1-L3)-FST12-MKM of the University of Macau. We also thank the reviewers for their insightful comments and corrections.

Supplementary Material

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fncom.2015.00054/abstract

References

Aguirre, G., Zarahn, E., and D'Esposito, M. (1998). The variability of human, BOLD hemodynamic responses. *Neuroimage* 8, 360–369.

Bellgowan, P., Saad, Z., and Bandettini, P. (2003). Understanding neural system dynamics through task modulation and measurement of functional mri amplitude, latency, and width. *Proc. Natl. Acad. Sci. U.S.A.* 100, 1415–1419. doi: 10.1073/pnas.0337747100

Birn, R., Saad, Z., and Bandettini, P. (2001). Spatial heterogeneity of the nonlinear dynamics in the fMRI BOLD response. *Neuroimage* 14, 817–826. doi: 10.1006/nimg.2001.0873

Boynton, G., Engel, S. A., Glover, G., and Heeger, D. (1996). Linear systems analysis of functional magnetic resonance imaging in human v1. *J. Neurosci.* 16, 4207–4221.

Buracas, G., and Boynton, G. (2002). Efficient design of event-related fMRI experiments using msequences. *Neuroimage* 16(3 Pt 1), 801–813. doi: 10.1006/nimg.2002.1116

Buxton, R., Uluda, K., Dubowitz, D., and Liu, T. (2004). Modeling the hemodynamic response to brain activation. *Neuroimage* 23, 220–233. doi: 10.1016/j.neuroimage.2004.07.013

Buxton, R., Wong, E., and Frank, L. (1998). Dynamics of blood flow and oxygenation changes during brain activation: the balloon model. *Magn. Reson. Med.* 39, 855–864. doi: 10.1002/mrm.1910390602

Cohen, E., and Ugurbil, K. (2002). Effect of basal conditions on the magnitude and dynamics of the blood oxygenation level-dependent fMRI response. *J. Cereb. Blood Flow Metab.* 22, 1042–1053. doi: 10.1097/00004647-200209000-00002

D'Esposito, M., Deouell, L., and Gazzaley, A. (2003). Alterations in the BOLD fMRI signal with ageing and disease: a challenge for neuroimaging. *Nat. Rev. Neurosci.* 4, 863–872. doi: 10.1038/nrn1246

Dale, A. (1999). Optimal experimental design for event-related fMRI. *Hum. Brain Mapp.* 8, 109–114.

David, O., Guillemain, I., Saillet, S., Reyt, S., Deransart, C., Segebarth, C., et al. (2008). Identifying neural drivers with functional mri: an electrophysiological validation. *PLoS Biol.* 6:2683–2697. doi: 10.1371/journal.pbio.0060315

Deco, G., and Jirsa, V. K. (2012). Ongoing cortical activity at rest: criticality, multistability, and ghost attractors. *J. Neurosci.* 32, 3366–3375. doi: 10.1523/JNEUROSCI.2523-11.2012

Deneux, T., and Faugeras, O. (2006). Using nonlinear models in fMRI data analysis: model selection and activation detection. *Neuroimage* 32, 1669–1689. doi: 10.1016/j.neuroimage.2006.03.006

Friston, K. (2002). Bayesian estimation of dynamical systems: an application to fMRI. *Neuroimage* 16, 513–530. doi: 10.1006/nimg.2001.1044

Friston, K., Fletcher, P., Josephs, O., Holmes, A., Rugg, M., and Turner, R. (1998). Event-related fMRI: characterizing differential responses. *Neuroimage* 7, 30–40. doi: 10.1006/nimg.1997.0306

Friston, K., Harrison, L., and Penny, W. (2003). Dynamic causal modelling. *Neuroimage* 19, 1273–1302. doi: 10.1016/S1053-8119(03)00202-7

Friston, K., Jezzard, P., and Turner, R. (1994). Analysis of functional mri time-series. *Hum. Brain Mapp.* 1, 153–171. doi: 10.1002/hbm.460010207

Friston, K., Mechelli, A., Turner, R., and Price, C. (2000). Nonlinear responses in fMRI: the Balloon model, Volterra kernels and other hemodynamics. *Neuroimage* 12, 466–477. doi: 10.1006/nimg.2000.0630

Friston, K., Trujillo-Barreto, N., and Daunizeau, J. (2008). DEM: a variational treatment of dynamic systems. *Neuroimage* 41, 849–885. doi: 10.1016/j.neuroimage.2008.02.054

Glover, G. H. (1999). Deconvolution of impulse response in event-related BOLD fMRI. *Neuroimage* 9, 416–429. doi: 10.1006/nimg.1998.0419

Grewal, M., and Glover, K. (1976). Identifiability of linear and nonlinear dynamical systems. *IEEE Trans. Autom. Control* 21, 833–837. doi: 10.1109/TAC.1976.1101375

Handwerker, D. A., Gonzalez-Castillo, J., D'Esposito, M., and Bandettini, P. A. (2012). The continuing challenge of understanding and modeling hemodynamic variation in fMRI. *Neuroimage* 62, 1017–1023. doi: 10.1016/j.neuroimage.2012.02.015

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

Havlicek, M., Friston, K. J., Jan, J., Brazdil, M., and Calhoun, V. D. (2011). Dynamic modeling of neuronal responses in fMRI using cubature kalman filtering. *Neuroimage* 56, 2109–2128. doi: 10.1016/j.neuroimage.2011.03.005

Huppert, T., Hoge, R., Diamond, S., 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

Iadecola, C. (2004). Neurovascular regulation in the normal brain and in Alzheimer's disease. *Nat. Rev. Neurosci.* 5, 347–360. doi: 10.1038/nrn1387

Jezzard, P., Matthews, P. M., and Smith, S. M. (Eds). (2001). *Functional MRI: an Introduction to Methods*, Vol. 61. Oxford: Oxford University Press.

Johnston, L. A., Duff, E., Mareels, I., and Egan, G. F. (2008). Nonlinear estimation of the BOLD signal. *Neuroimage* 40, 504–514. doi: 10.1016/j.neuroimage.2007.11.024

Jorge, J. P., van der Zwaag, W., and Figueiredo, P. (2014). EEG-fMRI integration for the study of human brain function. *NeuroImage* 102(Pt1), 24–34. doi: 10.1016/j.neuroimage.2013.05.114

Josephs, O., Turner, R., and Friston, K. (1997). Event-related fMRI. *Hum. Brain Mapp.* 5, 243–248.

Lindquist, M. A., Meng Loh, J., Atlas, L. Y., and Wager, T. D. (2009). Modeling the hemodynamic response function in fMRI: efficiency, bias and mis-modeling. *Neuroimage* 45, S187–S198. doi: 10.1016/j.neuroimage.2008.10.065

Lindquist, M. A., and Wager, T. D. (2007). Validity and power in hemodynamic response modeling: a comparison study and a new approach. *Hum. Brain Mapp.* 28, 764–784. doi: 10.1002/hbm.20310

Liu, T. T., Behzadi, Y., Restom, K., Uludag, K., Lu, K., Buracas, G. T., et al. (2004). Caffeine alters the temporal dynamics of the visual BOLD response. *Neuroimage* 23, 1402–1413. doi: 10.1016/j.neuroimage.2004.07.061

Liu, T. T., Frank, L. R., Wong, E. C., and Buxton, R. B. (2001). Detection power, estimation efficiency, and predictability in event-related fMRI. *Neuroimage* 13, 759–773. doi: 10.1006/nimg.2000.0728

Logothetis, N. K., and Wandell, B. A. (2004). Interpreting the BOLD signal. *Annu. Rev. Physiol.* 66, 735–769. doi: 10.1146/annurev.physiol.66.082602.092845

Maus, B., van Breukelen, G. J., Goebel, R., and Berger, M. P. (2012). Optimal design for nonlinear estimation of the hemodynamic response function. *Hum. Brain Mapp.* 33, 1253–1267. doi: 10.1002/hbm.21289

Miezin, F. M., MacCotta, L., Ollinger, J. M., Petersen, S. E., and Buckner, R. L. (2000). Characterizing the hemodynamic response: effects of presentation rate, sampling procedure, and the possibility of ordering brain activity based on relative timing. *Neuroimage* 11, 735–759. doi: 10.1006/nimg.2000.0568

Murta, T., Leal, A., Garrido, M. I., and Figueiredo, P. (2012). Dynamic causal modelling of epileptic seizure propagation pathways: a combined EEG–fMRI study. *Neuroimage* 62, 1634–1642. doi: 10.1016/j.neuroimage.2012.05.053

Riera, J., Aubert, E., Iwata, K., Kawashima, R., Wan, X., and Ozaki, T. (2005). Fusing EEG and fMRI based on a bottom-up model: inferring activation and effective connectivity in neural masses. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360, 1025–1041. doi: 10.1098/rstb.2005.1646

Riera, J. J., Watanabe, J., Kazuki, I., Naoki, M., Aubert, E., Ozaki, T., et al. (2004). A state-space model of the hemodynamic approach: nonlinear filtering of BOLD signals. *Neuroimage* 21, 547–567. doi: 10.1016/j.neuroimage.2003.09.052

Rosa, P., and Silvestre, C. (2011). "On the distinguishability of discrete linear time-invariant dynamic systems," in *Proceedings of the 50th IEEE Conference on Decision and Control*. (Orlando, FL).

Shan, Z. Y., Wright, M. J., Thompson, P. M., McMahon, K. L., Blokland, G. G., de Zubicaray, G. I., et al. (2014). Modeling of the hemodynamic responses in block design fMRI studies. *J. Cereb. Blood Flow Metab.* 34, 316–324. doi: 10.1038/jcbfm.2013.200

Silvestre, C., Figueiredo, P., and Rosa, P. (2010a). "Multiple-model set-valued observers: a new tool for HRF model selection in fMRI," in *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, (Buenos Aires), 5704–5707.

Silvestre, C., Figueiredo, P., and Rosa, P. (2010b). "On the distinguishability of HRF models in fMRI," in *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, (Buenos Aires), 5677–5680.

Smith, S. M., Miller, K. L., Salimi-Khorshidi, G., Webster, M., Beckmann, C. F., Nichols, T. E., et al. (2011). Network modelling methods for fMRI. *Neuroimage* 54, 875–891. doi: 10.1016/j.neuroimage.2010.08.063

Sotero, R. C., and Trujillo-Barreto, N. J. (2007). Modelling the role of excitatory and inhibitory neuronal activity in the generation of the BOLD signal. *Neuroimage* 35, 149–165. doi: 10.1016/j.neuroimage.2006.10.027

Sotero, R. C., Trujillo-Barreto, N. J., Jiménez, J. C., Carbonell, F., and Rodríguez-Rojas, R. (2009). Identification and comparison of stochastic metabolic/hemodynamic models (sMHM) for the generation of the BOLD signal. *J. Comput. Neurosci.* 26, 251–269. doi: 10.1007/s10827-008-0109-3

Tagliazucchi, E., Balenzuela, P., Fraiman, D., Montoya, P., and Chialvo, D. R. (2011). Spontaneous BOLD event triggered averages for estimating functional connectivity at resting state. *Neurosci. Lett.* 488, 158–163. doi: 10.1016/j.neulet.2010.11.020

Vincent, T., Badillo, S., Risser, L., Chaari, L., Bakhous, C., Forbes, F., et al. (2014). Flexible multivariate hemodynamics fMRI data analyses and simulations with PyHRF. *Front. Neurosci.* 8:67. doi: 10.3389/fnins.2014.00067

Wager, T. D., and Nichols, T. E. (2003). Optimization of experimental design in fMRI: a general framework using a genetic algorithm. *Neuroimage* 18, 293–309. doi: 10.1016/S1053-8119(02)00046-0

Walter, E., Lecourtier, Y., and Happel, J. (1984). On the structural output distinguishability of parametric models, and its relations with structural identifiability. *IEEE Trans. Autom. Control* 29, 56–57. doi: 10.1109/TAC.1984.1103379

Woolrich, M. W., Behrens, T. E., and Smith, S. M. (2004). Constrained linear basis sets for HRF modelling using variational bayes. *Neuroimage* 21, 1748–1761. doi: 10.1016/j.neuroimage.2003.12.024

Wu, G. R., Liao, W., Stramaglia, S., Ding, J. R., Chen, H., and Marinazzo, D. (2013). A blind deconvolution approach to recover effective connectivity brain networks from resting state fMRI data. *Med. Image Anal.* 17, 365–374. doi: 10.1016/j.media.2013.01.003

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.

frontiers in Computational Neuroscience

ORIGINAL RESEARCH
published: 24 March 2015
doi: 10.3389/fncom.2015.00038

Detection of epileptiform activity in EEG signals based on time-frequency and non-linear analysis

Dragoljub Gajic[1,2], Zeljko Djurovic[1], Jovan Gligorijevic[3], Stefano Di Gennaro[2] and Ivana Savic-Gajic[4]*

[1] Department of Signals and Systems, School of Electrical Engineering, University of Belgrade, Belgrade, Serbia, [2] Center of Excellence DEWS, University of L'Aquila, L'Aquila, Italy, [3] Faculty of Engineering, University of Kragujevac, Kragujevac, Serbia, [4] Faculty of Technology, University of Nis, Leskovac, Serbia

We present a new technique for detection of epileptiform activity in EEG signals. After preprocessing of EEG signals we extract representative features in time, frequency and time-frequency domain as well as using non-linear analysis. The features are extracted in a few frequency sub-bands of clinical interest since these sub-bands showed much better discriminatory characteristics compared with the whole frequency band. Then we optimally reduce the dimension of feature space to two using scatter matrices. A decision about the presence of epileptiform activity in EEG signals is made by quadratic classifiers designed in the reduced two-dimensional feature space. The accuracy of the technique was tested on three sets of electroencephalographic (EEG) signals recorded at the University Hospital Bonn: surface EEG signals from healthy volunteers, intracranial EEG signals from the epilepsy patients during the seizure free interval from within the seizure focus and intracranial EEG signals of epileptic seizures also from within the seizure focus. An overall detection accuracy of 98.7% was achieved.

Keywords: seizure detection, epileptiform activity, non-linear analysis, scatter matrices, quadratic classifiers

OPEN ACCESS

Edited by:
Tobias Alecio Mattei, Brain and Spine Center - InvisionHealth - Kenmore Mercy Hospital, USA

Reviewed by:
Peter König, University of Osnabrück, Germany
Germán Mato, Centro Atomico Bariloche, Argentina

****Correspondence:***
Dragoljub Gajic, Department of Signals and Systems, School of Electrical Engineering, University of Belgrade, Bulevar Kralja Aleksandra 73, Belgrade 11000, Serbia
dragoljubgajic@gmail.com

Received: *01 January 2015*
Accepted: *08 March 2015*
Published: *24 March 2015*

Citation:
Gajic D, Djurovic Z, Gligorijevic J, Di Gennaro S and Savic-Gajic I (2015) Detection of epileptiform activity in EEG signals based on time-frequency and non-linear analysis. Front. Comput. Neurosci. 9:38. doi: 10.3389/fncom.2015.00038

Introduction

According to the estimations of the World Health Organization around 50 million people worldwide suffer from epilepsy as the most common disorder of the brain activity (World Health Organization, 2012). It is characterized by sudden and recurrent seizures which are the result of an excessive and synchronous electrical discharge of a large number of neurons. Epileptic seizures can be divided by their clinical manifestation into two main classes, partial and generalized (Tzallas et al., 2007). Partial or focal epileptic seizures involve only a circumscribed region of the brain (epileptic focus) and remain restricted to this region while generalized epileptic seizures involve almost the entire brain. Both classes of epileptic seizures can occur at all ages. An epileptiform activity in EEG signals including spikes, sharp waves, or spike-and-wave complexes can be evident not only during a seizure (the ictal period) but also a short time before (the preictal period) as well as between seizures (the interictal period). Consequently, EEG signals have been the most utilized in clinical assessments of the brain state including both prediction and detection of epileptic seizures (Waterhouse, 2003; Casson et al., 2010). However, the detection of epileptiform activity in EEG signals by visual scanning of EEG recordings usually collected over a few days is a tedious and time-consuming process. In addition, it requires a team of experts to analyze the entire length of the EEG recordings in order to detect epileptiform activity. A reliable technique for detection

of epileptiform activity in EEG signals would ensure an objective and facilitating treatment of patients and thus improve the diagnosis of epilepsy. Furthermore, it would also enable an automated prediction and/or detection of epileptic seizures in real time by a system to be implanted in head of epileptic patients (Jerger et al., 2001). Such a system would significantly improve quality of life of people suffering from epilepsy. Most of the techniques for automated detection of epileptiform activity that have emerged in recent years consist of two key successive steps: extraction of features from EEG signals and then classification of the extracted features for detection of epileptiform activity.

The feature extraction, as the first step, has a direct influence on both precision and complexity of the entire technique. Most common statistical features in time domain, such as the mean, the variance, the coefficient of variation and the total variation, by themselves are not sufficient for a reliable detection of epileptiform activity, and thus are mostly used as statistical measures for features in other domains. The variance and the total variation are considered to have better discriminatory capabilities than the mean, since they are able to detect magnitude of change in a signal over time. Even though we can note a certain periodicity and synchronization between EEG signals from different electrodes, neither the autocorrelation nor the cross-correlation have proved to be reliable features for detection of epileptiform activity. This is especially true in the case of the cortical EEG where the recording electrodes are so close to each other that the synchronization could be noted even when there was no seizure. However, in the literature we can still find several applications of these two features (Niederhauser et al., 2003; Jerger et al., 2005).

Unlike the previous features, the spectral features of EEG signals obtained through the Fourier transform have found wide applications in the field (Polat and Gunes, 2007; Mousavi et al., 2008). Namely, all the research carried out to date clearly indicates that it is much better to identify and extract the features of interest in frequency domain than in time domain, even though the both domains contain identical information. The analysis in time-frequency domain gives even better results considering that it contains, in addition to frequency, also the temporal component of signal which is lost during the Fourier transform. The literature mainly contains techniques based on wavelet transform (Subasi, 2007a,b; Wang et al., 2011; Gajić et al., 2014) which has also been used in the research related to other brain disorders, such as schizophrenia (Hazarika et al., 1997) and Alzheimer's disease (Adeli and Ghosh-Dastidar, 2010). The detection of epileptiform activity based on non-linear analysis, i.e., extraction of the correlational dimension and the Lyapunov exponents as non-linear features can also be noted in some research studies (Iasemidis et al., 2003; Srinivasan et al., 2007; Adeli and Ghosh-Dastidar, 2010).

A precise classification as the second key step directly depends on the previously extracted features. That is, there is no classifier which could in any way make up for the shortcomings which are consequence of the information lost during the feature extraction. Like in the case of the feature extraction, we can come across a very wide range of classifiers starting from the most simple ones with thresholds (Altunay et al., 2010) or rule-based (Gotman, 1999), to linear classifiers (Liang et al., 2010; Iscan et al., 2011) and all the way to those more complex ones based on fuzzy logic and artificial neural networks (Gajić, 2007; Subasi, 2007a; Tzallas et al., 2007). We can also note the use of other techniques for classification based on k nearest neighbors (Guo et al., 2011; Orhan et al., 2011), decision trees (Tzallas et al., 2009), expert models (Ubeyli, 2007; Ubeyli and Guler, 2007) as well as Bayes classifiers (Tzallas et al., 2009; Iscan et al., 2011). Considering that the feature extraction as a process of higher priority can be computationally very demanding it is always more desirable to use simpler classifiers so that the entire decision-making system could ideally work in real time.

In this paper we present an automated technique for detection of epileptiform activity in EEG signals. In contrast with the existing techniques which are mainly based on features from one domain of interest, our new technique optimally integrates features from a few domains and frequency sub-bands of clinical interest in order to increase its robustness and accuracy. We extract features in both time and frequency domain as well as time-frequency domain using discrete wavelet transform which has already been recognized as a very good linear technique for analysis of non-stationary signals such as EEG signals. In addition, by non-linear analysis we extract the correlation dimension and the largest Lyapunov exponent as much better measures of EEG signal non-linearity which is only approximated by other linear techniques such as fast Fourier transform (FFT) and discrete wavelet transform (DWT). After the feature extraction we optimally reduce the feature space dimension to two using scatter matrices and then perform classification in the reduced feature space by quadratic classifiers which have already been known as very robust solutions for classification of random feature vectors.

Materials and Methods

Materials

The EEG signals used to design and test the new technique were recorded at the University Hospital Bonn, Germany with the same 128-channel amplifier system (Andrzejak et al., 2001). After 12 bit analog-to-digital conversion the EEG signals were saved in a data acquisition system at a sampling rate of 173.61 Hz. The amplifier range was adjusted well so that the recordings could be made with 12 bits. The recorded EEG signals were further passed through a low pass filter with the finite impulse response and bandwidth of 0–60 Hz. The frequencies higher than 60 Hz mostly present noise and are a very small part of the signal total energy in the frequency band up to 86.8 Hz saved by the acquisition system. We used 100 segments of epileptic and 200 segments of non-epileptic EEG signals to design and test our new technique. The epileptic EEG signals were recorded using cortical electrodes from 5 epileptic patients during seizure from within the seizure focus, i.e., the region of unhealthy brain tissue that was later removed by surgery. The first 100 segments of non-epileptic EEG signals were also recorded using cortical electrodes from the same epileptic patients and the same unhealthy brain tissue but during seizure-free interval. The remaining 100 segments of non-epileptic EEG signals were recorded using scalp electrodes from 5 healthy volunteers and of course their healthy brain tissue.

So, there was a total of three groups with 100 segments of the EEG signals. All the segments have duration of 4096 samples, i.e., 23.6 s, and were additionally tested on the weak stationarity (Andrzejak et al., 2001) in order to perform non-linear analysis. Since the EEG signals were recorded from different patients and with different electrodes, all extracted EEG signal segments were also additionally normalized in order to have the same zero mean and unit variance as shown in **Figure 1**. In this way, we wanted to design a detection technique that is not dependent on patient and the EEG recording system either.

Methods

There are five broad sub-bands of the EEG signal which are generally of clinical interest: delta (0–4 Hz), theta (4–8 Hz), alpha (8–16 Hz), beta (16–32 Hz), and gamma waves (32–64 Hz). Higher frequencies are often more common in abnormal brain states such as epilepsy, i.e., there is a shift of EEG signal energy from lower to higher frequency bands before and during a seizure (Gajić et al., 2014). These five frequency sub-bands provide more accurate information about neuronal activities underlying the problem. Consequently, some changes in the EEG signal, which are not so obvious in the original full-spectrum signal, can be amplified when each sub-band is considered independently. Thus, we extract features from each sub-band separately and also in time, frequency and time-frequency domain as well as by non-linear analysis. After the feature extraction we reduce dimension of the feature space to two. Finally, two quadratic classifiers able to separate all three groups of the EEG signals from each other are designed. The entire structure of the technique is shown in **Figure 2**.

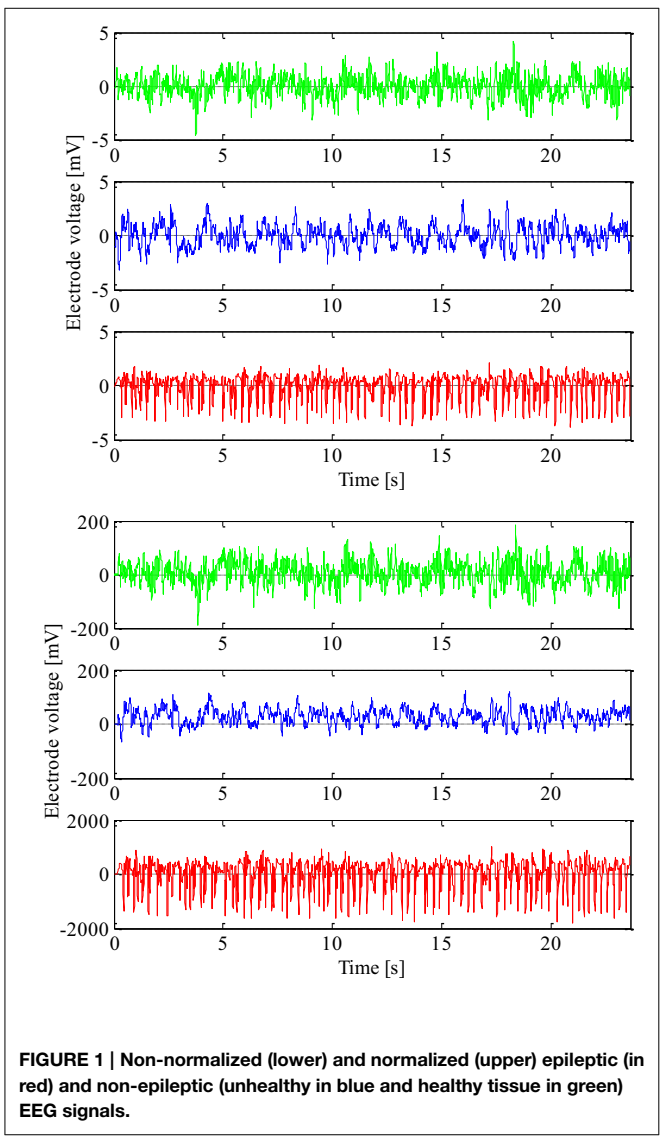


FIGURE 1 | Non-normalized (lower) and normalized (upper) epileptic (in red) and non-epileptic (unhealthy in blue and healthy tissue in green) EEG signals.

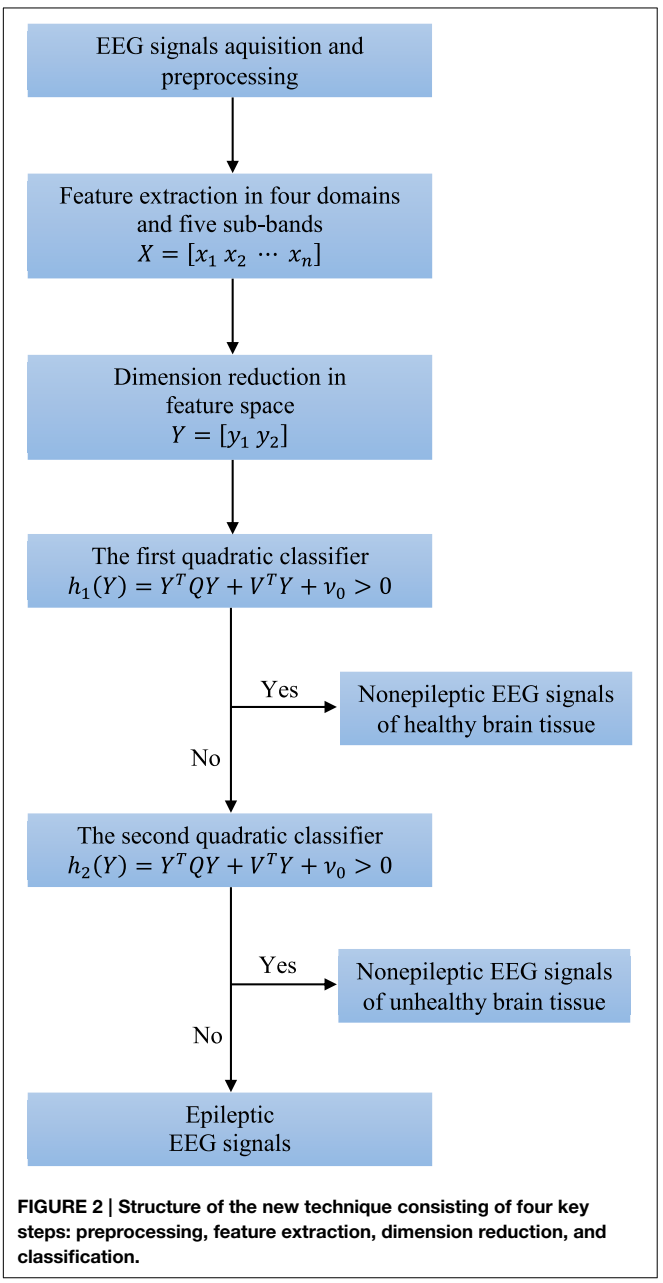


FIGURE 2 | Structure of the new technique consisting of four key steps: preprocessing, feature extraction, dimension reduction, and classification.

Time-Frequency Domain Analysis

Since the segments of the EEG signals have already been normalized and all have zero mean and unit variance, additional extraction of these two features as well as coefficient of variation as function of mean value and variance, does not make any sense. However, we extracted the total variation as another measure of signal variability in the time domain even after normalization since it counts number of signal sign changes or signal polarity. In the case of a signal segment $x\,[n]$ of N samples, i.e., $n = 1,\ 2 \cdots N$, the total variation is given by:

$$v_x = \frac{1}{N-1} \frac{\sum_{n=2}^{N} |x\,[n] - x[n-1]|}{(max_x - min_x)} \tag{1}$$

where the signal is essentially normalized by the difference between its maximum and minimum values in the segment of interest. Obviously, the value of the total variation is located in the range between $1/(N-1)$ for slower signals and 1 for signals with very high and frequent changes.

EEG signals, as the outcome of events with different repetition periods, contain signals whose different frequencies cannot be identified in the time domain, since all these signals are shown together. Thus, signal transformation from the time domain to the frequency domain is necessary, which in the case of a signal segment $x[n]$ of N samples is achieved using the fast Fourier transform (FFT) defined by:

$$\mathit{fft}\,[\omega] = \sum_{n=1}^{N} x[n] e^{-i\omega n},\ \omega = \frac{2\pi m}{N},\ 0 \le m \le N-1 \tag{2}$$

where $\omega = 2\pi f/f_s$ represents the angular frequency discretized in N samples (Proakis and Manolakis, 1996). In order to avoid discontinuities between the end and beginning of the segments and thus spurious spectral frequency components the beginning of each segment was chosen in such a way that the amplitude difference of the last and first data points was within the range of amplitude differences of consecutive data points, and the slopes at the end and beginning of each segment had the same sign. This procedure reduces edge effects that result in spectral leakage in the FFT spectrum. In order to further minimize spectral leakage windowing of signal segments by the Hamming window (the sum of a rectangle and a Hanning window) is used before application of the FFT. Considering the fact that by transforming the signal into the frequency domain we do not lose any original information from the time domain, the signal can completely be reconstructed using the inverse Fourier transform by:

$$x[n] = \frac{1}{N} \sum_{\omega=0}^{2\pi(N-1)/N} \mathit{fft}[\omega] e^{i\omega n},\ 1 \le n \le N \tag{3}$$

Clearly, the longer the segment $x[n]$, i.e., the larger N, the greater the frequency resolution.

Power spectral density is also one of the most important features of the signal in the frequency domain and represents the contribution of each individual frequency component to the power of the whole signal segment $x[n]$. In practice, power spectral density is usually estimated using the coefficients of the fast Fourier transform, i.e., the periodogram (Welch, 1967) given by:

$$per\,[\omega] = \frac{1}{N} \left|\mathit{fft}[\omega]\right|^2 \tag{4}$$

which is an unbiased and inconsistent estimator. Thus, with the increase in the length of the signal segment, the mean of the estimation tends toward the actual value of power spectral density, which is actually an advantage, unlike variance estimation, which is not reduced, i.e., which does not have a tendency toward zero with the increase in segment length. A periodogram can be further normalized by the total signal power, i.e.,:

$$per_{norm}\,[\omega] = \frac{1}{N} \left|\mathit{fft}\,[\omega]\right|^2 \Big/ \sum_{\omega=0}^{2\pi(N-1)/N} per\,[\omega] \tag{5}$$

where we obtain the relative contribution of each frequency component to the total power of the signal. If the original signal segment $x[n]$ is further divided into P sub-segments of the N/P samples, the periodogram can be calculated as follows:

$$per\,[\omega] = \frac{1}{P} \sum_{p=0}^{P-1} \frac{P}{N} \left|\mathit{fft}_p[\omega]\right|^2 \tag{6}$$

where $\mathit{fft}_p\,[\omega]$ is the fast Fourier transform of each of the sub-segments of the N/P sample. In this way, the periodogram is actually an averaged one with a smaller variance, but clearly with a lower resolution in the frequency domain. Based on the periodogram we extracted relative power of all five previously mentioned sub-bands, i.e., delta (0–4 Hz), theta (4–8 Hz), alpha (8–12 Hz), beta (12–30 Hz), and gamma (30–60 Hz), as features of interest in frequency domain.

By analyzing the EEG signals solely in the time domain, extracted features do not contain any information on frequencies, which are, as we will later show, also very important for the proper detection of epileptic EEG signals. On the other hand, by transforming the signals from the time into the frequency domain, any information on time is completely lost, except of course in the case of sequential application on sufficiently short and stationary sub-segments, which also has its disadvantage in terms of the correct choice of the length of these sub-segments which would enable the simultaneous achievement of the desired resolution in both domains. In addition, once selected, the sub-segment length, i.e., the resolution in the time domain, remains fixed throughout the entire frequency bands and cannot be adjusted to the dominant signal frequencies at a specific time. Signal processing using wavelets very accurately resolves this deficiency and results in sufficient information on non-stationary signals, both in the time and frequency domain. We are already familiar with the fact that a signal can be presented as a linear combination of its basic functions. A unit impulse function whose power is limited and whose mean differs from zero is the basic function of the signal in the time domain, whereas in the frequency domain, this role is assigned to the sinusoidal function

that has infinite power, and a zero mean. In the time-frequency domain, the basic function is the wavelet, which is actually a function of limited power, i.e., duration, and a zero mean (Rao and Bopardikar, 1998), and for which the following is valid:

$$\sum_{n=-\infty}^{\infty} |\psi[n]|^2 < \infty, \quad \sum_{n=-\infty}^{\infty} \psi[n] = 0. \tag{7}$$

The wavelet that is moved, or translated, in time for b samples and scaled by the so-called dilation parameter a is given by:

$$\psi_{ab}[n] = \frac{1}{\sqrt{a}}\psi\left[\frac{n-b}{a}\right]. \tag{8}$$

By changing the dilation parameter, the basic wavelet ($a = 1$) changes its width, that is, it spreads ($a > 1$) and contracts ($0 \leq a < 1$) in the time domain. In the analysis of non-stationary signals, the possibility of changing the width of the wavelet represents a significant advantage of this analysis technique, considering the fact that wider wavelets can be used to extract slower changes, i.e., lower signal frequencies, and narrower wavelets can be used to extract faster changes, i.e., higher frequencies. Following the selection of the values of parameters a and b it is possible to transform segments of the signal $x\left[k\right]$ of N samples, that is, to calculate the wavelet transform coefficients in the following way:

$$w_{ab}[n] = \sum_{\tau=1}^{N} x[\tau]\psi_{ab}[n-\tau], \ 1 \leq n \leq N \tag{9}$$

Thus, what is actually being extracted from the signal are only those frequencies that are within the wavelet frequency band $\psi_{ab}[n]$, i.e., the signals are filtrated by the wavelet $\psi_{ab}[n]$. As previously indicated, based on the coefficients obtained in this way, the original signal can be reconstructed using an inverse wavelet transform. Of course, if necessary, it is possible to also independently reconstruct the part of the signal which is filtered, as well as the part that was rejected by the wavelet $\psi_{ab}[n]$ on the basis of the so-called detail coefficients and approximation coefficients respectively, which are of course a function of the transformation coefficients $\psi_{ab}[n]$.

Parameters a and b can continuously change, which is not so practical especially bearing in mind that the signal can be completely and accurately transformed and reconstructed by using a smaller and finite number of wavelets, that is, by using a limited number of discrete values of parameters a and b, which is also known as the discrete wavelet transform (DWT). In this case, parameters a and b are the powers of 2, which gives us the dyadic orthogonal wavelet network with frequency bands which do not overlap each other. The dilation parameter a, as the power of 2, at each subsequent higher level of transformation, doubles in value in comparison to the value from the previous level, which means that the wavelet becomes twice as wide in the time domain, and has a frequency band that is half as narrow and twice as low. This actually decreases the resolution of the transformed signal in the time domain two-fold, increasing it twice as much in the frequency domain. Thus, the signal frequency band from the previous level is split into two halves at every next level, into a higher band which contains higher frequencies and describes the finer changes, or details, and a lower band that contains lower frequencies and actually represents an approximation of the signal from the previous level. This technique is also known as wavelet decomposition of the signal.

Before the application of DWT, it is necessary to choose the type of the basic wavelet as well as the number of levels into which the signal will be decompose. After analysis of several types of the basic wavelets, the fourth-order Daubechies wavelet (Rao and Bopardikar, 1998) was selected for further analysis within this work since it has good localizing properties both in the time and frequency domains (Kalayci and Özdamar, 1995; Petrosian et al., 2000) Due to its shape and smoothing feature this type of the basic wavelet has already shown good capabilities in the field of EEG signal processing. The discrete wavelet decomposition was performed at four levels that resulted into five sub-bands of clinical interest. The standard deviation and the average relative power of the DWT coefficients in each of the sub-bands were extracted as representative features in time-frequency domain.

Non-linear Analysis

EEG signals, as the result of the activities of an extremely complex and non-linear system, in addition to the fairly well-known and previously described linear techniques, can also be analyzed using some of the non-linear techniques. By using linear techniques, any non-linearity that can be found in the signal is only approximated, which can result in the loss of certain pieces of potentially relevant information. If that is the case, the use of non-linear techniques is preferred since they are more reliable for non-linear analyses, despite the fact that they imply weak signal stationarity (Varsavsky et al., 2011), and the fact that they need somewhat longer segments, which leads to their being computationally more demanding than linear techniques.

Let $x[n]$ again represent the signal segment which is to be analyzed, where $n = 1 \cdots N$. Also, let m denote the lag for which we can define two new sub-segments $x[n]$, the first x_k containing samples starting from k up to $N - m$ and the second x_{k+m} with samples starting from $k + m$ to N. Both of these sub-segments contain $N - k - m + 1$ samples and can be represented opposite one another in the phase space with a lag m and the so-called embedding dimension 2. In case of three sub-segments: x_{k+2m}, x_{k+m} and x_k, the embedding dimension of the phase space would be 3. The lagged phase space provides a completely different view of signal evolution in time, where we can note that the signal gravitates to a certain part of the phase space, known as the attractor. With the aim of constructing lagged phase space, i.e., the signal attractor, it is necessary to previously define the values of the lag and the embedding dimension, which although significantly smaller than the real dimension of the non-linear system space, provides an approximation of the signal complexity and non-linearity (Andrzejak et al., 2001). The lag m should be large enough so that these sub-segments would overlap as little as possible, that is, share as little mutual information as possible, but at the same time sufficiently small so that the sub-segments could be long enough for any further useful analysis. An optimal lag is obtained by determining the mutual information coefficient the sub-segments for different values of the lag m. The mutual

information coefficient is defined by Williams (1997):

$$info_m = \sum_{i=1}^{N_s}\sum_{j=1}^{N_s} p\left(x_k\left[i\right], x_{k+m}[j]\right) log_2 \frac{p\left(x_k[i], x_{k+m}[j]\right)}{p\left(x_k\left[i\right]\right) p\left(x_{k+m}[j]\right)} \quad (10)$$

where N_s represents the number of areas in which the signal is discretized based on the amplitude and p is the corresponding probability that the sub-segment belongs to a certain area. The first local minimum shown in the graph representing the dependence of the mutual information coefficient on lag determines the optimal lag m_o.

After determining the optimal lag, the minimum embedding dimension of the lagged phase space is estimated using Cao's technique (Cao, 1997). In the phase space with a lag m_o and embedding dimension d, the original segment is represented by its phase portraits, which all together make up the attractor defined by the following points in the lagged phase space:

$$y_d[i] = \left[x[i]\, x\left[i + m_o\right] \cdots x\left[i + m_o(d-1)\right]\right] \quad (11)$$

where $i = 1,\ 2,\ \cdots,\ N - m_o(d-1)$. According to the technique developed by Cao, if d is the right dimension, then the two points are also close to each other in phase space dimension d, as well as in the phase space of dimension $d+1$ and are referred to as real neighbors (Cao, 1997). Dimension increases gradually until the number of false neighbors reaches zero, that is, until the Cao's embedding function defined by:

$$e_d = \frac{1}{N - m_o d} \sum_{i=1}^{N - m_o d} \frac{\left\| y_{d+1}[i] - y_{d+1}[n_{i,\,d}] \right\|}{\left\| y_d[i] - y_d[n_{i,\,d}] \right\|} \quad (12)$$

becomes constant, where $i = 1,\ 2,\ \cdots,\ N - m_o d$ and $y_d[n_{i,\,d}]$ represents the nearest neighbor of $y_d[i]$ in the d-dimensional phase space with a lag m_o. In fact, the minimum embedding dimension d_{min} is determined when the ratio between the e_{d+1}/e_d approaches the value of 1. Since this ratio may approach 1 in some other cases, e.g., for completely random signals, an additional check is also carried out where the Cao's embedding function is redefined and given by:

$$e_d^* = \frac{1}{N - m_o d} \sum_{i=1}^{N - m_o d} \left| x\left[i + m_o d\right] - x\left[n_{i,\,d} + m_o d\right] \right| \quad (13)$$

where $x\left[n_{i,\,d} + m_o d\right]$ is the nearest neighbor of $x\left[i + m_o d\right]$. The constant value of the ratio e_{d+1}^*/e_d^* for different values of the embedding dimension indicates that we are dealing with a random signal. The signal is not random, i.e., it is deterministic if this ratio differs from 1 for at least one value of the embedding dimension, which in that case is also the minimum value.

The correlation dimension is a measure of the complexity of the signal attractor in the lagged phase space. This dimension, unlike most others better known dimensions, may have a fractional value and could thus characterize the dimension, that is, the complexity of the attractors with more precision than the embedding dimension; however, it is always less than or equal to the embedding dimension.

Let C_ε be the correlational sum of the signal segment with N samples within the radius ε in its phase space with a lag m_o and minimum embedding dimension d_{min}, i.e., $M = N - m_o d_{min}$ points $y_{d_{min}}$ given by Williams (1997):

$$C_\varepsilon = \lim_{M\to\infty} \frac{1}{M^2} \sum_{i=1}^{M}\sum_{j=1}^{M} H(\varepsilon - \left\| y_{d_{min}}[i] - y_{d_{min}}[j] \right\|) \quad (14)$$

where H is the Heaviside step function that results in 1 if $y_{d_{min}}[j]$ is within the radius ε of $y_{d_{min}}[i]$, i.e.,:

$$\varepsilon - \left\| y_{d_{min}}[i] - y_{d_{min}}[j] \right\| > 0 \quad (15)$$

otherwise it is 0. The correlation dimension d_{corr} is the approximated slope of the natural logarithm of the correlation sum as a function of ε. Given that the total number of possible distances between two points in a lagged phase space equals $M(M-1)/2$, the correlation dimension could directly be obtained by the Takens estimator (Takens, 1981; Cao, 1997) using:

$$d_{corr} = -\left[\frac{2}{M\,(M-1)} \sum_{i=1}^{M}\sum_{j=1}^{M} log\left(\frac{\left\| y_{d_{min}}\left[i\right] - y_{d_{min}}\left[j\right] \right\|}{\varepsilon}\right)\right] \quad (16)$$

The largest Lyapunov exponent λ_{max} represents a measure of both chaotic behavior of the attractor and the divergence of the trajectories in phase space, i.e., the predictability of the signal. Attractor divergence is the distance between two closely positioned points in a phase space after a certain period of time of k samples, which is also known as the prediction length. Based on chaos theory, i.e., the so-called butterfly effect, two points close in the phase space of a chaotic system may have completely different trajectories. Thus, the divergence of the trajectories implies a chaotic system, and vice versa. The Lyapunov exponent actually characterizes the exponential growth of that divergence. The number of Lyapunov exponents is equal to the embedding dimension, and each of these Lyapunov exponents represents the rate of a contracting ($\lambda < 0$) or expanding attractor ($\lambda > 0$) in a certain direction of the phase space. In the case of a chaotic system, the trajectories must diverge in at least one dimension, which means that at least one Lyapunov exponent must be greater than zero, when it is, at the same time, the largest Lyapunov exponent. If several Lyapunov exponents are positive, then the largest among them indicates the direction of the maximum expansion of the attractor and its chaotic behavior. The mean of the trajectory divergence after k samples and a sampling period T_s can be calculated by the Wolf's technique (Wolf et al., 1985; Rosenstein et al., 1993) using:

$$d_T = \frac{1}{(M-k)} \sum_{i=1}^{M-k} \frac{\left\| y_{d_{min}}[i+k] - y_{d_{min}}[n_i + k] \right\|}{\left\| y_{d_{min}}\left[i\right] - y_{d_{min}}[n_i] \right\|} \quad (17)$$

where $y_{d_{min}}[i]$ and $y_{d_{min}}[n_i]$ represent two close points on different trajectories in the phase space. The largest Lyapunov exponent λ_{max} is in this case an approximation of the slope of the natural logarithmic trajectory divergence as a function of the

number of samples k, i.e., $d_T = d_0 e^{kT_s\lambda_{max}}$ where d_0 stands for the initial divergence. In addition, there is another very similar more practical technique for the evaluation of the largest Lyapunov exponent proposed by Sato et al. where we first calculate the prediction error for several different values of the number of samples k using:

$$p_k = \frac{1}{(M-k)} \sum_{i=1}^{M-k} log_2 \frac{\left\| y_{d_{min}}\left[i+k\right] - y_{d_{min}}[n_i+k] \right\|}{\left\| y_{d_{min}}[i] - y_{d_{min}}[n_i] \right\|} \quad (18)$$

after which the λ_{max} is determined as the slope of the middle and approximately linear part of the prediction error p_k as a function of kT_s.

We extract both the correlation dimension and the largest Lyapunov exponent as features that describe complexity and chaotic behavior of the attractor in the lagged phase space. By choosing the radius ε, the phase space is divided into parts of the dimension ε. While the correlation dimension shows how many points can be found in the surrounding areas of the phase space, the Lyapunov exponent describes the distance between each of the trajectories that terminate in different parts of the phase space but start from the same one. In other words, both of these features give us an idea of how complex and predictable EEG signal is, which, of course, they both interpret and quantify in their own characteristic way.

Dimension Reduction in Feature Space

Let an n-dimensional random vector X be transformed through the application of a certain linear transformation into an n-dimensional random vector $Y = A^T X$ where A is the transformational square matrix of the dimension n. Then the mean vector and the covariance matrix of the random vector Y are $M_Y = A^T M_X$ and $\Sigma_Y = A^T \Sigma_X A$. Based on that, the distance function is:

$$\begin{aligned} d_Y^2(Y) &= (Y - M_Y)^T \Sigma_Y^{-1} (Y - M_Y) = (X - M_X)^T \Sigma_X^{-1} \\ &\quad (X - M_X) = d_X^2(X) \end{aligned} \quad (19)$$

that is, the distance function does not change with the linear transformation. If we were to perform the translation of the coordinate system for the mean vector M_X we would obtain the random vector $Z = X - M_X$ whose mean vector is zero and its covariance matrix is the same as Σ_X. If we wanted to determine the random vector Z which maximizes the distance function $d_Z^2(Z) = Z^T \Sigma^{-1} Z$ under the condition that $Z^T Z = 1$, it is necessary to minimize the following criterion:

$$J = Z^T \Sigma^{-1} Z - \mu \left(Z^T Z - 1 \right) \quad (20)$$

where μ is the Lagrange multiplier. By using a partial derivate $\partial J / \partial Z$ and by equating it with zero, we obtain the following:

$$\partial J / \partial Z = 2\Sigma^{-1} Z - 2\mu Z \implies \Sigma Z = \lambda Z \quad (21)$$

where $\lambda = 1/\mu$. With the aim of obtaining a non-zero solution which satisfies the equation:

$$\Sigma Z = \lambda Z \iff (\Sigma - \lambda I) Z = 0 \quad (22)$$

it is further necessary to find such a parameter λ which satisfies the following so-called characteristic equation of a matrix Σ:

$$|\Sigma - \lambda I| = 0 \quad (23)$$

Every λ which satisfies this characteristic equation is known as eigenvalue of the matrix Σ while the vector Z related to specific eigenvalue is known as an eigenvector. When Σ is a symmetric $n \times n$ matrix, then there are n real eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ and n real eigenvectors $\Phi_1, \Phi_2, \ldots, \Phi_n$ which are mutually orthogonal and for which $\Sigma\Phi = \Phi\Lambda$ and $\Phi^T\Phi = I$ where $\Phi = [\Phi_1\ \Phi_2 \cdots \Phi_n]$ is the square matrix of the eigenvectors, Λ the diagonal matrix of the eigenvalues:

$$\Lambda = \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{bmatrix} \quad (24)$$

while I is the identity matrix.

If the matrix Φ is used as a transformation matrix during the linear transformation $Y = \Phi^T X$, then the covariance matrix of the random vector Y will be $\Sigma_Y = \Phi^T \Sigma_X \Phi = \Lambda$. This kind of transformation is orthonormal since for the transformation matrix Φ holds $\Phi^T\Phi = I$. In addition, during all these orthonormal transformations, the Euclidean distance does not change, that is $\|Y\|^2 = Y^T Y = X^T \Phi^T \Phi X = X^T X = \|X\|^2$.

Let X be an n-dimensional random vector of the extracted features which could be represented using n linear independent vectors in the following way:

$$X = \sum_{i=1}^{n} y_i \Phi_i = \Phi Y \quad (25)$$

where $\Phi = [\Phi_1\ \Phi_2\ \cdots\ \Phi_n]$ and $Y = \left[y_1\ y_2\ \cdots\ y_n\right]$ that is Φ_i are the basis vectors of the new n-dimensional space, and the new coordinates y_i are the scalar products of the basis vectors Φ_i and the random vector X. Assuming that the columns of the matrix Φ or in other words the basis vectors Φ_i are orthogonal, the coordinates of the random vector X in the new space can be obtained in the following way:

$$y_i = \Phi_i^T X. \quad (26)$$

Thus, Y represents a mapped random vector and the orthonormal transformation of the original random vector X. The random vector X approximated using only the m ($m < n$) basis vectors, i.e., the mapped features, could be represented in the following way:

$$\widehat{X}(m) = \sum_{i=1}^{m} y_i \Phi_i + \sum_{i=m+1}^{n} b_i \Phi_i \quad (27)$$

where the approximation error becomes:

$$\Delta X(m) = X - \hat{X}(m) = \sum_{i=m+1}^{n} (y_i - b_i)\Phi_i \quad (28)$$

and the mean squared error:

$$\bar{\varepsilon}^2(m) = E\left\{\left\|\Delta X(m)\right\|^2\right\} = \sum_{i=m+1}^{n} E\left\{(y_i - b_i)^2\right\} \tag{29}$$

has its own minimal value for $b_i = E\{y_i\} = \Phi_i^T E\{X\}$. The optimal mean squared error can then be presented in the following form:

$$\begin{aligned}\bar{\varepsilon}^2_{opt}(m) &= \sum_{i=m+1}^{n} E\left\{(y_i - E\{y_i\})^2\right\} \\ &= \sum_{i=m+1}^{n} \Phi_i^T E\left\{(X - E\{X\})(X - E\{X\})^T\right\}\Phi_i \\ &= \sum_{i=m+1}^{n} \Phi_i^T \Sigma_X \Phi_i = \sum_{i=m+1}^{n} \lambda_i \end{aligned} \tag{30}$$

where Σ_X is the covariance matrix of the random vector X and λ_i are its eigenvalues. Thus, the minimal mean squared error of approximation is also equal to the sum of the eigenvalues of the leftout coordinates, which actually means that we should leave out coordinates with the smallest eigenvalues. The mapping of the random vector X into the space made up by the eigenvectors of its covariance matrix Σ_X is known as the Karhunen-Loeve (KL) expansion. When reducing the dimension of the feature space using the KL expansion technique we should bear in mind that the performance of each feature is characterized by its eigenvalue. Thus, by rejecting features we should first reject those with the smallest eigenvalue, i.e., with the smallest variance in the new feature space. For example, in the case of dimension reduction from two to one shown in **Figure 3** the feature y_2 would be rejected as less informative even though it has better discriminatory potential than y_1. Also the coordinates y_i are mutually uncorrelated considering that the covariance matrix of the random vector Y is diagonal, i.e.,:

$$\Sigma_Y = \Phi^T \Sigma_X \Phi = \Lambda = diag\{\lambda_1 \lambda_2 \cdots \lambda_n\}. \tag{31}$$

Unlike the previously outlined method, the reduction of dimension based on scatter matrices (Fukunaga, 1990; Djurovic, 2006) is of special significance for the new detection technique since it takes into consideration the very purpose of the reduction, that is, the classification of the random vectors. Let L be the number of classes which should be classified and M_i and Σ_i, $i = 1 \cdots L$ the mean vectors and the covariance matrices of these classes, respectively. Then the within-class scatter matrix can be defined by:

$$S_W = \sum_{i=1}^{L} P_i E\left\{(X - M_i)(X - M_i)^T/\omega_i\right\} = \sum_{i=1}^{L} P_i \Sigma_i \tag{32}$$

and the between-class scatter matrix as:

$$S_B = \sum_{i=1}^{L} P_i (M_i - M_0)(M_i - M_0)^T \tag{33}$$

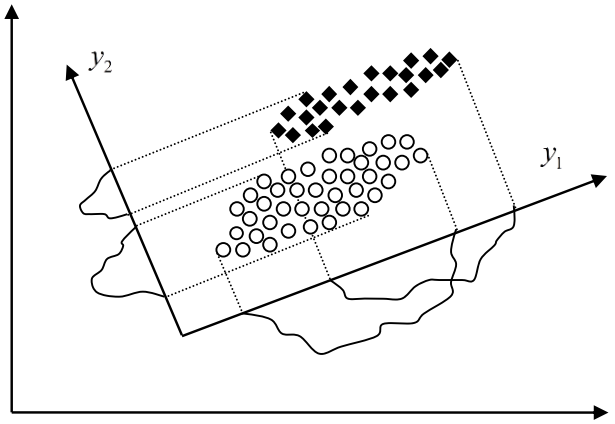


FIGURE 3 | Different approaches to dimension reduction in feature space, the KL expansion technique which rejects the feature y_2 and the technique based on scatter matrices which rejects the feature y_1.

where M_0 is the joint vector of mathematical expectation for all the classes together, that is:

$$M_0 = E\{X\} = \sum_{i=1}^{L} P_i M_i. \tag{34}$$

In addition the mixed scatter matrix can be defined by:

$$S_M = E\left\{(X - M_0)(X - M_0)^T\right\} = S_W + S_B. \tag{35}$$

Then the problem of dimension reduction is reduced to the identification of the $n \times m$ transformation matrix A which maps the random vector X of dimension n onto the random vector $Y = A^T X$ of dimension m and at the same time maximizes the criteria $J = tr(S_W^{-1} S_B)$. This criteria is invariant to non-singular linear transformations and results into transformation matrix that takes the following form:

$$A = [\Psi_1\ \Psi_2\ \cdots\ \Psi_m] \tag{36}$$

where Ψ_i, $i = 1, \ldots, m$ are the eigenvectors of the matrix $S_2^{-1} S_1$ which correspond to the greatest eigenvalues, i.e., $(S_W^{-1} S_B)\Psi_i = \lambda_i \Psi_i$, $i = 1, \ldots, n$, $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. Dimension reduction based on scatter matrices applied to the case shown in **Figure 3** would result into selection of the feature y_2 that is much better choice than the feature y_1 selected by the KL expansion technique, of course in terms of more accurate classification.

Design of Quadratic Classifiers

Quadratic classifiers are already known to be very good robust solutions to the problems of classification of random vectors whose statistical features are either unknown or change over time. Additionally, quadratic classifiers allow visual insight into the classification results. We design a piecewise quadratic classifier for detection of epileptiform activity, i.e., two quadratic

classifiers, able to separate all three classes of the EEG signals of interest as shown in **Figure 2**. The quadratic classifiers have the same structure defined by the following equation:

$$\begin{aligned} h(Y) &= Y^T Q Y + V^T Y + \nu_0 \\ &= \begin{bmatrix} y_1 & y_2 \end{bmatrix} \begin{bmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} + \begin{bmatrix} \nu_1 & \nu_2 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} + \nu_0 \end{aligned} \quad (37)$$

where y_1 and y_2 are two features in the reduced feature space. The matrix Q, the vector V and scalar ν_0 are the unknowns which are also need to be determined optimally. The quadratic equation (37) can be represented in a linear form as:

$$h(Y) = \begin{bmatrix} q_{11} & q_{12} & q_{22} & \nu_1 & \nu_2 \end{bmatrix} \begin{bmatrix} y_1^2 \\ 2y_1y_2 \\ y_2^2 \\ y_1 \\ y_2 \end{bmatrix} + \nu_0 = V_z^T Z + \nu_0. \quad (38)$$

In order to also achieve the largest possible between-class and shortest within-class scattering during the dimension reduction in the feature space, for the optimization criterion we have selected the following function (Fukunaga, 1990):

$$f = \frac{P_1 {\eta_1}^2 + P_2 {\eta_2}^2}{P_1 {\sigma_1}^2 + P_2 {\sigma_2}^2} \quad (39)$$

where P_1 and P_2 are probabilities and

$$\eta_l = E\left\{h(Z)/\omega_l\right\} = E\left\{V_z^T Z + \nu_0/\omega_l\right\} = V_z^T M_l + \nu_0 \quad (40)$$

$$\sigma_l^2 = var\left\{h(Z)/\omega_l\right\} = var\left\{V_z^T Z + \nu_0/\omega_l\right\} = V_z^T \Sigma_l V_z. \quad (41)$$

M_l and Σ_l are the mean vectors and covariance matrices, respectively, of the random vector Z for each of the two classes l that need to be classified. By optimizing the function f, for the optimal vector V_z, i.e., matrix Q and vector V from Equation (37), we have:

$$V_z = \begin{bmatrix} q_{11} \\ q_{12} \\ q_{22} \\ \nu_1 \\ \nu_2 \end{bmatrix} = \left[P_1 \Sigma_1 + P_2 \Sigma_2\right]^{-1} (M_2 - M_1) \quad (42)$$

and for the optimal scalar:

$$\nu_0 = - V_z^T (P_1 M_1 + P_2 M_2) \quad (43)$$

which finishes the design of the quadratic classifiers as well as the new technique for detection of epileptiform activity.

Statistical performances such as sensitivity, specificity and accuracy of the designed piecewise quadratic classifier, i.e., the new technique for detection of epileptiform activity, is estimated based on the classification results. The sensitivity is defined as a ratio between the number of correctly classified segments and the total number of the segments for each of the classes separately. The specificity is also calculated for each of these three classes separately and represents the ratio between the number of correctly classified features of the other two classes and the total number of the segments of these two classes. The accuracy is calculated as the ratio between the total number of correctly classified segments and the total number of the segments in all three classes together.

Results

Feature Extraction

In total 30 features for each of 300 analyzed segments of the EEG signals were extracted. All the features together with their mean values and standard deviations for all three different classes of EEG signals of interest are presented in **Table 1**. The extracted features refer to the adequate clinical sub-bands since these sub-bands had better discrimination characteristics compared with the whole frequency band between 0 and 60 Hz. The separability index as a measure of the discriminatory potential was also calculated for all the extracted features. In this case, the separability index is the criteria $J = tr(S_W^{-1} S_B)$ where S_W and S_B are earlier defined within- and between-class scatter matrices, respectively. Based on these matrices, a higher separability index corresponds to better separability between different classes of the EEG signals. Based on these 30 features, each original segment of the EEG signals from time domain can be presented now by its feature vector $X = [x_1 x_2 \cdots x_{30}]^T$, i.e., by the point in the feature space with dimension of 30.

The total variation is the only one feature that we extracted in the time domain. In **Table 1**, it can be noticed that the total variation has a certain potential for the detection of epileptiform activity in EEG signals. However, the total variation is not that much reliable despite the fact that is a pretty well estimated having in mind the duration of each of the analyzed segments.

The periodogram represents a very important feature of the signal in the frequency domain given that based on it we can get a relative contribution of either any individual frequency or a specific frequency band to the total power of the analyzed signal. The periodograms of one epileptic and two non-epileptic (from both unhealthy and healthy tissue) segments of the EEG signals are shown in **Figure 4** where it can be noticed that the EEG signal power of is shifting from lower to higher frequencies in the presence of epileptiform activity.

Using the discrete wavelet transform (DWT) we can completely and independently extract higher and lower frequencies from the signal. All that can be done with different resolution in the time domain, i.e., higher resolution in the time domain for higher frequencies and lower resolution in the time domain for lower frequencies. The EEG signal segments were analyzed at four levels, i.e., the discrete wavelet decomposition was performed at four levels as presented in **Figure 5**. At the first level of decomposition, the original frequency band of the EEG signals (0–60 Hz) was divided into its higher (30–60 Hz) and lower part (0–30 Hz), i.e., the details and the approximation of the signals at the first decomposition level, respectively. Then at the second decomposition level, the frequency band of the approximation from the first level was additionally divided into its higher (15–30 Hz) and lower (0–15 Hz) part, i.e., the

TABLE 1 | Normalized features extracted from different frequency sub-bands.

Index	Feature	Non-epileptic of healthy tissue		Non-epileptic of unhealthy tissue		Epileptic		Separa-bility
		μ	σ	μ	σ	μ	σ	J
x_1	Total variation—delta	0.011	0.002	0.011	0.003	0.019	0.005	1.253
x_2	Total variation—theta	0.027	0.004	0.022	0.006	0.028	0.006	0.300
x_3	Total variation—àlpha	0.044	0.005	0.034	0.011	0.042	0.011	0.215
x_4	Total variation—beta	0.075	0.008	0.057	0.024	0.062	0.023	0.150
x_5	Total variation—gamma	0.149	0.019	0.102	0.047	0.103	0.041	0.335
x_6	Relative power FFT—delta	0.446	0.090	0.628	0.147	0.267	0.220	0.720
x_7	Relative power FFT—theta	0.159	0.049	0.236	0.119	0.390	0.224	0.417
x_8	Relative power FFT—alpha	0.162	0.043	0.086	0.066	0.134	0.057	0.316
x_9	Relative power FFT—beta	0.221	0.075	0.046	0.024	0.205	0.151	0.641
x_{10}	Relative power FFT—gamma	0.012	0.010	0.004	0.003	0.004	0.005	0.264
x_{11}	St. dev. coeff. DWT—delta	2.825	0.275	3.362	0.290	2.507	0.549	0.810
x_{12}	St. dev. coeff. DWT—theta	1.795	0.180	1.709	0.366	2.181	0.505	0.300
x_{13}	St. dev. coeff. DWT—alpha	1.266	0.140	0.766	0.175	1.275	0.288	1.276
x_{14}	St. dev. coeff. DWT—beta	0.556	0.122	0.267	0.072	0.466	0.146	1.057
x_{15}	St. dev. coeff. DWT—gamma	0.154	0.039	0.085	0.028	0.115	0.040	0.596
x_{16}	Relative power DWÒ—delta	0.501	0.097	0.708	0.118	0.408	0.175	0.873
x_{17}	Relative power DWÒ—theta	0.203	0.039	0.190	0.081	0.311	0.132	0.347
x_{18}	Relative power DWÒ—alpha	0.202	0.043	0.077	0.035	0.213	0.097	0.913
x_{19}	Relative power DWÒ—beta	0.081	0.038	0.020	0.011	0.060	0.039	0.613
x_{20}	Relative power DWÒ—gamma	0.013	0.007	0.005	0.003	0.008	0.006	0.291
x_{21}	Correlation dimension—delta	6.979	3.443	6.494	1.605	5.763	1.489	0.045
x_{22}	Correlation dimension—theta	4.621	0.594	4.288	0.925	4.206	0.884	0.048
x_{23}	Correlation dimension—alpha	4.184	0.442	3.701	0.886	3.230	0.833	0.272
x_{24}	Correlation dimension—beta	3.635	0.359	3.097	0.940	2.348	0.832	0.490
x_{25}	Correlation dimension—gamma	6.729	1.248	6.374	1.838	4.003	1.994	0.493
x_{26}	Largest Lyapunov exp.—delta	3.282	0.873	2.910	0.856	4.203	1.102	0.327
x_{27}	Largest Lyapunov exp.—theta	8.213	1.935	8.188	1.914	8.286	1.933	0.000
x_{28}	Largest Lyapunov exp.—alpha	17.58	2.165	17.57	2.160	17.58	2.377	0.000
x_{29}	Largest Lyapunov exp.—beta	32.91	5.991	32.65	5.977	33.04	5.091	0.001
x_{30}	Largest Lyapunov exp.—gamma	11.71	2.985	11.62	2.965	11.89	5.210	0.001

details and the approximation of the signals at the second decomposition level, respectively. After all four decomposition levels, the original band was divided into its five sub-bands, i.e., four sub-bands with the details and one sub-band with the approximation. All these five sub-bands approximately correspond to the earlier defined clinical sub-bands. Power distribution of the EEG signals in the time-frequency domain is quite well described by the DWT coefficients. However, in order to reduce the dimension of the problem and make easier further classification we calculated certain statistics of these coefficients in each sub-band such as the standard deviation and the average relative power, i.e., the square of the absolute values of the DWT coefficients.

Given that the EEG signal also roughly represents a dynamics of a very complex non-linear system such as the brain, the non-linear analysis based on the chaos theory was used in order to extract the information that could not been extracted by any of previously described linear techniques. It is interesting to see that unlike the other feature extraction techniques in the field, a complete agreement about if at all and how to perform a non-linear analysis of the EEG signals has not been achieved yet. Thus, quite often it is possible to find contradictory results of such experiments in the literature. For example, the correlation dimension and the largest Lyapunov exponent have completely different values in Hively et al. (1999), Adeli and Ghosh-Dastidar (2010) and Iasemidis and Sackellares (1991). The feature extraction techniques and non-linear analysis implemented and used in this research are exclusively based on the chaos theory described in the methods part. In addition, there are no any further subjective adjustments applied to the EEG signals, which provides a high level of reproductivity of the obtained results at any time.

At first, the optimal lag and the embedding dimension were determined in order to reconstruct a segment of the EEG signals in its own lagged phase space. The optimal lag m_o was obtained as the first local minimum of the function of the mutual information

coefficients. The value of the optimal lag of the most of analyzed segments varied between 5 and 7. The minimum embedding dimension d_{min} was determined using Cao's technique, i.e., based on the saturation of the embedding function e_d, for example as presented in **Figure 6** in the case of one segment. In other words when a further increase in the embedding dimension does not result in more than 5% of increase in the embedding function. The value of the embedding function of all 300 segments processed approached 1. In fact, this confirms that there is a certain level of chaos present in the segments of the EEG signals. That chaos is not random but deterministic given that the value of the redefined embedding function e_d^* is not constant for all values of the embedding dimension as it can be seen in **Figure 6**. The value of the minimum embedding dimension varied between 4 and 10.

After reconstruction of the EEG signals in the lagged phase space, the correlation dimension of attractor was estimated using

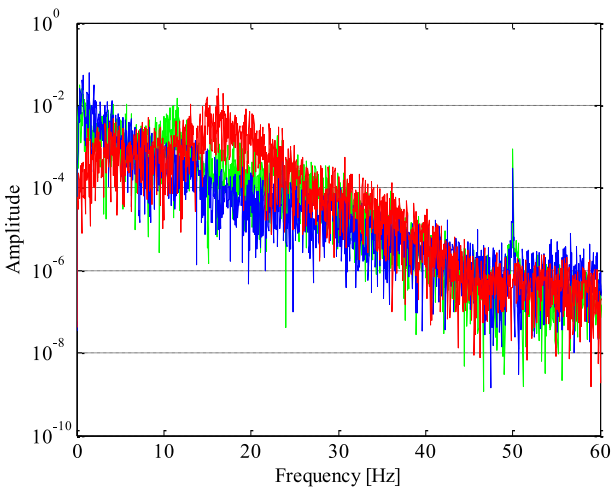


FIGURE 4 | Periodogram of epileptic (in red) and non-epileptic (unhealthy in blue and healthy tissue in green) segments of EEG signals where a shift in the EEG signal power from lower to higher frequencies in the presence of epileptiform activity is evident.

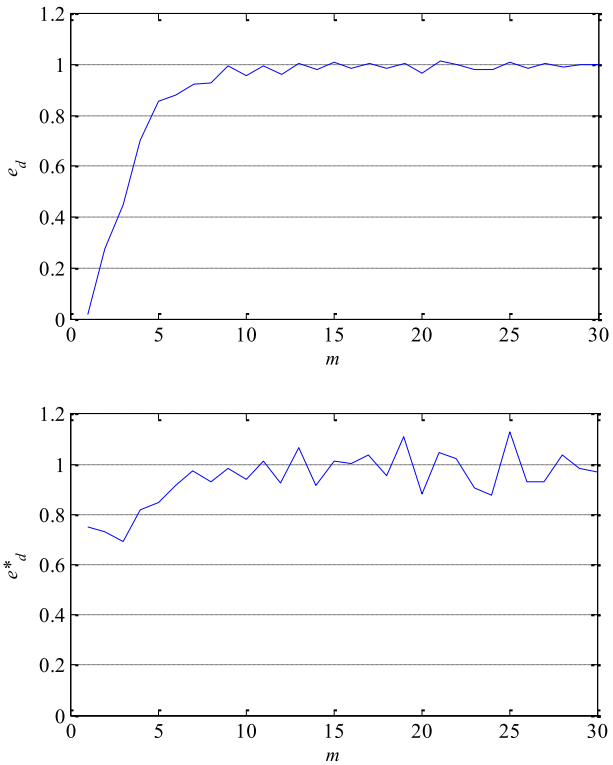


FIGURE 6 | Embedding function e_d (upper) which approaches 1 and thus confirms a presence of a certain level of chaos in EEG signals and redefined embedding function e_d^* (lower) which is not constant for all values of the embedding dimension *m* confirming that chaos is not random but deterministic.

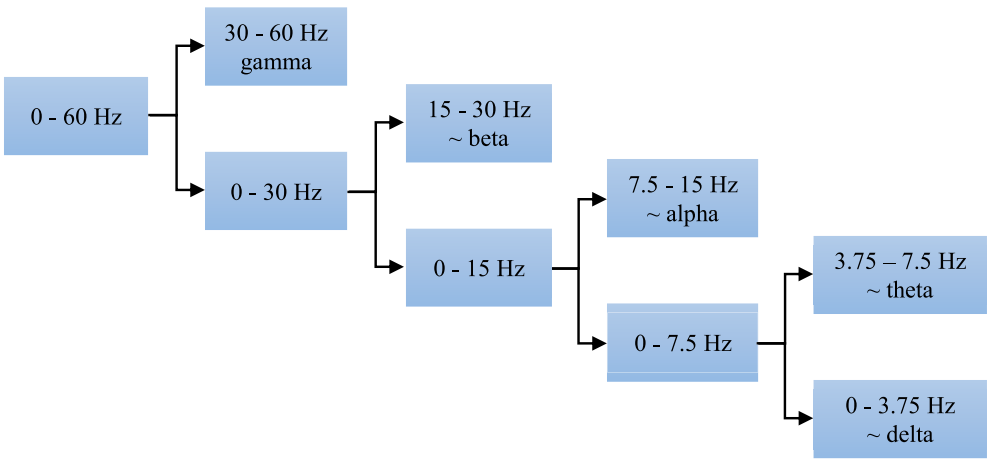


FIGURE 5 | Four-level decomposition of EEG signal that corresponds to five sub-bands of clinical interest which have better discriminatory characteristics compared with the entire frequency band of 0–60 Hz.

the Taken's estimator. After a few tests the value of radius ε in the phase space was set to 5% of the total size of the attractor since the higher values resulted into to many points, and the smaller ones into insufficient number of points for a good estimation of the correlation dimension. From **Table 1**, it can be concluded that the correlation dimension as a non-linear feature has a potential for detection of epileptiform activity in EEG signals. It is also obvious that the attractor complexity, i.e., the chaotic behavior of the EEG signals, is lower in presence of epileptiform activity. The values of the correlation dimension in all cases were higher than the embedding dimension of the lagged phase space, which is in accordance with the chaos theory.

The largest Lyapunov exponent as a measure of signal predictability was estimated using Sato's technique. At first, the prediction error as a function of number of samples k was determined as shown in **Figure 7** in the case of one segment. Then, the largest Lyapunov exponent was estimated based on the function's slope in its medium part. As it can be seen in **Table 1**, the largest Lyapunov exponent has smaller discrimination ability compared with the correlation dimension. Additionally, it can be also noticed that the presence of epileptiform activity reduces the predictability of the EEG signals since the largest Lyapunov exponent is slightly higher in that case.

Dimension Reduction in Feature Space

After the feature extraction from all the segments of the EEG signals, obviously none of the individually extracted features is sufficiently reliable for detection of epileptiform activity in EEG signals. This fact represents the main reason to perform the feature extraction in a few different domains of interest, i.e., time, frequency, time-frequency domain and non-linear analysis. The assumption is that the each of them contains some new information about the EEG signal, i.e., the information which is not present in any other domain and thus later contributes to more accurate classification and detection. Therefore, a better separability between the classes of epileptic and non-epileptic segments is expected after an optimal combination of the features from different domains than in the case of using only features from one domain as it is the case with almost all the literature in the field.

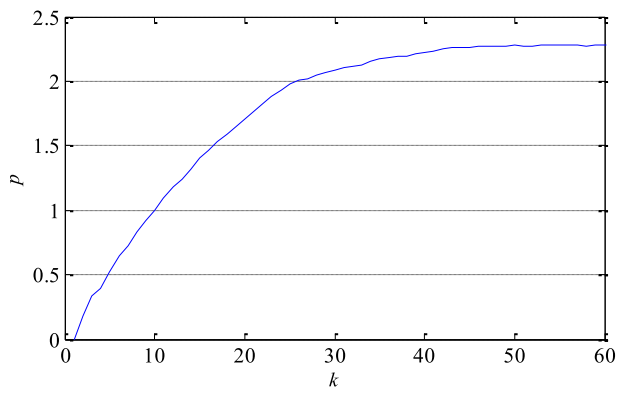


FIGURE 7 | Prediction error *p* of one segment of EEG signal as a function of the number of samples *k*. Its slope in the middle part determines the largest Lyapunov exponent as a measure of the exponential divergence of nearby phase space trajectories.

Both the KL expansion technique and the dimension reduction technique based on the scatter matrices were tested on the features from all the domains. The obtained results, i.e., adequate separability indexes before and after the dimension reduction in the feature space are presented in **Table 2**. The reduction technique based on the scatter matrices gives better results in all the domains of interest and also results into the separability index that is, as expected, greater than any individual separability index given in **Table 1**.

In **Table 2**, one can see that out of all the analyzed features, the highest separability index and the best discrimination characteristics between epileptic and non-epileptic segments have the features obtained in time-frequency domain after the DWT. However, the other features despite their lower separability indexes are also useful for later classification that is concluded based on an additional analysis whose results are presented in **Table 3**. It can be noticed that starting from the features in time domain the separability index increases by a gradual inclusion of the features from other domains.

Unlike the previous figures, **Figure 8** shows 50 original nineteen-dimensional feature vectors X, which correspond to 50 segments from each of the three classes of the EEG signals, mapped into their new reduced two-dimensional feature space. All these 150 two-dimensional vectors Y will be later used in the next section for the design of appropriate classifiers while the rest of 150 segments and their corresponding feature vectors will be used to test the performance of the designed classifiers as well as the total accuracy of the new technique for detection of epileptiform activity in EEG signals.

TABLE 2 | Separability indexes after application of two different techniques for dimension reduction in feature space.

Features analyzed	Dimension		Separability index	
	Before	After	KL expansion	By the scatter matrices
Time domain (x_{1-5})	5	2	1.93	2.13
Frequency domain (x_{6-10})	5	2	1.25	2.16
Time-frequency domain (x_{11-15})	10	2	1.40	4.78
Non-linear analysis (x_{16-20})	10	2	1.07	1.15

TABLE 3 | Separability indexes after the reduction based on the scatter matrices and gradual involvement of features from different domains.

Features analyzed	Dimension		Separability index
	Before	After	
Time domain (x_{1-5})	5	2	2.13
Including frequency domain (x_{1-10})	10	2	3.52
Including time-frequency domain (x_{1-20})	20	2	6.74
Including non-linear analysis (x_{1-30})	30	2	8.78

Classification

After the reduction of the feature space dimension to two, the next step is the design of appropriate classifiers that can separate epileptic from non-epileptic segments of the EEG signals in the reduced feature space shown in **Figure 8**. This represents the last step in design of the new technique for detection of epileptiform activity in EEG signals. Having in mind the nature of the EEG signals and possible changes in their statistical properties it is very desired to use robust classifiers. Based on **Figure 8** it can be concluded that quadratic classifiers represent quite logical choice for classification even though these three classes of the EEG signals are also piecewise linearly separable but with a much higher classification error. In total two quadratic classifiers were designed following the procedure described in Section Design of Quadratic Classifiers.

As it can be seen in **Figure 9**, the first classifier separates the non-epileptic segments of the EEG signals of healthy brain tissue (in green) from the non-epileptic segments of unhealthy tissue (in blue) as well as from the epileptic segments (in red). This classifier is defined using the following equation:

$$h(Y) = \sum_{i=1}^{2}\sum_{j=1}^{2} q_{ij} y_i y_j + \sum_{i=1}^{2} \nu_i y_i + \nu_0 \qquad (44)$$

where the unknown parameters are $q_{11} = -4870.8$, $q_{12} = q_{21} = -239.9$, $q_{22} = -174.9$, $\nu_1 = -29.2$, $\nu_2 = -174.9$ and $\nu_0 = -2.3$. After that, the second classifier which separates the remaining two unseparated classes of the EEG signals segments, i.e., the epileptic and the non-epileptic segments of unhealthy brain tissue, was designed. The parameters of the Equation (44) for this classifier are $q_{11} = -436.7$, $q_{12} = q_{21} = -128.2$, $q_{22} = 444.6$, $\nu_1 = -237.9$, $\nu_2 = -57.2$ and $\nu_0 = 0.5$ while the classifier itself is shown in **Figure 10**.

The performance of the designed classifiers and thus the new technique for detection of epileptiform activity in EEG signals was tested by classification of the remaining 150 segments which were not previously used during the design procedure. The obtained results are presented in **Figure 11**, where the piecewise quadratic classifier is just a combination of two quadratic classifiers.

The classification results can also be represented by a confusion matrix that is given in **Table 4**, where its each cell contains number of classified features for each combination of three classes of the EEG signals segments. Based on the confusion matrix and **Figure 11**, it can be concluded that all the non-epileptic segments of healthy tissue were correctly classified.

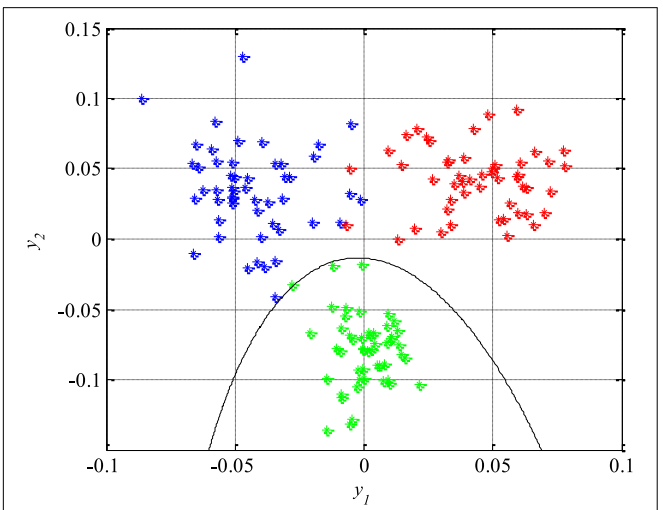


FIGURE 9 | The first quadratic classifier which separates non-epileptic EEG signals of healthy tissue (in green) from non-epileptic (in blue) and epileptic EEG signals of unhealthy tissue (in red) during the design and training phase.

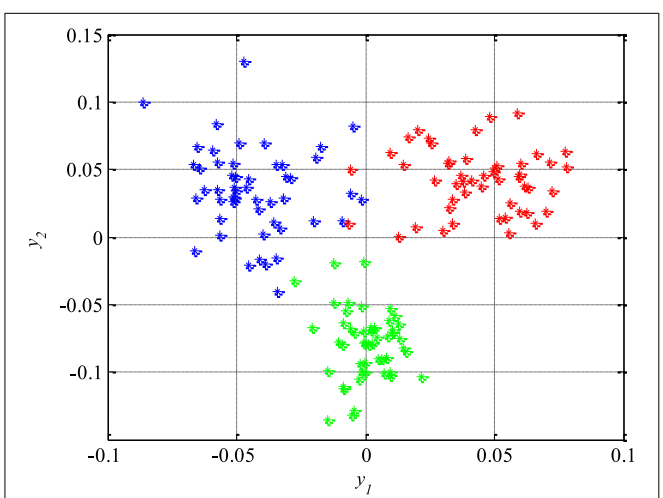


FIGURE 8 | Epileptic (in red) and non-epileptic (unhealthy in blue and healthy tissue in green) EEG signals in a new two-dimensional feature space after dimension reduction based on scatter matrices.

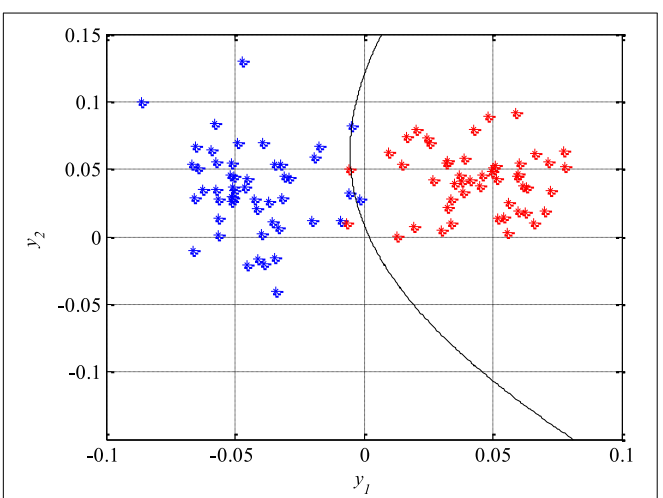


FIGURE 10 | The second quadratic classifier which separates epileptic (in red) from non-epileptic EEG signals of unhealthy tissue (in blue) during the design and training phase.

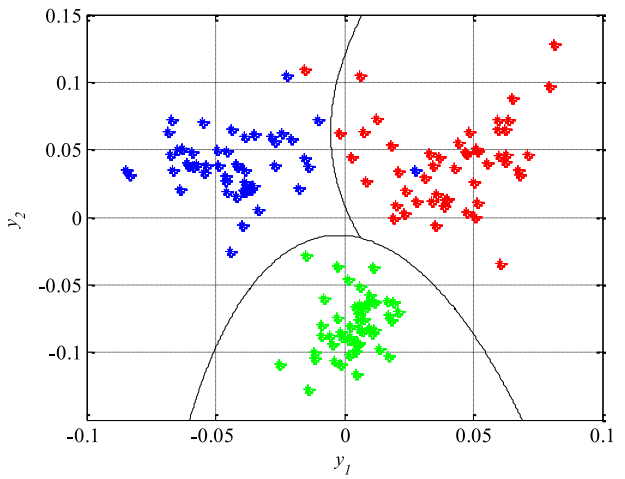


FIGURE 11 | Piecewise quadratic classifier which separates epileptic (in red) from non-epileptic (unhealthy in blue and healthy in green) EEG signals of the test set.

TABLE 4 | Confusion matrix.

EEG signals (input/output)	Non-epileptic		Epileptic
	Healthy	Unhealthy	
Non-epileptic of healthy brain tissue	50	0	0
Non-epileptic of unhealthy brain tissue	0	49	1
Epileptic	0	1	49

TABLE 5 | Statistical performances.

EEG signals	Statistical performances [%]		
	Sensitivity	Specificity	Accuracy
Non-epileptic of healthy brain tissue	100	100	98.7
Non-epileptic of unhealthy brain tissue	98	99	
Epileptic	98	99	

However, the remaining two classes contained one segment each which was incorrectly classified, i.e., classified as it belongs to the other class. The statistical performances such as sensitivity, specificity and accuracy, of the designed piecewise quadratic classifiers are presented in **Table 5**. As it can be seen, the total accuracy of the new technique for detection of epileptiform activity in EEG signals is 98.7%. Typically, quadratic classifiers are robust and do not exhibit overtraining when the number of parameters to be estimated is much less than the number of samples as in this case. Anyway, it is a good practice to cross validate this piecewise classifier in order to ensure its stability. A five-fold cross validation was performed and it resulted in the cross-validation loss, i.e., the error of the out-of-fold samples, of 1.7%. Even though it is slightly higher than the classification error of 1.3% it gives a confidence that the classifier is reasonably accurate.

Discussion

Having in mind the results of other techniques available in the literature, presented in **Table 6** and tested on the identical segments of the EEG signals, the new technique demonstrated a very good performance. The accuracy of the other techniques varied between 85 and 99%. In addition to high accuracy achieved, it should also be emphasized that all the segments of the analyzed EEG signals were normalized before the feature extraction. In that way we managed to overcome one of the main disadvantages of the techniques from **Table 6** in terms of real clinical application, i.e., those techniques rely on the amplitude of the EEG signals as one of the key discriminatory features. However, the EEG signal amplitude has been found as unreliable in real clinical applications since it varies significantly even with healthy individuals, depending on other brain activities as well as other activities of human body. Also, some other undesired effects, e.g., different electrodes used for recording, different patients and their brain tissues, on the detection technique has also been removed by normalization. Unlike the techniques from **Table 6**, which are mainly based only on features from one of the domains, the new technique relies on carefully extracted features from all the domains of interest including non-linear analysis as well. Because of that, this technique is more robust and less sensitive on changes in the EEG signals that dominantly impact the features from one or two domains while at the same time are invisible in other domains and do not have any relation with a presence of epileptiform activity in EEG signals to be detected.

In order to further increase the detection accuracy of the new technique during its real clinical application, a previous elimination of artifacts is very desirable immediately after acquisition of the EEG signals, i.e., before any further processing and feature extraction. The artifacts removal can be performed very reliably using some of already developed and available techniques (Hyvarinen et al., 2001; Rosso et al., 2002). In addition, it is also necessary to make a certain compromise in terms of duration of the segments to be sequentially analyzed in real time. The segment duration should be subsequently adjusted depending on both application and patient. Not only during the feature extraction and the dimension reduction in the feature space, but also during the design of classifiers, a special attention has been paid to the robustness of the detection technique. This resulted in the choice of quadratic classifiers which in addition to their simplicity are known for a high level of robustness in the applications of this type. Quadratic classifiers have also one more important feature that is possibility of visualization of the classification results in two-dimensional space. Namely, despite the fact that the mapped features y_1 and y_2 as a linear combination of the original features x_i extracted from the different domains cannot be anymore associated to certain properties of the EEG signals, they still can provide some further useful insights. For example, in **Figure 11** it can be noticed that the feature y_1 can help during determination of the damage level of the brain tissue, while the feature y_2 indicates either presence or absence of epileptic EEG signal.

As part of our future work we plan an additional testing on other bigger and mainly commercially available data bases of the EEG signals (e.g., http://epilepsy-database.eu) containing much

TABLE 6 | Other techniques for detection of epileptic EEG signals.

Authors and year	Feature extraction	Classification	Accuracy
Nigam and Graupe, 2004	Non-linear filter	Diagnostic neural networks	97.2
Kannathal et al., 2005a	Non-linear analysis	Surrogate data analysis	90.0
Kannathal et al., 2005b	Entropy	Adaptive neuro-fuzzy inference system	92.2
Guler and Ubeyli, 2005	Lyapunov exponents	Recurrent neural networks	96.8
Ubeyli, 2006	Lyapunov exponents	Artificial neural networks	95.0
Sadati et al., 2006	Wavelet transform	Adaptive neuro-fuzzy network	85.9
Subasi, 2007b	Wavelet transform	Expert models	95.0
Tzallas et al., 2007	Time-frequency domain analysis	Artificial neural networks	99.3
Chua et al., 2008	Power spectral density	Gaussian mixture model	93.1
Ghosh-Dastidar et al., 2008	Principal component analysis	Artificial neural networks	99.3
Ocak, 2008	Wavelet transform, approximate entropy and genetic algorithm	Learning vector quantization	98.0
Mousavi et al., 2008	Wavelet transform and autoregressive model	Artificial neural networks	96.0
Ubeyli, 2008	Wavelet transform	Expert models	93.2
Chandaka et al., 2009	Crosscorrelation	Support vectro machines	96.0
Ocak, 2009	Wavelet transform and approximate entropy	Surrogate data analysis	96.7
Guo et al., 2009	Wavelet transform and relative wavelet energy	Artificial neural networks	95.2
Naghsh-Nilchi and Aghashahi, 2010	Eigenvector methods	Artificial neural networks	97.5
Guo et al., 2011	Genetic programming	*K*-nearest neighbor classifier	93.5
Orhan et al., 2011	Wavelet transform	Cauterization and artificial neural networks	96.7
Gajić et al., 2014	Wavelet transform and dimension reduction based on scatter matrices	Quadratic classifiers	99.0

more interictal, preictal and ictal EEG data with the aim of further development and adaptation of the new technique for use in a real clinical environment. We will also try to access its potential in the field of emotion detection (e.g., happiness, sadness, depression, alertness, etc.) as well as detection of abnormal activities associated with some other brain disorders such as Alzheimer's disease and schizophrenia.

Acknowledgments

The support from the Marie Curie FP7-ITN InnHF, Contract No: PITN-GA-2011- 289837 and the Erasmus Mundus Action II EUROWEB Project, Contract No: 204625-1-2011-1-SE-ERA MUNDUS-EMA21 is gratefully acknowledged.

References

Adeli, H., and Ghosh-Dastidar, S. (2010). *Automated EEG-Based Diagnosis of Neurological Disorders*. Boca Raton, FL: CRC Press.

Altunay, S., Telatar, Z., and Erogul, O. (2010). Epileptic EEG detection using the linear prediction error energy. *Expert. Syst. Appl.* 37, 5661–5665. doi: 10.1016/j.eswa.2010.02.045

Andrzejak, R., Lehnertz, K., Mormann, F., Rieke, C., David, P., and Elger, C. (2001). Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: dependence on recording region and brain state. *Phys. Rev. E* 64:061907. doi: 10.1103/PhysRevE.64.061907

Cao, L. (1997). Practical method for determining the minimum embedding dimension of a scalar time series. *Phys. D* 110, 43–50. doi: 10.1016/S0167-2789(97)00118-8

Casson, A., Yates, D., Smith, S., Duncan, J., and Rodriguez-Villegas, E. (2010). Wearable electroencephalography. What is it, why is it needed, and what does it entail? *IEEE Eng. Med. Biol. Mag.* 29, 44–56. doi: 10.1109/MEMB.2010.936545

Chandaka, S., Chatterjee, A., and Munshi, S. (2009). Cross-correlation aided support vector machine classifier for classification of EG signals. *Expert Syst. Appl.* 36, 1329–1336. doi: 10.1016/j.eswa.2007.11.017

Chua, K. C., Chandran, V., Acharya, R., and Lim, C. M. (2008). Automatic identification of epilepsy by HOS and power spectrum parameters using EEG signals: a comparative study. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 2008, 3824–3827. doi: 10.1109/IEMBS.2008.4650043

Djurovic, Z. (2006). *Practicum on Pattern Recognition*. Belgrade: School of Electrical Engineering.

Fukunaga, K. (1990). *Introduction to Statistical Pattern Recognition*. Boston, MA: Academic Press.

Gajić, D. (2007). *M.Sc. Thesis: Detection of Epileptic Seizures using Wavelet Transform and Fuzzy Logic (in Serbian)*. Belgrade: School of Electrical Engineering.

Gajić, D., Djurovic, Z., Di Gennaro, S., and Gustafsson, F. (2014). Classification of EEG signals for detection of epileptic seizures based on wavelets and statistical pattern recognition. *Biomed. Eng. Appl. Basis. Commun.* 26, 1450021. doi: 10.4015/S1016237214500215

Ghosh-Dastidar, S., Adeli, H., and Dadmehr, N. (2008). Principal component analysisenhanced cosine radial basis function neural network for robust epilepsy and seizure detection. *IEEE Trans. Biomed. Eng.* 55(2 Pt 1), 512–518. doi: 10.1109/TBME.2007.905490

Gotman, J. (1999). Automatic detection of seizures and spikes. *J. Clin. Neurophysiol.* 16, 130–140. doi: 10.1097/00004691-199903000-00005

Guler, I., and Ubeyli, E. D. (2005). Adaptive neuro-fuzzy inference system for classification of EEG signals using wavelet coefficients. *J. Neurosci. Methods* 148, 113–121. doi: 10.1016/j.jneumeth.2005.04.013

Guo, L., Rivero, D., Dorado, J., Munteanu, C. R., and Pazos, A. (2011). Automatic feature extraction using genetic programming: an application to epileptic EEG classification. *Expert Syst. Appl.* 38, 10425–10436. doi: 10.1016/j.eswa.2011.02.118

Guo, L., Rivero, D., Seoane, J. A., and Pazos, A. (2009). "Classification of EEG signals using relative wavelet energy and artificial neural networks," in *Proceedings of the 1st ACM/SIGEVO Summit on Genetic and Evolutionary Computation* (Shanghai).

Hazarika, N., Chen, J., Tsoi, A., and Sergejew, A. (1997). Classification of EEG signals using the wavelet transform. *Signal Process* 59, 61–72. doi: 10.1016/S0165-1684(97)00038-8

Hively, L. M., Gailey, P. C., and Protopopescu, V. A. (1999). Detecting dynamical change in nonlinear time series. *Phys. Lett. A* 258, 103–114. doi: 10.1016/S0375-9601(99)00342-4

Hyvarinen, A., Karhunen, J., and Oja, E. (2001). *Independent Component Analysis*. New York, NY: John Wiley and Sons.

Iasemidis, L. D., and Sackellares, J. C. (1991). "The temporal evolution of the largest Lyapunov exponent on the human epileptic cortex," in *Measuring Chaos in the Human Brain*, eds D.W. Duke and W.S. Pritchard (Singapore: World Scientific), 49–82.

Iasemidis, L., Shiau, D., Chaovalitwongse, W., Sackellares, J., Pardalos, P., Principe, J., et al. (2003). Adaptive epileptic seizure prediction system. *IEEE Trans. Biomed. Eng.* 50, 616–627. doi: 10.1109/TBME.2003.810689

Iscan, Z., Dokur, Z., and Tamer, D. (2011). Classification of electroencephalogram signals with combined time and frequency features. *Expert Syst. Appl.* 38, 10499–10505. doi: 10.1016/j.eswa.2011.02.110

Jerger, K., Netoff, T., Francis, J., Sauer, T., Pecora, L., Weinstein, S., et al. (2001). Early seizure detection. *J. Clin. Neurophysiol.* 18, 259–268. doi: 10.1097/00004691-200105000-00005

Jerger, K., Weinstein, S., Sauer, T., and Schiff, S. (2005). Multivariate linear discrimination of seizures. *Clin. Neurophysiol.* 116, 545–551. doi: 10.1016/j.clinph.2004.08.023

Kalayci, T., and Özdamar, Ö. (1995). Wavelet preprocessing for automated neural network detection of EEG spikes. *IEEE Eng. Med. Biol.* 14, 160–166. doi: 10.1109/51.376754

Kannathal, N., Acharya, U. R., Lim, C. M., and Sadasivan, P. K. (2005a). Characterization of EEG-a comparative study. *Comput. Methods Programs Biomed.* 80, 17–23. doi: 10.1016/j.cmpb.2005.06.005

Kannathal, N., Choo, M. L., Acharya, U. R., and Sadasivan, P. K. (2005b). Entropies for detection of epilepsy in EEG. *Comput. Methods Programs Biomed.* 80, 187–194. doi: 10.1016/j.cmpb.2005.06.012

Liang, S. F., Wang, H. C., and Chang, W. L. (2010). Combination of EEG complexity and spectral analysis for epilepsy diagnosis and seizure detection. *EURASIP J. Adv. Signal Process* 2010:853434. doi: 10.1155/2010/853434

Mousavi, S. R., Niknazar, M., and Vahdat, B. V. (2008). "Epileptic seizure detection using AR model on EEG signals," in *Proceedings of the International Biomedical Engineering Conference* (Cairo).

Naghsh-Nilchi, A. R., and Aghashahi, M. (2010). Epilepsy seizure detection using eigensystem spectral estimation and Multiple Layer Perceptron neural network. *Biomed. Signal Process* 5, 147–157. doi: 10.1016/j.bspc.2010.01.004

Niederhauser, J., Esteller, R., Echauz, J., Vachtsevanos, G., and Litt, B. (2003). Detection of seizure precursors from depth EEG using a sign periodogram transform. *IEEE Trans. Biomed. Eng.* 51, 449–458. doi: 10.1109/TBME.2003.809497

Nigam, V. P., and Graupe, D. (2004). A neural-network-based detection of epilepsy. *Neurol. Res.* 26, 55–60. doi: 10.1179/016164104773026534

Ocak, H. (2008). Optimal classification of epileptic seizures in EEG using wavelet analysis and genetic algorithm. *Signal Process* 88, 1858–1867. doi: 10.1016/j.sigpro.2008.01.026

Ocak, H. (2009). Automatic detection of epileptic seizures in EEG using discrete wavelet transform and approximate entropy. *Expert Syst. Appl.* 36, 2027–2036. doi: 10.1016/j.eswa.2007.12.065

Orhan, U., Hekim, M., and Ozer, M. (2011). EEG signals classification using the K-means clustering and a multilayer perceptron neural network model. *Expert Syst. Appl.* 38, 13475–13481. doi: 10.1016/j.eswa.2011.04.149

Petrosian, A., Prokhorov, D., Homan, R., Dasheiff, R., and Wunsch, D. (2000). Recurrent neural network based prediction of epileptic seizures in intra- and extracranial EEG. *Neurocomputing* 30, 201–218. doi: 10.1016/S0925-2312(99)00126-5

Polat, K., and Gunes, S. (2007). Classification of epileptiform EEG using a hybrid system based on decision tree classier and fast fourier transform. *Appl. Math. Comp.* 187, 1017–1026. doi: 10.1016/j.amc.2006.09.022

Proakis, J., and Manolakis, D. (1996). *Digital Signal Processing: Principles, Algorithms, and Applications*. Upper Saddle River, NJ: Prentice Hall.

Rao, R., and Bopardikar, A. (1998). *Wavelet Transforms: Introduction to Theory and Applications*. Boston, MA: Addison-Wesley Longman.

Rosenstein, M. T., Collins, J. J., and De Luca, C. J. (1993). A practical method for calculating largest Lyapunov exponents from small data sets. *Phys. D* 65, 117–134. doi: 10.1016/0167-2789(93)90009-P

Rosso, O., Martin, M., and Plastino, A. (2002). Brain electrical activity analysis using wavelet-based informational tools. *Phys. A* 313, 587–608. doi: 10.1016/S0378-4371(02)00958-5

Sadati, N., Mohseni, H. R., and Magshoudi, A. (2006). "Epileptic seizure detection using neural fuzzy networks," in *Conference Proceeding IEEE International Conference on Fuzzy System* (Vancouver, BC). doi: 10.1109/FUZZY.2006.1681772

Srinivasan, V., Eswaran, C., and Sriraam, N. (2007). Approximate entropy-based epileptic EEG detection using artificial neural networks. *IEEE Trans. Inf. Technol. Biomed.* 11, 288–295. doi: 10.1109/TITB.2006.884369

Subasi, A. (2007a). Application of adaptive neuro-fuzzy inference system for epileptic seizure detection using wavelet feature extraction. *Comput. Biol. Med.* 37, 227–244. doi: 10.1016/j.compbiomed.2005.12.003

Subasi, A. (2007b). EEG signal classification using wavelet feature extraction and a mixture of expert model. *Expert Syst. Appl.* 32, 1084–1093. doi: 10.1016/j.eswa.2006.02.005

Takens, F. (1981). *Lecture Notes in Mathematics*. Berlin: Springer.

Tzallas, A. T., Tsipouras, M. G., and Fotiadis, D. I. (2007). Automatic seizure detection based on time-frequency analysis and artificial neural networks. *Comput. Intell. Neurosci.* 2007, 80510–80523. doi: 10.1155/2007/80510

Tzallas, A. T., Tsipouras, M. G., and Fotiadis, D. I. (2009). Epileptic seizure detection in EEGs using time-frequency analysis. *IEEE Trans. Inf. Technol. Biomed.* 13, 703–710. doi: 10.1109/TITB.2009.2017939

Ubeyli, E. D. (2006). Analysis of EEG signals using Lyapunov exponents. *Neural Netwk. World* 16, 257–273.

Ubeyli, E. D. (2007). Modified mixture of experts for analysis of EEG signals. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 2007, 1546–1549. doi: 10.1109/IEMBS.2007.4352598

Ubeyli, E. D. (2008). Wavelet/mixture of experts network structure for EEG classification. *Expert Syst. Appl.* 37, 1954–1962. doi: 10.1016/j.eswa.2007.02.006

Ubeyli, E. D., and Guler, I. (2007). Features extracted by eigenvector methods for detection variability of EEG signals. *Pattern Recogn. Lett.* 28, 592–603. doi: 10.1016/j.patrec.2006.10.004

Varsavsky, A., Mareels, I., and Cook, M. (2011). *Epileptic Seizures and the EEG: Measurement, Models, Detection and Prediction*. Boca Raton, FL: CRC Press.

Wang, D., Miao, D., and Xie, C. (2011). Best basis-based wavelet packet entropy feature extraction and hierarchical EEG classification for epileptic detection. *Expert Syst. Appl.* 38, 14314–14320. doi: 10.1016/j.eswa.2011.05.096

Waterhouse, E. (2003). New horizons in ambulatory electroencephalography. *IEEE Eng. Med. Biol. Mag.* 22, 74–80. doi: 10.1109/MEMB.2003.1213629

Welch, P. D. (1967). The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Trans. Audio Electroacoust.* AU-15, 70–73. doi: 10.1109/TAU.1967.1161901

Williams, G. P. (1997). *Chaos Theory Tamed*. Washington, DC: National Academy Press.

Wolf, A., Swift, J. B., Swinney, H. L., and Vastano, J. A. (1985). Determining Lyapunov exponents from a time series. *Phys. D* 16, 285–317. doi: 10.1016/0167-2789(85)90011-9

World Health Organization. (2012). *Epilepsy. Fact Sheet N 999*. Available online at: http://www.who.int/mediacentre/factsheets/fs999/en

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.

frontiers in Computational Neuroscience

ORIGINAL RESEARCH
published: 27 May 2015
doi: 10.3389/fncom.2015.00062

Input-output relation and energy efficiency in the neuron with different spike threshold dynamics

Guo-Sheng Yi[1], *Jiang Wang*[1]*, *Kai-Ming Tsang*[2], *Xi-Le Wei*[1] and *Bin Deng*[1]

[1] School of Electrical Engineering and Automation, Tianjin University, Tianjin, China, [2] Department of Electrical Engineering, The Hong Kong Polytechnic University, Hong Kong, China

Neuron encodes and transmits information through generating sequences of output spikes, which is a high energy-consuming process. The spike is initiated when membrane depolarization reaches a threshold voltage. In many neurons, threshold is dynamic and depends on the rate of membrane depolarization (dV/dt) preceding a spike. Identifying the metabolic energy involved in neural coding and their relationship to threshold dynamic is critical to understanding neuronal function and evolution. Here, we use a modified Morris-Lecar model to investigate neuronal input-output property and energy efficiency associated with different spike threshold dynamics. We find that the neurons with dynamic threshold sensitive to dV/dt generate discontinuous frequency-current curve and type II phase response curve (PRC) through Hopf bifurcation, and weak noise could prohibit spiking when bifurcation just occurs. The threshold that is insensitive to dV/dt, instead, results in a continuous frequency-current curve, a type I PRC and a saddle-node on invariant circle bifurcation, and simultaneously weak noise cannot inhibit spiking. It is also shown that the bifurcation, frequency-current curve and PRC type associated with different threshold dynamics arise from the distinct subthreshold interactions of membrane currents. Further, we observe that the energy consumption of the neuron is related to its firing characteristics. The depolarization of spike threshold improves neuronal energy efficiency by reducing the overlap of Na^+ and K^+ currents during an action potential. The high energy efficiency is achieved at more depolarized spike threshold and high stimulus current. These results provide a fundamental biophysical connection that links spike threshold dynamics, input-output relation, energetics and spike initiation, which could contribute to uncover neural encoding mechanism.

Keywords: spike threshold dynamic, input-output relation, energy efficiency, biophysical connection, spike initiation

OPEN ACCESS

Edited by:
Tobias Alecio Mattei, Kenmore Mercy Hospital, USA

Reviewed by:
Abdelmalik Moujahid, University of the Basque Country, Spain
Ramesh Kandimalla, Texas Tech University, USA

***Correspondence:**
Jiang Wang, School of Electrical Engineering and Automation, Tianjin University, No. 92 Weijin Road, Nankai District, Tianjin 300072, China
jiangwang@tju.edu.cn

Received: *14 March 2015*
Accepted: *08 May 2015*
Published: *27 May 2015*

Citation:
Yi G-S, Wang J, Tsang K-M, Wei X-L and Deng B (2015) Input-output relation and energy efficiency in the neuron with different spike threshold dynamics. Front. Comput. Neurosci. 9:62. doi: 10.3389/fncom.2015.00062

Introduction

Neurons, as the basic information-processing unit of the nervous system, can accurately represent and transmit various spatiotemporal patterns of sensory input in the form of sequences of output spikes (Koch, 1999; Dayan and Abbott, 2005; Klausberger and Somogyi, 2008). The generation and conduction of action potentials need to consume a lot of energy, which would have a great impact on neural codes and circuits (Niven and Laughlin, 2008; Alle et al., 2009; Sengupta et al., 2010, 2013, 2014; Moujahid et al., 2011). Characterizing energy efficiency associated with different

input-output relations is an essential step toward capturing the full strategies used by the neuron to encode stimulus. Previous experimental and modeling studies (Koch, 1999; Dayan and Abbott, 2005; Klausberger and Somogyi, 2008; Niven and Laughlin, 2008; Prescott et al., 2008a; Alle et al., 2009; Carter and Bean, 2009; Sengupta et al., 2010, 2013, 2014) have reported that both of the input-output relation and energy efficiency of neurons depend not only on input spatiotemporal properties but also on neuronal intrinsic characteristics.

One basic intrinsic property for all spiking neurons is the spike threshold, which is a special membrane potential that distinguishes subthreshold responses from spikes (Izhikevich, 2005; Goldberg et al., 2008). The small depolarization of membrane potential below this special value is subthreshold and decays to resting potential, while large depolarization above this value is suprathreshold and results in an action potential (Izhikevich, 2005; Prescott et al., 2008a; Wester and Contreras, 2013). That is, a spike is initiated only when membrane depolarization reaches this threshold potential. *In vivo*, the spike threshold is dynamic, and varies with input properties as well as spiking history. Especially, it is inversely correlated with the preceding rate of membrane depolarization (i.e., dV/dt) prior to spike initiation (Azouz and Gray, 2000, 2003; Henze and Buzsáki, 2001; Ferragamo and Oertel, 2002; Escabí et al., 2005; Wilent and Contreras, 2005; Kuba et al., 2006; Goldberg et al., 2008; Priebe and Ferster, 2008; Cardin et al., 2010; Higgs and Spain, 2011; Platkiewicz and Brette, 2011; Wester and Contreras, 2013; Fontaine et al., 2014). A dynamic threshold plays a critically important role in spike generation, which would participate in and produce profound influences on neuronal input-output properties (Azouz and Gray, 2000, 2003; Henze and Buzsáki, 2001; Ferragamo and Oertel, 2002; Escabí et al., 2005; Wilent and Contreras, 2005; Kuba et al., 2006; Priebe and Ferster, 2008; Cardin et al., 2010; Platkiewicz and Brette, 2011). For instance, the neuron with a dynamic threshold is more capable of filtering out synaptic inputs (Higgs and Spain, 2011) and regulating its response sensitivity (Azouz and Gray, 2000, 2003; Ferragamo and Oertel, 2002; Wilent and Contreras, 2005; Cardin et al., 2010). Further, the dynamic threshold could also effectively enhance feature selectivity (Azouz and Gray, 2003; Escabí et al., 2005; Wilent and Contreras, 2005; Priebe and Ferster, 2008), contribute to coincidence detection and gain modulation (Azouz and Gray, 2000, 2003; Platkiewicz and Brette, 2011), as well as facilitate precise temporal coding (Kuba et al., 2006; Higgs and Spain, 2011).

The spike threshold dynamics could be modulated by the biophysical properties of intrinsic membrane currents (Hodgkin and Huxley, 1952; Azouz and Gray, 2000, 2003; Wilent and Contreras, 2005; Guan et al., 2007; Goldberg et al., 2008; Higgs and Spain, 2011; Platkiewicz and Brette, 2011; Wester and Contreras, 2013; Fontaine et al., 2014). Two especially relevant biophysical mechanisms are Na^+ inactivation and K^+ activation, which are originally recognized by Hodgkin and Huxley (1952). Because Na^+ inactivation specifically affects spike initiation (Platkiewicz and Brette, 2011), it is usually regarded as the fundamental mechanism of regulating threshold (Azouz and Gray, 2000, 2003; Henze and Buzsáki, 2001; Wilent and Contreras, 2005; Platkiewicz and Brette, 2011; Wester and Contreras, 2013; Fontaine et al., 2014). Recently, more and more studies find that the outward K^+ channels, especially those activated at the subthreshold potentials, could also powerfully regulate spike threshold (Storm, 1988; Bekkers and Delaney, 2001; Dodson et al., 2002; Guan et al., 2007; Goldberg et al., 2008; Higgs and Spain, 2011; Wester and Contreras, 2013). Blocking them (Storm, 1988; Bekkers and Delaney, 2001; Dodson et al., 2002; Guan et al., 2007; Goldberg et al., 2008) or depolarizing their activation voltage to make them unactivated prior to spike initiation (Wester and Contreras, 2013) could both result in a loss of the inverse correlation between spike threshold and dV/dt.

In addition to modulating threshold dynamic, the biophysical properties of membrane currents could also control neuronal spike initiation (Koch, 1999; Izhikevich, 2005; Prescott and Sejnowski, 2008; Prescott et al., 2008a,b; Yi et al., 2014a,b). It is shown that if the K^+ current that flows out of the cell is absent or unactivated at the potentials around spike threshold, i.e., perithresholds, the neuron generates a continuous frequency-current curve through a saddle-node on invariant circle (SNIC) bifurcation, i.e., Hodgkin class 1 excitability (Izhikevich, 2005; Prescott et al., 2008a,b; Yi et al., 2014a). On the contrary, if the outward K^+ current has already activated at the perithresholds, the neuron generates a discontinuous frequency-current curve through a Hopf bifurcation, i.e., Hodgkin class 2 excitability (Izhikevich, 2005; Prescott et al., 2008a,b; Yi et al., 2014a). Furthermore, Rothman and Manis (2003a,b,c) find that a high density of low-threshold K^+ current in ventral cochlear nucleus is responsible for phasic firing of class 2 excitability, while a lower density promotes regular firing of class 1 excitability. These reports suggest that membrane biophysics is able to further determine neuronal input-output relations. Then, the dynamics of the spike threshold should also be dependent on input-output properties. Uncovering the biophysical connection between them is crucial for explaining how biophysical properties contribute to neural coding. Meanwhile, it could also provide a deeper insight into the mechanism of neural coding than a purely phenomepological description of input-output relation. However, the relevant studies are still lacking.

In fact, the biophysical properties of membrane currents not only affect spike threshold dynamic and input-output relation, but also influence neuronal energetics. During the generation of action potential, there is flux of different ions across their voltage-gated ionic channels, such as, influx of Na^+ and efflux of K^+. In this process, the ions need to expand significant quantities of energy to permeate cell membrane against their concentration gradient (Attwell and Laughlin, 2001; Niven and Laughlin, 2008; Alle et al., 2009; Carter and Bean, 2009; Sengupta et al., 2010, 2013, 2014; Moujahid et al., 2011, 2014; Moujahid and D'Anjou, 2012). The influx or efflux of ions, i.e., inward or outward ionic currents, dominate and make a significant contribution to neuronal energy consumption (Attwell and Laughlin, 2001; Alle et al., 2009; Sengupta et al., 2010, 2013, 2014). Previous studies (Alle et al., 2009; Carter and Bean, 2009; Sengupta et al., 2010, 2013; Moujahid and D'Anjou, 2012; Moujahid et al., 2014) have shown that adjusting the biophysical properties of voltage-gated Na^+ and K^+ currents, such as, channel conductance or

activation/inactivation time constant, could modulate the energy efficiency of neuron. Then, a critical question arises as to how the spike threshold dynamic, a basic property of neuron, influences its energy consumption. Until now, there is still no relevant research on this issue.

Here, we systematically characterize the input-output property and energy efficiency of the neuron with different spike threshold dynamics. To achieve this goal, we first adopt a two-dimensional biophysical model and vary its parameter that controls the voltage-dependency of K^+ current to produce different relationships between spike threshold and dV/dt. Then, we investigate how the minimal neuron responds to external stimulus as well as its relevant biophysical mechanism in the case of different threshold dynamics. Finally, we deduce the energy functions involved in the dynamics of neuron model, and determine the energy efficiency associated with each threshold dynamic.

TABLE 1 | Parameters in two-dimensional model (Prescott et al., 2008a).

Symbol	Value	Description
C	$2\mu F/cm^2$	Membrane capacitance
$\bar{g}_{Na}$	$20mS/cm^2$	Na^+ maximal conductance
$\bar{g}_K$	$20mS/cm^2$	K^+ maximal conductance
g_L	$2mS/cm^2$	Leak maximal conductance
V_{Na}	50 mV	Na^+ reversal potential
V_K	−100 mV	K^+ reversal potential
V_L	−70 mV	Leak reversal potential
β_m	−1.2 mV	Controlling the half-activation voltage of Na^+ current
γ_m	18 mV	Slope factor of activation curve $m_\infty(V)$
β_n	−5, −7, −9, −11, −13, or −15 mV	Controlling the half-activation voltage of K^+ current
γ_n	10 mV	Slope factor of activation curve $n_\infty(V)$
φ_n	0.15 (unitless)	Scaling factor for K^+ activation variable n

Materials and Methods

Two-Dimensional Neuron Model

A two-dimensional biophysical model proposed by Prescott et al. (2008a) is adopted to explore how spike threshold dynamic modulates neuronal input-output relation and metabolic energy in present study. It is a modified version of Morris-Lecar model, which incorporates three ionic currents, i.e., a fast Na^+ current I_{Na}, a delayed rectifying K^+ current I_K, as well as a leak current I_L. The model is given by the following differential equations (Prescott et al., 2008a)

$$C\frac{dV}{dt} = I_{in} + I_{noise} - \bar{g}_K n(V - V_K) - \bar{g}_{Na} m_\infty (V)(V - V_{Na}) - g_L(V - V_L) \quad (1)$$

$$\frac{dn}{dt} = \varphi_n \frac{n_\infty(V) - n}{\tau_n(V)} \quad (2)$$

where V is the membrane voltage and n is the activation gating variable for I_K. The three terms on the right side of Equation (1), i.e., $\bar{g}_K n(V - V_K)$, $\bar{g}_{Na} m_\infty(V)(V - V_{Na})$ and $g_L(V - V_L)$, respectively denote slow outward I_K, fast inward I_{Na} and outward I_L. $m_\infty(V) = 0.5\left\{1 + \tanh\left[(V - \beta_m)/\gamma_m\right]\right\}$ and $n_\infty(V) = 0.5\left\{1 + \tanh\left[(V - \beta_n)/\gamma_n\right]\right\}$ are the steady-state voltage-dependent activation functions for I_{Na} and I_K, and $\tau_n(V) = 1/\cosh\left[(V - \beta_n)/2\gamma_n\right]$ is the K^+ voltage-dependent time constant function. The kinetics of inward I_{Na} are controlled by parameter β_m and γ_m, and the kinetics of outward I_K are controlled by β_n and γ_n. In previous modeling study, Wester and Contreras (2013) have shown that hyperpolarizing K^+ activation voltage, even in the absence of Na^+ inactivation, is sufficient to produce a dynamic spike threshold that is inverse to the preceding dV/dt. Then, we vary parameter β_n from −5 to −15 mV in steps of −2 mV to produce different sensitivity of spike threshold to dV/dt in our stimulation. These values of β_n can span different spike initiation dynamics of the model (Prescott et al., 2008a). **Table 1** gives the numerical values and corresponding neural functions of the parameters in two-dimensional model, which are the same as those described in Prescott et al. (2008a).

I_{in} is the injected current used to stimulate neuron, which can be either steps or ramps in our study. I_{noise} is used to replicate synaptic noise, and is modeled as an Ornstein-Uhlenbeck process (Uhlenbeck and Ornstein, 1930)

$$\frac{dI_{noise}}{dt} = -\frac{I_{noise}}{\tau_{noise}} + \sigma N(t) \quad (3)$$

where $N(t)$ is a random number drawn from a Gaussian distribution with average 0 and unit variance. The amplitude of weak noise I_{noise} is controlled by the scaling parameter σ (Destexhe et al., 2001; Prescott and Sejnowski, 2008; Prescott et al., 2008a,b), which could vary from $0\mu A/cm^2$ to $3\mu A/cm^2$ in our study. The time constant is $\tau_{noise} = 5ms$ (Prescott and Sejnowski, 2008; Prescott et al., 2008b). When we determine spike threshold, phase response curve (PRC) and bifurcation patterns, the noisy current is removed from the neuron.

Method to Calculate Spike Threshold

The spike threshold for different values of dV/dt is determined by a novel approach proposed by Wester and Contreras (2013). According to their description, we use I_{in} to produce a cluster of ramps to stimulate the neuron, so

$$I_{in} = \begin{cases} Kt & (0 \le t \le t_0) \\ 0 & (t > t_0) \end{cases} \quad (4)$$

The ramp slope K controls the values of dV/dt leading to the spike initiation. With a larger value of K, the membrane potential V is forced to approach the threshold potential at a faster speed, which corresponds to a bigger value of dV/dt. The stimulation duration is controlled by t_0. For a given slope K, the membrane potential V will gradually approach the threshold as t_0 increases. When membrane potential V is around the threshold potential, we stepwise extend ramp duration t_0 to make each step result in about additional 0.1 mV depolarization in V until an action

potential is initiated in the neuron. In this way, if V is driven to cross spike threshold at the time of ramp offset, there will be a spike generated after removing ramp (i.e., $t > t_0$). Conversely, the neuron fails to initiate a spike if V does not reach the threshold potential at the time of ramp offset. Then, we empirically increase ramp duration t_0 to seek such a special membrane potential V^*: 0.1 mV hyperpolarized to V^* is subthreshold and neuron fails to initiate a spike at the time of ramp offset, whereas 0.1 mV depolarized to V^* is suprathreshold and neuron could initiate a spike at the time of ramp offset. We define this special membrane potential V^* as the spike threshold of the neuron. In this manner, the upstroke of the spike is purely due to the sufficient activation of Na^+ current, which has nothing to do with the current ramp. This method allows us to measure the spike threshold with a high precision less than 0.1 mV.

Phase Response Curve Calculation

The PRC measures the phase shift of a periodically oscillating neuron in response to a brief current pulse delivered at different phases of the oscillation cycle (Ermentrout, 1996; Izhikevich, 2005; Smeal et al., 2010; Fink et al., 2011; Schultheiss et al., 2012). The PRC of the neuron can be defined as (Ermentrout, 1996; Izhikevich, 2005; Smeal et al., 2010; Schultheiss et al., 2012)

$$PRC(\vartheta) = 1 - T'(\vartheta)/T \tag{5}$$

where T is the oscillation period of the neuron without perturbation (i.e., $1/T$ represents natural oscillation frequency), and $T'(\vartheta)$ is the oscillation period when the neuron is stimulated at phase ϑ. A positive value of PRC indicates there is a phase advance, and a negative value indicates a phase delay. If the amplitude of current pulse is sufficiently small and its duration is sufficiently brief, the PRC becomes the infinitesimal PRC, which could reflect the intrinsic dynamics of the oscillator (Ermentrout, 1996; Smeal et al., 2010; Fink et al., 2011; Schultheiss et al., 2012). In the following, we use "PRC" to refer to the infinitesimal PRC. Further, the PRCs of neural oscillator have often been classified into two categories: Type I that respond with only phase advances to excitatory stimuli, and Type II that display both phase advances and delays (Hansel et al., 1995; Smeal et al., 2010; Fink et al., 2011).

Method to Determine Energy Consumption in Two-Dimensional Model

We use the method proposed by Moujahid et al. (2011, 2014) and Moujahid and D'Anjou (2012) to determine the electrochemical energy involved in the modified Morris-Lecar model. The model in Equation (1) can be regarded as an electrical circuit, which consists of membrane capacitance C, Na^+, K^+ and leak ionic channels. According to the description by Moujahid et al. (2011, 2014) and Moujahid and D'Anjou (2012), the total electrical energy accumulated in this circuit at a given time can be expressed by

$$E(t) = \frac{1}{2}CV^2 + E_{Na} + E_K + E_L \tag{6}$$

Here, $\frac{1}{2}CV^2$ is the electrical energy accumulated in the membrane capacitance. E_{Na}, E_K, and E_L are the energies in the batteries needed to create the concentration jumps in Na^+, K^+ and chloride, respectively. These energies could be supplied by external stimuli, i.e., I_{in} or I_{noise}. The first-order derivative with respect to time of the Equation (6) is

$$\frac{dE}{dt} = CV\frac{dV}{dt} + I_{Na}V_{Na} + I_K V_K + I_L V_L \tag{7}$$

Substituting $\frac{dV}{dt}$ with Equation (1), the energy rate δ (i.e., $\frac{dE}{dt}$) in the circuit can be written as

$$\delta = (I_{in}+I_{noise})V - I_{Na}(V-V_{Na}) - I_K(V-V_K) - I_L(V-V_L) \tag{8}$$

where $(I_{in}+I_{noise})V$ is the energy power supplied by stimulus. The last three terms on the right hand of Equation (8) represent the energy consumption rate of the ionic channels. If we substitute I_{Na}, I_K, and I_L with their expressions, we can deduce the energy rate of each ionic channel

$$\delta_{Na} = \overline{g}_{Na} m_\infty(V)(V - V_{Na})^2 \tag{9}$$

$$\delta_K = \overline{g}_K n(V - V_K)^2 \tag{10}$$

$$\delta_L = g_L(V - V_L)^2 \tag{11}$$

It is easy to see that this method is not based on the stoichiometry of the ions. Thus, it requires no hypothesis about the overlapping between Na^+ and K^+ ions, and then avoids the overestimate values of energy (Moujahid et al., 2011, 2014; Moujahid and D'Anjou, 2012).

Numerical Stimulation

The differential equations of the entire system are numerically integrated with MATLAB. The bifurcation analysis is performed with XPPAUT (Ermentrout, 2002) following the standard procedures. In bifurcation analysis, we use I_{in} to produce step currents to stimulate the neuron and systematically vary its intensity to determine at what point the neuron qualitatively changes its dynamical behavior, such as, starting or ceasing repetitive spiking. This special point corresponds to a bifurcation. Further, the PRC is also calculated by XPPAUT.

Results

In this section, we first adjust parameter β_n that controls the half-activation voltage of K^+ channel to produce the spike threshold that has different sensitivity to the preceding dV/dt, as shown in **Figures 1A,B**. One can find that the spike threshold becomes more depolarized as we shift β_n alone from -5 to -15mV in steps of -2mV (**Figure 1B**). For three cases of $\beta_n = -5, -7$, and -9mV, the spike thresholds are all insensitive to dV/dt, and there is always no inverse relationship between spike threshold and dV/dt. On the contrary, the spike threshold shows relatively large variations and becomes sensitive to dV/dt with $\beta_n = -11, -13$, and -15mV. In these three cases, the spike threshold varies inversely with the preceding dV/dt, and

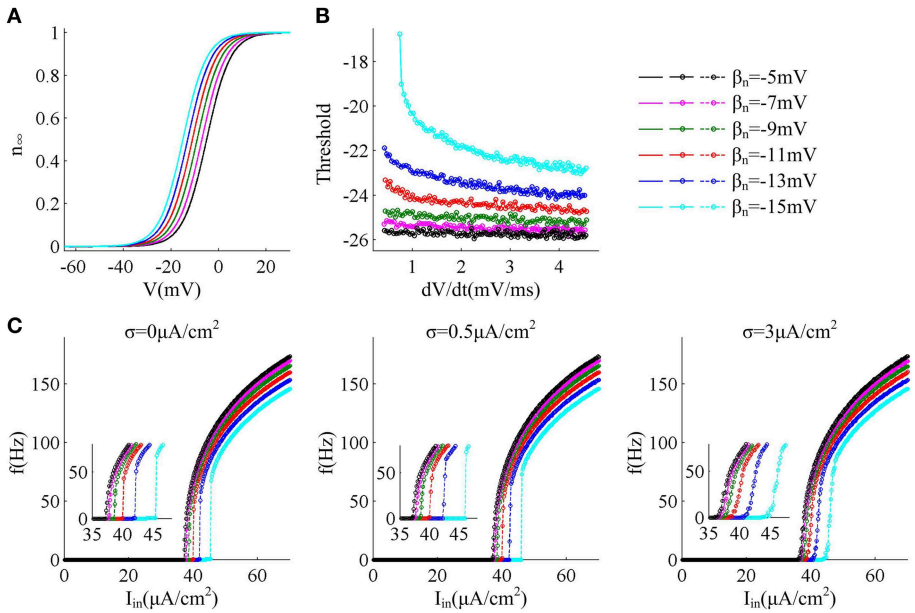


FIGURE 1 | $f - I_{in}$ curves associated with different threshold dynamics induced by adjusting β_n. (A) The half-activation voltage β_n of activation variable n is hyperpolarized from −5 to −15mV with a step of −2mV. **(B)** Spike threshold as a function of dV/dt with different values of β_n. The range of dV/dt is from 0.45 to 4.5mV/ms. **(C)** $f - I_{in}$ curves generated by the neuron with different threshold dynamics for three levels of noise. The noise amplitude is $\sigma = 0, 0.5$, and $3\mu A/cm^2$, respectively.

simultaneously the inverse relationship becomes more significant as β_n decreases. The range of dV/dt in **Figure 1B** is from 0.45 to 4.5 mV/ms, which is achieved by increasing ramp slope K in Equation (4). This range is selected in accordance with previous modeling (Wester and Contreras, 2013) and experimental (Wilent and Contreras, 2005) studies. In the following, we respectively explore neuronal input-output relation and energy efficiency in these six cases.

Input-Output Property of the Neuron with Different Threshold Dynamics

For different sensitivity of spike threshold to dV/dt, we respectively investigate how neuron responds to constant current in the cases of no noise ($\sigma = 0\mu A/cm^2$), low noise ($\sigma = 0.5\mu A/cm^2$) and high noise ($\sigma = 3\mu A/cm^2$). To achieve this goal, we use I_{in} to produce step current to stimulate the neuron and systematically alter its intensity to determine neuronal spike frequency f.

Figure 1C gives neuronal spike frequency f as a function of input current I_{in} (i.e., $f - I_{in}$ curve) in six cases of threshold dynamic. For three levels of noise, one can observe that the depolarization of spike threshold slightly reduces the slope of $f - I_{in}$ curve at the low firing rates and obviously shifts the curve to the right, which corresponds to increasing the minimal current intensity used for triggering repetitive spike (i.e., current threshold). If spike threshold is insensitive to dV/dt (i.e., $\beta_n = -5$, -7, and -9mV), the neuron could spike repetitively at very low frequencies in all levels of noise, which endows it with a continuous $f - I_{in}$ curve. However, when spike threshold is sensitive to dV/dt (i.e., $\beta_n = -11$, -13, and -15mV), the neuron is unable to maintain repetitive spike at low rates and produces a discontinuous $f - I_{in}$ curve in the cases of no or low noise levels (**Figure 1C**). This discontinuous $f - I_{in}$ curve could be switched to continuous by high level of noise.

Since noise is another ubiquitous feature of the nervous system with myriad effects on neural coding (Tuckwell, 1989; Gerstner and Kistler, 2002; Tuckwell et al., 2009; Tuckwell and Jost, 2010), we further investigate how noise modulates spike trains of the neuron with different spike threshold dynamics, as shown in **Figures 2**, **3**. It is observed that no matter there is an inverse relationship between spike threshold and dV/dt or not, the spike number always increases monotonically from 0 as noise amplitude σ increases when I_{in} is less than the bifurcation value I_{in}^*. For I_{in} just beyond I_{in}^*, the noise could inhibit or even terminate the repetitive spiking of neuron when its spike threshold is sensitive to dV/dt (**Figures 2D–F**). In this case, the neuron is able to generate repetitive spike without noise (i.e., $\sigma = 0\mu A/cm^2$), since I_{in} has already exceeded bifurcation value I_{in}^*. Introducing synaptic noise makes the spike trains become irregular. Unexpectedly, weak noise (such as, $\sigma = 0.2\mu A/cm^2$) has an obvious inhibitory effect on neuronal spiking behavior, which even terminates repetitive spiking for a long time. When noise amplitude is increased to $\sigma = 1.5\mu A/cm^2$ or even higher, there will be more spikes evoked again. That is, when I_{in} is in the vicinity of I_{in}^*, small noise could noticeably inhibit neuronal spiking and there is a minimum in the mean spike number as σ goes up (**Figures 3D–F**). Meanwhile, as the inverse relationship between spike threshold and dV/dt gets pronounced, the inhibitory effect induced by small noise becomes stronger. However, this inhibitory effect does not appear in the

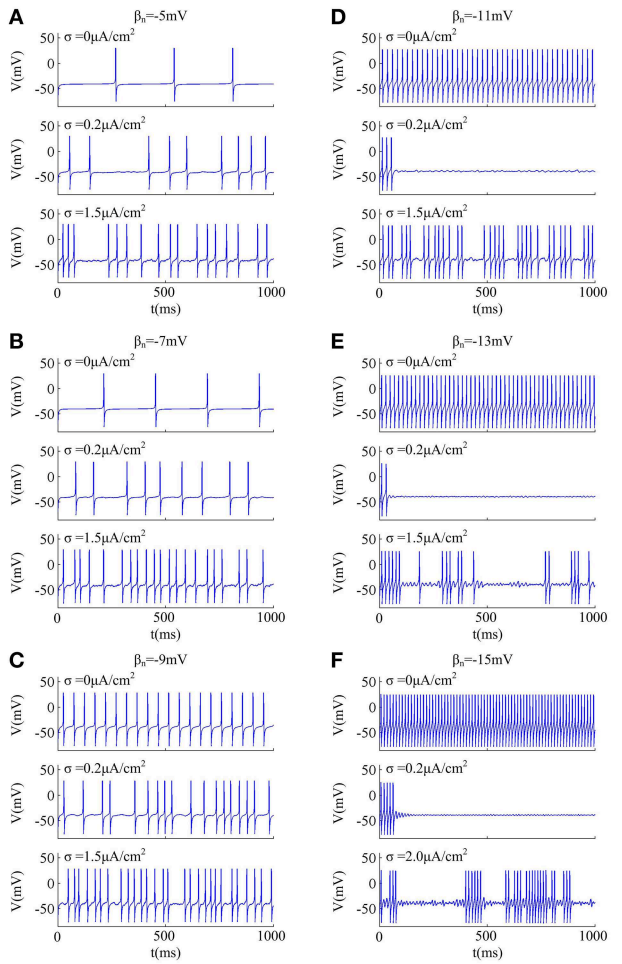


FIGURE 2 | Effects of weak noise on spiking trains around the bifurcation. The input current is **(A)** $I_{in} = 37.3\mu A/cm^2$, **(B)** $I_{in} = 37.8\mu A/cm^2$, **(C)** $I_{in} = 38.72\mu A/cm^2$, **(D)** $I_{in} = 40.2\mu A/cm^2$, **(E)** $I_{in} = 42.2\mu A/cm^2$, and **(F)** $I_{in} = 45.8\mu A/cm^2$. The values of noise amplitude σ have been indicated in each panel.

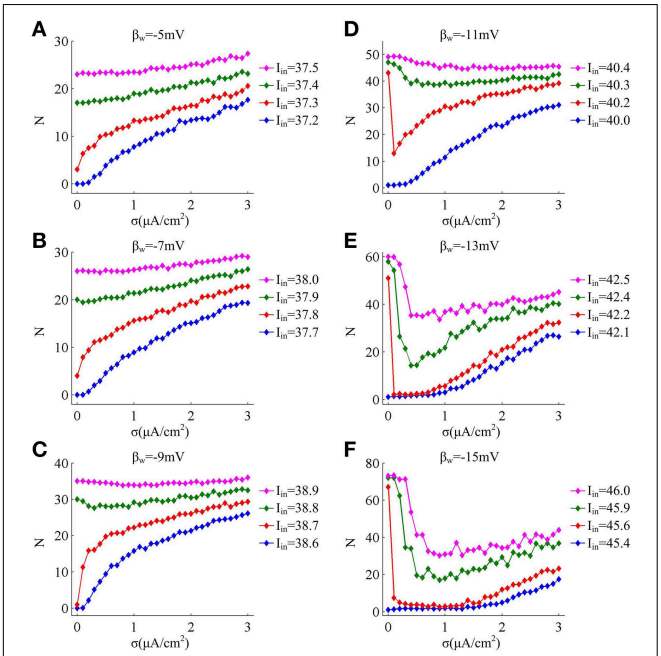


FIGURE 3 | Mean numbers of spikes as a function of noise amplitude for each threshold dynamic. (A–F) respectively give the mean spike number *N* (40 trials) as noise amplitude σ is increased in the neuron for 1000 ms time interval with different values of β_n. The value of I_{in} indicated by blue line is below the bifurcation point I_{in}^* and there is no repetitive spiking generated in the neuron without noise, while the values of I_{in} indicated by three other colors are above the bifurcation point I_{in}^*.

neuron with an insensitive spike threshold to dV/dt (left panels, **Figures 2**, **3**). In this case, the noise only disturbs its spike trains and makes them become irregular, which is unable to terminate repetitive spiking (**Figures 2A–C**).

Phase Response Curves of the Neuron with Different Threshold Dynamics

In previous section, we have found that different sensitivity of spike threshold to dV/dt could result in distinct (i.e., discontinuous or continuous) $f - I_{in}$ curves in the case of no or low noise. In this section, we use PRC theory to further characterize neuronal response properties in the case of different threshold dynamics.

Figure 4 displays the PRCs of the neuron model in six cases of spike threshold dynamic. It is found that the PRC is dependent on the natural oscillation frequency of neuron, and increasing it could attenuate the amplitude of phase shift. When spike threshold is insensitive to dV/dt, the neuron generates type I PRC, which exclusively displays phase advances (i.e., positive values) to excitatory brief pulse (**Figure 4A**). However, when spike threshold has an obvious inverse relation with dV/dt, the neuron shows phase delays (i.e., negative values) at earlier phases and phase advances at later phases (**Figure 4B**), which is manifested as a type II PRC. It has been proposed that type I PRC corresponds to a continuous $f - I_{in}$ curve and type II PRC corresponds to a discontinuous $f - I_{in}$ curve (Ermentrout, 1996; Izhikevich, 2005; Smeal et al., 2010; Fink et al., 2011). Our simulation results in **Figures 1C**, **4** are in accordance with this proposal. Further, it is worth pointing out that there are very small negative regions at the earlier phases of type I PRCs (**Figure 4A**). This is because the action potentials generated in Morris-Lecar like model consume a much larger portion of interspike interval than other models (Rinzel and Ermentrout, 1998; Fink et al., 2011). But according to the descriptions of Fink et al. (2011), we could ignore these early small phase delays in type I PRCs.

Biophysical Basis of the Spike Initiation Associated with Different Threshold Dynamics

By varying parameter β_n, we have identified the input-output property associated with each spike threshold dynamic. Our next step is to explore why the neuron with distinct threshold dynamics produces different input-output properties. It has been known that the membrane currents with opposite directions play different roles in spike generation. The currents flowing into

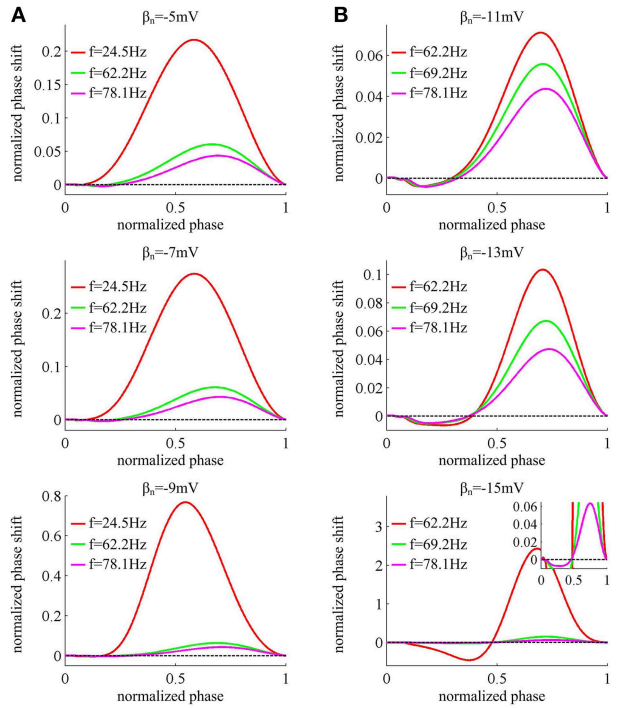


FIGURE 4 | PRCs of the neuron with different threshold dynamics. The neurons in **(A)** have an insensitive spike threshold to dV/dt, and in **(B)** have a sensitive threshold to dV/dt. For each threshold dynamic, we compute neuronal PRC at three different natural firing frequencies. The corresponding stimulation current is: I_{in} = 37.52, 39.31, and 40.87μA/cm^2 for $\beta_n = -5$mV; I_{in} = 37.966, 39.70, and 41.33μA/cm^2 for $\beta_n = -7$mV; I_{in} = 38.74, 40.28, and 41.94μA/cm^2 for $\beta_n = -9$mV; I_{in} = 41.17, 41.81, and 42.87μA/cm^2 for $\beta_n = -11$mV; I_{in} = 42.63, 43.20, and 44.25μA/cm^2 for $\beta_n = -13$mV; I_{in} = 45.478, 45.688, and 46.5μA/cm^2 for $\beta_n = -15$ mV. All PRCs are computed in the case of no noise, i.e., $\sigma = 0$μA/cm^2.

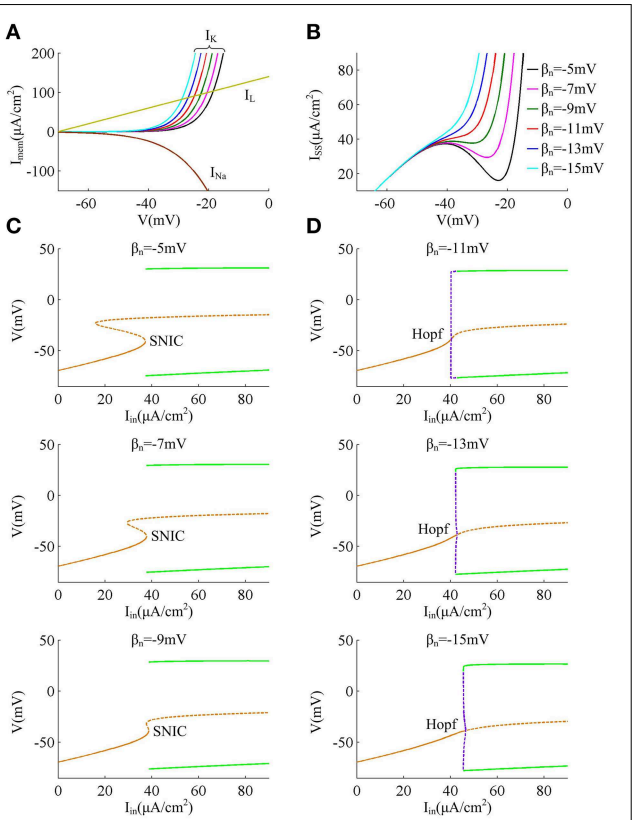


FIGURE 5 | Biophysical basis of the spike initiation for different threshold dynamics. (A) shows the individual steady-state membrane currents at the subthreshold potentials. Decreasing β_n has no effects on the activations of inward I_{Na} and outward I_L, while it causes outward I_K to be more strongly activated by perithreshold depolarization. **(B)** gives the relationship between steady-state net membrane current I_{SS} and membrane potential V (i.e., $I_{SS} - V$ curve). I_{SS} is computed as the sum of three individual currents, i.e., $I_{SS} = I_{Na} + I_K + I_L$. **(C,D)** summarize the bifurcation diagram associated with each spike threshold dynamic. The stable equilibrium is indicated by orange solid line and unstable is orange dotted line. The stable limit cycle is indicated by green solid line and unstable is purple dotted line.

the cell mainly depolarize membrane voltage to produce the rapid upstroke of the spike (i.e., positive feedback), whereas the currents flowing out of the cell mainly hyperpolarize membrane voltage which are responsible for the repolarization and produce the downstroke of the spike (i.e., negative feedback) (Izhikevich, 2005; Prescott et al., 2008a,b; Yi et al., 2014a). Here, we investigate how the opposite currents interact at the perithreshold potentials to determine neuronal response property in six cases of spike threshold dynamic.

Reducing parameter β_n from −5 to −15mV results in a hyperpolarizing shift in the half-activation voltage of outward K^+ current I_K (**Figure 1A**), which causes I_K to be more strongly activated by the perithreshold depolarization (**Figure 5A**). For three cases that the spike threshold is insensitive to dV/dt (i.e., $\beta_n = -5, -7$, and −9mV), the outward I_K activates at a higher potential than inward I_{Na} (**Figure 5A**), which indicates that the slow outward current I_K does not become activated until after the spike is initiated. In these three cases, the relationship between steady-state net membrane current I_{SS} and membrane voltage V (i.e., $I_{SS} - V$ curve) is always non-monotonic (**Figure 5B**), which has a region of negative slope. At the local maximum of $I_{SS} - V$ curve, the inward I_{Na} balances outward unactivated I_K and outward I_L. Then, any further depolarization could result in the progress activation of I_{Na} and make it become self-sustaining to generate the upstroke of the spike. In other words, the bifurcation occurs at this voltage, i.e., $\partial I_{SS}/\partial V = 0$. Since the depolarizing current I_{Na} faces no restraint of hyperpolarizing current at the perithreshold potentials, the membrane potential V could be driven to slowly pass through spike threshold. Thus, the neuron is able to spike repetitively at low frequencies and produce a continuous $f - I_{in}$ curve. This continuous input-output property is generated through a SNIC bifurcation (**Figure 5C**), which corresponds to a non-monotonic $I_{SS} - V$ curve (Izhikevich, 2005; Prescott et al., 2008a,b; Yi et al., 2014a). Further, because inward I_{Na} dominates spike initiation without the restraint of I_K at the perithresholds, a brief, excitatory stimulus only leads to advances in oscillation cycle and positive values of phase shift, which corresponds to a type I PRC.

For the other three cases that the spike threshold is sensitive to dV/dt (i.e., $\beta_n = -11, -13$, and −15mV), the outward

I_K activates at roughly the same V with inward I_{Na} or at a slightly lower V than I_{Na} (**Figure 5A**). The activation of I_K at low potentials makes the outward currents become so strong that the inward I_{Na} is unable to balance them at the perithreshold potentials, which results in a monotonic $I_{SS} - V$ curve without local maximum (**Figure 5B**). To initiate action potentials, the inward I_{Na} must exploit its fast kinetic to activate faster than slow outward I_K, and drives V through threshold potential with a sufficient speed that the outward I_K cannot catch up. Only in this way can the positive feedback outrun negative feedback to produce the upstroke of the spike. Since the V trajectory between two spikes must be more rapid than I_K, the neuron is unable to spike repetitively at low frequencies, which endows it with a discontinuous $f - I_{in}$ curve. This discontinuous input-output property is generated through a Hopf bifurcation (**Figure 5D**), which corresponds to a monotonic $I_{SS} - V$ curve (Izhikevich, 2005; Prescott et al., 2008a,b; Yi et al., 2014a). Further, in this case there is a special subthreshold region where the activation of low-threshold I_K is greater than inward I_{Na}. When voltage trajectory pass through this region, an excitatory pulse will evoke a larger response from outward I_K than from inward I_{Na}, which leads to negative PRC values at early phases. At higher membrane potential later in this special subthreshold region, the fast activating I_{Na} dominates neuronal response to brief excitatory pulse, which leads to the positive PRC values at later phases. Then, the neuron generates a type II PRC that has both phase delays and advances in these three cases.

Further, as spike threshold gets depolarized, the outward I_K becomes more strongly activated at the perithreshold potentials, which increases the net current I_{SS} and makes it reach a higher outward level prior to spike initiation. Since the outward current hyperpolarizes membrane potential V and prohibits action potential, there should be stronger step current I_{in} to counteract outward current and activate inward I_{Na} to generate spike. Then, the current threshold for triggering neuronal repetitive spiking increases as spike threshold gets depolarized.

Finally, when Hopf bifurcation occurs (i.e., the spike threshold is sensitive to dV/dt), there is a narrow bistable region in the vicinity of bifurcation, where stable resting state and stable limit cycle coexist (**Figure 5D**). Then, synaptic noise could switch voltage trajectory from one attractor, a stable limit cycle, to another, a stable resting point (Tuckwell et al., 2009; Tuckwell and Jost, 2010, 2011, 2012; Guo, 2011). This is the basis of the inhibitory effects of weak noise on spiking behavior. Meanwhile, the bistable region widens as the relationship between spike threshold and dV/dt gets pronounced, which causes the inhibitory effects of weak noise on repetitive spiking to become stronger. On the contrary, there is no bistable region in the case of SNIC bifurcation (**Figure 5C**), so the noise is unable to inhibit or terminate neuronal spiking in this case, i.e., the spike threshold is insensitive to dV/dt.

Energy Efficiency in the Neuron with Different Threshold Dynamics

We have identified the input-output property and spike initiation mechanism associated with each threshold dynamic. Here, we characterize the energy efficiency consumed by the neuron in six cases of threshold dynamic.

We first describe how ionic currents and their energy consumption evolve during the generation of a spike. **Figure 6A** shows an action potential generated in the neuron with $\beta_n = -5$mV to $I_{in} = 37.5\mu A/cm^2$ in the case of no noise (i.e., $\sigma = 0\mu A/cm^2$). At this value of I_{in} and σ, the neuron spikes repetitively at about 23.5 Hz. **Figure 6B** gives the Na^+, K^+ and leak currents corresponding to the spike waveform described in **Figure 6A**. The Na^+ current flows into the cell and has a negative sign, but we plot it with a positive sign for a better visualization of the overlap between Na^+ and K^+ currents. During the upstroke, the Na^+ current first activates and drives membrane voltage to quickly depolarize. Then, the outward K^+ current activates which hyperpolarizes membrane voltage and leads to the downstroke. The energy consumption rates of the three ions are shown in **Figure 6C**, which are computed according to Equations (9)–(11). They represent the instantaneous energy consumption per second by corresponding ionic channel, which are all positive. One can observe that there are overlaps between Na^+ and K^+ energy, especially during the downstroke (**Figure 6C**). **Figure 6D** gives the total energy rate δ consumed by all the ionic currents, which is used to generate the action potential in **Figure 6A**. In order to maintain the spiking activity of the neuron, this energy consumption must be replenished by the ion pumps and metabolically supplied by the hydrolysis of ATP molecules.

The left panels in **Figure 7** give the average energy consumption rate $\overline{\delta}$ as a function of input current I_{in} (i.e., $\overline{\delta} - I_{in}$ curve) in six cases of spike threshold dynamics for three levels of noise. It can be found that the energy consumption rate $\overline{\delta}$ in quiescent state is much lower than that in spiking state. This is because the increase of supplied energy to the neuron, i.e., increasing step current, promotes the ionic to pass through cell membranes, and makes them consume more energy. When spike threshold is insensitive to dV/dt (i.e., $\beta_n = -5, -7$, and -9mV), the $\overline{\delta} - I_{in}$ curve is always continuous for three levels of noise. However, if there is an obvious inverse relation between threshold and dV/dt (i.e., $\beta_n = -11, -13$, and -15mV), the $\overline{\delta} - I_{in}$ curve is discontinuous in the cases of no or low noise and continuous for high level of noise. Thus, the energy consumption rate of the neuron during the transition from quiescent state to spiking regime is dependent on its firing rates, which is displayed in **Figure 1C**. As spike threshold gets depolarized, the $\overline{\delta} - I_{in}$ curve in firing regime shifts to the right and the corresponding average energy consumption rate $\overline{\delta}$ of the neuron decreases.

The right panels in **Figure 7** show the total energy consumption in nJ per cm^2 calculated as the integral over long period of time of the area under the instantaneous ionic channel energy curve [i.e., the sum of the energy rates given by Equations (9)–(11)] divided by the number of spikes, which gives the energy consumption of a single spike. As step current I_{in} increases, the energy consumed in one spike first quickly decreases, and then has a very slight increase (about $0.1nJ/cm^2$ per $1\mu A/cm^2$). As threshold gets depolarized, the energy consumption in one action potential becomes larger with some low I_{in} values, and the synaptic noise obviously increases this consumption. However, with high values of I_{in}, the energy demand for a spike gets smaller

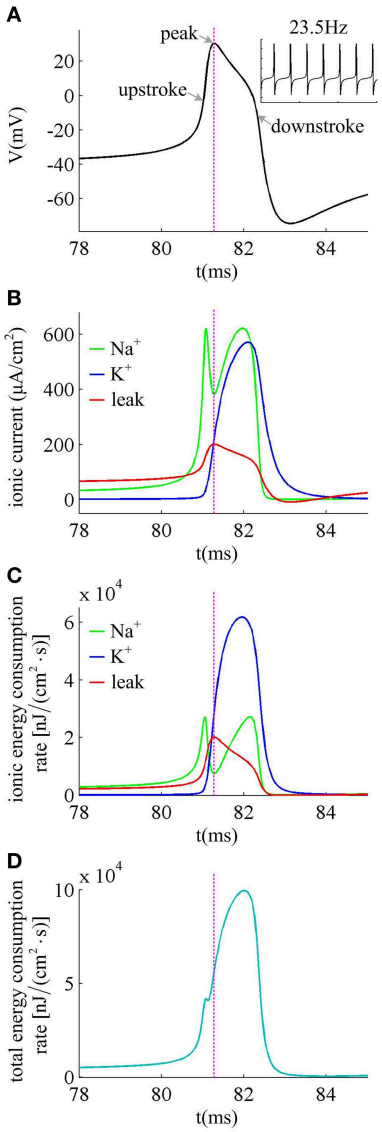


FIGURE 6 | Ionic currents and energy consumption involved in a spike. **(A)** shows an action potential generated in the neuron with $\beta_n = -5$mV. **(B)** gives the Na^+, K^+ and leak currents in this action potential. The Na^+ current is negative but we plot it with a positive sign. **(C)** shows the energy consumption rate for each ionic current, and **(D)** gives the total energy consumption rate of the action potential. The stimulus is $I_{in} = 37.5\mu A/cm^2$ and $\sigma = 0\mu A/cm^2$. In this case, the neuron generates repetitive spiking at about 23.5 Hz.

as spike threshold depolarizes, and increasing synaptic noise produces little effects on this consumption. That is, depolarizing spike threshold increases the energy utilization efficiency of the neuron in high firing rates. The lower values of energy consumption in one spike are achieved at more depolarized spike threshold and high stimulus current.

From the results in **Figures 6B,C**, it can be found that there are overlaps between Na^+ and K^+ currents in an action potential. These two positive charges flow in opposite directions as they pass through cell membrane, so that they can neutralize each other during the overlap. The overlap charge could be computed as the integral of Na^+ current during the hyperpolarized phase of the spike (Moujahid et al., 2011, 2014; Moujahid and D'Anjou, 2012), which is the inward Na^+ that is counterbalanced by outward K^+. Previous studies (Alle et al., 2009; Carter and Bean, 2009; Sengupta et al., 2010, 2013; Moujahid and D'Anjou, 2012; Moujahid et al., 2014) have shown that reducing this overlap load could decrease the energy demands for spike generation. From **Figure 8A**, one can find that the overlap Na^+ indeed undergoes a reduction as spike threshold gets depolarized in the case of high I_{in} values. The efficient use of inward Na^+ could decrease the energy consumption in an action potential and enhance the energy efficiency of the neuron (**Figure 8B**).

Discussion

Our results demonstrate there is a fundamental connection between spike threshold dynamics and neuronal input-output properties. When spike threshold is insensitive to dV/dt, the $f - I_{in}$ curve is continuous and weak noise is unable to produce inhibitory effects on spiking rhythms. In this case, the neuron generates a type I PRC that exclusively displays phase advances. However, when spike threshold is sensitive to dV/dt, the neuron generates a discontinuous $f - I_{in}$ curve and a type II PRC in the cases of no or low noise. Increasing noise amplitude switches the $f - I_{in}$ curve from discontinuous to continuous. Simultaneously, weak synaptic noise obviously prohibits spiking rhythms when I_{in} is near and above the bifurcation point I_{in}^*. In this case, as the inverse relationship between spike threshold and dV/dt gets pronounced, the inhibitory effects of weak noise on spiking rhythms and the discontinuity of $f - I_{in}$ curve both become more significant. Further, the depolarization of the spike threshold shifts the $f - I_{in}$ curve to the right, alters the slope of $f - I_{in}$ curve at low spike rates, and increases the current threshold for evoking neuronal repetitive spiking. These results indicate that the spike threshold properties, such as, whether it is sensitive to dV/dt, the inverse degree of it depends on dV/dt, or even the values of threshold potential could all obviously influence neuronal input-output relations.

All these input-output properties associated with each spike threshold dynamic are derived from the distinct nonlinear interactions between inward (depolarizing) and outward (hyperpolarizing) currents at the perithreshold potentials. When spike threshold is insensitive to dV/dt, the outward I_K does not activate prior to spike threshold, which leads inward I_{Na} to dominate spike initiation without the restraint of I_K. Due to the absent of outward I_K, the inward I_{Na} is able to balance weak outward currents at the perithreshold potentials, which results in a non-monotonic $I_{SS} - V$ curve, a type I PRC, and a SNIC bifurcation. Under these conditions, V could be forced to slowly pass through threshold potential and the neuron is able to spike at low frequencies, thus producing a continuous $f - I_{in}$ curve. Since the SNIC bifurcation does not have the bistable region, the inhibitory effects of weak noise on spiking rhythms is missing in this case. When spike threshold is sensitive to dV/dt, the outward I_K is able to activate at the subthresholds, and could become sufficiently strong prior to spike initiation.

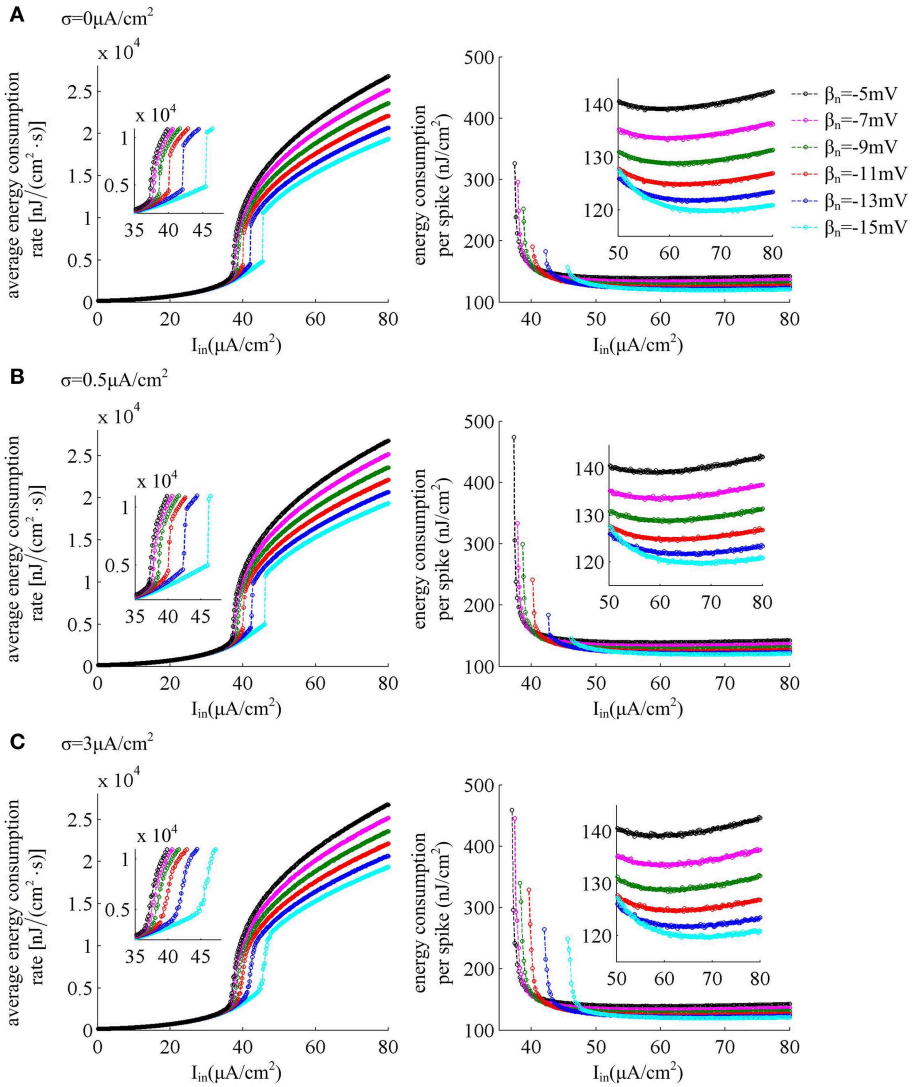


FIGURE 7 | Energy consumption as a function of I_{in} associated with each threshold dynamic. Left panels give the average energy consumption rate of the neuron with different spike threshold dynamics for three levels of noise. The energy consumption rate is averaged over the 7000 ms time interval. Right panels are the total electrochemical energy consumed by an action potential related to each spike threshold dynamic and input current I_{in}. The noise amplitude is **(A)** $\sigma = 0\mu A/cm^2$, **(B)** $\sigma = 0.5\mu A/cm^2$, and **(C)** $\sigma = 3\mu A/cm^2$.

Then, inward I_{Na} is unable to balance it at the perithreshold potentials, which leads to a monotonic $I_{SS} - V$ curve, a type II PRC and a Hopf bifurcation. The action potential could be successfully initiated because inward I_{Na} activates quickly to drive V through threshold with a sufficient speed that slow outward I_K cannot overtake. This means the neuron is unable to spike at low rates, which corresponds to a discontinuous $f - I_{in}$ curve. Since the neuron generates a narrow bistable region when Hopf bifurcation occurs, the weak noise could convert its state from stable limit cycle to resting and then prohibit repetitive spiking. Further, the increase of current threshold for evoking repetitive spiking is also due to the intensity of net outward current becomes stronger as threshold gets depolarized.

The biophysical explanation about how the activation properties of intrinsic membrane currents contribute to the spike threshold dynamic with the preceding dV/dt has been reported in many experimental and modeling studies (Hodgkin and Huxley, 1952; Storm, 1988; Azouz and Gray, 2000, 2003; Bekkers and Delaney, 2001; Henze and Buzsáki, 2001; Dodson et al., 2002; Wilent and Contreras, 2005; Guan et al., 2007; Goldberg et al., 2008; Higgs and Spain, 2011; Wester and Contreras, 2013; Fontaine et al., 2014). Meanwhile, the biophysical basis of how different dynamical mechanisms of spike initiation (i.e., SNIC and Hopf bifurcation) generate distinct input-output relations, such as Hodgkin class 1 and class 2 excitability (Koch, 1999; Izhikevich, 2005; Prescott and Sejnowski, 2008; Prescott et al., 2008a,b; Yi et al., 2014a) or type I and type II PRC (Ermentrout,

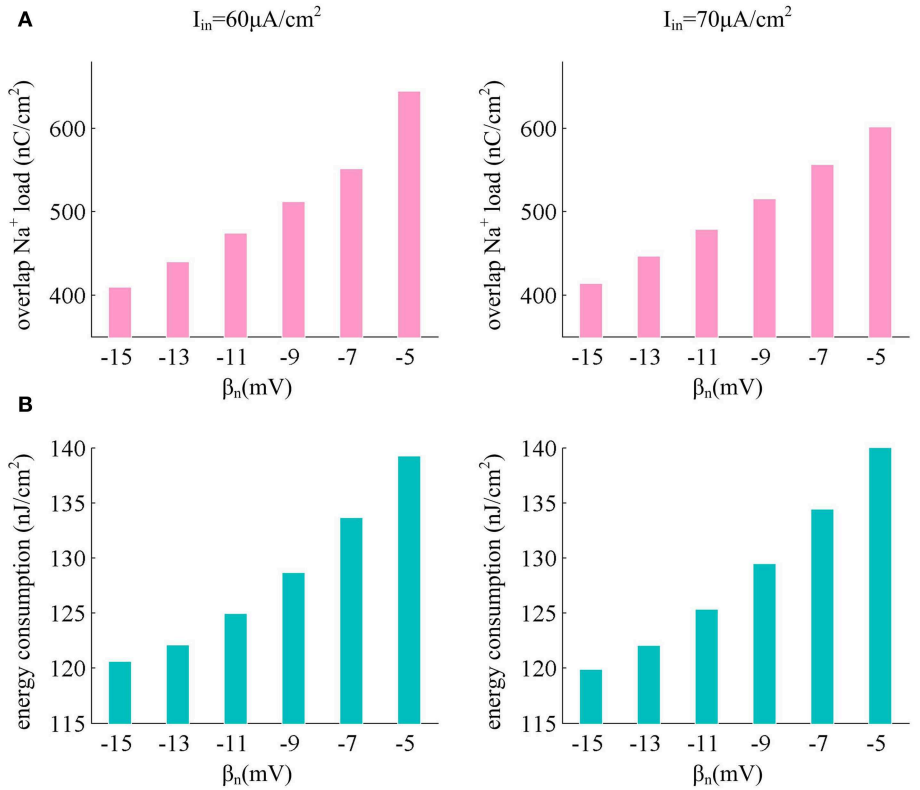


FIGURE 8 | Depolarizing spike threshold increases energy efficiency by reducing overlaps between Na^+ and K^+ currents. (A) shows the overlap Na^+ load for different spike threshold dynamics in the case of high stimulus. **(B)** gives the corresponding total energy required by a spike for each spike threshold dynamic. The stimulus current is I_{in} = 60μA/cm^2 and I_{in} = 70μA/cm^2, the noise amplitude is σ = 0μA/cm^2.

1996; Smeal et al., 2010; Fink et al., 2011), has also been well established. However, none of them has explored how spike threshold dynamic modulates neuronal input-output relation. With a simple biophysical model, we have successfully identified a fundamental connection between spike threshold dynamic and input-output property in this study. We also provided a biophysical interpretation about how the nonlinear interactions between inward and outward currents at the perithersholds contribute to such connection. The powerful predictive ability of subthreshold biophysical properties is further attested in our work, which may be conducive to increase its future applications in neural coding.

Since the stochasticity is a prominent feature of neural system (Tuckwell, 1989; Gerstner and Kistler, 2002; Tuckwell and Jost, 2010), much effort has been devoted to exploring what effects of noise may produce on neuronal activity. A lot of modeling and experimental studies have reported that noise is able to enrich neuronal stochastic dynamics and trigger many complex behaviors near different bifurcation points. For example, it may induce stochastic firing patterns and enhance neuronal information transmission capability through coherence resonance near SNIC bifurcation (Gu et al., 2011; Jia et al., 2011; Jia and Gu, 2012), inhibit repetitive spiking through inverse stochastic resonance near Hopf bifurcation (Paydarfar et al., 2006; Tuckwell et al., 2009; Tuckwell and Jost, 2010, 2011, 2012; Guo, 2011), or completely destroy bifurcation scenarios and make neuronal response present a reliable feature (Tateno and Pakdaman, 2004). However, most of these studies focus on the phenomenological description of how noise impacts spiking behavior, while do not provide a satisfying explanation about the relation between neuronal intrinsic property and noisy effects. Unlike them, the present study associates noisy effects on spiking rhythms with neuronal intrinsic threshold dynamic. What is more, we provide a plausible biophysical interpretation for the observed noisy effects by relating them to the dynamical mechanism of spike initiation. All these investigations could provide a great insight into how noise participates in neural coding.

In addition, we adopt a novel approach proposed by Moujahid et al. (2011, 2014) and Moujahid and D'Anjou (2012) to characterize the electrochemical energy of the neuron with different spike threshold dynamics. This approach is based on the biophysical considerations about the nature of neuron model, which allows one to deduce an analytical expression of the electrochemical energy involved in the dynamics of the model. Contrary to the ion counting approach, this method does not need to calculate the number of Na^+ required to depolarize membrane when estimating energy consumption, and also it

requires no hypothesis about the extent of the overlapping between Na^+ and K^+ (Moujahid et al., 2011, 2014; Moujahid and D'Anjou, 2012). Thus, it could avoid the overestimate value of energy that results from the ionic-counting based method (Attwell and Laughlin, 2001; Alle et al., 2009; Hertz et al., 2013). With this approach, we have found a basic link between spike threshold, energy efficiency, and spiking frequency. It is shown that the average energy consumption rate increases with spiking frequency and could detect the transition of the neuron from quiescence to firing state, whereas the energy demand of a single spike decreases with spiking frequency. This relation between energy consumption and spiking frequency is consistent with that observed in the neocortex, hippocampus, thalamus, and squid axon (Moujahid and D'Anjou, 2012; Moujahid et al., 2014). As spike threshold gets depolarized, the average energy consumption rate gets smaller. Meanwhile, the energy demand for generating an action potential in the case of high stimulus also decreases. This demonstrates that depolarizing spike threshold could increase the energy efficiency of the neuron. We further show that the more efficient use of electrochemical energy in the case of more depolarized threshold is mainly due to the reduced overlap load between inward Na^+ and outward K^+ currents. Previous reports (Alle et al., 2009; Carter and Bean, 2009; Sengupta et al., 2010, 2013; Moujahid and D'Anjou, 2012; Moujahid et al., 2014) have proposed that if the Na^+ and K^+ currents have the substantially reduced overlap, the corresponding action potential is more energy efficient. Our stimulation results are consistent with this proposal. All these experimental and modeling observations suggest that the interactions between inward and outward currents could also determine the electrochemical energy required by the neuron to generate action potentials.

Conclusion

A dynamic spike threshold dependent on dV/dt plays a vital role in neural coding and spike initiation, which requires a number of metabolic energy. In this work, we have used a modified Morris-Lecar model to systematically investigate the input-output property and energy efficiency of the neuron with different spike threshold dynamics. To the best of our knowledge, this is the first study that links spike threshold dynamics, biophysical properties, spike initiation, input-output relations and energy efficiency together. The predictions and relevant mechanistic explanations could be tested by intracellular recording *in vivo*, and simultaneously more biophysically realistic simulations will be required if we want to replicate these biological effects more accurately. The systematic investigation about how spike threshold dynamics modulates neural input-output properties and energy efficiency is a useful stepwise method for exploring how spike threshold participates in neural coding. Moreover, translating the phenomenological descriptions into biophysical interpretation is crucial for revealing how membrane biophysics impacts neural coding. Thus, our stimulations could contribute to uncover the functional significance of spike threshold as well as biophysical properties in neural coding mechanism.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grants 61471265, 61401312, 61372010, and 61172009, and Tianjin Municipal Natural Science Foundation under Grants 12JCZDJC21100, 13JCZDJC27900, and 13JCQNJC03700.

References

Alle, H., Roth, A., and Geiger, J. R. P. (2009). Energy-efficient action potentials in hippocampal mossy fibers. *Science* 325, 1405–1408. doi: 10.1126/science.1174331

Attwell, D., and Laughlin, S. B. (2001). An energy budget for signaling in the grey matter of the brain. *J. Cereb. Blood Flow Metab.* 21, 1133–1145. doi: 10.1097/00004647-200110000-00001

Azouz, R., and Gray, C. M. (2000). Dynamic spike threshold reveals a mechanism for synaptic coincidence detection in cortical neurons *in vivo*. *Proc. Natl. Acad. Sci. U.S.A.* 97, 8110–8115. doi: 10.1073/pnas.130200797

Azouz, R., and Gray, C. M. (2003). Adaptive coincidence detection and dynamic gain control in visual cortical neurons *in vivo*. *Neuron* 37, 513–523. doi: 10.1016/S0896-6273(02)01186-8

Bekkers, J. M., and Delaney, A. J. (2001). Modulation of excitability by alpha-dendrotoxin-sensitive potassium channels in neocortical pyramidal neurons. *J. Neurosci.* 21, 6553–6560.

Cardin, J. A., Kumbhani, R. D., Contreras, D., and Palmer, L. A. (2010). Cellular mechanisms of temporal sensitivity in visual cortex neurons. *J. Neurosci.* 30, 3652–3662. doi: 10.1523/JNEUROSCI.5279-09.2010

Carter, B. C., and Bean, B. P. (2009). Sodium entry during action potentials of mammalian neurons: incomplete inactivation and reduced metabolic efficiency in fast-spiking neurons. *Neuron* 64, 898–909. doi: 10.1016/j.neuron.2009.12.011

Dayan, P., and Abbott, L. F. (2005). *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*. London: The MIT Press.

Destexhe, A., Rudolph, M., Fellous, J. M., and Sejnowski, T. J. (2001). Fluctuating synaptic conductances recreate *in vivo*-like activity in neocortical neurons. *Neuroscience* 107, 13–24. doi: 10.1016/S0306-4522(01)00344-X

Dodson, P. D., Barker, M. C., and Forsythe, I. D. (2002). Two heteromeric Kv1 potassium channels differentially regulate action potential firing. *J. Neurosci.* 22, 6953–6961.

Ermentrout, B. (1996). Type I membranes, phase resetting curves, and synchrony. *Neural Comput.* 8, 979–1001. doi: 10.1162/neco.1996.8.5.979

Ermentrout, B. (2002). *Simulating, Analyzing, and Animating Dynamical Systems: a Guide to Xppaut for Researchers and Students*. Philadelphia, PA: SIAM. doi: 10.1137/1.9780898718195

Escabí, M. A., Nassiri, R., Miller, L. M., Schreiner, C. E., and Read, H. L. (2005). The contribution of spike threshold to acoustic feature selectivity, spike information content, and information throughput. *J. Neurosci.* 25, 9524–9534. doi: 10.1523/JNEUROSCI.1804-05.2005

Ferragamo, M. J., and Oertel, D. (2002). Octopus cells of the mammalian ventral cochlear nucleus sense the rate of depolarization. *J. Neurophysiol.* 87, 2262–2270. doi: 10.1152/jn.00587.2001

Fink, C. G., Booth, V., and Zochowski, M. (2011). Cellularly-driven differences in network synchronization propensity are differentially modulated by firing frequency. *PLoS Comput. Biol.* 7:e1002062. doi: 10.1371/journal.pcbi.1002062

Fontaine, B., Peña, J. L., and Brette, R. (2014). Spike-threshold adaptation predicted by membrane potential dynamics *in vivo*. *PLoS Comput. Biol.* 10:e1003560. doi: 10.1371/journal.pcbi.1003560

Gerstner, W., and Kistler, W. M. (2002). *Spiking Neuron Models*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511815706

Goldberg, E. M., Clark, B. D., Zagha, E., Nahmani, M., Erisir, A., and Rudy, B. (2008). K+ channels at the axon initial segment dampen near-threshold excitability of neocortical fast-spiking GABAergic interneurons. *Neuron* 58, 387–400. doi: 10.1016/j.neuron.2008.03.003

Gu, H., Zhang, H., Wei, C., Yang, M., Liu, Z., and Ren, W. (2011). Coherence resonance induced stochastic neural firing at a saddle-node bifurcation. *Int. J. Mod. Phys. B* 25, 3977–3986. doi: 10.1142/S0217979211101673

Guan, D., Lee, J. C. F., Higgs, M. H., Spain, W. J., and Foehring, R. C. (2007). Functional roles of Kv1 channels in neocortical pyramidal neurons. *J. Neurophysiol.* 97, 1931–1940. doi: 10.1152/jn.00933.2006

Guo, D. (2011). Inhibition of rhythmic spiking by colored noise in neural systems. *Cogn. Neurodyn.* 5, 293–300. doi: 10.1007/s11571-011-9160-2

Hansel, D., Mato, G., and Meunier, C. (1995). Synchrony in excitatory neural networks. *Neural Comput.* 7, 307–337. doi: 10.1162/neco.1995.7.2.307

Henze, D. A., and Buzsáki, G. (2001). Action potential threshold of hippocampal pyramidal cells *in vivo* is increased by recent spiking activity. *Neuroscience* 105, 121–130. doi: 10.1016/S0306-4522(01)00167-1

Hertz, L., Junnan, X., Dan, S., Enzhi, Y., Li, G., and Liang, P. (2013). Astrocytic and neuronal accumulation of elevated extracellular K(+) with a 2/3 K(+)/Na(+) flux ratio-consequences for energy metabolism, osmolarity and higher brain function. *Front. Comput. Neurosci.* 7:114. doi: 10.3389/fncom.2013.00114

Higgs, M. H., and Spain, W. J. (2011). Kv1 channels control spike threshold dynamics and spike timing in cortical pyramidal neurons. *J. Physiol.* 589, 5125–5142. doi: 10.1113/jphysiol.2011.216721

Hodgkin, A. L., and Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol.* 117, 500–544. doi: 10.1113/jphysiol.1952.sp004764

Izhikevich, E. M. (2005). *Dynamical Systems in Neuroscience: the Geometry of Excitability and Bursting*. Cambridge, MA: The MIT Press.

Jia, B., and Gu, H. (2012). Identifying type I excitability using dynamics of stochastic neural firing patterns. *Cogn. Neurodyn.* 6, 485–497. doi: 10.1007/s11571-012-9209-x

Jia, B., Gu, H., and Li, Y. (2011). Coherence-resonance-induced neuronal firing near a saddle-node and homoclinic bifurcation corresponding to type-I excitability. *Chinese Phys. Lett.* 28, 090507. doi: 10.1088/0256-307X/28/9/090507

Klausberger, T., and Somogyi, P. (2008). Neuronal diversity and temporal dynamics: the unity of hippocampal circuit operations. *Science* 321, 53–57. doi: 10.1126/science.1149381

Koch, C. (1999). *Biophysics of Computation: Information Processing in Single Neurons*. New York, NY: Oxford University Press.

Kuba, H., Ishii, T. M., and Ohmori, H. (2006). Axonal site of spike initiation enhances auditory coincidence detection. *Nature* 444, 1069–1072. doi: 10.1038/nature05347

Moujahid, A., and D'Anjou, A. (2012). Metabolic efficiency with fast spiking in the squid axon. *Front. Comput. Neurosci.* 6:95. doi: 10.3389/fncom.2012.00095

Moujahid, A., D'Anjou, A., and Graña, M. (2014). Energy demands of diverse spiking cells from the neocortex, hippocampus, and thalamus. *Front. Comput. Neurosci.* 8:41. doi: 10.3389/fncom.2014.00041

Moujahid, A., D'Anjou, A., Torrealdea, F. J., and Torrealdea, F. (2011). Energy and information in Hodgkin-Huxley neurons. *Phys. Rev. E* 83:031912. doi: 10.1103/PhysRevE.83.031912

Niven, J. E., and Laughlin, S. B. (2008). Energy limitation as a selective pressure on the evolution of sensory systems. *J. Exp. Biol.* 211, 1792–1804. doi: 10.1242/jeb.017574

Paydarfar, D., Forger, D. B., and Clay, J. R. (2006). Noisy inputs and the induction of on-off switching behavior in a neuronal pacemaker. *J. Neurophysiol.* 96, 3338–3348. doi: 10.1152/jn.00486.2006

Platkiewicz, J., and Brette, R. (2011). Impact of fast sodium channel inactivation on spike threshold dynamics and synaptic integration. *PLoS Comput. Biol.* 7:e1001129. doi: 10.1371/journal.pcbi.1001129

Prescott, S. A., De Koninck, Y., and Sejnowski, T. J. (2008a). Biophysical basis for three distinct dynamical mechanisms of action potential initiation. *PLoS Comput. Biol.* 4:e1000198. doi: 10.1371/journal.pcbi.1000198

Prescott, S. A., Ratté, S., De Koninck, Y., and Sejnowski, T. J. (2008b). Pyramidal neurons switch from integrators *in vitro* to resonators under *in vivo*-like conditions. *J. Neurophysiol.* 100, 3030–3042. doi: 10.1152/jn.90634.2008

Prescott, S. A., and Sejnowski, T. J. (2008). Spike-rate coding and spike-time coding are affected oppositely by different adaptation mechanisms. *J. Neurosci.* 28, 13649–13661. doi: 10.1523/JNEUROSCI.1792-08.2008

Priebe, N. J., and Ferster, D. (2008). Inhibition, spike threshold, and stimulus selectivity in primary visual cortex. *Neuron* 57, 482–497. doi: 10.1016/j.neuron.2008.02.005

Rinzel, J., and Ermentrout, G. B. (1998). *Analysis of Neural Excitability and Oscillations*. Cambridge, MA: The MIT Press.

Rothman, J. S., and Manis, P. B. (2003a). Differential expression of three distinct potassium currents in the ventral cochlear nucleus. *J. Neurophysiol.* 89, 3070–3082. doi: 10.1152/jn.00125.2002

Rothman, J. S., and Manis, P. B. (2003b). Kinetic analyses of three distinct potassium currents in the ventral cochlear nucleus. *J. Neurophysiol.* 89, 3083–3096. doi: 10.1152/jn.00126.2002

Rothman, J. S., and Manis, P. B. (2003c). The roles potassium currents play in regulating the electrical activity of ventral cochlear nucleus neurons. *J. Neurophysiol.* 89, 3097–3113. doi: 10.1152/jn.00127.2002

Schultheiss, N. W., Prinz, A. A., and Butera, R. J. (2012). *Phase Response Curves in Neuroscience*. New York, NY: Springer-Verlag. doi: 10.1007/978-1-4614-0739-3

Sengupta, B., Faisal, A. A., Laughlin, S. B., and Niven, J. E. (2013). The effect of cell size and channel density on neuronal information encoding and energy efficiency. *J. Cereb. Blood Flow Metab.* 33, 1465–1473. doi: 10.1038/jcbfm.2013.103

Sengupta, B., Laughlin, S. B., and Niven, J. E. (2014). Consequences of converting graded to action potentials upon neural information coding and energy efficiency. *PLoS Comput. Biol.* 10:e1003439. doi: 10.1371/journal.pcbi.1003439

Sengupta, B., Stemmler, M., Laughlin, S. B., and Niven, J. E. (2010). Action potential energy efficiency varies among neuron types in vertebrates and invertebrates. *PLoS Comput. Biol.* 6:e1000840. doi: 10.1371/journal.pcbi.1000840

Smeal, R. M., Ermentrout, G. B., and White, J. A. (2010). Phase-response curves and synchronized neural networks. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 365, 2407–2422. doi: 10.1098/rstb.2009.0292

Storm, J. F. (1988). Temporal integration by a slowly inactivating K+ current in hippocampal neurons. *Nature* 336, 379–381. doi: 10.1038/336379a0

Tateno, T., and Pakdaman, K. (2004). Random dynamics of the Morris-Lecar neural model. *Chaos* 14, 511–530. doi: 10.1063/1.1756118

Tuckwell, H. C. (1989). *Stochastic Processes in the Neurosciences*. Philadelphia, PA: SIAM. doi: 10.1137/1.9781611970159

Tuckwell, H. C., and Jost, J. (2010). Weak noise in neurons may powerfully inhibit the generation of repetitive spiking but not its propagation. *PLoS Comput. Biol.* 6:e1000794. doi: 10.1371/journal.pcbi.1000794

Tuckwell, H. C., and Jost, J. (2011). The effects of various spatial distributions of weak noise on rhythmic spiking. *J. Comput. Neurosci.* 30, 361–371. doi: 10.1007/s10827-010-0260-5

Tuckwell, H. C., and Jost, J. (2012). Analysis of inverse stochastic resonance and the long-term firing of Hodgkin-Huxley neurons with Gaussian white noise. *Phys. A Stat. Mech. Appl.* 391, 5311–5325. doi: 10.1016/j.physa.2012.06.019

Tuckwell, H. C., Jost, J., and Gutkin, B. S. (2009). Inhibition and modulation of rhythmic neuronal spiking by noise. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 80:031907. doi: 10.1103/PhysRevE.80.031907

Uhlenbeck, G. E., and Ornstein, L. S. (1930). On the theory of brownian motion. *Phys. Rev.* 36, 823–841. doi: 10.1103/PhysRev.36.823

Wester, J. C., and Contreras, D. (2013). Biophysical mechanism of spike threshold dependence on the rate of rise of the membrane potential by sodium channel inactivation or subthreshold axonal potassium current. *J. Comput. Neurosci.* 35, 1–17. doi: 10.1007/s10827-012-0436-2

Wilent, W. B., and Contreras, D. (2005). Stimulus-dependent changes in spike threshold enhance feature selectivity in rat barrel cortex neurons. *J. Neurosci.* 25, 2983–2991. doi: 10.1523/JNEUROSCI.4906-04.2005

Yi, G. S., Wang, J., Wei, X. L., Tsang, K. M., Chan, W. L., and Deng, B. (2014a). Neuronal spike initiation modulated by extracellular electric fields. *PLoS ONE* 9:e97481. doi: 10.1371/journal.pone.0097481

Yi, G. S., Wang, J., Wei, X. L., Tsang, K. M., Chan, W. L., Deng, B., et al. (2014b). Exploring how extracellular electric field modulates neuron activity through dynamical analysis of a two-compartment neuron model. *J. Comput. Neurosci.* 36, 383–399. doi: 10.1007/s10827-013-0479-z

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.

frontiers in
COMPUTATIONAL NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 04 February 2014
doi: 10.3389/fncom.2014.00008

Linear stability in networks of pulse-coupled neurons

***Simona Olmi*[1,2], *Alessandro Torcini*[1,2]* *and Antonio Politi*[1,3]**

[1] *Consiglio Nazionale delle Ricerche, Istituto dei Sistemi Complessi, Sesto Fiorentino, Italy*
[2] *INFN—Sezione di Firenze and CSDC, Sesto Fiorentino, Italy*
[3] *SUPA and Institute for Complex Systems and Mathematical Biology, King's College, University of Aberdeen, Aberdeen, UK*

Edited by:
Tobias A. Mattei, Ohio State University, USA

Reviewed by:
Marc Timme, Max Planck Institute for Dynamics and Self Organization, Germany
Tobias A. Mattei, Ohio State University, USA

***Correspondence:**
Alessandro Torcini, Consiglio Nazionale delle Ricerche, Istituto dei Sistemi Complessi, Via Madonna del Piano 10, I-50019 Sesto Fiorentino, Italy
e-mail: alessandro.torcini@cnr.it

In a first step toward the comprehension of neural activity, one should focus on the stability of the possible dynamical states. Even the characterization of an idealized regime, such as that of a perfectly periodic spiking activity, reveals unexpected difficulties. In this paper we discuss a general approach to linear stability of pulse-coupled neural networks for generic phase-response curves and post-synaptic response functions. In particular, we present: (1) a mean-field approach developed under the hypothesis of an infinite network and small synaptic conductances; (2) a "microscopic" approach which applies to finite but large networks. As a result, we find that there exist two classes of perturbations: those which are perfectly described by the mean-field approach and those which are subject to finite-size corrections, irrespective of the network size. The analysis of perfectly regular, asynchronous, states reveals that their stability depends crucially on the smoothness of both the phase-response curve and the transmitted post-synaptic pulse. Numerical simulations suggest that this scenario extends to systems that are not covered by the perturbative approach. Altogether, we have described a series of tools for the stability analysis of various dynamical regimes of generic pulse-coupled oscillators, going beyond those that are currently invoked in the literature.

Keywords: linear stability analysis, splay states, synchronization, neural networks, pulse coupled neurons, Floquet spectrum

1. INTRODUCTION

Networks of oscillators play an important role in both biological (neural systems, circadian rhythms, population dynamics) (Pikovsky et al., 2003) and physical contexts (power grids, Josephson junctions, cold atoms) (Hadley and Beasley, 1987; Filatrella et al., 2008; Javaloyes et al., 2008). It is therefore comprehensible that many studies have been and are still devoted to understanding their dynamical properties. Since the development of sufficiently powerful tools and the resulting discovery of general laws is an utterly difficult task, it is convenient to start from simple setups.

The first issue to consider is the model structure of the single oscillators. Since phases are typically more sensitive than amplitudes to mutual coupling, they are likely to provide the most relevant contribution to the collective evolution (Pikovsky et al., 2003). Accordingly, here we restrict our analysis to oscillators characterized by a single, phase-like, variable. This is typically done by reducing the neuronal dynamics to the evolution of the membrane potential and introducing the corresponding *velocity field* which describes the single-neuron activity. Equivalently, one can map the membrane potential onto a phase variable and simultaneously introduce a phase-response curve (PRC) [Upon changing variables, the velocity field can be made independent of the local variable (as intuitively expected for a true phase). When this is done, the phase dependence of the velocity field is moved to the coupling function, i.e., to the PRC] to take into account the dependence of the neuronal response on the current value of the membrane potential (i.e., the phase). In this paper we adopt the first point of view, with a few exceptions, when the second one is mathematically more convenient.

As for the coupling, two mechanisms are typically invoked in the literature, diffusive and pulse-mediated. While the former mechanism is pretty well understood [see e.g., the very many papers devoted to Kuramoto-like models (Acebrón et al., 2005)], the latter one, more appropriate in neural dynamics, involves a series of subtleties that have not yet been fully appreciated. This is why here we concentrate on pulse-coupled oscillators.

Finally, for what concerns the topology of the interactions, it is known that they can heavily influence the dynamics of the neural systems leading to the emergence of new collective phenomena even in weakly connected networks (Timme, 2006), or of various types of chaotic behavior, ranging from weak chaos for diluted systems (Popovych et al., 2005; Olmi et al., 2010) to extensive chaos in sparsely connected ones (Monteforte and Wolf, 2010; Luccioli et al., 2012). We will, however, limit our analysis to globally coupled identical oscillators, which provide a much simplified, but already challenging, test bed. The high symmetry of the corresponding evolution equations simplifies the identification of the stationary solutions and the analysis of their stability properties. The two most symmetric solutions are: (1) the fully synchronous state, where all oscillators follow exactly the same trajectory; (2) the splay state (also known as "ponies on a merry-go-round," antiphase state or rotating waves) (Hadley and Beasley, 1987; Ashwin et al., 1990; Aronson et al., 1991), where the oscillators still follow the same periodic trajectory, but with different (evenly distributed) time shifts. The former solution is

the simplest representative of the broad class of clustered states (Golomb and Rinzel, 1994), where several oscillators behave in the same way, while the latter is the prototype of asynchronous states, characterized by a smooth distribution of phases (Renart et al., 2010).

In spite of the many restrictions on the mathematical setup, the stability of the synchronous and splay states still depend significantly on additional features such as the synaptic response-function, the velocity field, and the presence of delay in the pulse transmission. As a result, one can encounter splay states that are either strongly stable along all directions, or that present many almost-marginal directions, or, finally, that are marginally stable along various directions (Nichols and Wiesenfeld, 1992; Watanabe and Strogatz, 1994). Several analytic results have been obtained in specific cases, but a global picture is still missing: the goal of this paper is to recompose the puzzle, by exploring the role of the velocity field (or, equivalently, of the phase response curve) and of the shape of the transmitted post-synaptic potentials. Although we are neither going to discuss the role of delay nor that of the network topology, it is useful to recall the stability analysis of the synchronous state in the presence of delayed δ-pulses and for arbitrary topology, performed by Timme and Wolf in Timme and Wolf (2008). There, the authors show that even the complete knowledge of the spectrum of the linear operator does not suffice to address the stability of the synchronized state.

The stability analysis of the fully synchronous regime is far from being trivial even for a globally coupled network of oscillators with no delay in the pulse transmission: in fact, the pulse emission introduces a discontinuity which requires separating the evolution before and after such event. Moreover, when many neurons spike at the same time, the length of some interspike intervals is virtually zero but cannot be neglected in the mathematical analysis. In fact, the first study of this problem was restricted to excitatory coupling and δ-pulses (Mirollo and Strogatz, 1990). In that context, the stability of the synchronous state follows from the fact that when the phases of two oscillators are sufficiently close to one another, they are instantaneously reset to the same value (as a result of a non-physical lack of invertibility of the dynamics). The first, truly linear stability analyses have been performed later, first in the case of two oscillators (van Vreeswijk et al., 1994; Hansel et al., 1995) and then considering δ-pulses with continuous PRCs (Goel and Ermentrout, 2002). Here, we extend the analysis to generic pulse-shapes and discontinuous PRCs [such as for leaky integrate and fire (LIF) neurons].

As for the splay states, their stability can be assessed in two ways: (1) by assuming that the number of oscillators is infinite (i.e., taking the so called thermodynamic limit) and thereby studying the evolution of the distribution of the membrane potentials—this approach is somehow equivalent to dealing with (macroscopic) Liouville-type equations in statistical mechanics; (2) by dealing with the (microscopic) equations of motion for a large but finite number N of oscillators. As shown in some pioneering works (Kuramoto, 1991; Treves, 1993), the former approach corresponds to develop a mean field theory. The resulting equations have been first solved in Abbott and van Vreeswijk (1993) for pulses composed of two exponential functions, in the limit of a small effective coupling [A small effective coupling can arise also when PRC has a very weak dependence on the phase (see section 3)]. Here, following Abbott and van Vreeswijk (1993), we extend the analysis to generic pulse-shapes, finding that substantial differences exist among δ, exponential and the so-called α-pulses (see the next section for a proper definition).

Direct numerical studies of the linear stability of finite networks suggest that the eigenfunctions of the (Floquet) operator can be classified according to their wavelength ℓ (where ℓ refers to the neuronal phase—see section 4.1 for a precise definition). In finite systems, it is convenient to distinguish between long (LW) and short (SW) wavelengths. Upon considering that $\ell = n/N$ ($1 \leq n \leq N$), LW can be identified as those for which $n \ll N$, while SW correspond to larger n values. Numerical simulations suggest also that the time scale of a LW perturbation typically increases upon increasing its wavelength, starting from a few milliseconds (for small n values) up to much longer values (when n is on the order of the network size N) which depend on "details" such as the continuity of the velocity field, or the pulse shape. On the other hand, SW are characterized by a slow size-dependent dynamics.

For instance, in LIF neurons coupled via α-pulses, it has been found (Calamai et al., 2009) that the Floquet exponents of LW decrease as $1/\ell^2$ (for large ℓ), while the time scale of the SW component is on the order of N^2. In practice the LW spectral component as determined from the finite N analysis coincides with that one obtained with the mean field approach (i.e., taking first the thermodynamic limit). As for the SW component, it cannot be quantitatively determined by the mean-field approach, but it is nevertheless possible to infer the correct order of magnitude of this time scale. In fact, upon combining the $1/\ell^2$ decay (predicted by the mean-field approach) with the observation that the minimal wavelength is $1/N$, it naturally follows that the SW time scale is N^2, as analytically proved in Olmi et al. (2012). Furthermore, it has been found that the two spectral components smoothly connect to each other and the predictions of the two theoretical approaches coincide in the crossover region.

It is therefore important to investigate whether the same agreement extends to more generic pulse shapes and velocity fields. The finite-N approach can, in principle, be generalized to arbitrary shapes, but the analytic calculations would be quite lengthy, due to the need of distinguishing between fast and slow scales and the need of accounting for higher order terms. For this reason, here we limit ourselves to give a positive answer to this question with the help of numerical studies.

The only, important, exception to this scenario is obtained for quasi δ-like pulses (Zillmer et al., 2007), i.e., for pulses whose width is smaller than the average time separation between any two consecutive spikes, in which case all the SW eigenvalues remain finite for increasing N.

In section 2 we introduce the model and derive the corresponding event-driven map, a necessary step before undertaking the analytic calculations. Section 3 is devoted to a perturbative stability analysis of the splay state in the infinite-size limit for generic velocity fields and pulse shapes. The following section 4 reports a discussion of the stability in finite networks. There we briefly recall the main results obtained in Olmi et al. (2012) for

the splay state and we extensively discuss the method to quantify the stability of the fully synchronous regime. The following two sections are devoted to a numerical analysis of various setups. In section 5 we study splay states in finite networks for generic velocity fields and three different classes of of pulses, namely, with finite, vanishing ($\approx 1/N$), and zero width. In section 6 we study periodically forced networks. Such studies show that the scaling relations derived for the splay states apply also to such a microscopically quasi-periodic regime. A brief summary of the main results together with a recapitulation of the open problem is finally presented in section 7. In the first appendix we derive the Fourier components needed to assess the stability of a splay state for a generic PRC. In the second appendix the evaporation exponent is determined for the synchronous state in LIF neurons.

2. THE MODEL

The general setup considered in this paper is a network of N identical pulse-coupled neurons (rotators), whose evolution is described by the equation

$$\dot{X}^j = F(X^j) + gE(t), \quad j = 1, \ldots, N \tag{1}$$

where X^j represents the membrane potential, g is the coupling constant and the *mean field* $E(t)$ denotes to the synaptic input, common to all neurons in the network. When X^j reaches the threshold value $X^j = 1$, it is reset to $X^j = 0$ and a spike contributes to the mean field E in a way that is described here below. The resetting procedure is an approximation of the discharge mechanism operating in real neurons. The function $F(X)$ (the velocity field) is assumed to be everywhere positive, thus ensuring that the neuron is repetitively firing. For $F_0(X) = a - X$ the model reduces to the well-known case of LIF neurons.

The mean field $E(t)$ arises from the linear superposition of the pulses emitted by the single neurons. In order to describe its time evolution, it is sufficient to introduce a suitable ordinary differential equation (ODE), such that its Green function reproduces the expected pulse shape,

$$E^{(L)} = \sum_{i}^{L-1} a_i E^{(i)} + \frac{K}{N} \sum_{n|t_n<t} \delta(t - t_n), \tag{2}$$

where the superscript (i) denotes the ith time derivative, L the order of the differential equation and $K = \prod_i \alpha_i$, ($-\alpha_i$ being the poles of the differential equation), so as to ensure that the single pulses have unit area (for $N = 1$). The δ-functions appearing on the right hand side of Equation (2) correspond to the spikes emitted at times $\{t_n\}$: each time a spike is emitted, the term $E^{(L-1)}$ has a finite jump of amplitude K/N. Therefore L controls the smoothness of the pulses: $L - 1$ is the order of the lowest derivative that is discontinuous. $L = 0$ corresponds to the extreme case of δ-pulses with no field dynamics; $L = 1$ corresponds to discontinuous exponential pulses; $L = 2$ (with $\alpha_1 = \alpha_2$) to the so-called α-pulses ($E_s(t) = \alpha^2 t e^{-\alpha t}$). Since α-pulses will be often referred to, it is worth being a little more specific. In this case, Equation (2) reduces to

$$\ddot{E}(t) + 2\alpha\dot{E}(t) + \alpha^2 E(t) = \frac{\alpha^2}{N} \sum_{n|t_n<t} \delta(t - t_n), \tag{3}$$

and it is convenient to transform this equation into a system of two ODEs, namely

$$\dot{E} = P - \alpha E, \quad \dot{P} + \alpha P = \frac{\alpha^2}{N} \sum_{n|t_n<t} \delta(t - t_n), \tag{4}$$

where we have introduced, for the sake of simplicity, the auxiliary variable $P \equiv \alpha E + \dot{E}$.

2.1. EVENT-DRIVEN MAP

By following Zillmer et al. (2006) and Calamai et al. (2009), it is convenient to pass from a continuous—to a discrete-time evolution rule, by deriving the event-driven map which connects the network configuration at consecutive spike times. For the sake of simplicity, in the following part of this section we refer to α-pulses, but there is no conceptual limitation in extending the approach to $L > 2$.

By integrating Equation (4), we obtain

$$E_{n+1} = E_n e^{-\alpha \mathcal{T}_n} + P_n \mathcal{T}_n e^{-\alpha \mathcal{T}_n} \tag{5}$$

$$P_{n+1} = P_n e^{-\alpha \mathcal{T}_n} + \frac{\alpha^2}{N}, \tag{6}$$

where we have taken into account the effect of the incoming pulse (see the term α^2/N in the second equation) while $\mathcal{T}_n = t_{n+1} - t_n$ is the interspike interval; t_{n+1} corresponds to the time when the neuron with the largest membrane potential reaches the threshold.

Since all neurons follow the same first-order differential equation (this is a mean-field model), the ordering of their membrane potentials is preserved [neurons "rotate" around the circle [0, 1] without overtaking each other (Jin, 2002)]. It is, therefore, convenient to order the potentials from the largest to the smallest one and to introduce a co-moving reference frame, i.e., to shift backward the label j, each time a neuron reaches the threshold. By formally integrating Equation (1),

$$X^j_{n+1} = \mathcal{F}(X^{j+1}_n, \mathcal{T}_n) + g\frac{e^{-\mathcal{T}_n} - e^{-\alpha \mathcal{T}_n}}{\alpha - 1}\left(E_n + \frac{P_n}{\alpha - 1}\right) - g\frac{\mathcal{T}_n e^{-\alpha \mathcal{T}_n}}{(\alpha - 1)} P_n. \tag{7}$$

Moreover, since X^1_n is always the largest potential, the interspike interval is defined by the threshold condition

$$X^1_n(\mathcal{T}_n, E_n, P_n) \equiv 1. \tag{8}$$

Altogether, the model now reads as a discrete-time map, involving $N + 1$ variables, E_n, P_n, and X^j_n ($1 \leq j < N$), since one degree of freedom has been eliminated as a result of having taken the Poincaré section ($X^N_n \equiv 0$ due to the resetting mechanism). The advantage of the map description is that we do not have to deal any longer with δ-like discontinuities, or with formally infinite sequences of past events.

In this framework, the splay state is a fixed point of the event-driven map. Its coordinates can be determined in the following

way. From Equation (5), one can express $\tilde{P}$ and $\tilde{E}$ as a function of the yet unknown interspike interval $\mathcal{T}$,

$$\tilde{P} = \frac{\alpha^2}{N}(1 - e^{-\alpha\mathcal{T}})^{-1} \quad \tilde{E} = \mathcal{T}\tilde{P}(e^{\alpha\mathcal{T}} - 1)^{-1}. \tag{9}$$

The value of the membrane potentials $\tilde{X}^k$ are then obtained by iterating backward in j Equation (7) (the n dependence is dropped for the fixed point) starting from the initial condition $\tilde{X}^N = 0$. The interspike interval $\mathcal{T}$ is finally obtained by imposing the condition $\tilde{X}^0 = 0$. In practice the computational difficulty amounts to finding the zero of a one dimensional function and, even though $\mathcal{F}(X^{j+1}, \mathcal{T})$ can, in most cases, be obtained only through numerical integration, the final error can be very well kept under control.

3. THEORY ($N = \infty$)

The stability of a dynamical state can be assessed by either first taking the infinite-time limit and then the thermodynamic limit, or vice versa. In general it is not obvious whether the two methods yield the same result and this is particularly crucial for the splay state, as many eigenvalues tend to 0 for $N \to \infty$. In this section we discuss the scenarios that have to be expected when the thermodynamic limit is taken first. We do that by following Abbott and van Vreeswijk (1993).

As a first step, it is convenient to introduce the phase-like variable

$$y^i = \int_0^{X^i} \frac{dx}{G(x)}, \qquad 0 \le y^i \le 1 \tag{10}$$

where, for later convenience, we have defined $G(X) \equiv g + T_0 F(X)$, $T_0 = N\mathcal{T}$ being the period of the splay state (i.e., the single-neuron interspike interval). The phase y^i evolves according to the equation

$$\frac{dy^i}{dt} = \tilde{E} + \frac{g\varepsilon(t)}{G(X(y^i))} \tag{11}$$

where $\tilde{E} = 1/T_0$ is the amplitude of the field in the splay state, $\varepsilon(t) = E(t) - \tilde{E}$. In the splay state, since $\varepsilon = 0$, y^i grows linearly in time, as indeed expected for a well-defined phase. In the thermodynamic limit, the evolution is ruled by the continuity equation

$$\frac{\partial \rho}{\partial t} = -\frac{\partial J}{\partial y} \tag{12}$$

where $\rho(y, t)dy$ is the fraction of neurons whose phase y^i lies in $(y, y + dy)$ at time t, and

$$J(y, t) = \left[\tilde{E} + \frac{g\varepsilon(t)}{G(X(y))}\right]\rho(y, t) \tag{13}$$

is the corresponding flux. As the resetting implies that the outgoing flux $J(1, t)$ (which coincides with the firing rate) equals the incoming flux at the origin, the above equation has to be complemented with the boundary condition $J(0, t) = J(1, t)$. Finally, in this macroscopic representation, the field equation writes

$$\varepsilon^{(L)} = \sum_i^{L-1} a_i \varepsilon^{(i)} + K(J(1, t) - \tilde{E}), \tag{14}$$

while the splay state corresponds to the fixed point $\rho = 1$, $\varepsilon = 0$, $J = \tilde{E}$. The smoothness of the splay state justifies the use of a partial differential equation such as (Equation 12). Its stability can be studied by introducing the perturbation $j(y, t)$

$$j(y, t) = J(y, t) - \tilde{E}, \tag{15}$$

and linearizing the continuity equation,

$$\frac{\partial j}{\partial t} = \frac{g}{G(X(y))}\frac{\partial \varepsilon}{\partial t} - \tilde{E}\frac{\partial j}{\partial y}. \tag{16}$$

while the field equation simplifies to

$$\varepsilon^{(L)} = \sum_i^{L-1} a_i \varepsilon^{(i)} + Kj(1, t). \tag{17}$$

By now introducing the Ansatz

$$j(y, t) = j_f(y)e^{\lambda t} \qquad \varepsilon(y, t) = \varepsilon_f(y)e^{\lambda t}, \tag{18}$$

in Equations (16) and (17) and, thereby solving the resulting ODE, one can obtain an implicit expression for $j_f(y)$,

$$j_f(y) = e^{-\lambda y/\tilde{E}}\left[1 + \frac{gK\lambda\ j_f(1)}{\tilde{E}\prod_{k=1}^{L}(\lambda + \alpha_k)}\int_0^y dz \frac{e^{\lambda z/\tilde{E}}}{G(X(z))}\right],$$

where $-\alpha_k$ and K are defined as below Equation (2). By imposing the boundary condition for the flux, $j_f(1) = j_f(0) = 1$, one finally obtains the eigenvalue equation (Abbott and van Vreeswijk, 1993),

$$\left(e^{\lambda/\tilde{E}} - 1\right)\prod_{k=1}^{L}(\lambda + \alpha_k) = \frac{gK\lambda}{\tilde{E}}\int_0^1 dy \frac{e^{\lambda y/\tilde{E}}}{G(X(y))}. \tag{19}$$

In the case of a constant $G(X(y)) = \sigma$, L eigenvalues correspond to the zeroes of the following polynomial equation

$$\prod_{k=1}^{L}(\lambda + \alpha_k) = \frac{gK}{\sigma}. \tag{20}$$

For $g = 0$ such solutions are the poles which define the field dynamics, while for $g = \sigma$, $\lambda = 0$ is a solution: this corresponds to the maximal value of the (positive) coupling strength beyond which the model does no longer support stationary states, as the feedback induces an unbounded growth of the spiking rate.

Besides such L solution, the spectrum is composed of an infinite set of purely imaginary eigenvalues,

$$\lambda = 2\pi i n\tilde{E} = \frac{2\pi i n}{T_0} \quad n \neq 0. \tag{21}$$

The existence of such marginally stable directions reflects the fact that all y^i phases experience the same velocity field, independently of their current value (see Equation 11), so that no effective interaction is present among the oscillators. In the limit of small variations of $G(X(y))$, one can develop a perturbative approach. Here below, we proceed under the more restrictive assumption that the coupling constant g is itself small: we have checked that this restriction does not change the substance of our conclusions, while requiring a simpler algebra.

A small g value implies that λ is close to $2\pi i n\tilde{E}$ and thereby expand the exponential in Equation (19). Up to first order, we find

$$\lambda_n = 2\pi i n\tilde{E}\left[1 + \frac{gK(A_n + iB_n)}{\prod_{k=1}^{L}(2\pi i n\tilde{E} + \alpha_k)}\right] \tag{22}$$

where

$$(A_n + iB_n) = \int_0^1 dy \frac{e^{i2\pi n y}}{G(X(y))} \tag{23}$$

are the Fourier components of the phase-response curve $1/G(X(y))$.

In order to estimate the leading terms of the real part of λ_n in the large n limit, let us rewrite Equation (22) as

$$\lambda_n = i\gamma_n + gK\gamma_n \frac{-B_n + iA_n}{\prod_{k=1}^{L}(\alpha_k^2 + \gamma_n^2)} \prod_{k=1}^{L}(\alpha_k - i\gamma_n) \tag{24}$$

where $\gamma_n = 2\pi n\tilde{E} = (2\pi n)/T_0$. Since γ_n is proportional to n, the leading terms in the product at numerator of Equation (24) are

$$\prod_{k=1}^{L}(\alpha_k - i\gamma_n) \sim (-i)^L \gamma_n^L + S(-i)^{L-1}\gamma_n^{L-1}, \tag{25}$$

where $S = \sum_{k=1}^{L}\alpha_k$ while the leading term in the product at denominator in Equation (24) is γ_n^{2L}. Accordingly, the main contribution to the real part of the eigenvalues is, in the case of even L,

$$Re\{\lambda_n\} \sim gK(-1)^{L/2}\left[\frac{SA_n}{\gamma_n^L} - \frac{B_n}{\gamma_n^{L-1}}\right] \tag{26}$$

and, for odd L,

$$Re\{\lambda_n\} \sim gK(-1)^{(L+3)/2}\left[\frac{A_n}{\gamma_n^{L-1}} + \frac{SB_n}{\gamma_n^L}\right]. \tag{27}$$

An exact expression for the Fourier components A_n and B_n appearing in Equation (23) can be derived in the large n limit. In particular, the integral over the interval $[0, 1]$ appearing in Equation (23) can be rewritten as a sum of integrals, each performed on a sub-interval of vanishingly small length $1/n$. Furthermore, since the phase-response $1/G$ has a limited variation within each sub-interval, it can be replaced by its polynomial expansion up to second order. Finally, as shown in Appendix A, the following expression are obtained at the leading order in $1/n$ for a discontinuous $F(X)$

$$A_n \simeq \frac{-T_0}{4\pi^2 n^2}\left[\frac{F'(1)}{G(1)^2} - \frac{F'(0)}{G(0)^2}\right], \tag{28}$$

$$B_n \simeq \frac{T_0}{2\pi n}\left[\frac{F(1) - F(0)}{G(1)G(0)}\right]. \tag{29}$$

Therefore, for even L, the leading term for $n \to \infty$ is

$$Re\{\lambda_n\} = \frac{g\mathrm{K}\mathrm{T}_0^L(-1)^{L/2}\,(F(0) - F(1))}{(2\pi n)^L G(1)G(0)}. \tag{30}$$

For even L, the stability of the short-wavelength modes (large n) is controlled by the sign of $(F(0) - F(1))$: for even (odd) $L/2$ and excitatory coupling, i.e., $g > 0$, the splay state is stable whenever $F(1) > F(0)$ $(F(1) < F(0))$. Obviously the stability is reversed for inhibitory coupling.

Notice that for $L = 0$, i.e., δ-spikes, the eigenvalues do not decrease with n, as previously observed in Zillmer et al. (2007). This is the only case where all modes exhibit a finite stability even in the thermodynamic limit.

For odd L, the real part of the eigenvalues is

$$Re\{\lambda_n\} = \frac{g\mathrm{K}\mathrm{T}_0^L(-1)^{(L+1)/2}}{(2\pi n)^{(L+1)}} \times \left\{\frac{F'(1)}{G(1)^2} - \frac{F'(0)}{G(0)^2} - ST_0\frac{F(1) - F(0)}{G(1)G(0)}\right\}, \tag{31}$$

in this case the value of $F(X)$ and of its derivative $F'(X)$ at the extrema mix up in a non-trivial way.

Finally, as for the scaling behavior of the leading terms we observe that

$$Re\{\lambda_n\} \sim n^{-q}, \quad q = 2\left\lfloor\frac{L+1}{2}\right\rfloor \tag{32}$$

where $\lfloor\cdot\rfloor$ stays for the integer part of the number. Therefore the scaling of the short-wavelength modes for discontinuous $F(X)$ is dictated by the post-synaptic pulse profile.

For a continuous but non-differentiable $F(X)$, (i.e., $F'(1) \neq F'(0)$), if L is even, it is necessary to go two orders beyond in the estimate of the Fourier coefficients (see Appendix A). As a result, the eigenvalues scale as

$$Re\{\lambda_n\} \propto n^{-(L+2)}. \tag{33}$$

For odd L, it is instead sufficient to assume $F(0) = F(1)$ in Equation (31).

Altogether, we have seen that the non-smoothness of both the post-synaptic pulse and of the velocity field (or, equivalently,

of the phase response curve) play a crucial role in determining the degree of stability of the splay state. The smoother are such functions and the slower short-wavelength perturbations decay, although the changes occur in steps which depend on the parity of the order of the discontinuity (at least for the pulse structure). Moreover, the overall stability of the spectral components depends in a complicate way on the sign of the discontinuity itself.

4. THEORY (FINITE N)

4.1. THE SPLAY STATE

The stability for finite N can be investigated by linearizing Equations (5–7). A thorough analysis has been developed in Olmi et al. (2012); here we limit ourselves to review the key ideas as a guide for the numerical analysis.

We start by introducing the vector $W = (\{x^j\}, \epsilon, p)$ $(j = 1, N-1)$, whose components represent the infinitesimal perturbations of the solution $\{X^j\}$, E, P. The Floquet spectrum can be determined by constructing the matrix $\mathbf{A}$ which maps the initial vector $W(0)$ onto $W(\mathcal{T})$,

$$W(\mathcal{T}) = \mathbf{A}W(0) \tag{34}$$

where $\mathcal{T}$ corresponds to the time separation between two consecutive spikes. This is done in two steps, the first of which corresponds to evolving the components of a Cartesian basis according to the equations obtained from the linearization of Equations (1, 4) (in the comoving reference frame),

$$\begin{aligned} &\dot{x}^j = \frac{dF}{dx_{j+1}} x^{j+1} + g\epsilon, \quad j = 2, \ldots, N \quad \dot{x}^N \equiv 0 \\ &\dot{\epsilon} = p - \alpha\epsilon, \quad \dot{p} = -\alpha p. \end{aligned} \tag{35}$$

The second step consists in accounting for the spike emission, which amounts to add the vector

$$U = [\{\dot{X}^j(\mathcal{T})\}, \dot{E}(\mathcal{T}), \dot{P}(\mathcal{T})]\tau \tag{36}$$

where τ is obtained from the linearization of the threshold condition (8),

$$\tau = -\left(\frac{\partial X^1}{\partial E}\epsilon + \frac{\partial X^1}{\partial P}p\right)\frac{1}{\dot{X}^1} \tag{37}$$

The diagonalization of the resulting matrix $\mathbf{A}$, gives $N+1$ Floquet eigenvalues μ_k, which we express as

$$\mu_k = e^{i\phi_k} e^{T_0(\lambda_k + i\omega_k)/N}, \tag{38}$$

where $\phi_k = \frac{2\pi k}{N}$, $k = 1, \ldots, N-1$, and $\phi_N = \phi_{N-1} = 0$, while λ_k and ω_k are the real and imaginary parts of the Floquet exponents. The variable ϕ_k plays the role of the wavenumber k in the linear stability analysis of spatially extended systems.

Previous studies (Olmi et al., 2012) have shown that the spectrum can be decomposed into two components: (1) $k \sim \mathcal{O}(1)$; (2) $k/N \sim \mathcal{O}(1)$. The former one is the LW component and can be directly obtained in the thermodynamic limit (see the previous section). For $L = 2$ and $\alpha_1 = \alpha_2$ (i.e., for α pulses), it has been found that the results reported in Abbott and van Vreeswijk (1993) match does obtained for $1 \ll k \ll N$ in Olmi et al. (2012). The latter one corresponds to the SW component: it depends on the system size and cannot, indeed, be derived from the mean field approach discussed in the previous section. In the next section, we illustrate some examples that go beyond the analytic studies carried out in Olmi et al. (2012).

4.2. THE SYNCHRONIZED STATE

In this section we address the problem of measuring the stability of the fully synchronized state for a generic oscillator dynamics $F(x)$. The task is non-trivial, because of the resetting mechanism, which acts simultaneously on all neurons. On the one side, we extend the results obtained in Goel and Ermentrout (2002) which are restricted to a continuous PRC, on the other side we extend the results of Mirollo and Strogatz (1990) which refer to excitatory coupling and δ pulses. In order to make the analysis easier to understand we start considering α-pulses. Other cases are discussed afterward.

The starting point amounts to writing the event driven map in a comoving frame,

$$X^j_{n+1} = \mathcal{F}\left(X^{j+1}_n, E_n, P_n, \mathcal{T}_n\right) \tag{39}$$

$$E_{n+1} = E_n e^{-\alpha\mathcal{T}_n} + P_n \mathcal{T}_n e^{-\alpha\mathcal{T}_n}, \tag{40}$$

$$P_{n+1} = P_n e^{-\alpha\mathcal{T}_n} + \frac{\alpha^2}{N}, \tag{41}$$

where the function $\mathcal{F}$ is obtained by formally integrating the equations of motion over the time interval $\mathcal{T}_n$. Notice that the field dynamics has been, instead, explicitly obtained from the exact integration of the equations of motion [compare with Equations (3, 4)]. The interspike time interval $\mathcal{T}_n$ is finally determined by solving the implicit equation

$$\mathcal{F}(X^1_n, E_n, P_n, \mathcal{T}_n) = 1. \tag{42}$$

In order to determine the stability of the synchronized state, it is necessary to assume that the neurons have an infinitesimally different membrane potentials, even though they coincide with one another. As a result, the full period must be broken into N steps. In the first one, of length T, all neurons start in $X = 0$ and arrive at 1, but only the "first" reaches the threshold; in the following $N-1$ steps, of 0-length, one neuron after the other passes the threshold and it is accordingly reset in 0.

With this scheme in mind we proceed to linearize the equations, writing the evolution equations for the infinitesimal perturbations x^j_n, ϵ_n, p_n, and τ_n around the synchronous solution. From Equations (39–41) we obtain,

$$\begin{aligned} x^j_{n+1} = {} & \mathcal{F}_X(j+1)x^{j+1}_n + \mathcal{F}_E(j+1)\epsilon_n + \\ & \mathcal{F}_P(j+1)p_n + \mathcal{F}_{\mathcal{T}}(j+1)\tau_n \quad 1 \le j < N \end{aligned} \tag{43}$$

$$\begin{aligned} \epsilon_{n+1} = {} & e^{-\alpha\mathcal{T}}\epsilon_n + \mathcal{T}e^{-\alpha\mathcal{T}}p_n - \\ & \left(\alpha\tilde{E} - P_n e^{-\alpha\mathcal{T}}\right)\tau_n \end{aligned} \tag{44}$$

$$p_{n+1} = e^{-\alpha\mathcal{T}}p_n - \alpha P_n e^{-\alpha\mathcal{T}}\tau_n. \tag{45}$$

with the boundary condition $x^N_{n+1} = 0$ (due to the reset mechanism) and where the subscripts X, E, P, and $\mathcal{T}$ denote a partial derivative with respect to the given variable. Moreover, the dependence on $j+1$ is a shorthand notation to remind that the various derivatives depend on the membrane potential of the $(j+1)$st neuron. Finally, we have left the n-dependence in the variable P as it changes (in α^2/N steps, when the neurons progressively cross the threshold), while $\tilde{E}$ refers to the field amplitude, which, instead, stays constant.

The above equations must be complemented by the condition

$$\tau_n = -\mathcal{T}_X x^1_n + \mathcal{T}_E \epsilon_n + \mathcal{T}_P p_n, \tag{46}$$

where $\mathcal{T}_Z = \mathcal{F}_Z(1)/\mathcal{F}_\mathcal{T}(1)$ $(Z = X,\ E,\ P)$. Equation (46) is obtained by differentiating Equation (42) which defines the period of the splay state.

We now proceed to build the Jacobian for each of the N steps, starting from the first one. In order not to overload the notations, from now on, the time index n corresponds to the step of the procedure. It is convenient to order all the variables, starting from x^j $(j = 1, N-1)$, and then including ϵ and p, into a single vector, so that the evolution is described by an $(N+1) \times (N+1)$ matrix with the following structure,

$$\mathcal{N}(n) = \begin{pmatrix} \Gamma(n) & \mathbf{0} \\ \Psi(n) & \Omega(n) \end{pmatrix}, \tag{47}$$

where $\mathbf{0}$ is an $(N-1) \times 2$ null matrix; $\Gamma(n)$ is a quadratic $(N-1) \times (N-1)$ matrix, whose only non-zero elements are those in the first column and along the supradiagonal; $\Psi(n)$ is a $2 \times (N-1)$ matrix whose elements are all zero except for the first column; finally $\Omega(n)$ is a 2×2 matrix.

Since in the first step all neurons start from the same position $X = 0$, one can drop the j dependence in $\mathcal{F}$. With the help of Equations (46, 43)

$$\begin{aligned} \Gamma(1)_{j,1} &= -\mathcal{F}_X \\ \Gamma(1)_{j,j+1} &= \mathcal{F}_X \end{aligned} \tag{48}$$

Moreover, with the help of Equations (44–46)

$$\begin{aligned} \Psi(1)_{11} &= -\left(\alpha\tilde{E} - \tilde{P}e^{-\alpha T}\right)\mathcal{T}_X \\ \Psi(1)_{12} &= -\alpha P e^{-\alpha T}\mathcal{T}_X, \end{aligned} \tag{49}$$

where we have also made use that $P_1 = \tilde{P}$. Finally,

$$\begin{aligned} \Omega(1)_{11} &= e^{-\alpha T} - \left(\alpha\tilde{E} - \tilde{P}e^{-\alpha T}\right)\mathcal{T}_E, \\ \Omega(1)_{12} &= T e^{-\alpha T} - \left(\alpha\tilde{E} - \tilde{P}e^{-\alpha T}\right)\mathcal{T}_P, \\ \Omega(1)_{21} &= -\alpha\tilde{P}e^{-\alpha T}\mathcal{T}_E, \\ \Omega(1)_{22} &= e^{-\alpha T} - \alpha\tilde{P}e^{-\alpha T}\mathcal{T}_P, \end{aligned} \tag{50}$$

In the next steps, $\mathcal{T}_n$ vanishes, so that $\mathcal{F}_E = \mathcal{F}_P = 0$, while $\mathcal{F}_X = 1$ and $\mathcal{F}_\mathcal{T}(1) = F(1) + g\tilde{E} := V^1$. Moreover, $\mathcal{F}_\mathcal{T}(j)$ depends on whether the jth neuron has passed the threshold or not. In the former case $\mathcal{F}_\mathcal{T}(j+1) = F(0) + g\tilde{E} := V_0$, otherwise $\mathcal{F}_\mathcal{T}(j+1) = V^1$. As a result,

$$\begin{aligned} \Gamma(n)_{j,1} &= -V^j/V^1 \\ \Gamma(n)_{j,j+1} &= 1 \end{aligned} \tag{51}$$

where $V^j = V^0$ if $j < n$ and $V^j = V^1$, otherwise. At the same time, from the equations for the field variables, we find that

$$\begin{aligned} \Psi(n)_{11} &= \frac{\alpha\tilde{E} - (\tilde{P} + (n-1)\frac{\alpha^2}{N})}{V^1} \\ \Psi(n)_{12} &= \frac{\alpha(\tilde{P} + (n-1)\frac{\alpha^2}{N})}{V^1}, \end{aligned} \tag{52}$$

while $\Omega(n)$ reduces to the identity matrix.

From the multiplication of all matrices, we find that the structure is preserved, namely

$$\mathcal{N}(N)\cdots\mathcal{N}(2)\mathcal{N}(1) = \begin{pmatrix} \Lambda & \mathbf{0} \\ \bar{\Psi} & \Omega(1) \end{pmatrix}, \tag{53}$$

where $\bar{\Psi}(n)$ is a $2 \times (N-1)$ matrix, whose elements are all zero except for those of the first column, namely

$$\begin{aligned} \bar{\Psi}_{11} &= \Psi(1)_{11} + \Psi(n)_{11} \\ \bar{\Psi}_{12} &= \Psi(1)_{12} + \Psi(n)_{12} \end{aligned}$$

Furthermore, Λ is a diagonal matrix, with

$$\Lambda_{jj} = \mathcal{F}_X \frac{V^0}{V^1} = \frac{F(0) + g\tilde{E}}{F(1) + g\tilde{E}} \exp\left[\int_0^T dt F'(X(t))\right] \tag{54}$$

Therefore, it is evident that the stability of the orbit is measured by the diagonal elements Λ_{jj} together with the eigenvalues of Ω which are associated to the pulse structure. In practice, $\mathcal{F}_X$ corresponds to the expansion rate from $X = 0$ to $X = 1$ under the action of the mean field E and we recover a standard result in globally coupled identical oscillators: the spectrum is degenerate, all eigenvalues being equal and independent of the network size. The result is, however, not obvious in this context, due to the care that is needed in taking into account the various discontinuities. We have separately verified that the same conclusion holds for exponential spikes.

The stability of the synchronized state can be also addressed by determining the evaporation exponent Λ_e (van Vreeswijk, 1996; Pikovsky et al., 2001), which measures the stability of a probe neuron subject to the mean field generated by the synchronous neurons with no feedback toward them. By implementing this approach for a negative perturbation, van Vreeswijk found that Λ_e is equal to Λ_{jj} (for α-functions). By further assuming that $F' < 0$, he was able to prove that the synchronized state is stable for inhibitory coupling and sufficiently small α-values. The situation is more delicate for exponential pulse-shapes. As shown in di Volo et al. (2013), $\Lambda_e > 0$ $(\Lambda_e < 0)$ depending whether the

perturbation is positive (negative). In this case, the Floquet exponent reported in Equation (54) coincides with the evaporation exponent estimated for negative perturbations. In Appendix B. we show that the difference between the left and right stability is to be attributed to the discontinuous shape of the pulse: no anomaly is expected for α pulses.

5. NUMERICAL ANALYSIS

The theoretical approaches discussed in the previous sections allow determining: (1) the SW components of the Floquet spectrum for discontinuous velocity fields; (2) the leading LW exponents directly in the thermodynamic limit for generic velocity fields and pulse shapes, in the weak coupling limit. It would be possible to extend the finite N results to other setups, but we do not think that the effort is worth, given the huge amount of technicalities. We thus prefer to illustrate the expected behavior with the help of some simulations which, incidentally, cover a wider range than possibly accessible to the analytics.

More precisely, in this and the following section we study the models listed in **Table 1** in a standard set up (splay states) and under the effect of periodic external perturbations.

5.1. FINITE PULSE WIDTH

Here, we discuss the stability of the splay state for different degrees of smoothness of the velocity field at the borders of the unit interval for post-synaptic pulses of α-function type.

We start from discontinuous velocity fields. They have been the subject of an analytic study which proved that the SW component scales as $1/N^2$ (Olmi et al., 2012). The data reported in **Figure 1A** for $F_1(X)$ confirms the expected scaling: the agreement with the theoretical curve derived in Olmi et al. (2012) is impressive over the entire spectral range, while the mean field Equation (30) gives a very good estimation of the spectrum except for the shortest wavelengths, where it overestimates the numerical data. The mean field approximation turns out to be more accurate for continuous velocity fields (with a discontinuity of the first derivative at the borders of the definition interval). Indeed the agreement between the theoretical expression Equation (A10) and the numerical data is very good for the entire range [see **Figure 1B** which refers to $F_4(X)$].

The numerical Floquet spectra for fields that are $\mathcal{C}^{(0)}$, but not $\mathcal{C}^{(1)}$ ($F(0) = F(1)$, $F'(0) \neq F'(1)$), are reported in **Figure 2** [the curves in panels (**A**, **B**) refer to $F_2(X)$ and F_4, respectively]. For these velocity fields, we have also verified that the spectra scale as $1/N^4$, confirming the observation reported in Calamai et al. (2009) for a different velocity field with the same analytical properties. The data displayed in **Figures 2A,B** refer to the LW components: they indeed confirm to be independent of the system size and scale as $1/k^4$ (see the dashed line) as predicted by the perturbative theory discussed in section 3.

The spectra reported in the other two panels refer to analytic velocity fields: in all cases the initial part of the Floquet spectra is again independent of N and scales approximately exponentially with k, confirming that the scaling behavior of the exponents is related to the analyticity of the velocity field. The fluctuating background with approximate height 10^{-12} is just a consequence of the finite numerical accuracy. This is the reason why we did not dare to estimate the SW components that would be exceedingly small.

5.2. VANISHING PULSE-WIDTH

Here, we analyze the intermediate case between finite pulse-width and δ-like impulses. Similarly to what done in Zillmer et al. (2007) for the LIF, we consider α pulses, where $\alpha = \beta N$, with β independent of N.

In **Figure 3A** we report the spectra for a discontinuous velocity field, $F_1(x)$. In this case the Floquet spectra remain finite, so that the corresponding states remain robustly stable even in the thermodynamic limit. Also in this case the agreement with the theoretical expression reported in Equation (7) in Olmi et al. (2012) is extremely good, while Equation (30) overestimates the spectra for large phases. The field considered in panel (b) ($F_2(X)$) is $\mathcal{C}^{(0)}$ but not $\mathcal{C}^{(1)}$. In this case, the Floquet spectra scale as $1/N$: this scaling is predicted by the analysis reported in section 3 and the whole spectrum is very well reproduced by Equation (A10).

Table 1 | In the first column is reported the list of the velocity fields $F(X)$ analyzed in the paper. All the considered fields are everywhere positive within the definition interval $X \in [0,1]$, thus ensuring that the neuron is supra-threshold. The second column refers to the continuity properties of the fields within the interval [0,1].

Velocity field	Continuity properties
$F_0(X) = a - X$	Discontinuous
$F_1(X) = a - X(X - 0.7)$	Discontinuous
$F_2(X) = a - 0.25\sin(\pi X)$	$\mathcal{C}^{(0)}$
$F_3(X) = a + X(X - 1)$	$\mathcal{C}^{(0)}$
$F_4(X) = a - 0.25\sin(\pi X)\cos^2(\pi X)$	$\mathcal{C}^{(0)}$
$F_5(X) = a - 0.25\sin(2\pi X)\cos^2(2\pi X)$	$\mathcal{C}^{(\infty)}$
$F_6(X) = a - 0.25\sin(2\pi X)e^{\cos(2\pi X)}$	$\mathcal{C}^{(\infty)}$
$F_7(X) = a - 1 + e^{2\sin(2\pi X)}$	$\mathcal{C}^{(\infty)}$

The function is labeled as discontinuous if $F(0) \neq F(1)$; it is $\mathcal{C}^{(0)}$ if $F(0) = F(1)$ but $F'(0) \neq F'(1)$ and $\mathcal{C}^{(1)}$ if $F(0) = F(1)$ and $F'(0) = F'(1)$. $F(X)$ is $\mathcal{C}^{(\infty)}$ if it is infinitely differentiable and each derivative is continuous at the extrema of the definition interval.

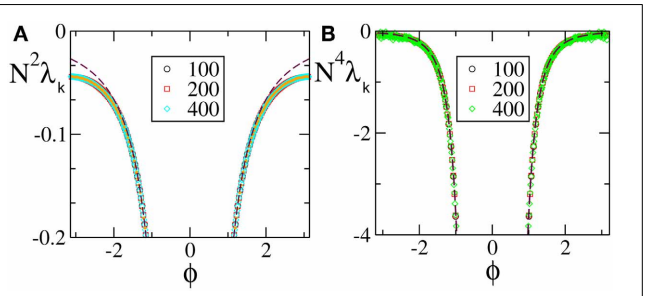


FIGURE 1 | Floquet spectra for α-pulses for (A) a discontinuous field $F_1(X)$ and (B) a continuous field $F_4(X)$. The orange dotted line in **(A)** represents the theoretical curve estimated by using Equation (7) in Olmi et al. (2012), while the dashed maroon curve represents the theoretical curve estimated by using Equation (30) in section 3. In **(B)** the dashed maroon curve is calculated by using Equation (A10). All data refer to $a = 1.3$ and $\alpha = 3$.

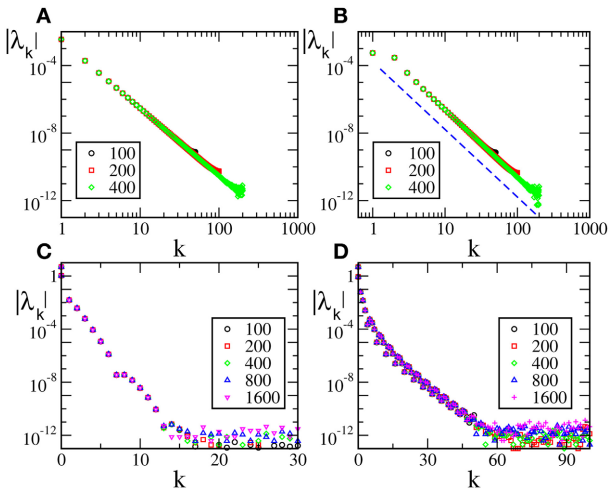


FIGURE 2 | Floquet spectra for α-pulses for two continuous sinusoidal fields, namely $F_2(X)$ (A) and $F_4(X)$ (B); and two analytic fields, namely $F_6(X)$ (C) and $F_7(X)$ (D). The dashed blue line in **(B)** indicates a scaling $1/k^4$. All data refer to $a = 1.3$ and $\alpha = 3$.

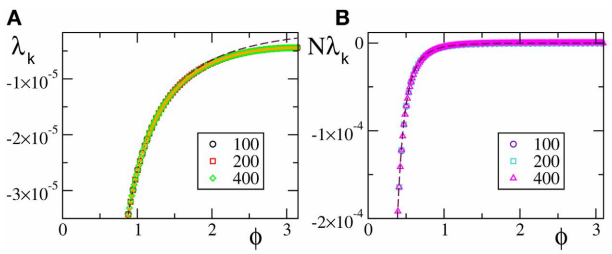


FIGURE 3 | Floquet spectra for β-pulses with a discontinuous field [$F_1(X)$] (A) and a $C^{(0)}$ field [$F_2(X)$] (B). The orange dotted line in **(A)** represents the theoretical curve estimated by using Equation (7) in Olmi et al. (2012). The dashed line in **(A)** [resp. **(B)**] represents the theoretical curve computed by using Equation (30) [resp. Equation (A10)] for β-pulses. The data refer to $a = 1.3$ and $\beta = 0.03$.

Last but not least, we have studied an analytic field, namely $F_7(X)$. In this case the Floquet spectra appear to scale exponentially to zero with the wavevector k, similarly to what observed for the finite pulse width, as shown in **Figure 4**.

5.3. δ PULSES

Finally we considered the case of δ-pulses: whenever the potential X^j reaches the threshold value, it is reset to zero and a spike is sent to and *instantaneously* received by all neurons. We studied just two cases: (1) the analytic field $F_7(X)$; (2) a leaky integrate-and fire neuron model with $F_0(X)$. The results, obtained for inhibitory coupling [since the splay state is known to be stable only in such a case (van Vreeswijk, 1996; Zillmer et al., 2006)] are consistent with the expectation for the β model.

In particular we found, in the analytic case (1), that the Floquet spectra decay exponentially to zero. The exponential scaling is not altered if a phase shift ζ is introduced in the velocity field (i.e., for $F(X) = a - 1 + e^{2\sin(2\pi X+\zeta)}$). In the case of the LIF model (F_0),

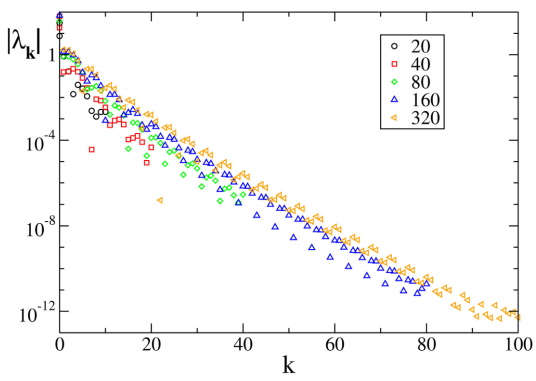


FIGURE 4 | Floquet spectra for β-pulses for the analytic field $F_7(X)$. The data refer to $a = 1.3$ and $\beta = 0.03$.

we already know that the Lyapunov spectrum tends, in the δ-pulse limit, to Zillmer et al. (2007)

$$\lim_{\beta\to\infty} \lambda_\pi = -1 + \frac{1}{T_0} \ln\left(\frac{a}{a-1}\right). \tag{55}$$

This result is confirmed by our simulations which also reveal that the splay state is stable even for small, excitatory coupling values, extending previous results limited to inhibitory coupling (Zillmer et al., 2006).

6. PERIODIC FORCING

In this section we numerically investigate the scaling behavior of the Floquet spectrum in the presence of a periodic forcing, to test the validity of the previous analysis in a more general context. We have restricted our studies to splay-state-like regimes, where it is important to predict the behavior of the many almost marginally stable directions. Moreover, we have considered only the smooth α-pulses. In this case, the dynamical equations read

$$\begin{aligned} \dot{X}^j &= F(X^j) + gE + A\cos(\varphi), \qquad j = 1, \ldots, N, \\ \dot{E} &= P - \alpha E, \\ \dot{P} &= -\alpha P, \\ \dot{\varphi} &= \omega. \end{aligned} \tag{56}$$

They have been written in an autonomous form, since it is more convenient to perform the Poincaré section according to the spiking times, rather than introducing a stroboscopic map. The interspike interval is determined by the equation

$$\mathcal{T} = \int_{X_{\text{old}}}^{1} \frac{dX^1}{F(X^1) + gE + A\cos(\varphi)}. \tag{57}$$

where X^1 is the membrane potential of the first neuron (the closest to threshold), and X_{old} is its initial value.

We analyzed only those setups where the unperturbed splay state is stable. More precisely: the two discontinuous fields $F_0(X)$ and $F_1(X)$, the two $C^{(0)}$ fields ($F_2(X)$ and $F_3(X)$), and the analytic

field $F_7(X)$. In all cases the external modulation induces a periodic modulation of the mean field E with a period $T_a = 2\pi/\omega$ equal to the period of the modulation. At the same time, we have verified that, although the forcing term has zero average (i.e., it does not change the average input current), the average interspike interval is slightly self-adjusted and, what is more important, there is no evidence of locking between the modulation and the frequency of the single neurons. In other words, the behavior is similar to the spontaneous partial synchronization observed in van Vreeswijk (1996) (where the modulation is self-generated).

Because of the unavoidable oscillations of the interspike intervals, it is necessary to identify the spike times with great care. In practice we integrate Equation (56) with a fixed time step Δt, by employing a standard fourth-order Runge–Kutta integration scheme. At each time step we check if $X^1 > 1$, in which case we go one step back and adopt the Hénon trick, which amounts to exchanging t and X^1 in the role of independent variable (Henon, 1982).

The linear stability analysis can be performed by linearizing the system (56), to obtain

$$\begin{aligned}
\dot{x}^j &= \frac{dF(X^j)}{dX^j}x^j + g\epsilon - A\sin(\varphi)\delta\varphi, \qquad j = 1, \ldots, N,\\
\dot{\epsilon} &= p - \alpha\epsilon,\\
\dot{p} &= -\alpha p,\\
\delta\dot{\varphi} &= 0;
\end{aligned}$$

and by thereby estimating the corresponding Lyapunov spectrum.

In the case of F_0 and F_1, we have always found that the Lyapunov spectrum scales as $1/N^2$ as theoretically predicted in the absence of external modulation (see **Figure 5** for one instance of each of the two velocity fields).

A similar agreement is also found for F_3, where the Lyapunov spectrum scales as $1/N^4$, exactly as in the absence of external forcing (see **Figure 6**). Analogous results have been obtained for the other velocity fields (data not shown), which confirm that the validity of the previous analysis extends to more complex dynamical regimes, as long as the membrane potentials are smoothly distributed.

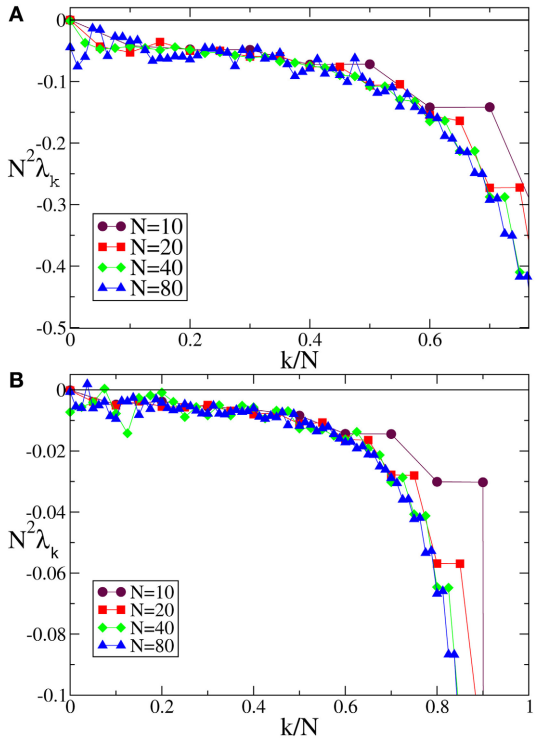


FIGURE 5 | Lyapunov spectra for neurons forced by an external periodic signal, we observe the scaling $1/N^2$ for the discontinuous velocity fields (A) $F_0(X)$ and (B) $F_1(X)$. In both cases $A = 0.1$, $T_a = 2$.

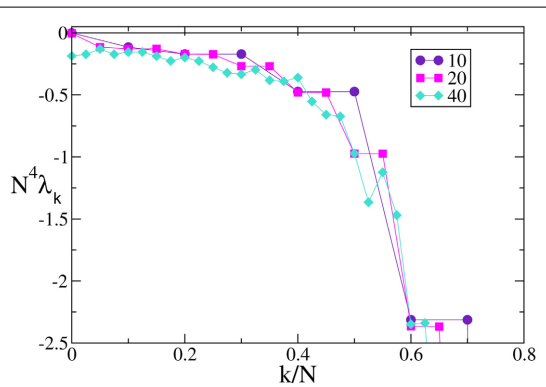


FIGURE 6 | Lyapunov spectra for neurons forced by an external periodic signal, we observe the scaling $1/N^4$ for the continuous velocity field $F_3(X)$. The data refer to $A = 0.1$, $T_a = 2$.

7. SUMMARY AND OPEN PROBLEMS

In this paper we have discussed the linear stability of both fully synchronized and splay states in pulse-coupled networks of identical oscillators. By following Abbott and van Vreeswijk (1993), we have obtained analytic expressions for the long-wavelength components of the Floquet spectra of the splay state for generic velocity fields and post synaptic potential profiles. The structure of the spectra depends on the smoothness of both the velocity field and the transmitted pulses. The smoother they are and the faster the eigenvalues decrease with the wavelength of the corresponding eigenvectors. In practice, while splay states arising in LIF neurons with δ-pulses have a finite degree of (in)stability along all directions, those emerging in analytic velocity fields have many exponentially small eigenvalues. These results have been derived in the mean field framework, where the system is assumed to be infinite. Although realistic neural networks are finite, the present analysis predicts correctly, even for finite systems, the stability of the eigenmodes associated to the fastest scales and the order of magnitude of the eigenvalues corresponding to slower time scales. Interestingly, the scaling behavior of the eigenvalues carries over to that of the Lyapunov exponents, when the network is periodically forced, suggesting that our results have a relevance that goes beyond the highly symmetric solutions studied in this paper.

Finally, we derived an analytic expression for the Floquet spectra for the fully synchronous state. In this case the exponents associated to the dynamics of the membrane potentials are all

identical, as it happens for the diffusive coupling, but here the result is less trivial, due to the fact that one must take into account that arbitrarily close to the solution, the ordering of the neurons may be different. Moreover, the value of the (degenerate) Floquet exponent coincides with the evaporation exponent (van Vreeswijk, 1996; Pikovsky et al., 2001) whenever the pulses are sufficiently smooth, while for discontinuous pulses (like exponential and δ-spikes) the equivalence is lost (see also di Volo et al., 2013).

For discontinuous velocity fields, another important property that has been confirmed by our analysis is the role of the ratio $R = N/(T_0\alpha)$ between the width of the single pulse ($1/\alpha$) and the average interspike interval of the whole network ($\mathcal{T} = T_0/N$). In fact, it turns out that the asynchronous regimes can be strongly stable along all directions only when R remains finite in the thermodynamic limit (and is possibly small). This includes the idealized case of δ-like pulses, but also setups where the single pulses are so short that they can be resolved by the single neurons. Mathematically speaking, this result implies that the thermodynamic limit does not commute with the limit of a zero pulse-width. It would be interesting to check to what extent this property extends to more realistic models. A first confirmation result is contained in Pazó and Montbrió (2013), where the authors find a similar property in a network of Winfree oscillators.

Among possible extensions of our analysis, one should definitely mention the inclusion of delay in the pulse transmission. This generalization is far from trivial as it modifies the phase diagram of the possible states (see Bär et al., 2012 for a recent brief overview of the possible scenarios) and it complicates noticeably the stability analysis of the synchronized phase. An analytic treatment of this latter case is reported in Timme et al. (2002) for generic velocity fields and excitatory δ-pulses.

ACKNOWLEDGMENTS

We thank David Angulo Garcia for the help in the use of symbolic algebra software. Alessandro Torcini acknowledges financial support from the European Commission through the Marie Curie Initial Training Network "NETT," project N. 289146, as well as from the Italian Ministry of Foreign Affairs for the activity of the Joint Italian-Israeli Laboratory on Neuroscience. Simona Olmi and Alessandro Torcini thanks the Italian MIUR project CRISIS LAB PNR 2011–2013 for economic support and the German Collaborative Research Center SFB 910 of the Deutsche Forschungsgemeinschaft for the kind hospitality at Physikalisch-Technische Bundesanstalt in Berlin during the final write up of this manuscript.

REFERENCES

Abbott, L. F., and van Vreeswijk, C. (1993). Asynchronous states in networks of pulse-coupled oscillators. *Phys. Rev. E* 48, 1483. doi: 10.1103/PhysRevE.48.1483

Acebrón, J. A., Bonilla, L. L., Pérez Vicente, C. J., Ritort, F., and Spigler, R. (2005). The Kuramoto model: a simple paradigm for synchronization phenomena. *Rev. Mod. Phys.* 77, 137–185. doi: 10.1103/RevModPhys.77.137

Aronson, D. G., Golubitsky, M., and Krupa, M. (1991). Coupled arrays of Josephson junctions and bifurcation of maps with S_N symmetry. *Nonlinearity* 4, 861–902. doi: 10.1088/0951-7715/4/3/013

Ashwin, P., King, G. P., and Swift, J. W. (1990). Three identical oscillators with symmetric coupling. *Nonlinearity* 3, 585–601. doi: 10.1088/0951-7715/3/3/003

Bär, M., Schöll, E., and Torcini, A. (2012). Synchronization and complex dynamics of oscillators with delayed pulse-coupling. *Angew. Chem. Int. Ed.* 51, 9489–9490. doi: 10.1002/anie.201205214

Calamai, M., Politi, A., and Torcini, A. (2009). Stability of splay states in globally coupled rotators. *Phys. Rev. E* 80:036209. doi: 10.1103/PhysRevE.80.036209

di Volo, M., Livi, R., Luccioli, S., Politi, A., and Torcini, A. (2013). Synchronous dynamics in the presence of short-term plasticity. *Phys. Rev. E.* 87:032801. doi: 10.1103/PhysRevE.87.032801

Filatrella, G., Nielsen, A. H., and Pedersen, N. F. (2008). Analysis of a power grid using a Kuramoto-like model. *Eur. Phys. J. B* 61, 485–491. doi: 10.1140/epjb/e2008-00098-8

Goel, P., and Ermentrout, B. (2002). Synchrony, stability and firing patterns in pulse-coupled oscillators. *Physica D* 163, 191–216. doi: 10.1016/S0167-2789(01)00374-8

Golomb, D., and Rinzel, J. (1994). Clustering in globally coupled neurons. *Physica D* 72, 259–282. doi: 10.1016/0167-2789(94)90214-3

Hadley, P., and Beasley, M. R. (1987). Dynamical states and stability of linear arrays of Josephson junctions. *App. Phys. Lett.* 50, 621–623 . doi: 10.1063/1.98100

Hansel, D., Mato, G., and Meunier, C. (1995). Synchrony in excitatory neural networks. *Neural Comput.* 7, 307. doi: 10.1162/neco.1995.7.2.307

Henon, M. (1982). On the numerical computation of Poincaré maps. *Physica D* 5, 412–414. doi: 10.1016/0167-2789(82)90034-3

Javaloyes, J., Perrin, M., and Politi, A. (2008). Collective atomic recoil laser as a synchronization transition. *Phys. Rev. E* 78:011108. doi: 10.1103/PhysRevE.78.011108

Jin, D. Z. (2002). Fast convergence of spike sequences to periodic patterns in recurrent networks. *Phys. Rev. Lett.* 89, 208102. doi: 10.1103/PhysRevLett.89.208102

Kuramoto, Y. (1991). Collective synchronization of pulse-coupled oscillators and excitable units. *Physica D* 50, 15–30. doi: 10.1016/0167-2789(91)90075-K

Luccioli, S., Olmi, S., Politi, A., and Torcini, A. (2012). Collective dynamics in sparse networks. *Phys. Rev. Lett.* 109:138103. doi: 10.1103/PhysRevLett.109.138103

Monteforte, M., and Wolf, F. (2010). Dynamical entropy production in spiking neuron networks in the balanced state. *Phys. Rev. Lett.* 105:268104. doi: 10.1103/PhysRevLett.105.268104

Mirollo, R. E., and Strogatz, S. H. (1990). Synchronization of pulse-coupled biological oscillators. *SIAM Journal on Applied Mathematics* 50, 1645–1662. doi: 10.1137/0150098

Nichols, S., and Wiesenfield, K. (1992). Ubiquitous neutral stability of splay-phase states. *Phys. Rev. A* 45, 8430–8345. doi: 10.1103/PhysRevA.45.8430

Olmi, S., Livi, R., Politi, A., and Torcini, A. (2010). Collective oscillations in disordered neural network. *Phys. Rev. E* 81:046119. doi: 10.1103/PhysRevE.81.046119

Olmi, S., Politi, A., and Torcini, A. (2012). Stability of the splay state in networks of pulse-coupled neurons. *J. Math. Neurosci.* 2, 12. doi: 10.1186/2190-8567-2-12

Pazó, and D. Montbrió, E. (2013). Low-dimensional dynamics of populations of pulse-coupled oscillators. arXiv:1305.4044 [nlin.AO].

Pikovsky, A., Popovych, O., and Maistrenko, Y. (2001). Resolving clusters in chaotic ensembles of globally coupled identical oscillators. *Phys. Rev. Lett.* 87:044102. doi: 10.1103/PhysRevLett.87.044102

Pikovsky, A., Rosenblum, M., and Kurths, J. (2003). *Synchronization: A Universal Concept in Nonlinear Sciences*. Cambridge: Cambridge University Press.

Popovych, O. V., Maistrenko, Y. L., and Tass, P. A. (2005). Phase chaos in coupled oscillators. *Phys. Rev. E.* 71:065201(R). doi: 10.1103/PhysRevE.71.065201

Renart, A., de la Rocha, J., Bartho, P., Hollender, L., Parga, N., Reyes, A., et al. (2010). The asynchronous state in cortical circuits. *Science* 327, 587–590. doi: 10.1126/science.1179850

Timme, M. (2006). Does dynamics reflect topology in directed networks? *Europhys. Lett.* 76, 367. doi: 10.1209/epl/i2006-10289-y

Timme, M., and Wolf, F. (2008). The simplest problem in the collective dynamics of neural networks: is synchrony stable? *Nonlinearity* 21, 1579. doi: 10.1088/0951-7715/21/7/011

Timme, M., Wolf, F., and Geisel, T. (2002). Coexistence of regular and irregular dynamics in complex networks of pulse-coupled oscillators. *Phys. Rev. Lett.* 89, 258701. doi: 10.1103/PhysRevLett.89.258701

Treves, A. (1993). Mean-field analysis of neuronal spike dynamics. *Network Comput. Neural Syst.* 4, 259–284. doi: 10.1088/0954-898X/4/3/002

van Vreeswijk, C. (1996). Partial synchronization in populations of pulse-coupled oscillators. *Phys. Rev. E* 54, 5522. doi: 10.1103/PhysRevE.54.5522

van Vreeswijk, C., Abbott, L. F., and Ermentrout, G. B. (1994). When inhibition not excitation synchronizes neural firing. *J. Comput. Neurosci.* 1, 313. doi: 10.1007/BF00961879

Watanabe, S., and Strogatz, S. H. (1994). Constants of motion for superconducting Josephson arrays. *Physica D* 74, 197–253. doi: 10.1016/0167-2789(94)90196-1

Zillmer, R., Livi, R., Politi, A., and Torcini, A. (2006). Desynchronization in diluted neural networks. *Phys. Rev. E* 75:036203. doi: 10.1103/PhysRevE.74.036203

Zillmer, R., Livi, R., Politi, A., and Torcini, A. (2007). Stability of the splay state in pulse–coupled networks. *Phys. Rev. E* 76:046102. doi: 10.1103/PhysRevE.76.046102

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.

Received: 12 November 2013; paper pending published: 13 December 2013; accepted: 13 January 2014; published online: 04 February 2014.

*Citation: Olmi S, Torcini A and Politi A (2014) Linear stability in networks of pulse-coupled neurons. Front. Comput. Neurosci. **8**:8. doi: 10.3389/fncom.2014.00008*

This article was submitted to the journal Frontiers in Computational Neuroscience.

APPENDICES

A. FOURIER COMPONENTS OF THE PHASE RESPONSE CURVE

In this appendix we briefly outline the way the explicit expression of A_n and B_n, defined in Equation (23), can be derived in the large n limit for a velocity field $F(X)$ that is either discontinuous, or continuous with discontinuous first derivatives at the border of the definition interval.

The integration interval [0, 1] appearing in Equation (23) is splitted in n sub-intervals of length $1/n$, and the original equation can be rewritten as

$$(A_n + iB_n) = \sum_{k=1}^{n} \int_{(k-1)/n}^{k/n} dy \frac{e^{i2\pi ny}}{G(y)}. \tag{A1}$$

For n sufficiently large we can assume that the variation of $1/G(y)$ is quite limited within each sub-interval, and we can approximate the function as follows, up to the second order

$$\frac{1}{G(y)} = \frac{1}{g + T_0F(y_0)} \left\{ 1 - \frac{T_0F'(y_0)}{g + T_0F(y_0)}(y - y_0) + \left[\left(\frac{T_0F'(y_0)}{g + T_0F(y_0)} \right)^2 - \frac{T_0F''(y_0)}{2(g + T_0F(y_0))} \right] (y - y_0)^2 \right\}$$

where $y_0 = (k-1)/n$ is the lower extremum of the nth sub-interval.

By inserting these expansions into Equation (A1) and by performing the integration over the n sub-intervals, we can determine an approximate expression for A_n and B_n. The estimation of A_n involves integrals containing $\cos(2\pi ny)$; it is easy to show that the integral over each sub-interval is zero if the integrand, which multiplies the cosinus term, is constant or linear in y; therefore the only non-zero terms are,

$$\int_{(k-1)/n}^{k/n} dy \cos(2\pi ny) y^2 = \frac{1}{2\pi^2 n^3}. \tag{A2}$$

This allows to rewrite

$$A_n = \frac{1}{2\pi^2 n^2} \sum_{k=1}^{n} H_2\left(\frac{k-1}{n}\right) \frac{1}{n} = \frac{1}{2\pi^2 n^2} \left[\int_0^1 dx H_2(x) \right] + \mathcal{O}\left(\frac{1}{n^3}\right) \tag{A3}$$

where

$$H_2(x) = \left[\frac{(T_0F'(x))^2}{(g + T_0F(x))^3} - \frac{T_0F''(x)}{2(g + T_0F(x)^2)} \right]. \tag{A4}$$

It is easy to verify that $H_2(x)$ admits an exact primitive and therefore to perform the integral appearing in Equation (A3) and to arrive at the expression reported in Equation (28).

The estimation of B_n is more delicate, since now integrals containing $\sin(2\pi ny)$ are involved. The only vanishing integrals over the sub-intervals are those with a constant integrand multiplied by the sinus term and therefore the estimation of B_n reduces to

$$B_n = \sum_{k=1}^{n} H_1\left(\frac{k-1}{n}\right) \int_{(k-1)/n}^{k/n} dy \sin(2\pi ny) y + \sum_{k=1}^{n} H_2\left(\frac{k-1}{n}\right) \int_{(k-1)/n}^{k/n} dy \sin(2\pi ny) \left(y^2 - 2y\frac{k-1}{n} \right)$$

where

$$H_1(x) = -\frac{T_0F'(x)}{(g + T_0F(x))^2}, \tag{A5}$$

and the non-zero integrals are

$$\int_{(k-1)/n}^{k/n} dy \sin(2\pi ny) y = -\frac{1}{2\pi n^2}, \tag{A6}$$

and

$$\int_{(k-1)/n}^{k/n} dy \sin(2\pi ny) y^2 = \frac{1-2k}{2\pi n^3}. \tag{A7}$$

This allows to rewrite B_n as

$$B_n = -\frac{1}{2\pi n} \sum_{k=1}^{n} H_1\left(\frac{k-1}{n}\right) \frac{1}{n} - \frac{1}{2\pi n^2} \sum_{k=1}^{n} H_2\left(\frac{k-1}{n}\right) \frac{1}{n}. \tag{A8}$$

We can then return to a continuous variable by rewriting (A8), up to the $\mathcal{O}(1/n^3)$, as

$$B_n = -\frac{1}{2\pi n} \left[\int_0^1 H_1(x) dx + \frac{H_1(1) - H_1(0)}{2n} \right] - \frac{1}{2\pi n^2} \int_0^1 H_2(x) dx. \tag{A9}$$

The expression Equation (29) is finally obtained by noticing that the primitive of $H_2(x)$ is $H_1(x)/2$, and that

$$\int_0^1 H_1(x) dx = \frac{1}{(g + T_0F(0))} - \frac{1}{(g + T_0F(1))}.$$

For continuous velocity fields, $B_n = 0$ so that, we can derive from Equation (26) an exact expression for the real part of the Floquet spectrum in the case of even L (for odd L the equivalent expression is given by Equation (31))

$$Re\{\lambda_n\} = \frac{g\mathrm{KST}_0^{L+1}(-1)^{L/2}}{(2\pi n)^{(L+2)}} \frac{F'(0) - F'(1)}{G(1)^2}. \tag{A10}$$

A rigorous validation of the above formula would require going one order beyond in the $1/n$ expansion of B_n, a task that is

utterly complicated. In the specific case of the Quadratic Integrate and Fire neuron (or Θ-neuron) $F(X) = a - X(X-1)$, it can be, however, analytically verified that B_n is exactly zero. Moreover, Equation (A10) is in very good agreement with the numerically estimated Floquet spectra for two other continuous velocity fields, namely $F_4(X)$ and $F_2(X)$ as shown in **Figures 1**, **3**, respectively. As a consequence, it is reasonable to conjecture that Equation (29) is correct up to order $\mathcal{O}(1/n^4)$.

B. EVAPORATION EXPONENT FOR THE LIF MODEL

In this appendix we determine the (left and right) evaporation exponent for a synchronous state of a network of LIF neurons. This is done by estimating how the potential of a probe neuron, forced by the mean field generated by the network activity, converges toward the synchronized state. The stability analysis is performed by following the evolution of a perturbed probe neuron. Let us first consider an initial condition, where the synchronized cluster has just reached the threshold ($X_c = 1$), while the probe neuron is lagging behind at a distance δ_i. Such a distance is equivalent to a delay t_d

$$t_d = \frac{\delta_i}{F^+(1)}, \tag{A11}$$

where the subscript "+" means that the velocity field is estimated just after the pulses have been emitted. Over the time t_d, the potential of the cluster increases from the reset value 0 to

$$\delta_c = F^+(0)t_d = \frac{F^+(0)}{F^+(1)}\delta_i. \tag{A12}$$

From now on (in LIF neurons), the distance decreases exponentially, reaching the value

$$\delta_f = \delta_c e^{-T}, \tag{A13}$$

after a period T. As a result,

$$\frac{\delta_f}{\delta_i} = \frac{F^+(0)}{F^+(1)}e^{-T} = \frac{a + gE^+}{a - 1 + gE^+}. \tag{A14}$$

The logarithm of the expansion factor gives the left evaporation exponent

$$\Lambda_e^l = \ln\left(\frac{a + gE^+}{a - 1 + gE^+}\right) - T. \tag{A15}$$

Let us now consider a probe neuron which precedes the synchronized cluster by an amount δ_i. After a time T the distance becomes

$$\delta_c = \delta_i e^{-T} \tag{A16}$$

since no reset event has meanwhile occurred. Such a distance corresponds to a delay

$$t_d = \frac{\delta_c}{F^-(1)}, \tag{A17}$$

where the subscript "−" means that the velocity has now to be estimated just before the pulse emission. By proceeding as before one obtains,

$$\frac{\delta_f}{\delta_i} = \frac{F^-(0)}{F^-(1)}e^{-T}. \tag{A18}$$

so that the right evaporation exponent writes

$$\Lambda_e^r = \ln\left(\frac{a + gE^-}{a - 1 + gE^-}\right) - T. \tag{A19}$$

It is easy to see that the left and right exponents differ if and only if $E^- \neq E^+$, i.e., if the pulses themselves are not continuous: this is, for instance, the case of exponential and δ pulses.

frontiers in
COMPUTATIONAL NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 18 November 2014
doi: 10.3389/fncom.2014.00145

Macroscopic complexity from an autonomous network of networks of theta neurons

Tanushree B. Luke, Ernest Barreto and Paul So*

School of Physics, Astronomy, and Computational Sciences and The Krasnow Institute for Advanced Study, George Mason University, Fairfax, VA, USA

Edited by:
Tobias Alecio Mattei, Ohio State University, USA

Reviewed by:
Jianbo Gao, Wright State University, USA
Yasuhiro Tsubo, Ritsumeikan University, Japan

****Correspondence:***
Paul So, George Mason University, Mail Stop 2A1, Fairfax, VA 22030, USA
e-mail: paso@gmu.edu

We examine the emergence of collective dynamical structures and complexity in a network of interacting populations of neuronal oscillators. Each population consists of a heterogeneous collection of globally-coupled theta neurons, which are a canonical representation of Type-1 neurons. For simplicity, the populations are arranged in a fully autonomous driver-response configuration, and we obtain a full description of the asymptotic macroscopic dynamics of this network. We find that the collective macroscopic behavior of the response population can exhibit equilibrium and limit cycle states, multistability, quasiperiodicity, and chaos, and we obtain detailed bifurcation diagrams that clarify the transitions between these macrostates. Furthermore, we show that despite the complexity that emerges, it is possible to understand the complicated dynamical structure of this system by building on the understanding of the collective behavior of a single population of theta neurons. This work is a first step in the construction of a mathematically-tractable network-of-networks representation of neuronal network dynamics.

Keywords: theta neuron, type-I neuron, hierarchical network, neural field, macroscopic behavior, coherence, synchrony, chaos

1. INTRODUCTION

The brain is a complex hierarchical network of networks (Zhou et al., 2006; Bullmore and Sporns, 2009; Meunier et al., 2010). Neurons are organized into different neuronal assemblies, and these neuronal assemblies interact with each other, forming larger assemblies (Sherrington, 1906; Hebb, 1949; Harris, 2005). But while there is a wealth of knowledge on the microscopic scale regarding the dynamics of individual neurons, the macroscopic behavior of such interacting populations of neurons is not well understood. Indeed, the functional and information-processing activity of the brain, from perception to consciousness, is thought to result from the emergent collective behavior of these assemblies.

In recent years, the mathematical study of networks of this kind, based on globally-coupled populations of simple phase oscillators, has advanced significantly. This is in large part due to new analytical techniques (Ott and Antonsen, 2008, 2009; Marvel et al., 2009; Ott et al., 2011; Pikovsky and Rosenblum, 2011). These techniques enable the derivation of low-dimensional dynamical systems that reveal the collective emergent behavior of the full discrete population (in the limit of an infinite number of interacting elements). In the context of computational neuroscience, these methods were applied to autonomous globally-coupled networks of canonical Type-I neurons (i.e., theta neurons) by Luke et al. (2013), and to non-autonomous theta neuron networks by So et al. (2014). More recently, Laing (2014) extended these results to include space-dependent coupling. A similar approach, based on phase-response curves, was pursued by Pazó and Montbrió (2014).

Of course, such networks lack the intricate connectivity found in real biological networks. Nevertheless, they are ideal building blocks for the construction of a more realistic, yet mathematically tractable, network-of-networks representation of the brain. In the current study, we consider the simplest hierarchical structure as a first step in this process. Using two globally-coupled networks of theta neurons, we arrange for the activity of one population to drive the second population. Thus, the overall network has an autonomous driver-response configuration. We demonstrate that even in this simplest network-of-networks, the collective behavior of the response network can exhibit a full range of complex behavior, from simple collective rhythms to temporally chaotic dynamics. Most importantly, we provide a complete non-linear dynamical analysis of this system, including predictive bifurcation diagrams for the behavior of the response population in terms of the driver's dynamics and the network characteristics.

2. RECAP OF SINGLE POPULATION RESULTS

2.1. THE THETA NEURON

Neurons are typically classified into two types, based on the nature of the onset of spiking as a constant injected current exceeds an effective threshold (Hodgkin, 1948; Ermentrout, 1996; Izhikevich, 2007). Type-I neurons begin to spike at an arbitrarily low rate, whereas Type-II neurons spike at a non-zero rate as soon as the threshold is exceeded. Neurophysiologically, excitatory pyramidal neurons are often of Type-I, and fast-spiking inhibitory interneurons are often of Type-II (Nowak et al., 2003; Tateno et al., 2004). Near the onset of spiking, Type-I neurons can be represented by a canonical phase model that features a

saddle-node bifurcation on an invariant cycle, or SNIC bifurcation (Ermentrout and Kopell, 1986; Ermentrout, 1996). This model has come to be known as the theta neuron, and is given by

$$\dot{\theta} = (1 - \cos\theta) + (1 + \cos\theta)\eta, \tag{1}$$

where θ is a phase variable on the unit circle and η is a bifurcation parameter related to the injected current. For $\eta < 0$, the neuron is attracted to a stable equilibrium which represents the resting state. An unstable equilibrium is also present, representing the threshold. If an external stimulus pushes the neuron's phase across the unstable equilibrium, θ will move around the circle and approach the resting equilibrium from the other side. When θ crosses $\theta = \pi$, the neuron is said to have spiked. Thus, for $\eta < 0$, the neuron is excitable. As the parameter η increases, these equilibria approach each other and merge via the SNIC bifurcation at $\eta = 0$. At this point, the equilibria disappear, leaving a limit cycle. The neuron spikes regularly for $\eta > 0$. In the following, we call η the "excitability parameter."

2.2. A NETWORK OF THETA NEURONS

We formulate a single population of N theta neurons as follows:

$$\dot{\theta}_j = \left(1 - \cos\theta_j\right) + \left(1 + \cos\theta_j\right)\left[\eta_j + I_{syn}\right], \tag{2}$$

where $j = 1, \ldots, N$ is the index for the j-th neuron. The neurons are coupled via a pulse-like synaptic current

$$I_{syn} = \frac{k}{N}\sum_{i=1}^{N} P_n(\theta_i), \tag{3}$$

where $P_n(\theta) = a_n\,(1 - \cos\theta)^n$, $n \in \mathbb{N}$, and a_n is a normalization constant[1] such that

$$\int_0^{2\pi} P_n(\theta)d\theta = 2\pi.$$

The parameter n defines the sharpness of the pulse-like synapse in that $P_n(\theta)$ becomes more and more sharply peaked as n increases. We assume that the synaptic strength k is the same for all neurons.

Note that the connectivity described by Equations (2) and (3) includes self-coupling terms. These have negligible effect on the collective network dynamics (data not shown), which is to be expected since they represent only one out of N inputs to any given neuron. Nevertheless, we note that these self-connections have real-world analogs in "autapses," which have been found in several regions of the brain (e.g., Bacci et al., 2003; Bekkers, 2003).

Neurons in real biological networks exhibit a range of different intrinsic dynamics. We model this by taking the excitability parameter η_j of each neuron to be different, with each η_j being drawn randomly from a distribution $g(\eta)$. In the following analysis, we assume a Lorentzian distribution,

$$g(\eta) = \frac{1}{\pi}\frac{\Delta}{(\eta - \eta_0)^2 + \Delta^2}, \tag{4}$$

where η_0 is the center of the distribution, and Δ, the half-width at half-maximum, describes the degree of heterogeneity in the population.

2.3. REDUCTION AND ASYMPTOTIC STATES OF THE SINGLE POPULATION

The macroscopic behavior of our network can be quantified by the "macroscopic mean field," or order parameter, defined as

$$\tilde{z}(t) = \sum_{j=1}^{N} e^{i\theta_j}, \tag{5}$$

where the tilde indicates that the sum is over a finite population of N oscillators. (Below we will drop the tilde in the case of an infinite network.) The magnitude of the order parameter $|\tilde{z}(t)| \in [0, 1]$ quantifies the degree of synchronization present at time t.

In Luke et al. (2013), we used the Ott-Antonsen method (Ott and Antonsen, 2008, 2009; Ott et al., 2011) to derive a low-dimensional dynamical system whose asymptotic dynamics can be shown to coincide with that of the order parameter of the single-population network defined above (Equations 2–4), in the limit $N \to \infty$. This reduced dynamical system is

$$\dot{z} = -i\frac{(z-1)^2}{2} + \frac{(z+1)^2}{2}\left\{-\Delta + i\left[\eta_0 + kH_n(z)\right]\right\}, \tag{6}$$

where

$$H_n(z) = I_{syn}/k = a_n\left(A_0 + \sum_{q=1}^{n} A_q(z^q + z^{*q})\right), \tag{7}$$

$$A_q = \sum_{j,m=0}^{n} \delta_{j-2m,q} Q_{jm}, \tag{8}$$

and

$$Q_{jm} = \frac{(-1)^{j-2m} n!}{2^j m!(n-j)!(j-m)!}. \tag{9}$$

In these equations, z^* denotes the complex conjugate of z, and $\delta_{i,j}$ is the Kronecker delta function on the indices (i, j). Note that $H_n(z) = H_n^*(z)$ is a real-valued function.

The analysis of Equations (6–9) reported in Luke et al. (2013) showed that the theta neuron network can exhibit three types of asymptotic states. These correspond to a node, a focus, and a limit cycle in the order parameter. A complete bifurcation analysis describing how these states change as the parameters k, η_0, and Δ change was also reported. For our purposes in the current work, we now briefly describe the three possible collective macroscopic states.

We called the node, focus, and limit cycle solutions the "Partially Synchronous Rest" (PSR), "Partially Synchronous Spiking" (PSS), and "Collective Periodic Wave" (CPW) states, respectively. In the PSR state, most neurons remain at rest, while in the PSS state, most neurons spike continuously. Nevertheless, in both these states, the macroscopic mean field (or order parameter) sits at an equilibrium. In contrast, the CPW state corresponds to periodic oscillations of the complex order parameter,

[1] $a_n = 2\pi / \int_{-\pi}^{\pi} (1 - \cos(x))^n = n!/(2n-1)!!$

and typically, both $|z(t)|$ and $\arg(z)$ oscillate in time indicating that the individual neurons clump together and spread apart in a periodic fashion. We refer the interested reader to Luke et al. (2013) for further details, including movies that illustrate both the microscopic and macroscopic behaviors of these collective states.

3. FORMULATION OF THE DRIVER-RESPONSE NETWORK

In this work, we are interested in the dynamics exhibited by a network of two coupled populations of theta neurons. We formulate the general case, but restrict analysis to the simplest such configuration: a driver-response network.

3.1. GENERAL TWO-POPULATION MODEL

Extending the model described above, a general formulation of a pair of interacting populations of theta neurons can be expressed as follows:

$$\dot{\theta}_{1,j} = 1 + \eta_{1,j} - (1-\eta_{1,j})\cos\theta_{1,j} + a_n(1+\cos\theta_{1,j}) \left[\frac{k_{11}}{N_1}\sum_{p=1}^{N_1}(1-\cos\theta_{1,p})^n + \frac{k_{12}}{N_2}\sum_{q=1}^{N_2}(1-\cos\theta_{2,q})^n\right],$$

$$\dot{\theta}_{2,j} = 1 + \eta_{2,j} - (1-\eta_{2,j})\cos\theta_{2,j} + a_n(1+\cos\theta_{2,j}) \left[\frac{k_{21}}{N_1}\sum_{p=1}^{N_1}(1-\cos\theta_{1,p})^n + \frac{k_{22}}{N_2}\sum_{q=1}^{N_2}(1-\cos\theta_{2,q})^n\right], \quad (10)$$

where $\theta_{1,j}$ and $\theta_{2,j}$ denote the jth neuron in the first and second populations, respectively, and the extension to any number of interacting populations is straightforward. The excitability parameters $\eta_{1,j}$ and $\eta_{2,j}$ are randomly drawn from two independent Lorentzian distributions as in Equation (4), with medians η_1, η_2 and widths Δ_1, Δ_2, respectively. We take the sharpness parameter of the pulse-like synaptic interaction, n, to be the same for both populations. Macroscopic mean field parameters $\tilde{z}_1(t)$, $\tilde{z}_2(t)$ can be defined for each population by analogy with Equation (5).

Adapting the procedures described in Luke et al. (2013), we derived the Ott-Antonsen reduction of the coupled networks of Equation (10). This resulted in the following dynamical system:

$$\dot{z}_1 = -i\frac{(z_1-1)^2}{2} + \frac{(z_1+1)^2}{2}\{-\Delta_1 + i[\eta_1 + k_{11}H_n(z_1) + k_{12}H_n(z_2)]\},$$

$$\dot{z}_2 = -i\frac{(z_2-1)^2}{2} + \frac{(z_2+1)^2}{2}\{-\Delta_2 + i[\eta_2 + k_{21}H_n(z_1) + k_{22}H_n(z_2)]\}. \quad (11)$$

with $H_n(z)$ defined as in Equations (7–9). As before, the asymptotic dynamics of Equation (11) can be shown to coincide with that of the order parameters of the populations in the network of Equation (10), in the limit $N_1, N_2 \to \infty$.

We showed in Luke et al. (2013) that the dynamical structure of the single population depends rather weakly on the synaptic sharpness parameter n. Furthermore, we argued that a modest sharpness is more biophysically plausible than the δ-function coupling obtained in the limit $n \to \infty$. Thus, from here on, we fix $n = 2$ and drop the subscript on H_n to ease notation.

3.2. THE DRIVER-RESPONSE SYSTEM

To put our network in the driver-response form, we set $k_{12} = 0$, so that population 1 receives no input from population 2. Therefore, the macrostates and bifurcations of population 1 are identical to those explored in Luke et al. (2013), described above. However, we allow $k_{21} \neq 0$. Our goal is to examine the consequences of the influence of population 1 on population 2. We call population 1 the "driver" and population 2 the "response" system. See **Figure 1**.

Writing the governing equation of population 2 as

$$\dot{z}_2 = -i\frac{(z_2-1)^2}{2} + \frac{(z_2+1)^2}{2}\left\{-\Delta_2 + i\left[\eta_{eff} + k_{22}H(z_2)\right]\right\} \quad (12)$$

with

$$\eta_{eff} \equiv \eta_2 + k_{21}H(z_1), \quad (13)$$

and comparing to Equation (6), we see that the behavior of population 2 is the same as that of a single population of theta neurons with an effective median excitability parameter η_{eff}. This effective parameter depends on the median excitability parameter intrinsic to population 2 η_2, the inter-population coupling k_{21}, and the state of the driver z_1.

Note that η_{eff} depends linearly on both η_2 and k_{21} and nonlinearly on the driver's state z_1 through $H(z_1)$. Additionally, η_{eff}

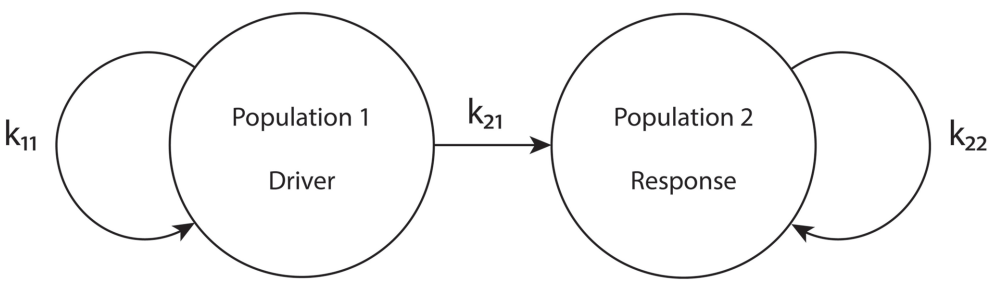


FIGURE 1 | The driver-response configuration. k_{11} and k_{22} are the intra-population coupling strengths for populations 1 and 2, respectively, and k_{21} is the uni-directional coupling strength between the driver population (1) and the response population (2).

may be time-dependent if population 1 exhibits a CPW state, since in that case z_1 oscillates periodically. In the following, we will examine all these cases.

4. RESULTS

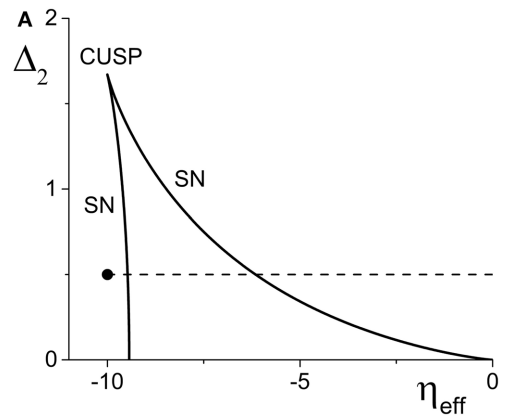
We will examine the behavior of population 2 as various parameters are varied. We organize the presentation of our results by first considering the case in which the driver population exhibits an equilibrium state. Later, we consider the case in which the driver population exhibits periodic behavior.

We will mainly consider two configurations of the response system. The "excitatorily coupled" response system has $k_{22} > 0$, and the "inhibitorily coupled" response system has $k_{22} < 0$. Other parameters are as noted below.

The bifurcation diagrams that appear below in **Figures 2**, **3**, **4B**, **5B**, **8C** were obtained using XPPAUT (Ermentrout, 2002). Data for all other figures were generated using custom-designed code.

4.1. DRIVER ON A MACROSCOPIC EQUILIBRIUM

We begin by fixing the driving population's parameters at $\eta_1 = -0.2$, $\Delta_1 = 0.1$, and $k_{11} = -2$, which corresponds to a PSR state. Thus, z_1 remains fixed at a constant value. We examine the behavior of the two response system configurations as we vary the



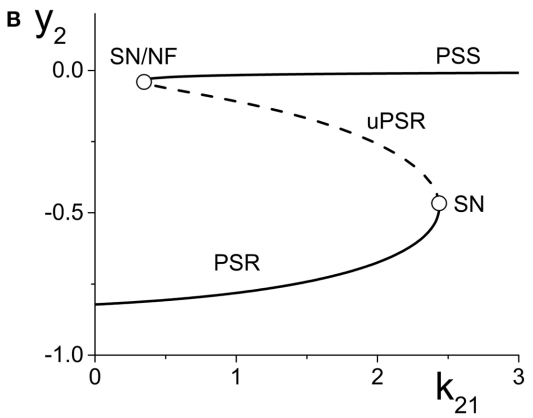
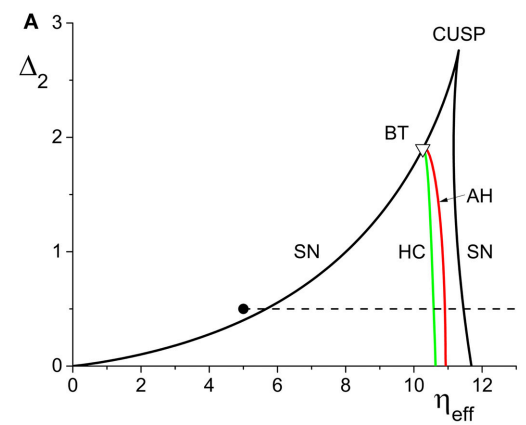


FIGURE 2 | (A) A two-dimensional bifurcation diagram of the excitatorily-coupled response system. The heavy black lines are saddle-node (SN) bifurcation curves, and the solid dot denotes the parameters of the response system when decoupled from the driver. In the cases considered in the main text, the driver causes η_{eff} to vary along the horizontal dotted line. The parameters are: $\eta_1 = -0.2$, $\Delta_1 = 0.1$, $k_{11} = -2$, and $k_{22} = 9$. **(B)** The one-dimensional bifurcation diagram showing the asymptotic values of $y_2 = \mathrm{Im}(z_2)$ vs. k_{21}. Solid and dashed curves indicate stable and unstable equilibria, respectively, corresponding to partially synchronous spiking (PSS) and partially synchronous resting (PSR) states. The parameters are as in **(A)**, with $\eta_2 = -10$ and $\Delta_2 = 0.5$.



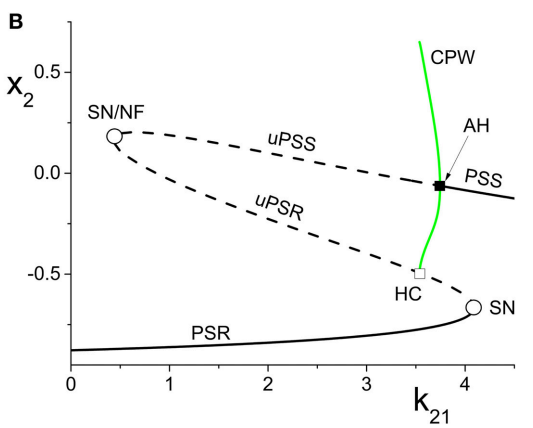


FIGURE 3 | (A) The two-dimensional bifurcation diagram of the inhibitorily-coupled response system. The heavy black lines are saddle-node (SN) bifurcation curves, green is a homoclinic (HC) bifurcation curve, and red is an Andronov-Hopf (AH) bifurcation curve. The latter two curves emerge from a Bogdanov-Takens (BT) point. The solid dot denotes the parameters of the response system when decoupled from the driver. In the cases considered in the main text, the driver causes η_{eff} to vary along the horizontal dotted line. The parameters are: $\eta_1 = -0.2$, $\Delta_1 = 0.1$, $k_{11} = -2$, and $k_{22} = -9$. **(B)** The one-dimensional bifurcation diagram showing the asymptotic value of $x_2 = \mathrm{Re}(z_2)$ vs. k_{21}. Solid curves denote stable equilibria; dashed black curves are unstable equilibria. Green represents the maxima and minima of a collective periodic wave (CPW) limit cycle. The parameters are as in **(A)**, with $\eta_2 = 5$ and $\Delta_2 = 0.5$.

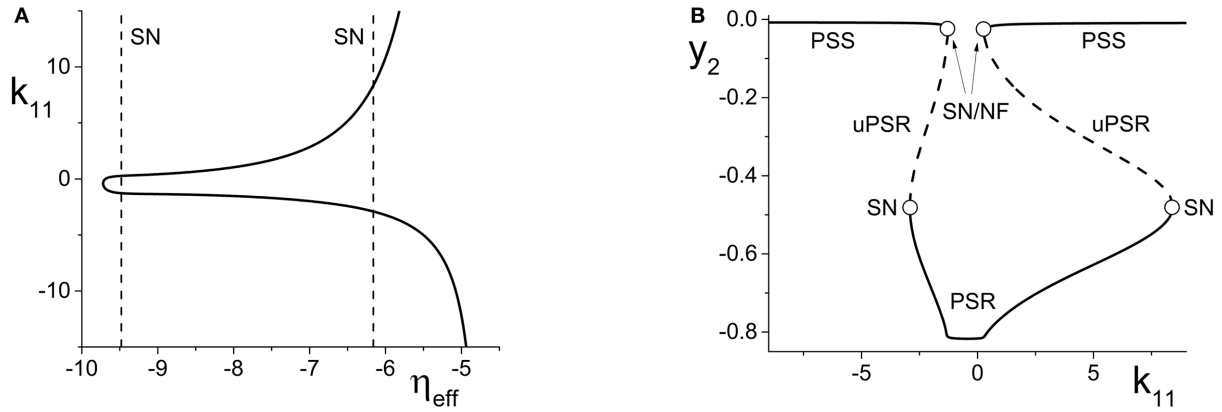


FIGURE 4 | (A) The non-linear behavior of η_{eff} as a function of k_{11} for the excitatorily-coupled response system. η_{eff} is plotted horizontally to facilitate comparison with **Figure 2A**. The parameters are: $\eta_1 = -0.05$, $\Delta_1 = 0.05$, $\eta_2 = -10$, with the inter-population coupling fixed at $k_{21} = 2.0$. **(B)** The one-dimensional bifurcation diagram showing the asymptotic value of $y_2 = \mathrm{Im}(z_2)$ vs. k_{11}. Solid and dashed curves indicate stable and unstable equilibria, respectively. The parameters are as in **(A)**, with $\Delta_2 = 0.5$ and $k_{22} = 9$.

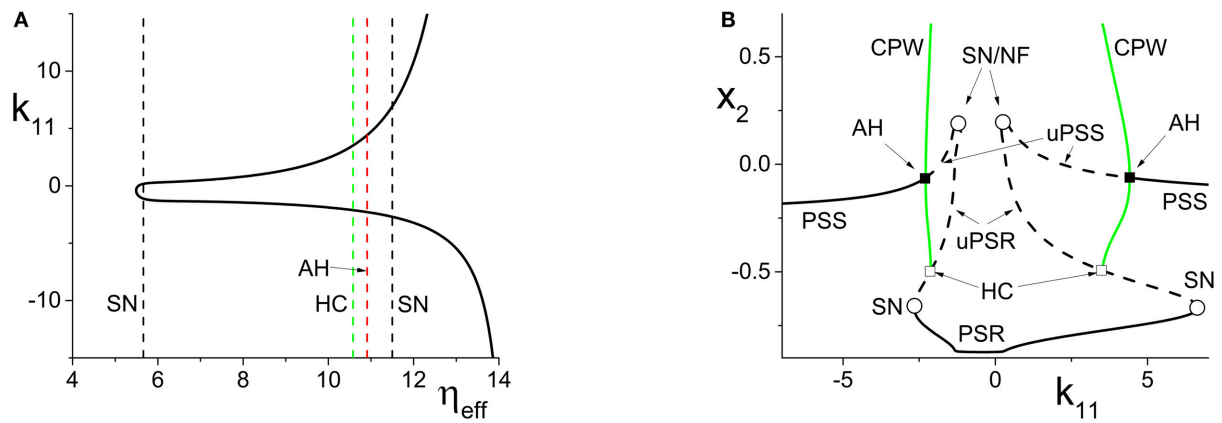


FIGURE 5 | (A) The non-linear behavior of η_{eff} as a function of k_{11} for the inhibitorily-coupled response system. η_{eff} is plotted horizontally to facilitate comparison with **Figure 3A**. **(B)** The one-dimensional bifurcation diagram showing the asymptotic value of $x_2 = \mathrm{Im}(z_2)$ vs. k_{11}. Solid and dashed black curves indicate stable and unstable equilibria, respectively, and green represents the maxima and minima of a CPW limit cycle state. The parameters are: $\eta_1 = -0.05$, $\Delta_1 = 0.05$, $\eta_2 = 5$, $\Delta_2 = 0.5$, and $k_{22} = -9$. The inter-population coupling is fixed at $k_{21} = 3.5$.

inter-population coupling parameter, k_{21}. From Equation (13), η_{eff} varies linearly with respect to k_{21}.

4.1.1. Excitatorily-coupled response system

We set the response system's internal coupling to $k_{22} = 9$, and show in **Figure 2A** the two-parameter bifurcation diagram of the response system with respect to Δ_2 and η_{eff}. Two saddle-node bifurcation curves which meet at a cusp are seen. To the left of these curves, the response network exhibits a PSR state, and to the right, a PSS state. These states coexist inside the approximately triangular region.

We set the remaining parameters of the response system to $\eta_2 = -10$ and $\Delta_2 = 0.5$. Thus, for $k_{21} = 0$, $\eta_{eff} = \eta_2$, and the response system is situated at the solid black point marked in **Figure 2A**. As k_{21} increases from zero, η_{eff} increases linearly along the dotted line in **Figure 2A**, starting from the black point. In so doing, it traverses the SN bifurcation curves. **Figure 2B** shows how the imaginary part of the response's asymptotic macroscopic mean field $[y_2 = \mathrm{Im}(z_2)]$ changes with respect to k_{21}, illustrating the coexistence of the stable PSR and PSS states, along with an unstable PSR state (uPSR).

The point marked "SN/NF" in **Figure 2B** indicates that as k_{21} increases, a saddle node bifurcation is encountered, corresponding to the left SN curve in **Figure 2A**. This creates a stable and an unstable PSS state. However, the unstable PSS state converts into an unstable PSR state at a value of k_{21} very slightly beyond the SN bifurcation. That is, the node corresponding to the unstable PSS state becomes a unstable PSR focus, a transition we called a Node-Focus (NF) transition in Luke et al. (2013). The distinction between these events is indistinguishable in the figure.

4.1.2. Inhibitorily-coupled response system

We performed a similar analysis for the case in which the response system's internal coupling is $k_{22} = -9$, i.e., inhibitory, and $\eta_2 =$

5. The remaining parameters were unchanged. The results are shown in **Figure 3**. In this case, the two-dimensional bifurcation diagram of the response system with respect to Δ_2 and η_{eff} (**Figure 3A**) shows a similar (but mirror-image) cusp of saddle-node curves. A new feature is the occurrence of a codimension-2 Bogdanov-Takens (BT) point on the left SN curve, and the emergence of homoclinic (HC; green) and Andronov-Hopf (AH; red) bifurcation curves from the BT point.

Figure 3B shows how the real part of the response's asymptotic macroscopic mean field $[x_2 = \mathrm{Re}(z_2)]$ changes with respect to k_{21}. As before, η_{eff} increases linearly as k_{21} increases, starting from the black solid point in **Figure 3A** and moving toward the right, traversing the various bifurcation curves along the dotted line. Note the presence of the attracting limit cycle CPW state in **Figure 3B**, which emerges at the HC bifurcation and terminates at the AH bifurcation as k_{21} increases.

It is interesting to note that in both cases described above, the same bifurcation structure would be encountered if, instead of varying k_{21} with a fixed value η_2, we varied η_2 with a fixed value of k_{21}. While this is obvious from Equation (13) since $H(z_1)$ is constant in these cases, this leads to the non-obvious conclusion that by modifying either the inter-population coupling or the intrinsic median excitability of the response population—two rather different system characteristics—one obtains identical transitions in the response network.

4.1.3. Variation of the driver's macroscopic equilibrium

In the cases we considered previously, η_{eff} changed linearly with respect to the inter-population coupling k_{21}. We now turn our attention to the effects incurred by altering the value of the driver influence function $H(z_1)$ in Equation (13). We do this by varying the driver's internal coupling strength k_{11}, thus causing the driver's asymptotic macroscopic mean field z_1 to change. This manipulation has the effect of changing η_{eff} *non-linearly* with respect to k_{11}.

For simplicity, we only consider a range of k_{11} such that the driver always remains on a macroscopic equilibrium state, and we fix the inter-population coupling at $k_{21} = 2$.

We begin with the case of the excitatorily-coupled response system considered above, with $\eta_2 = -10$, $\Delta_2 = 0.5$, and $k_{22} = 9$, and choose the remaining driver parameters to be $\eta_1 = -0.05$ and $\Delta_1 = 0.05$. **Figure 4A** shows the non-linear behavior of η_{eff} as k_{11} is varied. Even though we are considering k_{11} to be the independent parameter, we plot η_{eff} horizontally so that it may be easily compared to **Figure 2A**; recall that this shows the two-dimensional bifurcation diagram of the response system. Now, as k_{11} changes, η_{eff} moves back and forth along the dotted line non-linearly. In particular, **Figure 4A** shows that for very negative values of k_{11}, η_{eff} is near -5, which corresponds to a point in **Figure 2A** to the right of the SN curves. As k_{11} increases, η_{eff} decreases to approximately -10, thus crossing both SN curves in **Figure 2A** from right to left in the process. η_{eff} subsequently increases, and goes back across the SN curves from left to right. Note that **Figure 4A** includes vertical lines marking the position of the SN bifurcations (i.e., the values of η_{eff} at which the horizontal line at $\Delta_2 = 0.5$ in **Figure 2A** crosses the SN curves).

Figure 4B shows the behavior of the asymptotic state of the response system $[y_2 = \mathrm{Im}(z_2)]$ as a function of k_{11}. This shows that as k_{11} increases, the response system passes through two separate regions of bistability, corresponding to the two traversals of the triangular bistable region in **Figure 2A**. Thus, **Figure 4B** is qualitatively similar to two copies of **Figure 2B**, with the structure for $k_{11} < 0$ reversed. Note that the two regions are not symmetrical. This is due to the non-symmetric behavior of η_{eff} as k_{11} changes.

Next, we examine how the same manipulation of the driver system affects the inhibitorily-coupled response system. The parameters are as above, but with $\eta_2 = 5$ and $k_{22} = -9$. **Figure 5A** shows how η_{eff} changes as k_{11} is varied, again plotted with η_{eff} on the horizontal axis for ease of comparison with **Figure 3A**. Note the vertical lines in **Figure 5A** marking the SN, HC, and AH bifurcations.

The one-dimensional bifurcation diagram depicting the asymptotic state of the response system as a function of k_{11} is shown in **Figure 5B**. A situation similar to the previous case is observed. Two distorted versions of the structure of **Figure 3B**, with the features for $k_{11} < 0$ being reversed, are seen. Again, this is due to the non-linear and asymmetric behavior of η_{eff} as it traverses the bifurcations in **Figure 3A** twice: first right to left, and then left to right, as k_{11} is increased. Note also the presence of an attracting limit cycle CPW state in intervals of both positive and negative k_{11}.

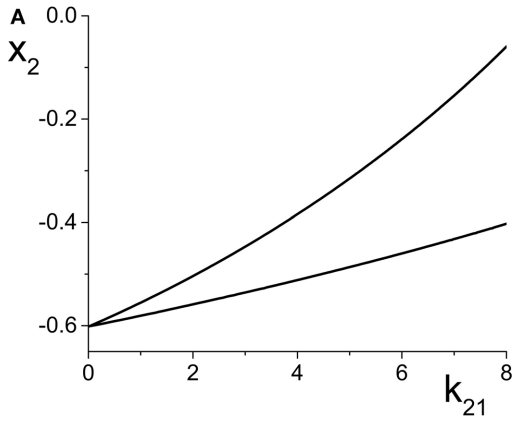
4.2. DRIVER ON A MACROSCOPIC LIMIT CYCLE

We now focus on the behavior of the response population when the driver is on a CPW state, which is a limit cycle of the driver's macroscopic mean field (or order parameter). Throughout this section, we fix the driver parameters at $\eta_1 = 10.75$, $k_{11} = -9$, and $\Delta_1 = 0.5$, which results in a CPW driver state for which $H(z_1)$ oscillates periodically in time. In particular, we have $H(z_1) > 0$ for all time. Thus, according to Equation (13), η_{eff} also oscillates periodically for $k_{21} \neq 0$, and both the centroid and the amplitude of the η_{eff} oscillation increase as k_{21} increases.

We show below that in this configuration, the response population can exhibit periodic, multistable, chaotic, and/or quasiperiodic behavior, depending on the response system's parameters and the interpopulation coupling strength k_{21}.

4.2.1. Periodic behavior in the response system

We begin by considering the excitatorily coupled response system, with $\Delta_2 = 0.5$ and $k_{22} = 9$, but with $\eta_2 = -20$. When decoupled from the driver, this places the response system at a point well to the left in the parameter space of **Figure 2A**. Thus, the response system in isolation asymptotes to a PSR state. As k_{21} is increased from zero to eight, η_{eff} oscillates back and forth along the horizontal line in **Figure 2A** at $\Delta_2 = 0.5$, but always stays to the left of the SN curves shown in that figure. Thus, the driver simply pushes the response system's PSR state back and forth, avoiding any bifurcations. The result is simple periodic behavior in the driven response system. **Figure 6A** shows a plot of the maximum and minimum of $x_2 = \mathrm{Re}(z_2)$ vs. k_{21}. As k_{21} increases, the amplitude of this simple periodic behavior increases. We observe that the frequency of the response system's oscillation is the same



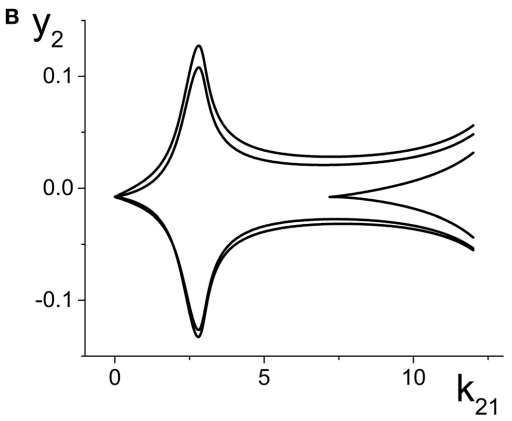


FIGURE 6 | (A) Simple periodic behavior in the response system driven by a CPW state of the driver as a function of the inter-population coupling strength k_{21}. The curves are local maxima and minima of $x_2 = \mathrm{Re}(z_2)$. The driver parameters are $\eta_1 = 10.75$, $\Delta_1 = 0.5$, and $k_{11} = -9$, and the response parameters are $\eta_2 = -20$, $\Delta_2 = 0.5$, and $k_{22} = 9$. **(B)** Slightly more complicated periodic behavior obtained at the same parameters, except with $\eta_2 = -5$. The curves are local maxima and minima of $y_2 = \mathrm{Im}(z_2)$.

as that of the driver throughout this range of interpopulation coupling.

We now change the response system such that $\eta_2 = -5$, and leave all other parameters the same as above. This change places the response system at a point to the right of the SN curves in **Figure 2A**, and for these parameters, the uncoupled response system asymptotes to a PSS state. Once again, as k_{21} increases, η_{eff} oscillates back and forth along the $\Delta_2 = 0.5$ line in **Figure 2A**, but this time it does so always staying to the right of the SN curves.

The result is multi-frequency periodic behavior in the response system that is more complicated than in the previous example. **Figure 6B** shows a plot of the *local* minima and maxima of $y_2 = \mathrm{Im}(z_2)$ vs. k_{21}. **Figure 7** shows y_2 vs. x_2 plots of the periodic orbits at $k_{21} = 6$ (upper panels) and $k_{21} = 10$ (lower panels). As k_{21} increases from zero, a periodic orbit with winding number two emerges (similar to that shown in **Figure 7A**) and grows in amplitude, peaking near $k_{21} \approx 2.5$. The amplitude subsequently decreases to a minimum near $k_{21} \approx 7.2$, and then slowly increases again. Note that the four curves in **Figure 6B** for $k_{21} \in [0, 7.2]$ correspond to two pairs of alternating local maxima and minima in the time series of y_2, as shown in **Figure 7B**.

Interestingly, near $k_{21} \approx 7.2$, an additional loop appears in the orbit, as shown in **Figure 7C**. This is reflected in the additional inner curves in **Figure 6B** that appear for $k_{21} \gtrapprox 7.2$, and the two additional local maxima and minima in the time series of y_2 in **Figure 7D**.

4.2.2. Multistability in the response system

Continuing with the excitatorily coupled response system (with $k_{22} = 9 > 0$), we set $\eta_2 = -10$ and leave all other parameters unchanged. In this case the uncoupled response system is at a point just to the left of the left SN curve in **Figure 2A**, and as k_{21} increases, η_{eff} again sweeps back and forth along the horizontal line at $\Delta_1 = 0.5$. However, now this sweeping cuts across both SN curves. Thus, the response system sweeps back and forth across the approximately triangular multistable region bounded by the SN curves.

Figure 8A shows the maxima and minima of y_2 vs. k_{21} for this case. The first feature to emerge as k_{21} increases from zero is a simple periodic orbit whose amplitude increases, similar to the example in **Figure 6A**. At $k_{21} \approx 0.5$, a new and separate coexisting limit cycle appears, as indicated by the upper curves that emerge in **Figure 8A**. **Figure 8B** shows the y_2 vs. x_2 plots of these two limit cycles at $k_{21} = 1.5$, where the larger orbit corresponds to the upper two curves in **Figure 8A**. In this bistable region, the macroscopic dynamics of the response system approaches one or the other of these periodic states, depending on the initial conditions.

Figure 8C shows, in black, the asymptotic states of y_2 vs. η_{eff} for *fixed* values of η_{eff}, with $k_{21} = 1.5$. These curves show that for a large interval of η_{eff}, a stable PSR coexists with a stable PSS and an unstable PSR state for the frozen (i.e., η_{eff} fixed) system. With the driver on the CPW state, η_{eff} sweeps from approximately -9.1 to -7.6 and back again–a range which is well within the bistable region. Superimposed in green in **Figure 8C** are projections of the two coexisting limit cycles onto this space, showing that the lower limit cycle is a simple periodic perturbation of the response system's underlying PSR state, and the upper limit cycle is a periodic perturbation of the underlying PSS state.

4.2.3. Chaos in the response system

We now switch to the inhibitorily coupled response system, with parameters $\eta_2 = 5$, $\Delta_2 = 0.5$, and $k_{22} = -9$. The parameter space of this system is shown in **Figure 3A**, and the uncoupled response system resides at the solid black dot in that figure, to the left of all the bifurcations. As the interpopulation coupling strength k_{21} increases, η_{eff} sweeps across the same horizontal line at $\Delta_2 = 0.5$ with increasing amplitude and centroid, initially crossing just the left SN bifurcation curve. At $k_{21} \approx 5.2$, η_{eff} begins sweeping across the homoclinic and the Andronov-Hopf bifurcation curves. Eventually, for sufficiently large k_{21}, η_{eff} sweeps across all four bifurcation curves (SN, AH, HC, and SN).

Figure 9A shows the local maxima and minima of $x_2 = \mathrm{Re}(z_2)$ vs. k_{21}. We initially see the emergence of a simple periodic orbit that grows slowly in amplitude. However, at $k_{21} \approx 5.2$, chaos

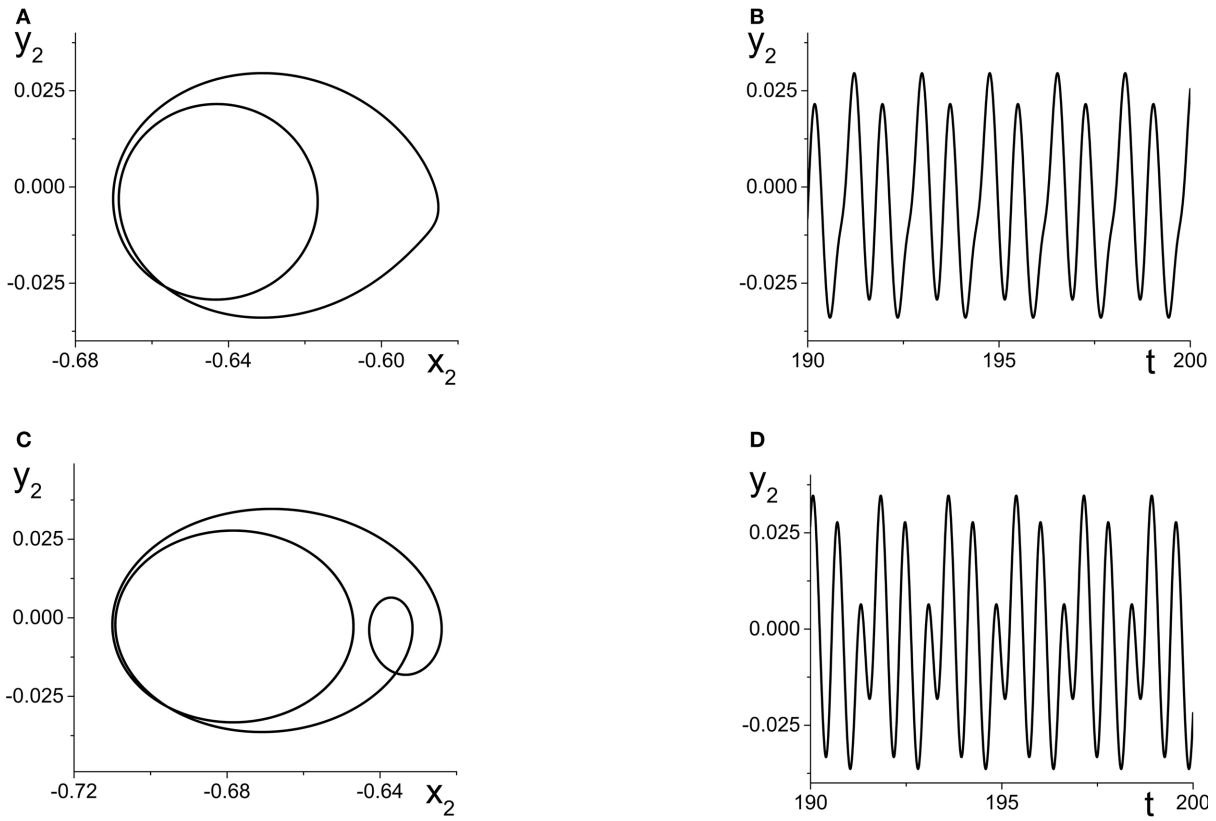


FIGURE 7 | The response system's behavior at parameters corresponding to Figure 6B at $k_{21} = 6$ (A,B) and $k_{21} = 10$ (C,D), with $z_2 = x_2 + iy_2$.

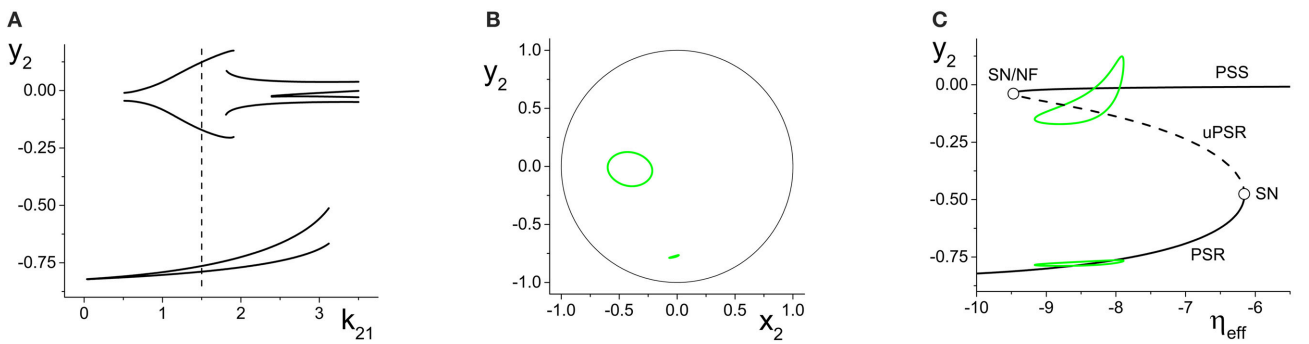


FIGURE 8 | Multistability in the response system driven by a CPW state of the driver. (A) Local maxima and minima of $y_2 = \mathrm{Re}(z_2)$ vs. the inter-population coupling k_{21}. **(B)** y_2 vs. x_2 plots showing two coexisting limit cycles of the response system at $k_{21} = 1.5$ (dotted vertical line in **A**). **(C)** The solid and dashed black curves show the asymptotic states of the response for fixed values of η_{eff}, with $k_{21} = 1.5$. Green curves are coexisting limit cycles of the response system when coupled to the driver. Parameters are: $\eta_1 = 10.75$, $\Delta_1 = 0.5$, $k_{11} = -9$; $\eta_2 = -10$, $\Delta_2 = 0.5$, $k_{22} = 9$.

suddenly emerges through a crisis. **Figure 9B** shows a magnification of this region, with a plot of the two largest Lyapunov exponents. We see that there are significant intervals of k_{21} for which there is a positive Lyapunov exponent, indicating the presence of macroscopic chaos.

As k_{21} increases, the first chaotic band, beginning at $k_{21} \approx 5.28$, coexists with the simple periodic loop that was present for smaller k_{21} (this coexistence is not apparent in the figure). Outside of this band, there is a window dominated by periodic behavior of rather high period. A second chaotic band appears at approximately $k_{21} = 5.48$. This second band terminates at approximately $k_{21} = 5.65$, after which a series of reverse period-doubling cascades are seen.

The y_2 vs. x_2 plot of the chaotic attractor present at $k_{21} = 5.296$, for which the largest Lyapunov exponent is approximately 0.2118, is shown in **Figure 10A**.

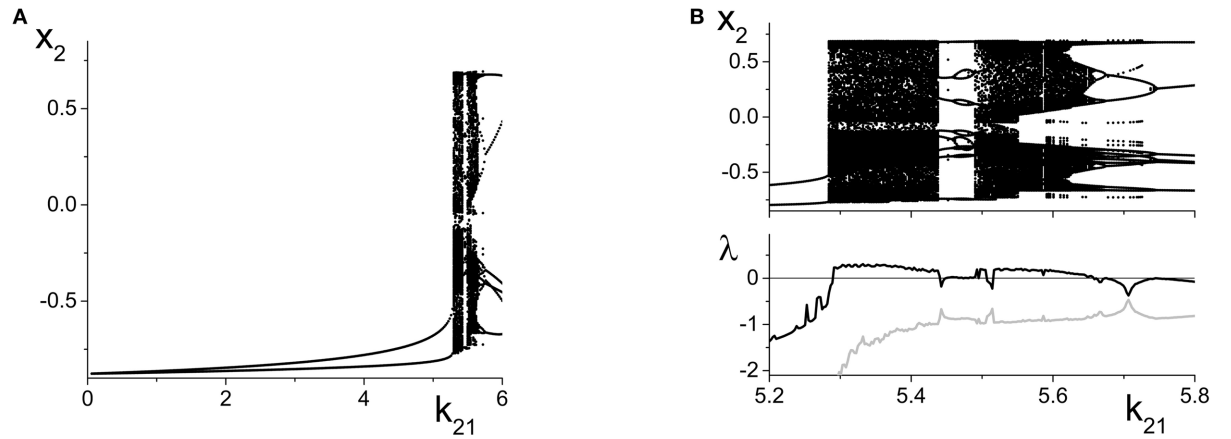


FIGURE 9 | Emergence of macroscopic chaos in the response system driven by a CPW state of the driver. (A) Local minima and maxima of $x_2 = \mathrm{Re}(z_2)$ vs. the inter-population coupling k_{21}. **(B)** Magnification of the chaotic region (top), with a plot of the largest two Lyapunov exponents. Parameters are: $\Delta_1 = 0.5$, $k_{11} = -9$, $\eta_1 = 10.75$; $\Delta_2 = 0.5$, $k_{22} = -9$, $\eta_2 = 5$.

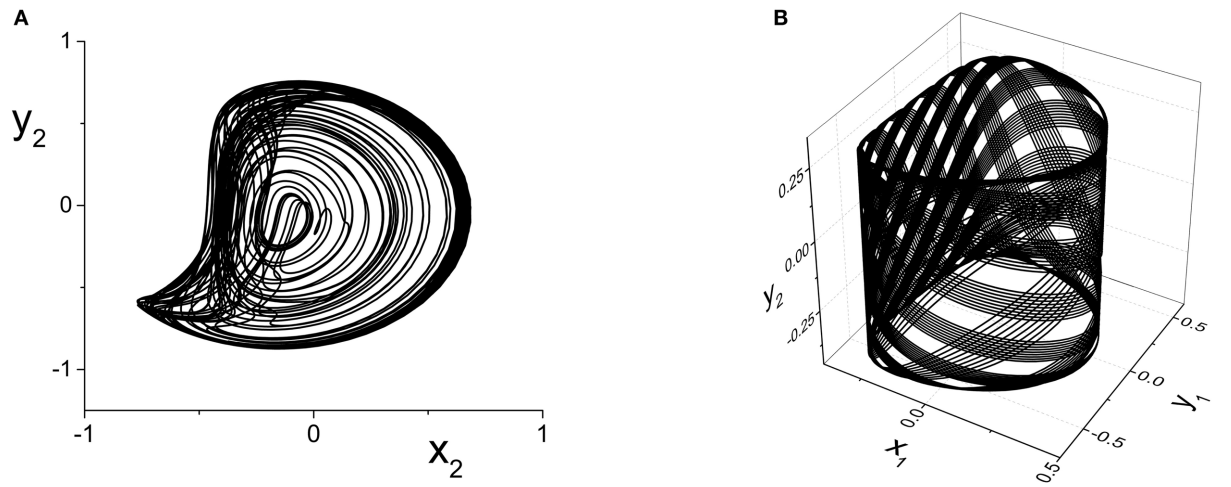


FIGURE 10 | (A) Chaotic (y_2 vs. x_2) and **(B)** Quasiperiodic (y_2 vs. x_1 vs. y_1) attractors in the response system driven by a CPW state of the driver. Parameters are: $k_{11} = k_{22} = -9$, with **(A)** $\eta_2 = 5$, $\Delta_1 = \Delta_2 = 0.5$ and $k_{21} = 5.296$, and **(B)** $\eta_2 = 10.75$, $\Delta_1 = 0.5$, $\Delta_2 = 0.3$, and $k_{21} = 0.1$.

4.2.4. Quasiperiodicity in the response system

Finally, we consider the case in which the response system exhibits a CPW state when uncoupled from the driver, and ask what happens when this is driven by another CPW state in the driver. We use the same drive system parameters as above, and set the response system's parameters to be the same except for $\Delta_2 = 0.3$. As the inter-population coupling strength k_{21} is increased, various phase-locked and quasiperiodic states are seen. An example of quasiperiodic behavior in the response system for $k_{21} = 0.1$ is shown in **Figure 10B**.

5. DISCUSSION

In this work, we have taken the first step toward designing a mathematically tractable modular network-of-networks representation of neuronal systems. Our approach is based on dynamical analysis techniques that enable a complete description of the emergent macroscopic behavior of large, heterogeneous discrete networks of globally-coupled phase oscillators. Building on previous results (Luke et al., 2013) in which we used these techniques to show that the collective dynamics of a single such population of theta neurons is relatively simple (exhibiting just equilibria and limit cycle states), we constructed the next simplest hierarchical structure: a driver-response configuration of theta neuron populations. Our results show that even in this simplest of configurations, the response system (and hence, the network as a whole) can exhibit a full range of dynamical behaviors and surprising complexity. A notable strength of our work is that despite the complexity that emerges from this arrangement, the behavior can be understood and explained in terms of what is known about a single population's dynamics and bifurcation structure.

With the driving system on a fixed equilibrium, we showed that the response system is equivalent to a single population with

a simple shift in one parameter. Specifically, this parameter is the median of the distribution of excitability parameters in the response system, which indicates whether the response population is dominated by excitable or intrinsically-spiking neurons. Although this arrangement does not introduce any new dynamical features, we showed that the response system can nevertheless still exhibit an interesting bifurcation structure involving macroscopic equilibria, limit cycles, and multistability as the strength of the inter-population coupling varies. More interestingly, we found that the inter-population coupling strength is effectively equivalent to the response system's median intra-population excitability. By this we mean that changes in either of these rather different network parameters lead to identical bifurcation scenarios. This surprising result follows from the drive-response network configuration in particular.

The first level of additional complication arose when modestly altering an internal parameter of the drive system. This effectively led to a *non-linear* change in the response system's median excitability, causing a dramatic change in the response's bifurcation structure. Such bifurcation structures might be difficult to understand if encountered blindly, as might be the case when studying the dynamics of a network without knowledge of its internal structure. Experimental studies of neuronal networks often take a similar "black box" approach out of necessity, since detailed knowledge of connectivity (i.e., the "connectome") is rarely available. In our case, however, we showed that knowledge of the non-linearity, along with knowledge of the bifurcation structure of a single network, leads to a natural explanation of the additional features that arise due to the network-of-networks structure. In our particular case studies, we observed multiple distorted and reversed copies of the bifurcation structure that is associated with a single population of theta neurons. We therefore speculate that in "black box" investigations, the observation of such repeated and/or distorted bifurcation structures might be indicative of driver-response-type connectivity in the network of study.

Finally, we investigated the consequences of placing the driver system on a collective rhythmic state (i.e., a macroscopic periodic orbit). Our results were consistent with previous results that studied non-autonomous phase oscillator (So and Barreto, 2011) and theta neuron systems (So et al., 2014). In those investigations, it was shown that networks of oscillators subjected to a sinusoidal variation of a network parameter led to complicated dynamics including quasiperiodicity and macroscopic chaos. Here, our driver-response arrangement of two separate interacting populations of theta neurons leads to an overall autonomous system, but with the response system being subjected to a periodic driving signal from the driver. Such arrangements might be found in real neuronal systems at the early stages of sensory input processing. For example, the lateral geniculate nucleus may be driven by a periodic visual signal delivered to the retina. Another candidate might be the trisynaptic circuit of the dentate gyrus and the CA3 and CA1 regions of the hippocampus (Kandel et al., 2000). More generally, the information-processing capabilities of the brain are thought to be regulated by collective rhythms, notably theta and gamma oscillations, which arise in various areas and periodically drive other areas (Buzsáki, 2006).

Our results may also have implications for populations of bursting neurons (So et al., 2014). Neuronal bursting in individual neurons is commonly understood to arise as the result of the interplay between a slowly oscillating neuronal parameter (or "slow variable") and the neuron's fast spiking dynamics. Bursting arises if the slow parameter sweeps back and forth across bifurcations, and (Rinzel and Ermentrout, 1989) classified bursters as square, parabolic, or elliptic based on the bifurcations encountered in this process. It has also been demonstrated that slowly oscillating intra- and extra-cellular ion concentrations can lead to wide range of neuronal bursting behaviors (Cressman et al., 2009, 2011; Barreto and Cressman, 2011).

Finally, we note that our explorations in this work were limited to cases in which the driver population's parameters were either fixed or were varied only modestly. In the latter case, we changed the driver's median excitability parameter only to the extent that its collective equilibrium state was displaced but not altered. Significantly greater complexity in the response's dynamics would arise if the collective state of the driver were pushed across its own bifurcations, possibly resulting in topological changes and hysteretic effects in the driver's macroscopic state. As discussed above, such complexity would be difficult to understand if encountered in a "black box"-type investigation. Nevertheless, if it is known that the network of interest has a driver-response structure, it may be possible to comprehend the origin of such complexity in the manner that we have outlined here.

This study constitutes an initial attempt at building a mathematically tractable model to understand the collective behavior of a hierarchical "network-of-networks" arrangement of model neurons. In future work we plan to consider networks of networks that include feedback connections and additional populations in an effort to understand the emergence of macroscopic dynamical complexity in more realistic networks.

AUTHOR CONTRIBUTIONS

Tanushree B. Luke, Ernest Barreto, and Paul So conceived and designed the investigation, analyzed the data, and wrote the paper. Tanushree B. Luke and Paul So performed the numerical computations.

ACKNOWLEDGMENT

Publication of this article was funded by the George Mason University Libraries Open Access Publishing Fund.

REFERENCES

Bacci, A., Huguenard, J. R., and Prince, D. A. (2003). Functional autaptic neurotransmission in fast-spiking interneurons: a novel form of feedback inhibition in the neocortex. *J. Neurosci.* 23, 859–866. Available online at: http://www.jneurosci.org/content/23/3/859.abstract

Barreto, E., and Cressman, J. (2011). Ion concentration dynamics as a mechanism for neuronal bursting. *J. Biol. Phys.* 37, 361–373. doi: 10.1007/s10867-010-9212-6

Bekkers, J. M. (2003). Synaptic transmission: functional autapses in the cortex. *Curr. Biol.* 13, R433–R435. doi: 10.1016/S0960-9822(03)00363-4

Bullmore, E., and Sporns, O. (2009). Complex brain network: graph theoretical analysis of structural and functional systems. *Nat. Rev. Neurosci.* 10, 186–198. doi: 10.1038/nrn2575

Buzsáki, G. (2006). *Rhythms of The Brain, 1 Edn*. New York, NY: Oxford University Press. doi: 10.1093/acprof:oso/9780195301069.001.0001

Cressman, J., Ullah, G., Ziburkus, J., Schiff, S., and Barreto, E. (2009). The influence of sodium and potassium dynamics on excitability, seizures, and the stability of persistent states: 1. Single neuron dynamics. *J. Comput. Neurosci.* 26, 159–170. doi: 10.1007/s10827-008-0132-4

Cressman, J., Ullah, G., Ziburkus, J., Schiff, S., and Barreto, E. (2011). Erratum to: the influence of sodium and potassium dynamics on excitability, seizures, and the stability of persistent states: I. single neuron dynamics. *J. Comput. Neurosci.* 30, 781. doi: 10.1007/s10827-011-0333-0

Ermentrout, G. (1996). Type I membranes, phase resetting curves and synchrony. *Neural Comput.* 8, 979–1001. doi: 10.1162/neco.1996.8.5.979

Ermentrout, G. (2002). *Simulating, Analyzing, and Animating Dynamical Systems (A Guide to XPPAUT for Researchers and Students*. Philadelphia, PA: SIAM. doi: 10.1137/1.9780898718195

Ermentrout, G., and Kopell, N. (1986). Parabolic bursting in an excitable system coupled with a slow oscillation. *SIAM J. Appl. Math.* 46, 233–253. doi: 10.1137/0146017

Harris, K. (2005). Neural signatures of cell assembly organization. *Nat. Rev. Neurosci.* 6, 399–407. doi: 10.1038/nrn1669

Hebb, D. (1949). *The Organization of Behavior: A Neuropsychological Theory, 1 Edn*. New York, NY: Wiley.

Hodgkin, A. (1948). The local electric charges associted with repetitive action in non-medulated axons. *J. Physiol.* 107, 165–181.

Izhikevich, E. (2007). *Dynamical Systems in Neuroscience: The Geometry of Excitability and Bursting*. Cambridge, MA: MIT Press.

Kandel, E., Schwartz, J., and Jessell, T. (2000). *Principles of Neural Science, 4 Edn*. New York, NY: McGraw-Hill.

Laing, C. (2014). Derivation of a neural field model from a network of theta neurons. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 90:010901. doi: 10.1103/PhysRevE.90.010901

Luke, T., Barreto, E., and So, P. (2013). Complete classification of the macroscopic behavior of a heterogeneous network of theta neurons. *Neural Comput.* 25, 3207–3234. doi: 10.1162/NECO_a_00525

Marvel, S. A., Mirollo, R. E., and Strogatz, S. H. (2009). Identical phase oscillators with global sinusoidal coupling evolve by mbius group action. *Chaos* 19:043104. doi: 10.1063/1.3247089

Meunier, D., Lambiotte, R., and Bullmore, E. (2010). Modular and hierarchically modular organization of brain networks. *Front. Neurosci.* 4:200. doi: 10.3389/fnins.2010.00200

Nowak, L., Azouz, R., Sanchez-Vives, M., Gray, C., and McCormick, D. (2003)). Electrophysiological classes of cat primary visual cortical neurons *in vivo* as revealed by quantitative analyses. *J. Neurophysiol.* 89, 1541–1566. doi: 10.1152/jn.00580.2002

Ott, E., and Antonsen, T. (2008). Low dimensional behavior of large systems of globally coupled oscillators. *Chaos* 18, 037113. doi: 10.1063/1.2930766

Ott, E., and Antonsen, T. (2009). Long time evolution of phase oscillator systems. *Chaos* 19, 023117. doi: 10.1063/1.3136851

Ott, E., Hunt, B. R., and Antonsen, T. M. (2011). Comment on long time evolution of phase oscillator systems [*Chaos* 19, 023117 (2009)]. *Chaos* 21, 025112. doi: 10.1063/1.3574931

Pazó, D., and Montbrió, E. (2014). Low-dimensional dynamics of populations of pulse-coupled oscillators. *Phys. Rev. X* 4:011009. doi: 10.1103/PhysRevX.4.011009

Pikovsky, A., and Rosenblum, M. (2011). Dynamics of heterogeneous oscillator ensembles in terms of collective variables. *Phys. D Nonlinear Phenom.* 240, 872–881. doi: 10.1016/j.physd.2011.01.002

Rinzel, J., and Ermentrout, G. (1989). "Analysis of neural excitability and oscillations," in *Methods in Neuronal Modeling*, eds C. Koch and I. Segev (Cambridge, MA: The MIT Press), 251–291.

Sherrington, C. (1906). *The Integrative Action of The Nervous System, 1 Edn*. New York, NY: Scribners.

So, P., and Barreto, E. (2011). Generating macroscopic chaos in a network of globally coupled phase oscillators. *Chaos* 21, 033127. doi: 10.1063/1.3638441

So, P., Luke, T., and Barreto, E. (2014). Networks of theta neurons with time-varying excitability: macroscopic chaos, multistability, and final-state uncertainty. *Phys. D* 267, 16–26. doi: 10.1016/j.physd.2013.04.009

Tateno, T., Harsch, A., and Robinson, H. (2004). Threshold firing frequency-current relationships of neurons in rat somatosensory cortex: type 1 and type 2 dynamics. *J. Neurophysiol.* 92, 2283–2294. doi: 10.1152/jn.00109.2004

Zhou, C., Zemanova, L., Zamora, G., Hilgetag, C., and Kurths, J. (2006). Hierarchical organization unveiled by functional connectivity in complex brain networks. *Phys. Rev. Lett.* 97:238103. doi: 10.1103/PhysRevLett.97.238103

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.

Received: 31 August 2014; accepted: 27 October 2014; published online: 18 November 2014.

Citation: Luke TB, Barreto E and So P (2014) Macroscopic complexity from an autonomous network of networks of theta neurons. Front. Comput. Neurosci. ***8****:145. doi: 10.3389/fncom.2014.00145*

This article was submitted to the journal Frontiers in Computational Neuroscience.

frontiers in Computational Neuroscience

ORIGINAL RESEARCH
published: 02 June 2015
doi: 10.3389/fncom.2015.00064

Multiscale entropy analysis of biological signals: a fundamental bi-scaling law

Jianbo Gao[1,2]*, *Jing Hu*[2], *Feiyan Liu*[1,3] and *Yinhe Cao*[1,2]

[1] Institute of Complexity Science and Big Data Technology, Guangxi University, Nanning, China, [2] PMB Intelligence LLC, Sunnyvale, CA, USA, [3] School of Management, University of Chinese Academy of Sciences, Beijing, China

Since introduced in early 2000, multiscale entropy (MSE) has found many applications in biosignal analysis, and been extended to multivariate MSE. So far, however, no analytic results for MSE or multivariate MSE have been reported. This has severely limited our basic understanding of MSE. For example, it has not been studied whether MSE estimated using default parameter values and short data set is meaningful or not. Nor is it known whether MSE has any relation with other complexity measures, such as the Hurst parameter, which characterizes the correlation structure of the data. To overcome this limitation, and more importantly, to guide more fruitful applications of MSE in various areas of life sciences, we derive a fundamental bi-scaling law for fractal time series, one for the scale in phase space, the other for the block size used for smoothing. We illustrate the usefulness of the approach by examining two types of physiological data. One is heart rate variability (HRV) data, for the purpose of distinguishing healthy subjects from patients with congestive heart failure, a life-threatening condition. The other is electroencephalogram (EEG) data, for the purpose of distinguishing epileptic seizure EEG from normal healthy EEG.

Keywords: scaling law, multiscale entropy analysis, fractal signal, heart rate variability (HRV), adaptive filtering

Edited by:
Tobias Alecio Mattei, Brain & Spine Center - InvisionHealth - Kenmore Mercy Hospital, USA

Reviewed by:
Guillaume Lajoie, Max Planck Institute for Dynamics and Self-Organization, Germany
Bailu Si, Chinese Academy of Sciences, China
Xiaoli Li, Beijing Normal University, China

***Correspondence:**
Jianbo Gao, Institute of Complexity Science and Big Data Technology, Guangxi University, 100 Daxue Road, Nanning, Guangxi 530005, China
jbgao.pmb@gmail.com

Received: *14 December 2014*
Accepted: *14 May 2015*
Published: *02 June 2015*

Citation:
Gao J, Hu J, Liu F and Cao Y (2015) Multiscale entropy analysis of biological signals: a fundamental bi-scaling law. Front. Comput. Neurosci. 9:64. doi: 10.3389/fncom.2015.00064

1. Introduction

Biological systems provide the definitive examples of highly integrated systems functioning at multiple time scales. Neurons function on a time scale of milliseconds. Circadian rhythms operate on time scale of hours, reproductive cycles occur on a time scale of weeks, and bone remodeling involves time scales of months. As an integrated system, each process interacts with faster and slower processes. Consequently, biosignals often are multiscaled (Gao et al., 2007)—depending upon the scale at which the signals are examined, they may exhibit different behaviors (e.g., nonlinearity, sensitive dependence on small disturbances, long memory, extreme variations, and nonstationarity), just as a great painting may exhibit various details and arouse a multitude of aesthetic feelings when appreciated at different distances, from different angles, under different illuminations, and under different moods.

With the rapid advance of sensing technology, complex data have been accumulating exponentially in all areas of life sciences. To better cope with such complex data, recently, Costa et al. (2005) have introduced an interesting method, the multiscale entropy (MSE) analysis. MSE has found numerous applications in various types of biosignal analysis, including fetal heart rate monitoring (Cao et al., 2006), assessment of EEG dynamical complexity in

Alzheimer's disease (Mizuno et al., 2010), classification of surface EMG of neuromuscular disorders (Istenic et al., 2010), heart rate analysis for predicting hospital mortality (Norris et al., 2008), and analysis of hear beat interval and blood flow for characterizing psychological dimensions in non-pathological subjects (Nardelli et al., 2015). MSE has also been extended to multivariate MSE (Ahmed and Mandic, 2011) and multiscale permutation entropy (Li et al., 2010). So far, however, no analytic analyses about MSE or multivariate MSE have been carried out. This has severely limited our basic understanding of MSE. For example, it has not been known whether MSE estimated using default parameter values and short data set is meaningful or not. Nor is it known whether MSE has any relation with other complexity measures, such as the Hurst parameter, which characterizes the correlation structure of the data.

To help gain insights into the above questions, and to guide more fruitful applications of MSE in diverse fields of life sciences, in this work, we report a fundamental bi-scaling law for MSE of the most popular model of biosignals, the fractal $1/f$ type time series. As example applications, we will analyze heart rate variability (HRV) and electroencephalogram (EEG) data. With HRV, we will focus on distinguishing healthy subjects from patients with congestive heart failure (CHF), a life-threatening condition, as well as resolving an interesting debate (Wessel et al., 2003; Nikulin and Brismar, 2004) regarding the usefulness of MSE in distinguishing HRV of healthy subjects from that of patients with certain cardiac disease. With EEG, we will focus on distinguishing epileptic seizure EEG from normal healthy EEG.

2. Materials and Methods

2.1. Data

To illustrate the use of scaling analysis of MSE, in this paper, we analyze two types of data, heart rate variability (HRV), for the purpose of distinguishing healthy subjects from patients with congestive heart failure (CHF), and EEG, for the detection of epileptic seizures.

We downloaded two types of HRV data from the PhysioNet (MIT-BIH Normal Sinus Rhythm Database and BIDMC Congestive Heart Failure Database available at http://www.physionet.org/physiobank/database/#ecg), one for healthy subjects, and the other for subjects with CHF. The latter includes long-term ECG recordings from 15 subjects (11 men, aged 22 to 71, and 4 women, aged 54 to 63) with severe CHF (NYHA class 3–4). This group of subjects was part of a larger study group receiving conventional medical therapy prior to receiving the oral inotropic agent, milrinone. Further details about the larger study group can be found at the PhysioNet. The individual recordings of ECG are each about 20 h in duration, and contain two ECG signals each sampled at 250 samples per second with 12-bit resolution over a range of ±10 millivolts. The other database are for 18 normal subjects. The individual recordings are each about 25 h in duration, each sampled at 128 samples per second. The HRV data analyzed here are the R-R intervals (in unit of second) derived from the ECG recordings.

The EEG database is downloaded at http://www.meb.unibonn.de/epileptologie/science/physik/eegdata.html. The database consists of three groups, H (healthy), E (epileptic subjects during a seizure-free interval), and S (epileptic subjects during seizure); each group contains 100 data segments, whose length is 4097 data points with a sampling frequency of 173.61 Hz. These data have been carefully examined by adaptive fractal analysis (Gao et al., 2011c) and scale-dependent Lyapunov exponent (Gao et al., 2006b, 2011b, 2012), for the same purpose of distinguishing epileptic seizure EEG from normal healthy EEG.

2.2. Methods

Entropy characterizes creation of information in a dynamical system. To facilitate derivation of a fundamental scaling law for MSE, we first rigorously define MSE and all related concepts.

Suppose that the F-dimensional phase space is partitioned into boxes of size ε^F. Suppose that there is an attractor in phase space and consider a transient-free trajectory $\vec{x}(t)$. The state of the system is now measured at intervals of time τ. Let $p(i_1, i_2, \cdots, i_d)$ be the joint probability that $\vec{x}(t=\tau)$ is in box i_1, $\vec{x}(t=2\tau)$ is in box $i_2, \cdots$, and $\vec{x}(t=d\tau)$ is in box i_d. Let us now introduce the block entropy,

$$H_d(\varepsilon, \tau) = -\sum_{i_1,\cdots,i_d} p(i_1, \cdots, i_d) \ln p(i_1, \cdots, i_d), \quad (1)$$

take the difference between $H_{d+1}(\varepsilon, \tau)$ and $H_d(\varepsilon, \tau)$, and normalize it by τ,

$$h_d(\varepsilon, \tau) = \frac{1}{\tau}[H_{d+1}(\varepsilon, \tau) - H_d(\varepsilon, \tau)]. \quad (2)$$

Let

$$h(\varepsilon, \tau) = \lim_{d\to\infty} h_d(\varepsilon, \tau) \quad (3)$$

It is called the (ε, τ)-entropy (Gaspard and Wang, 1993). Taking limits, we obtain the Kolmogorov-Sinai (K-S) entropy,

$$\begin{aligned} K &= \lim_{\tau\to 0} \lim_{\varepsilon\to 0} h(\varepsilon, \tau) \\ &= \lim_{\tau\to 0} \lim_{\varepsilon\to 0} \lim_{d\to\infty} \frac{1}{\tau}[H_{d+1}(\varepsilon, \tau) - H_d(\varepsilon, \tau)] \end{aligned} \quad (4)$$

We now consider computation of the (ε, τ)-entropy from a time series of length N, $x_1, x_2, \cdots, x_N$. As is well-known, the first step is to use the time delay embedding to construct vectors of the form:

$$V_i = [x_i, x_{i+L}, ..., x_{i+(m-1)L}], \quad (5)$$

where m, the embedding dimension, and L, the delay time, can be chosen according to certain optimization criterion (Gao et al., 2007). Then one can employ the Cohen-Procaccia algorithm (Cohen and Procaccia, 1985) to estimate the (ε, τ)-entropy. In particular, when it is evaluated at a fixed finite scale $\hat{\varepsilon}$, the resulting entropy is called the approximate entropy. To get better statistics from a finite time series, one may compute $K_2(\varepsilon)$

using the Grassberger-Procaccia's algorithm (Grassberger and Procaccia, 1983):

$$K_2(\varepsilon) = \lim_{m\to\infty} \frac{\ln C^{(m)}(\varepsilon) - \ln C^{(m+1)}(\varepsilon)}{mL\delta t} \quad (6)$$

where δt is the sampling time, $C^{(m)}(\varepsilon)$ is the correlation integral based on the $m-$dimensional reconstructed vectors V_i and V_j,

$$C^{(m)}(\varepsilon) = \lim_{N_v\to\infty} \frac{2}{N_v(N_v-1)} \sum_{i=1}^{N_v-1} \sum_{j=i+1}^{N_v} H(\varepsilon - ||V_i - V_j||), \quad (7)$$

where $N_v = N - (m-1)L$ is the number of reconstructed vectors, $H(y)$ is the Heaviside function (1 if $y \geq 0$ and 0 if $y < 0$). $C^{(m+1)}(\varepsilon)$ can be computed similarly based on the $m+1-$dimensional reconstructed vectors. When we evaluate $K_2(\varepsilon)$ at a finite fixed scale $\hat{\varepsilon}$, we obtain the sample entropy S_e (Richman and Moorman, 2000).

MSE analysis is based on the sample entropy S_e. The procedure is as follows. Let $X = \{x_t : t = 1, 2, \dots\}$ be a covariance stationary stochastic process with mean μ, variance σ^2, and autocorrelation function $r(k), k \geq 0$. Construct a new covariance stationary time series

$$X^{(b_s)} = \{x_t^{(b_s)} : t = 1, 2, 3, \dots\}, \; b_s = 1, 2, 3, \dots,$$

by averaging the original series X over non-overlapping blocks of size b_s,

$$x_t^{(b_s)} = (x_{tb_s-b_s+1} + \cdots + x_{tb_s})/b_s, \quad t \geq 1\,. \quad (8)$$

MSE analysis involves (i) choosing a finite scale $\hat{\varepsilon}$ in phase space, and (ii) computing S_e from the original and the smoothed data X and $X^{(b_s)}$ at the chosen scale $\hat{\varepsilon}$. For convenience of later discussion, we denote $K_2^{(b_s)}(\varepsilon)$ for the correlation entropy of the smoothed data. When $b_s = 1$, it is the correlation entropy of the original data, and can be simply denoted as $K_2(\varepsilon)$.

We emphasize that the length of the smoothed time series is only $1/b_s$ of the original one. To fully resolve the scaling behavior of $K_2(\varepsilon)$, the requirement on data length is quite stringent. A fundamental question is whether MSE calculated from short noisy data is meaningful or not.

3. Results

3.1. Scaling for the MSE of Fractal Time Series

Among the most widely used models for biological signals, including HRV, EEG, and posture (Gao et al., 2011a), is the fractal time series with long memory, the so-called $1/f^{\alpha}$, or $1/f^{2H-1}$, $\alpha = 2H-1$ processes, where $0 < H < 1$ is called the Hurst parameter, whose value determines the correlation structure of the data (Gao et al., 2006a, 2007): when $H = 1/2$, the process is like the independent steps of the standard Brownian-motion; when $H < 1/2$, the process has anti-persistent correlations; when $H > 1/2$, the process has persistent correlations. Two special cases, white noise with $H = 0.5$ and $1/f$ process with $H = 1$, have been extensively used for the development of multivariate MSE (Ahmed and Mandic, 2011). In this subsection, we derive fundamental scalings for MSE of the ubiquitous $1/f^{2H-1}$ noise.

A covariance stationary stochastic process $X = \{X_t : t = 0, 1, 2, \dots\}$, with mean μ, variance σ^2, and autocorrelation function $r(w), w \geq 0$, is said to have long range correlation if $r(w)$ is of the form Cox (1984)

$$r(w) \sim w^{2H-2}, \;\; as \;\; w \to \infty, \quad (9)$$

where $0 < H < 1$ is the Hurst parameter. When $1/2 < H < 1$, $\sum_w r(w) = \infty$, leading to the term long range correlation. Note the X time series has a power spectral density $1/f^{2H-1}$. Its integration, $\{y_t\}$, where $y_t = \sum_{i=1}^{t} x_i$, is called a random walk process which is nonstationary with power-spectral density (PSD) $1/f^{2H+1}$. Being $1/f$ processes, they cannot be aptly modeled by Markov processes or ARIMA models (Box and Jenkins, 1976), since the PSD for those processes are distinctly different from $1/f$. To adequately model $1/f$ processes, fractional order processes has to be used. The most popular is the fractional Brownian motion model Mandelbrot (1982), whose increment process is called the fractional Gaussian noise (fGn). The importance and popularity of fGn in modeling various types of noises in science and engineering motivates us to focus our analysis on it when deriving the bi-scaling law.

$1/f^{2H-1}$ noises are self-similar, with the autocorrelation for the original data and the smoothed data (defined by Equation 8) being the same (Gao et al., 2006a, 2007). This signifies that there must exist a simple relation between $K_2^{(b_s)}(\varepsilon)$ and $K_2(\varepsilon)$. To find this relation, we note that the variance, $var(X^{(b_s)})$, of the smoothed data, and the variance, σ^2, of the original data, are related by the following simple and elegant scaling law (Gao et al., 2006a, 2007),

$$var(X^{(b_s)}) = \sigma^2 {b_s}^{2H-2} \quad (10)$$

Equation (10) states that the scale ε for the original data is transformed to a smaller scale $b_s^{H-1}\varepsilon$ for the smoothed data. Using the self-similarity property of the $1/f^{2H-1}$ noise, we therefore obtain,

$$K_2^{(b_s)}\left(b_s^{H-1}\varepsilon\right) = K_2(\varepsilon) \quad (11)$$

Since for stationary random processes, $K_2(\varepsilon)$ diverges when $\varepsilon \to 0$, Equation (11) states that $K_2^{(b_s)}\left(b_s^{H-1}\varepsilon\right)$ can be obtained from $K_2(\varepsilon)$ by shifting downward the curve for $K_2(\varepsilon)$. How much $K_2(\varepsilon)$ should be shifted depends on the functional form for $K_2(\varepsilon)$, which we shall find out momentarily.

First we note that for 1-D independent random variables, which correspond to $H = 1/2$, $h(\varepsilon, \tau) \sim -\ln\varepsilon$ (Gaspard and Wang, 1993). Therefore, $K_2(\varepsilon) \sim -\ln\varepsilon$. In fact, for any stationary noise process, irrespective of its correlation structure, we always have $C^{(m)}(\varepsilon) \sim \varepsilon^{-m}$, $\varepsilon \to 0$, therefore,

$$K_2(\varepsilon) \sim -\ln\varepsilon, \;\; \varepsilon \to 0 \quad (12)$$

Equation (12) is, however, not adequate for us to understand the scaling of $K_2(\varepsilon)$ on finite scales. To gain more insights, we resort to the rate distortion function or the Shannon-Kolmogorov (SK) entropy (Berger, 1971; Gaspard and Wang, 1993). It is thought to diverge with ε in the same way as the (ε, τ)-entropy and $K_2(\varepsilon)$ (Gaspard and Wang, 1993).

Suppose we wish to approximate the random signal $X(t)$ by $Z(t)$ according to

$$\rho(X, Z) = \lim_{T\to\infty} \frac{1}{T}\int_0^T \Big\langle [X(t) - Z(t)]^2 \Big\rangle dt \le \varepsilon^2 \qquad (13)$$

where <> denotes averaging. Equation (13) may be considered a partition of the phase space containing the random signal $X(t)$ by centering around $X(t)$. Denote the conditional probability density for Z given x by $q(z|x)$. The mutual information $I(q)$ between X and Z is a functional of $q(z|x)$,

$$I(q) = \int\int dxdz\, p(x)q(z|x)\ln[q(z|x)/q(z)]. \qquad (14)$$

The SK (ε, τ)-entropy is

$$H_{SK}(\varepsilon, \tau, T) = \mathrm{Inf}_{q\in Q(\varepsilon)} I(q) \qquad (15)$$

where $Q(\varepsilon)$ is the set of all conditional probabilities $q(z|x)$ such that Condition (13) is satisfied. The SK (ε, τ)-entropy per unit time is then

$$h_{SK}(\varepsilon, \tau) = \lim_{T\to\infty} H_{SK}(\varepsilon, \tau, T)/T \qquad (16)$$

For stationary Gaussian processes, $h_{SK}(\varepsilon, \tau)$ can be readily computed by the Kolmogorov formula (Berger, 1971; Kolmogorov, 1956). In the case of a discrete-time process, it reads

$$\varepsilon^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi} \min[\theta, \Phi(\omega)]d\omega \qquad (17)$$

$$h_{SK}(\varepsilon, \tau) = \frac{1}{4\pi}\int_{-\pi}^{\pi} \max\{0, \ln[\Phi(\omega)/\theta]\}d\omega \qquad (18)$$

where $\Phi(\omega)$ is the PSD of the process and θ is an intermediate variable.

We now evaluate the SK entropy for a popular model of $1/f^{2H-1}$ noise, the fractional Gaussian noise (fGn). It is a stationary Gaussian process with PSD $1/\omega^{2H-1}$. Since we are primarily interested in small ε, we may choose the intermediate variable $\theta \le \Phi(\omega)$. Let us denote $\Phi(\omega) = B(H)\omega^{1-2H}$, where $B(H)$ is a factor depending on H. When $H = 1/2$, it equals the variance of the noise $\sigma^2_{H=1/2}$. Using Equations (17) and (18), we immediately have

$$h_{SK}(\varepsilon) = A(H) - \ln\varepsilon \qquad (19)$$

where

$$A(H) = \frac{1-2H}{2}(\ln\pi - 1) + \frac{1}{2}\ln B(H) \qquad (20)$$

If we assume fGn of different H to have the same variance, then $\int_0^\pi \Phi(\omega)d\omega$ is a constant independent of H. $A(H)$ can then be written as

$$A(H) = \frac{1}{2}\ln\sigma^2_{H=1/2} + \frac{1}{2}\Big[\ln(2-2H) - (1-2H)\Big] \qquad (21)$$

$A(H)$ is maximal when $H = 1/2$. However, when H is not close to 0 or 1, the term $\frac{1}{2}[\ln(2-2H) - (1-2H)]$ is negligibly small, signifying that $h_{SK}(\varepsilon)$ cannot readily classify fGn of different H.

Since $h_{SK}(\varepsilon)$ and $K_2(\varepsilon)$ diverge in the same fashion (Gaspard and Wang, 1993), using Equation (12) to determine the prefactor, we have a scaling for finite ε

$$K_2(\varepsilon) \sim -\ln\,\varepsilon \qquad (22)$$

Combining Equations (22) and (11), we arrive at a fundamental bi-scaling law for $K_2^{(b_s)}(\varepsilon)$ for fractal time series:

$$K_2^{(b_s)}(\varepsilon) \sim (H-1)\,\ln\,b_s - \ln\,\varepsilon \qquad (23)$$

To verify the above bi-scaling law, and more importantly, to gain insights into the relative importance of the two scale parameters b_s and ε in MSE analysis, we numerically perform MSE analysis of fGn processes with different H. A few examples are shown in **Figures 1**, **2**. The computations are done with 2^{14} points and $m = 2$. We observe excellent bi-scaling relations, thus verifying Equation (23). Recalling our earlier comment that $K_2(\varepsilon)$ itself is not very useful for distinguishing fGn of different H, **Figure 2** clearly shows that the scaling $K_2^{(b_s)}(\varepsilon) \sim (H-1)\,\ln b_s$ can aptly separate fGn processes of different H. In fact, H values estimated from **Figure 2** are fully consistent the values of H chosen in simulating the fGn processes. This analysis thus has demonstrated the major advantage of the scale parameter b_s over ε for the study of fGn processes using MSE. It has also made it clear that MSE is a highly non-trivial extension of the sample entropy, and more generally, the correlation entropy $K_2(\varepsilon)$.

While Equation (23) is fundamental for MSE, it can also help us better understand the behavior of multivariate MSE, which is shown in numerical simulations to be almost constant for $1/f$ processes with $H = 1$, and decays in a well-defined fashion for white noise, where $H = 1/2$, and some randomized data derived from experimental data possibly with correlations (Ahmed and Mandic, 2011). The reason is very clear. For $1/f$ process, $H = 1$, and therefore, MSE or multivariate MSE does not vary with the scale parameter b_s. For white noise or some derived randomized data, $H = 1/2$, and therefore, MSE or multivariate MSE decays with the scale parameter b_s in a well-defined fashion,

$$K_2^{(b_s)}(\varepsilon) \sim -\frac{1}{2}\ln b_s, \quad or \quad b_s \sim e^{-2K_2(\varepsilon)}. \qquad (24)$$

One can readily check that the MSE curve for white noise shown in Ahmed and Mandic (2011) is fully consistent with the formula derived here.

3.2. Heart Rate Variability Data Analysis

As an important application of MSE, we analyze HRV data for the purpose of distinguishing healthy subjects from patients with

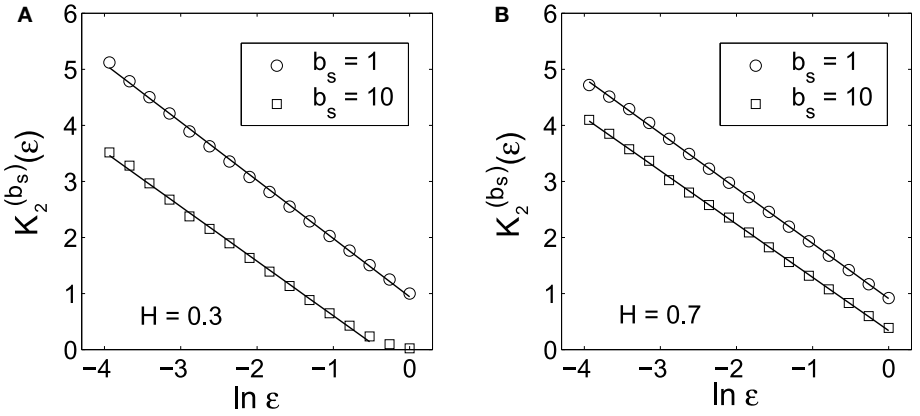


FIGURE 1 | $K_2^{(bs)}(\varepsilon)$ vs. $\ln \varepsilon$ curves corresponding to the original data ($b_S = 1$) and the smoothed data ($b_S = 10$) for fGn processes with (A) $H = 0.3$ and (B) $H = 0.7$. The slopes of the linear regression lines are very close to 1.

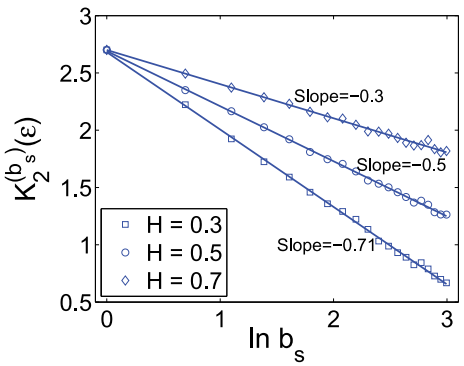


FIGURE 2 | $K_2^{(bs)}(\varepsilon)$ vs. $\ln b_S$ curves for fGn processes with different H values. The scale ε is chosen as 20% of the standard deviation of the corresponding fGn process. H value is estimated as 1 plus the slope of the curve.

CHF, a life-threatening condition. This is an important issue. We refer to (Hu et al., 2009, 2010) and references therein for the background. Note that part of the data examined here were analyzed in prior work (Ivanov et al., 1999; Barbieri and Brown, 2006), for the same purpose. We analyze all 33 datasets here. For ease of comparison, we take the first 3×10^4 points of both groups of HRV data for analysis. Note that based on different b_s parameter, MSE was not very good at separating the two groups (Hu et al., 2010). This instigated a debate on whether MSE was useful or not for analyzing HRV (Wessel et al., 2003; Nikulin and Brismar, 2004). To resolve this interesting debate, and more importantly, to satisfactorily separate the two groups of HRV data, we shall focus on the dependence of MSE on the scale parameter ε in the following discussions.

Since earlier studies find HRV data to be nonstationary, having $1/f$ spectrum with anti-persistent long-range correlations and multifractality (see Ivanov et al., 1999 and references therein), we analyze the increment processes of the HRV data. **Figure 3** shows $K_2(\varepsilon)$ vs. $\ln \varepsilon$ curves for the two groups of HRV data. We observe: (i) On small scales, $K_2(\varepsilon)$ vs. $\ln \varepsilon$ curves for both groups of HRV data show good scaling behavior. As a consequence, one can expect a scaling relation between $K_2^{(b_s)}(\varepsilon)$ and $\ln b_s$ (Equation 23). This is indeed so. The results, being very similar to that shown in **Figure 2**, are not shown here, however. (ii) The scaling of $K_2(\varepsilon)$ vs. $\ln \varepsilon$ is better and longer for the normal HRV data. (iii) As indicated by ε^* in the figure, the smallest scale resolvable by the HRV data of the healthy subjects is much larger than that of the diseased subjects.

We now discuss how to use MSE to distinguish the healthy subjects from patients with CHF. We have found (i) The curves $K_2^{(b_s)}(\varepsilon)$ vs. b_s averaged over all the subjects within the two groups are different, just as reported in Costa et al. (2005). However, such curves are not very useful for separating the two groups as a diagnostic tool, as pointed out in Nikulin and Brismar (2004). The fundamental reason is of course that the Hurst parameter H is not very effective in distinguishing healthy subjects from patients with HRV, as quantitatively analyzed in Hu et al. (2010). (ii) The smallest resolvable scale, ε^*, completely separates the healthy subjects from patients with CHF, as shown by **Figure 3**. Note the scale parameter ε is a generalization of the concept variance (or standard deviation). The observation made by Nikulin and Brismar (2004) that a variance-like parameter is better than MSE with varying block size parameter b_s in distinguishing healthy subjects from patients with HRV is most appropriately interpreted as the following: the parameter b_s is less important than the scale parameter ε. This is somewhat the opposite of the case for $1/f$ noise analyzed in the last section.

To more clearly see how much more advantageous ε is over b_s in distinguishing healthy subjects from patients with HRV, we examine how the scaling $K_2(\varepsilon) \sim -\ln \varepsilon$ can be used for this purpose. We have found that the errors obtained by linearly fitting the $K_2(\varepsilon)$ vs. $\ln \varepsilon$ curves of **Figure 3** are much smaller for the normal HRV data than for those of CHF patients and also can completely separate the healthy subjects from patients with CHF. This is shown in **Figure 4**. Therefore, the scale parameter ε is indeed more important than b_s.

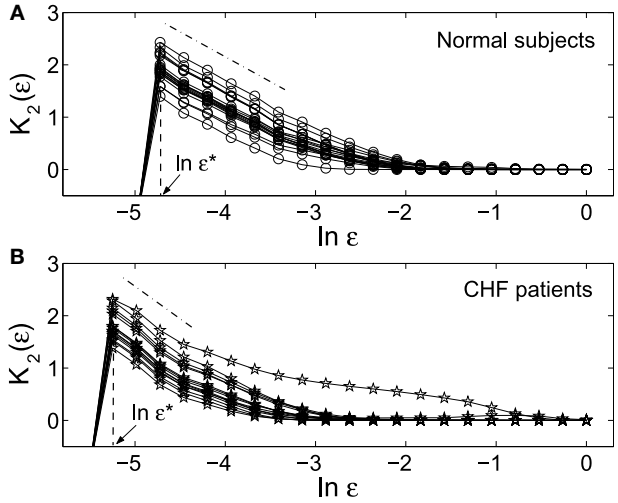


FIGURE 3 | $K_2(\varepsilon)$ vs. ln ε curves for the HRV data of (A) 18 normal subjects and (B) 15 patients with CHF. Each curve corresponds to one subject. The computations were done with 3×10^4 points and $m = 5$. ε^* indicates the smallest scale resolvable by the data.

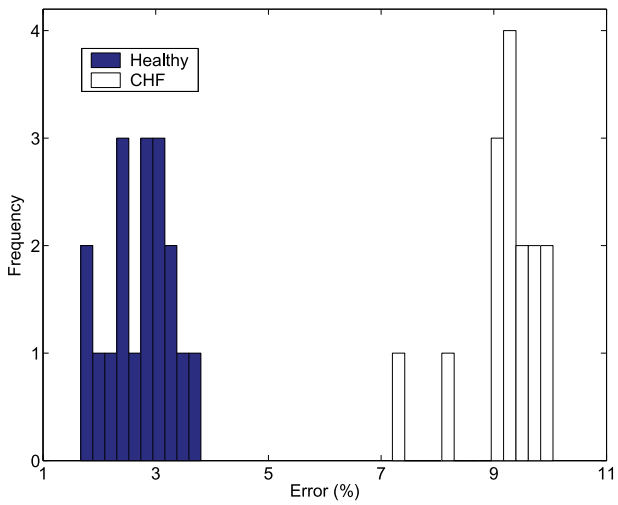


FIGURE 4 | The frequency of the percentage of errors obtained by linearly fitting the $K_2(\varepsilon)$ vs. ln ε curves in Figure 3 with 6 points starting from ε^* for the healthy and diseased subjects.

3.3. Epileptic Seizure Detection Through MSE of EEG

Epilepsy is a common and debilitating brain disorder. It is characterized by intermittent seizures. During a seizure, the normal activity of the central nervous system is disrupted. The concrete symptoms include abnormal running/bouncing fits, clonus of face and forelimbs, or tonic rearing movement as well as simultaneous occurrence of transient EEG signals such as spikes, spike and slow wave complexes or rhythmic slow wave bursts. Clinical effects may include motor, sensory, affective, cognitive, automatic and physical symptomatology. To make medications effective, timely detection of seizure is very important. In the

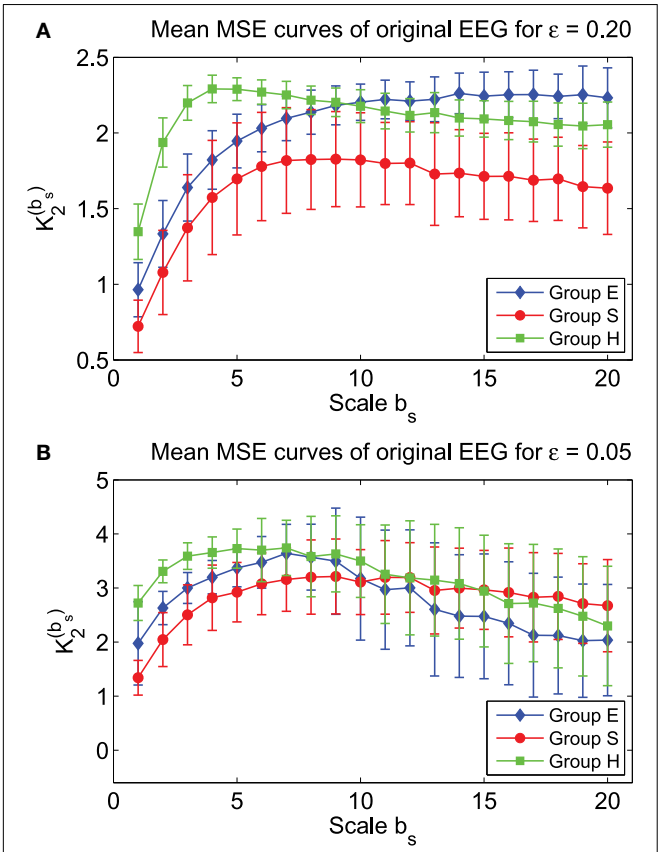


FIGURE 5 | Mean MSE curves for the 3 EEG groups with (A) ε = 0.2 and (B) ε = 0.05.

past several decades, considerable efforts have been made to detect/predict seizures through nonlinear analysis of EEGs. For a list of the major nonlinear methods proposed for seizure detection, we refer to Gao and Hu (2013) and references therein. In particular, the three groups of EEG data analyzed here, H (healthy), E (epileptic subjects during a seizure-free interval), and S (epileptic subjects during seizure), were examined by adaptive fractal analysis (Gao et al., 2011c) and scale-dependent Lyapunov exponent (Gao et al., 2012), and excellent classification was achieved.

To examine how well MSE characterizes the three groups of EEG data, we have plotted in **Figure 5** the mean MSE curves for the three groups, for two parameter values of the phase space scale, ε. We observe that they separate very well. Indeed, statistical test shows that the separations are significant. In particular, for the scale parameter in the phase space $\varepsilon = 0.2$, the MSE curve for the S group lies well below the other 2 curves. One may be tempted to equate this as smaller complexity of the seizure EEG. However, such an interpretation is informative only relative to the specific ε chosen here, which is 0.2. When $\varepsilon = 0.05$, the red curve for seizure EEG actually lie above the other 2 curves for larger b_s. In fact, if one can pause a moment and think twice, one would realize that such interpretations are not too helpful for clinical applications, since MSE can vary substantially within and across the groups.

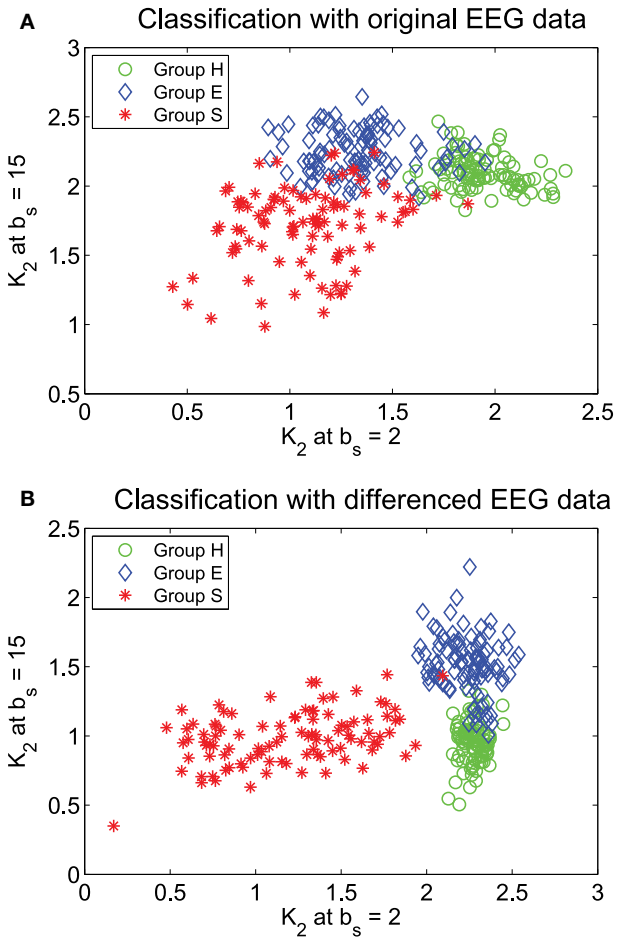


FIGURE 6 | Classification of the 3 EEG groups using features from the MSE curves: (A) the original data and (B) the differenced data.

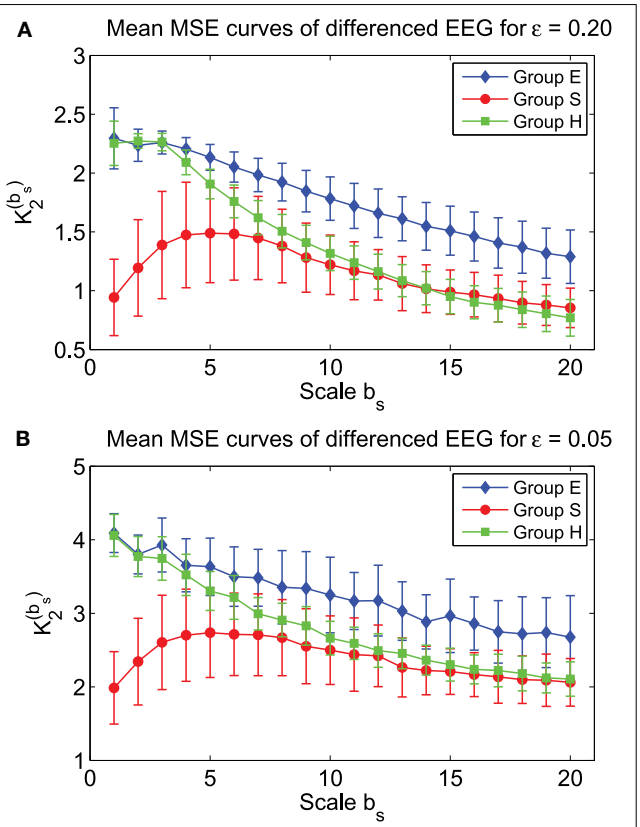


FIGURE 7 | Mean MSE curves for the differenced data of the 3 EEG groups with (A) ε = 0.2 and (B) ε = 0.05.

We have tried to use MSE at specific b_s values to classify the three groups of EEG. Guided by the mean MSE curves in **Figure 5**, we have found that when $\varepsilon = 0.2$, if only two b_s can be used, then $b_2 = 2$ and 15 are the optimal values. The result of the classification is shown in **Figure 6A**. We observe that there are some overlaps between groups H (healthy) and E (epileptic subjects during a seizure-free interval), as well as E and S (epileptic subjects during seizure). Intuitively, this is reasonable. Overall, the classification is not very satisfactory. How may we improve the accuracy of the classification?

Recall that in fractal scaling analysis of EEG, EEG data are found to be equivalent to random walk processes, but not noise or increment processes (Gao et al., 2011c). The latter amounts to a differentiation of the random walk processes. Since the basic scaling law derived here is for noise or increment process, but not for random walk processes, it suggests us to try to compute MSE from the differenced data of EEG, defined by $y_i = x_i - x_{i-1}$, where x_i is the original EEG signal. The mean MSE curves for the differenced data of EEG are shown in **Figure 7**, again for two ε values. We observe that the separation between the mean MSE curves becomes wider. Indeed, classification of the 3 EEG groups now is much improved, as shown in **Figure 6B**. It should be noted however that the accuracy of the classification is still slightly worse than using other methods, such as adaptive fractal analysis (Gao et al., 2011c) and scale-dependent Lyapunov exponent (Gao et al., 2012).

4. Conclusion and Discussion

To better understand MSE, we have derived a fundamental bi-scaling relation for the MSE analysis. While MSE analysis normally only focuses on the scale parameter b_s with ε more or less arbitrarily chosen, our analysis of fGn and HRV data clearly demonstrates that both scale parameters are important—in the case of HRV analysis, the ε is more important, while in the case of $1/f$ noise, the b_s parameter is more important. In fact, we have shown (Hu et al., 2010) that MSE, when used with ε fixed, is not very effective in distinguishing healthy subjects from patients with HRV. The accuracy achieved when we focus on the scaling of $K_2(\varepsilon) \sim -\ln \varepsilon$ is not only much higher, but also comparable to that using the scale-dependent Lyapunov exponent (SDLE) (Gao et al., 2006a, 2007, 2013), as reported by Hu et al. (Hu et al., 2010). The fundamental reason of course is that SDLE has a similar scaling as $K_2(\varepsilon) \sim -\ln \varepsilon$.

We have also computed MSE for the original as well as the differenced data of the three EEG groups, H (healthy), E (epileptic subjects during a seizure-free interval), and S (epileptic

subjects during seizure), and found that mean MSE curves for the three groups are well separated. The classification of the 3 EEG groups using MSE at two specific scale parameters b_s is reasonably good, and is better for the differenced data than for the original EEG data. This strongly suggests that EEG data are like random walk processes. However, even with the differenced data of EEG, the classification is still not as accurate as using adaptive fractal analysis (Gao et al., 2011c) and scale-dependent Lyapunov exponent (Gao et al., 2011a). One of the reasons for this inferiority lies in the difference in the range of scales covered by these three multiscale methods. Adaptive fractal analysis and scale-dependent Lyapunov exponent both cover the entire range of scales presented in the EEG data. However, with the length of the EEG data, which is only 4097 points for each data set, MSE can only cover a moderate range of scales, with the largest b_s only around 20, since with $b_s = 20$, the smoothed data is already only 200 points long. Our analysis here has raised an important question: how do we use MSE to analyze short data? We conjecture that it may be beneficial to focus on the scaling of $K_2(\varepsilon) \sim -\ln \varepsilon$, or develop new smoothing schemes, by introducing a parameter equivalent to $1/b_s$ but without sacrificing the length of the smoothed data.

Acknowledgments

One of the authors (JG) is grateful for the generous support by National Institute for Mathematical and Biological Synthesis (NIMBIOS) at the University of Tennessee to attend Heart Rhythm Disorders Investigative Workshop.

References

Ahmed, M. U., and Mandic, D. P. (2011). Multivariate multiscale entropy: a tool for complexity analysis of multichannel data. *Phys. Rev. E* 84:061918. doi: 10.1103/PhysRevE.84.061918

Barbieri, R., and Brown, E. N. (2006). Analysis of heartbeat dynamics by point process adaptive filtering. *IEEE Trans. Biomed. Eng.* 53, 4–12. doi: 10.1109/TBME.2005.859779

Berger, T. (1971). *Rate Distortion Theory*. Englewood Cliffs, NJ: Prentice-Hall.

Box, G. E. P., and Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control, 2nd Edn.* San Francisco, CA: Holden-Day.

Cao, H. Q., Lake, D. E., Ferguson, J. E., Chisholm, C. A., Griffin, M. P., and Moorman, J. R. (2006). Toward quantitative fetal heart rate monitoring. *IEEE Trans. Biomed. Eng.* 53, 111–118. doi: 10.1109/tbme.2005.859807

Cohen, A., and Procaccia, I. (1985). Computing the Kolmogorov entropy from time series of dissipative and conservative dynamical systems. *Phys. Rev. A* 31, 1872–1882. doi: 10.1103/PhysRevA.31.1872

Costa, M., Goldberger, A. L., and Peng, C. K. (2005). Multiscale entropy analysis of biological signals. *Phys. Rev. E* 71:021906. doi: 10.1103/PhysRevE.71.021906

Cox, D. R. (1984). "Long-range dependence: a review," in *Statistics: An Appraisal*, eds H. A. David and H. T. Davis (Ames: The Iowa State University Press), 55–74.

Gao, J. B., Cao, Y. H., Tung, W. W., and Hu, J. (2007). *Multiscale Analysis of Complex Time Series — Integration of Chaos and Random Fractal Theory, and Beyond*. Hoboken, NJ: Wiley.

Gao, J. B., Hu, J., Tung, W. W., Cao, Y. H., Sarshar, N., and Roychowdhury, V. P. (2006a). Assessment of long range correlation in time series: how to avoid pitfalls. *Phys. Rev. E* 73:016117. doi: 10.1103/PhysRevE.73.016117

Gao, J. B., Hu, J., Tung, W. W., and Cao, Y. H. (2006b). Distinguishing chaos from noise by scale-dependent Lyapunov exponent. *Phys. Rev. E* 74:066204. doi: 10.1103/PhysRevE.74.066204

Gao, J. B., Hu, J., Buckley, T., White, K., Hass, C. (2011a). Shannon and Renyi entropies To classify effects of mild traumatic brain injury on postural sway. *PLoS ONE* 6:e24446. doi: 10.1371/journal.pone.0024446

Gao, J. B., Hu, J., and Tung, W. W. (2011b). Complexity measures of brain wave dynamics. *Cogn. Neurodynamics* 5, 171–182. doi: 10.1007/s11571-011-9151-3

Gao, J. B., Hu, J., and Tung, W. W. (2011c). Facilitating joint chaos and fractal analysis of biosignals through nonlinear adaptive filtering. *PLoS ONE* 6:e24331. doi: 10.1371/journal.pone.0024331

Gao, J. B. Hu, J. and Tung, W. W. (2012). Entropy measures for biological signal analysis. *Nonlin. Dynamics* 68, 431–444. doi: 10.1007/s11071-011-0281-2

Gao, J. B., Gurbaxani, B. M., Hu, J., Heilman, K. J., Emauele, V. A., Lewis, G. F. et al. (2013). Multiscale analysis of heart rate variability in nonstationary environments. *Front. Comput. Physiol. Med.* 4:119. doi: 10.3389/fphys.2013.00119

Gao, J. B., and Hu, J. (2013). Fast monitoring of epileptic seizures using recurrence time statistics of electroencephalography. *Front. Comput. Neurosci.* 7:122. doi: 10.3389/fncom.2013.00122

Gaspard, P., and Wang, X. J. (1993). Noise, chaos, and (ε, τ)-entropy per unit time. *Phys. Rep.* 235, 291–343. doi: 10.1016/0370-1573(93)90012-3

Grassberger, P., and Procaccia, I. (1983). Estimation of the Kolmogorov entropy from a chaotic signal. *Phys. Rev. A* 28, 2591–2593. doi: 10.1103/PhysRevA.28.2591

Hu, J., Gao, J. B., and Tung, W. W. (2009). Characterizing heart rate variability by scale-dependent Lyapunov exponent. *Chaos* 19, 028506. doi: 10.1063/1.3152007

Hu, J., Gao, J. B., Tung, W. W., and Cao, Y. H. (2010). Multiscale analysis of heart rate variability: a comparison of different complexity measures. *Ann. Biom. Eng.* 38, 854–864. doi: 10.1007/s10439-009-9863-2

Istenic, R., Kaplanis, P. A., Pattichis, C. S., Zazula, D. (2010). Multiscale entropy-based approach to automated surface EMG classification of neuromuscular disorders. *Medi. Biol. Eng. Comput.* 48, 773–781. doi: 10.1007/s11517-010-0629-7

Ivanov, P. C., Amaral, L. A. N., Goldberger, A. L., Havlin, S., Rosenblum, M. G., Struzik, Z. R. (1999). Multifractality in human heartbeat dynamics. *Nature* 399, 461–465. doi: 10.1038/20924

Kolmogorov, A. N. (1956). On the Shannon theory of information transmission in the case of continuous signals. *IRE Trans. Inf. Theory* 2, 102–108. doi: 10.1109/TIT.1956.1056823

Li, D. A., Li, X. L., Liang, Z. H., Voss, L. J., and Sleigh, J. W. (2010). Multiscale permutation entropy analysis of EEG recordings during sevoflurane anesthesia. *J. Neural Eng.* 7:046010. doi: 10.1088/1741-2560/7/4/046010

Mandelbrot, B. B. (1982). *The Fractal Geometry of Nature*. San Francisco, CA: Freeman.

Mizuno, T., Takahashi, T., Cho, R. Y., Kikuchi, M., Murata, T., Takahashi, K., et al. (2010). Assessment of EEG dynamical complexity in Alzheimer's disease using multiscale entropy *Clin. Neurophysil.* 121, 1438–1446. doi: 10.1016/j.clinph.2010.03.025

Nardelli, M., Valenza, G., Cristea, I. A., Gentili, C., Cotet, C., David, D., et al. (2015). Characterizing psychological dimensions in non-pathological subjects through autonomic nervous system dynamics. *Front. Comput. Neurosci.* 9:37. doi: 10.3389/fncom.2015.00037

Nikulin, V. V., and Brismar, T. (2004). Comment on "Multiscale entropy analysis of complex physiologic time series." *Phys. Rev. Lett.* 92:089803. doi: 10.1103/PhysRevLett.92.089803

Norris, P. R., Anderson, S. M., Jenkins, J. M., Williams, A. E., and Morris, J. A. Jr. (2008). Heart rate multiscale entropy at three hours predicts hospital mortality in 3,154 Trauma patients. *Shock* 30, 17–22. doi: 10.1097/SHK.0b013e318164e4d0

Richman, J. S., and Moorman, J. R. (2000). Physiological time-series analysis using approximate entropy and sample entropy. *Am. J. Physiol. Heart Circ. Physiol.* 278, H2039–H2049. Available online at: http://ajpheart.physiology.org/content/278/6/H2039

Wessel, N., Schirdewan, A., and Kurths, J. (2003). Intermittently decreased beat-to-beat variability in congestive heart failure *Phys. Rev. Lett.* 91:119801. doi: 10.1103/PhysRevLett.91.119801

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.

frontiers in Computational Neuroscience

METHODS
published: 17 July 2015
doi: 10.3389/fncom.2015.00094

A three-dimensional mathematical model for the signal propagation on a neuron's membrane

*Konstantinos Xylouris * and Gabriel Wittum*

Department of Simulation and Modeling, Faculty of Informatics, Goethe Center for Scientific Computing, Goethe University Frankfurt, Frankfurt am Main, Germany

In order to be able to examine the extracellular potential's influence on network activity and to better understand dipole properties of the extracellular potential, we present and analyze a three-dimensional formulation of the cable equation which facilitates numeric simulations. When the neuron's intra- and extracellular space is assumed to be purely resistive (i.e., no free charges), the balance law of electric fluxes leads to the Laplace equation for the distribution of the intra- and extracellular potential. Moreover, the flux across the neuron's membrane is continuous. This observation already delivers the three dimensional cable equation. The coupling of the intra- and extracellular potential over the membrane is not trivial. Here, we present a continuous extension of the extracellular potential to the intracellular space and combine the resulting equation with the intracellular problem. This approach makes the system numerically accessible. On the basis of the assumed pure resistive intra- and extracellular spaces, we conclude that a cell's out-flux balances out completely. As a consequence neurons do not own any current monopoles. We present a rigorous analysis with spherical harmonics for the extracellular potential by approximating the neuron's geometry to a sphere. Furthermore, we show with first numeric simulations on idealized circumstances that the extracellular potential can have a decisive effect on network activity through ephaptic interactions.

Keywords: models, theoretical, ephaptic coupling, dipole effect, detailed 3D-modeling, 3D-modeling, cable equation

OPEN ACCESS

Edited by:
Tobias Alecio Mattei, Brain & Spine Center - InvisionHealth - Kenmore Mercy Hospital, USA

Reviewed by:
Ingo Bojak, University of Reading, UK
Le Wang, Boston University, USA
Xin Tian, Tianjin Medical University, China
Mikhail Katkov, Weizmann Institute of Science, Israel

****Correspondence:***
Konstantinos Xylouris, Department of Simulation and Modeling, Faculty of Informatics, Goethe Center for Scientific Computing, Goethe University Frankfurt, Kettenhofweg 139, Frankfurt am Main 60325, Germany
konstantinos.xylouris@gcsc.uni-frankfurt.de

Received: *17 March 2015*
Accepted: *02 July 2015*
Published: *17 July 2015*

Citation:
Xylouris K and Wittum G (2015) A three-dimensional mathematical model for the signal propagation on a neuron's membrane. Front. Comput. Neurosci. 9:94. doi: 10.3389/fncom.2015.00094

Introduction

The membrane potential belongs to the most important quantities of a neuron. Its function of time and space describes neuronal activity. It is a voltage across the membrane defined by the difference between the intra- and extracellular potential.

Since the neuron is embedded in ionic milieus, potential gradients in the off-membrane spaces result in electric fluxes, which are conserved according to the first principles. This conservation law is the basis of the standard cable equation which describes the unfolding and propagation of an action potential (Rall, 1962, 1964; Scott, 1975) very efficiently. The standard cable equation maps a neuron to a tree of lines, each of which corresponds to a cylindric compartment with mean diameter. On these structures, it computes the evolution of the membrane potential according to its diffusion equation.

The resulting extracellular potentials can be theoretically computed with the line source method (Holt and Koch, 1999; Gold et al., 2006), once the transmembrane currents have been determined with the aid of the cable equation's solution.

These extracellular potentials in turn can be exploited to examine ephaptic feedbacks on other neurons (Holt and Koch, 1999). Indeed, the distribution of the extracellular potential can elicit transmembrane currents which may have decisive effects on the membrane potential of neighboring cells (Anastassiou et al., 2011; Buzsáki et al., 2012).

The goal of the current paper is to develop and implement an integrated three-dimensional model which synchronously captures both quantities, the membrane potential and the extracellular potential, during activity and which uses the neuron's geometry as it is instead of reducing it to cylindric compartments. The aim of such a model is to deepen the knowledge in signal processing and to carry out simulations on small networks of realistic neurons while having all these influences in action.

The work of Voßen et al. (2007) did a first step in the development of a generalized cable equation. It was built on the principle of the continuity of electric fluxes. Although the core model with the intra-, extracellular and membrane potential was correctly derived, the subsequent approach used to couple these unknowns and to solve them numerically resulted in major difficulties. The limit case to the standard cable equation evoked greater challenges and the simulations themselves were restricted to a very small time period of hundreds of micro seconds on a small part of a passive membrane.

The study of Xylouris et al. (2010) used a more direct approach for the coupling and generalized the existing model to active membranes. Nonetheless, although it was capable to reproduce action potentials, it still lacked in many characteristics of the signal processing, like the width of the propagating signal, the waveform of extracellular potential at activity and the computation on more complicated geometries. Indeed, computations on more complicated geometries diverged numerically. Furthermore, the membrane potential's defining equation in Xylouris et al. (2010) was the transmembrane current, which contains the time derivative of the membrane's capacitive property as only differential operator. The membrane potential's propagation was provided indirectly through the difference between the intra- and extracellular potential- thus making it actually hard to expect correct results for the spacial distribution. Moreover, as consequence, it produced vanishing transmembrane currents causing zero extracellular potentials and zero ephaptic interactions. This is why, the solving procedure with this direct coupling was of little use.

This paper introduces a completely new coupling of the unknowns. Therein, the defining equation for the membrane potential contains its own spacial differential operator. For the first time, we could carry out simulations on three-dimensionally resolved ideal neurons and on a small network of cells. This description, furthermore, allows for a proof that the extracellular potential distributes in the extracellular space like a current multipole. It will show that the only current monopole for a neuron exists at rest.

Model

Three-Dimensional Cable Equation

Let Ω_{in} and Ω_{out} be domains in $\mathbb{R}^3$ denoting the neuron's intra- and extracellular space, respectively, and $\bar{\Omega}_{\text{in}} \cap \bar{\Omega}_{\text{out}} = \Gamma$ the membrane, a two dimensional manifold embedded in $\mathbb{R}^3$. Let $\Omega = \Omega_{\text{in}} \cup \Omega_{\text{out}} = \mathbb{R}^3$ be the whole space. Let, furthermore, Φ_{in}, Φ_{out}, and V_m be the intra-, extracellular, and membrane potential, respectively. Φ will represent either Φ_{in} or Φ_{out}.

The quantities σ_{in} and σ_{out} denote the intra- and extracellular conductivities, respectively. The normal $n_{\text{in}\to\text{out}}$ is the normal on the membrane Γ pointing from the intracellular space to the extracellular. We will need this quantities in order to define the fluxes. For the active transmembrane flux, we will just consider the Hodgkin–Huxley model for the sake of a simpler writing. There we have the sodium conductivity g_{Na^+}, the potassium conductivity g_{K^+} and the leakage conductivity g_{L}. The quantities E_{Na^+}, E_{K^+}, and E_{L} denote the reversal potentials of the indexed ions. The gating parameters n, m, h obey ordinary differential equations (Hodgkin and Huxley, 1952) and calibrate how much of the maximal possible ionic flux passes through the channel.

Considering the non-membrane conductivity ($\approx 3\frac{mS}{cm}$) (López-Aguado et al., 2001) and the dielectricity of water (≈ 1), Gary Holt demonstrated in his Ph.D. Thesis (Holt, 1997) that a possible non-membrane capacitor would discharge with a time constant of approximately 3 ns. Because this time scale is much faster than the one of the phenomena considered—the fast channel dynamics react on a μs-time scale—, it appears as good approximation to assume no capacitive properties for the non-membrane spaces ($\rho = 0$ in Ω_{in} and Ω_{out}). Indeed, this is the basis of the derivation for the three dimensional cable equation. In addition, we will assume to have time invariant magnetic fields ($\frac{d\vec{B}}{dt} = 0$). Then, Gauß's and Faraday's law satisfy root equations in the intra- and extracellular space, so that the conservative electric field can be expressed with the aid of a potential gradient. Combining this gradient with Gauß's law immediately leads to the Laplace equation for the potentials in the non-membrane spaces.

$$\nabla \cdot \vec{E} = \frac{\rho}{\epsilon\epsilon_0} \overset{(\rho=0)}{=} 0, \tag{1}$$

$$\nabla \times \vec{E} = -\frac{d\vec{B}}{dt} \overset{!}{=} 0, \tag{2}$$

$$\Rightarrow \vec{E} = -\nabla\Phi, \tag{3}$$

$$\Rightarrow -\Delta\Phi = 0. \tag{4}$$

The constants ϵ_0 and ϵ are the dielectricities in vacuum and material, respectively.

Because of flux continuity, the flux across the membrane is continuous and must correspond to the flux emerging from the membrane dynamics [denoted with $j_{\text{all}}(V_m)$]. Hence,

$$-\sigma_{\text{in}}\nabla\Phi_{\text{in}} \cdot n_{\text{in}\to\text{out}} = -\sigma_{\text{out}}\nabla\Phi_{\text{out}} \cdot n_{\text{in}\to\text{out}} = j_{\text{all}} \quad \text{on } \Gamma. \tag{5}$$

With this boundary condition in mind, we arrive at the three-dimensional cable equation (**Figure 1**):

$$-\Delta\Phi_{\text{out}} = 0 \qquad \text{in } \Omega_{\text{out}}, \tag{6}$$

$$-\Delta\Phi_{\text{in}} = 0 \qquad \text{in } \Omega_{\text{in}}, \tag{7}$$

$$V_m = \Phi_{\text{in}} - \Phi_{\text{out}} \qquad \text{on } \Gamma. \tag{8}$$

The flux j_{all} contains all fluxes passing the membrane. Considering just the Hodgkin–Huxley model and some additional stimulus, it looks like:

$$j_{\text{all}} = c_m \frac{dV_m}{dt} + m^3 h g_{\text{Na}^+}(V_m - E_{\text{Na}^+}) + n^4 g_{\text{K}^+}(V_m - E_{\text{K}^+}) + g_{\text{L}}(V_m - E_{\text{L}}) + j_{\text{Stm}}. \tag{9}$$

Since it is possible to have different dynamics on each region of the neuronal membrane, we furthermore introduce the following δ-functions

$$\delta_{\text{dend}}(x) = \left\{ \begin{array}{ll} 1 & \text{on the dendrite} \\ 0 & \text{else} \end{array} \right\},$$

$$\delta_{\text{active}}(x) = \left\{ \begin{array}{ll} 1 & \text{on the soma or nodes of Ranvier} \\ 0 & \text{else} \end{array} \right\},$$

$$\delta_{\text{syn}}(x) = \left\{ \begin{array}{ll} 1 & \text{on the postsynaptic density} \\ 0 & \text{else} \end{array} \right\},$$

$$\delta_{\text{stim}}(x) = \left\{ \begin{array}{ll} 1 & \text{on the stimulation area} \\ 0 & \text{else} \end{array} \right\}.$$

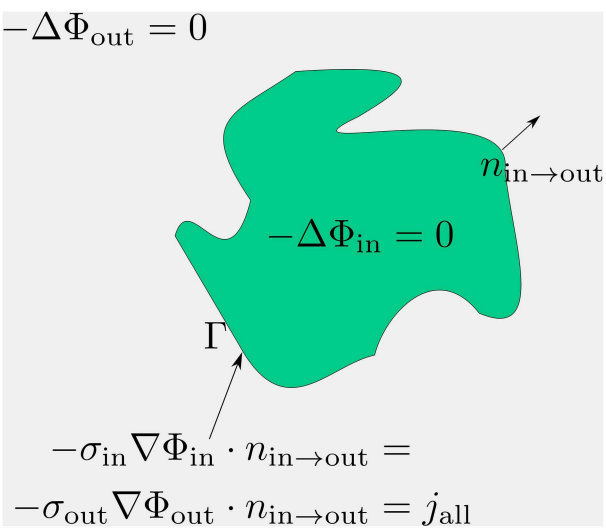


FIGURE 1 | Compact scheme of the three-dimensional cable equation. Assuming pure resistivity for the non-membrane spaces, the respective potentials distribute according to the Laplace equation therein. These Laplace problems fulfill at the membrane an interface-condition which complies with the conservation of fluxes. The emerging flux from the potential equals the total transmembrane flux, denoted with j_{all}. Within j_{all} all transmembrane currents are accumulated: capacitive, channel, any stimulation or synaptic currents.

With the help of these δ-functions, we can define a more refined transmembrane flux considering where it precisely occurs.

We define

$$j_{\text{HH}}(n, m, h, V_m) = m^3 h g_{\text{Na}^+}(V_m - E_{\text{Na}^+}) + n^4 g_{\text{K}^+}(V_m - E_{\text{K}^+}) + g_{\text{L}}(V_m - E_{\text{L}}). \tag{10}$$

The synaptic activity is simply modeled with the aid of a modified Heaviside function $H(x, t)$. This function should be one as soon as the membrane potential at the pre-synapse exceeds a certain value, say 2 mV, and it remains one for the time the synapse is active regardless of the presynaptic membrane potential. Additional activation at the pre-synapse should integrated by the synaptic function $\alpha(V_m|_{\text{pre}}, t)$

$$j_{\text{syn}}(V_m|_{\text{pre}}, t) = H(V_m|_{\text{pre}}, t) \cdot \alpha(V_m|_{\text{pre}}, t), \tag{11}$$

where $V_m|_{\text{pre}}$ is the membrane potential at the presynaptic terminal.

Then the refined total transmembrane current has the form:

$$j_{\text{all}}(x, V_m) = c_m \frac{dV_m}{dt} + \delta_{\text{active}}(x) j_{\text{HH}}(n, m, h, V_m) + \delta_{\text{stim}}(x) j_{\text{Stm}}(t) + \delta_{\text{syn}}(x) j_{\text{syn}}(V_m|_{\text{pre}}, t). \tag{12}$$

Numeric Model

The three dimensional cable equation (Equations 6–8) is a non-symmetric system (Φ_{in} does not couple with Φ_{out} the same way as Φ_{out} with Φ_{in}) of PDEs which couples two Laplace equations in the intra- and extracellular space with the transmembrane flux. This flux depends on the membrane potential. One difficulty in solving this system is the coupling of the membrane potential, which lives on a lower dimensional manifold, with the quantities, which live in full space. Since the discretization of this system is carried out with the help of integrals, the lower dimensional quantity cannot be measured the same way as the quantities in space (because the space integrals do not see it at all). In order to get rid of this particularity, we will extend the membrane potential, which is defined by the difference between the intra- and extracellular potential ($V_m = \Phi_{\text{in}} - \Phi_{\text{out}}$) on the membrane, to the intracellular space. To that end, we extend the extracellular potential to the intracellular space and combine its extension with the intracellular potential equation. So, we arrive at a problem for the membrane potential in the intracellular space.

Because $V_m = \Phi_{\text{in}} - \Phi_{\text{out}}$ on the membrane Γ, we will extend Φ_{out} to the intracellular space continuously so that the following identity holds. Let this extension be denoted with Φ_{out}^{IN}:

$$V_m = \Phi_{\text{in}} - \Phi_{\text{out}} = \Phi_{\text{in}} - \Phi_{\text{out}}^{IN} \qquad \text{on } \Gamma, \tag{13}$$

$$\Rightarrow \Phi_{\text{out}} = \Phi_{\text{out}}^{IN} \qquad \text{on } \Gamma. \tag{14}$$

At this point we have some freedom to choose the right hand side of the extracellular potential extension equation. We choose

it to be zero. Then it can be easily combined with the intracellular problem (Equation 7), which is a Lapalcian, too. We have

$$\begin{aligned} -\Delta\Phi_{\text{out}}^{IN} &= 0 && \text{in } \Omega_{\text{in}}, && (15)\\ \Phi_{\text{out}}^{IN} &= \Phi_{\text{out}} && \text{on } \Gamma, \\ \Rightarrow -\Delta(\Phi_{\text{in}} - \Phi_{\text{out}}^{IN}) = -\Delta V_m &= 0 && \text{in } \Omega_{\text{in}}, && (16)\\ -\sigma_{\text{in}}\nabla V_m \cdot n_{\text{in}\to\text{out}} &= j_{\text{all}}(V_m) \\ &\quad + \sigma_{\text{in}}\nabla\Phi_{\text{out}}^{IN} \cdot n_{\text{in}\to\text{out}} \\ & && \text{on } \Gamma. \end{aligned}$$

Thus, instead of solving the system (Equations 6–8) we solve (**Figure 2**):

$$\begin{aligned} -\Delta\Phi_{\text{out}} &= 0 && \text{in } \Omega_{\text{out}}, && (17)\\ -\sigma_{\text{out}}\nabla\Phi_{\text{out}} \cdot n_{\text{in}\to\text{out}} &= j_{\text{all}}(V_m) && \text{on } \Gamma, \\ -\Delta\Phi_{\text{out}}^{IN} &= 0 && \text{in } \Omega_{\text{in}}, && (18)\\ \Phi_{\text{out}}^{IN} &= \Phi_{\text{out}} && \text{on } \Gamma, \\ -\Delta V_m &= 0 && \text{in } \Omega_{\text{in}}, && (19)\\ -\sigma_{\text{in}}\nabla V_m \cdot n_{\text{in}\to\text{out}} &= j_{\text{all}}(V_m) \\ &\quad + \sigma_{\text{in}}\nabla\Phi_{\text{out}}^{IN} \cdot n_{\text{in}\to\text{out}} \\ & && \text{on } \Gamma. \end{aligned}$$

For referencing reasons, we will call the additional current, which is considered in the boundary condition of the membrane potential equation (Equation 19), as ephaptic current

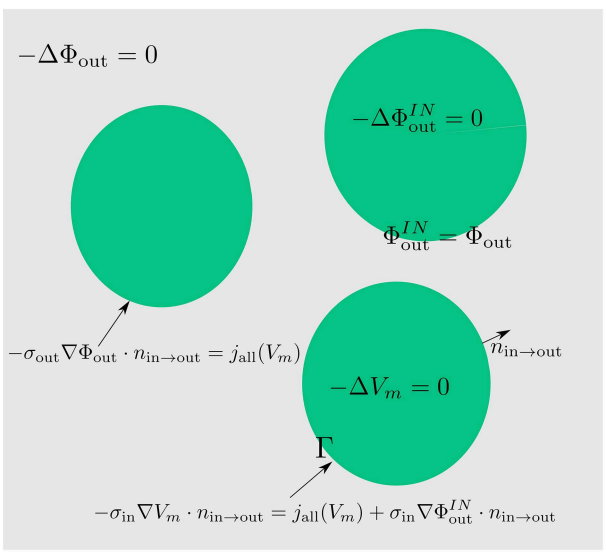


FIGURE 2 | By extending the extracellular potential to the intracellular space (green) continuously, an extension of the membrane potential into the intracellular space is established. By means of this trick we obtain a coupling between the extracellular and the membrane potential which can be directly used for numerics and simulations.

$$j_{\text{eph}} := \sigma_{\text{in}}\nabla\Phi_{\text{out}}^{IN} \cdot n_{\text{in}\to\text{out}}. \quad (20)$$

Numeric Discretization and Procedures

In space, we discretize this system (Equations 17–19) with the finite volume method (Versteeg and Malalasekera, 2007). This method guarantees the local conservation of fluxes. This is necessary, because the model has been derived on this principle. Furthermore, important characteristics of the solution, as we will see in the following section depend on this conservation. In time, an implicit method is used while the non-linearity is resolved with the Newton method.

Similarly to the finite element method, we discretize the domain Ω with volume elements, for example tetrahedrals, whose edge points and edges form the grid Ω_h, and we approximate the unknown functions (in our case V_m, Φ_{out}, and Φ_{out}^{IN}) with a linear combination of shape functions. Our shape functions $b_j(x)$ have the property to be continuous and linear on each elements ($j = 0, \ldots, \#\Omega_h = N$). They are as many as our grid points ($\#\Omega_h = N$) and are uniquely determined by the following defining conditions

$$b_j(x_k) = \delta_{jk} \quad x_j \in \Omega_h \quad (21)$$

$$b_j(x) \text{ is continuous and linear on each element} \quad (22)$$

We represent our unknown functions with these

$$V_m(x,t) = \sum_{j=0}^{N} v_{m_j}^t b_j(x), \quad (23)$$

$$\Phi_{\text{out}}(x,t) = \sum_{j=0}^{N} \phi_{\text{out}_j}^t b_j(x), \quad (24)$$

$$\Phi_{\text{out}}^{IN}(x,t) = \sum_{j=0}^{N} \phi_{\text{out}_j}^{IN,t} b_j(x). \quad (25)$$

Purpose of the discretization schema is to establish linear systems out of the differential Equations (17–19) which uniquely determine the unknowns coefficients $v_{m_j}^t, \phi_{\text{out}_j}^t, \phi_{\text{out}_j}^{IN,t}$ of these linear combinations. The upper index t should indicate that these coefficients are time dependent.

For the finite volume method, we need to construct a so called dual grid, which arises from the domain discretization and which is used in order to discretize the differential space operators. We call the elements of the dual grid control volumes. The volume elements of the dual grid are defined by the edge points which correspond to the barycenters of the initial tetrahedrals and the barycenters of its sides and edges. By this construction, we create as many control volumes as we have nodes in the grid Ω_h. Let B_k be the control volume of the k-th grid node. We integrate the differential equations over this control volume and apply Gauß' integral theorem:

$$-\Delta\Phi_{\text{out}}(x,t) = 0 \quad (26)$$

$$\int_{B_k} -\Delta \sum_{j=0}^{N} \phi^t_{\text{out}_j} b_j(x)dx = -\int_{B_k} \sum_{j=0}^{N} \phi^t_{\text{out}_j} \Delta b_j(x)dx \tag{27}$$
$$= \int_{\partial B_k} \sum_{j=0}^{N} \phi^t_{\text{out}_j} \nabla b_j(x) \cdot \vec{n}(x)dS(x)$$
$$= \sum_{j=0}^{N} \phi^t_{\text{out}_j} \int_{\partial B_k} \nabla b_j(x) \cdot \vec{n}(x)dS(x)$$
$$= \sum_{j=0}^{N} \phi^t_{\text{out}_j} a_{kj}.$$

Because ∂B_k is a polyhedron and $b_j(x)$ is analytically known, the integrals $\int_{\partial B_k} \nabla b_j(x) \cdot \vec{n}(x)dS(x) = a_{kj}$ can be analytically computed. Furthermore, on the membrane these integrals equal to the transmembrane flux (Equation 12) which in general can also depend on other unknowns, like the gating variables or the membrane potential. Furthermore, this is the term which includes the time operator $\frac{d}{dt}$. We discretize our equation fully implicit and because this flux is not linear, we apply Newton's method to solve the emerging equations for each time step. Therein, the Jacobian of the system needs to be inverted, which we accomplish with high efficient iterative solvers. More precisely, we use a parallel ILU-preconditioned BiCGstab method (Barrett et al., 1987). All of this has been implemented with the use of the C++-library ug4 (Vogel et al., 2012), providing flexible numerical tools for these purposes.

Results

The intracellular problem (Equation 7) is a Laplace problem with a Neumann boundary. We referred this to the approximation of purely resistive non-membrane spaces (i.e., the intra- and extracellular space do not contain any free charges). Thus, the driving force of the intracellular potential is given by its Neumann-flux on the boundary (i.e., the membrane). Now, integrating the Laplace equation over the whole neuron and applying Gauß's theorem yields an important constrain for the transmembrane currents: The fluxes are balanced out over the whole membrane at each point of time!

$$-\Delta \Phi_{\text{in}} = 0 \tag{28}$$
$$\Rightarrow \int_{\Omega_{\text{in}}} -\Delta \Phi_{\text{in}} dx = \int_{\Gamma} -\sigma_{\text{in}} \nabla \Phi_{\text{in}} \cdot n_{\text{in}\to\text{out}} dS(x)$$
$$= \int_{\Gamma} j_{\text{all}}(V_m)dS(x) \stackrel{!}{=} 0 \tag{29}$$

There are at least two important implications of this situation. First, an influx at some point of the membrane, necessarily leads to an out-flux at some other point of the membrane with the same total amount of current. Moreover, this must happen simultaneously, since otherwise the condition is violated.

Second, the extracellular potential distributes like a multipole in the extracellular space.

Dipole-like Distribution of the Extracellular Potential for a Idealized Sphere Neuron

Regardless of the neuron's shape, the extracellular potential equation (Equation 17) demonstrates that its only source is the transmembrane flux as expressed through its boundary condition. A current monopole of the extracellular potential would be defined by the overall transmembrane flux. Yet, this flux is always zero as shown before (Equation 29). Thus, there is no monopole component and the extracellular potential distributes in space like a current multipole. To get some quantitative idea of its distribution, we approximate the neuron's geometry to a sphere. Then, we are able to express the extracellular potential with a generalized Fourier series of spherical harmonics.

Let $\Omega_{\text{in}} = B_R$ be a sphere with radius R and $\Gamma = \partial B_R$ its boundary. The spherical harmonics $Y_l^m(\theta, \phi)$ satisfy the Laplace problem on this geometry:

$$-\Delta Y_l^m = 0 \tag{30}$$
$$\Phi_{\text{out}}(r, \theta, \phi) = \sum_{l\geq 0} \sum_{m\geq -l}^{l} (b_{lm} r^{-(l+1)}) Y_l^m(\theta, \phi) \tag{31}$$
$$\Rightarrow -\Delta \Phi_{\text{out}} = 0. \tag{32}$$

The solution Φ_{out} is concretized by the coefficients b_{lm}. These are determined by the transmembrane flux $j_{\text{all}}(V_m)$:

$$\frac{\partial \Phi_{\text{out}}}{\partial r}|_{r=R} = \sum_{l\geq 0} \sum_{m\geq -l}^{l} -(l+1)\frac{1}{R^{l+2}} b_{lm} Y_l^m(\theta, \phi) = j_{\text{all}}(V_m) \tag{33}$$
$$\Rightarrow b_{kn} = -\frac{R^{l+2}}{l+1} \int_0^{\pi} \int_0^{2\pi} sin(\theta) j_{\text{all}}(V_m) Y_k^n(\theta, \phi) d\theta d\phi. \tag{34}$$

Especially, we obtain for the first coefficient b_{00} which corresponds to the potential of a monopole:

$$b_{00} = -\frac{R^{l+2}}{l+1} \int_0^{\pi} \int_0^{2\pi} sin(\theta) j_{\text{all}}(V_m) \frac{1}{\sqrt{4\pi}} d\theta d\phi$$
$$= -\frac{R^{l+2}}{(l+1)\sqrt{4\pi}} \int_{\Gamma} j_{\text{all}}(V_m)dS(x) = 0. \tag{35}$$

Thus, the solution of the extracellular potential does not contain any monopole-part and behaves like a multipole falling in space with higher powers of the distance.

Numerical Error Analysis and Verification by a Comperison with NEURON

NEURON (Hines and Carnevale, 1997) is a highly sophisticated simulation environment for modeling a wide range of neuronal networks with the aid of the standard cable equation. Since the current three-dimensional model generalizes the one dimensional cable equation and since there are no non-trivial analytic solutions of an active neuron for our equations, we want to use this software environment in order verify both our model and our implementation. Our results should be very

similar with these of NEURON for comparable computational domains. In order to keep the three-dimensional computation fast and in order to be able to create suitable three-dimensional computational domains, we carry out this comparison on a very long cylinder $l = 9.9$ mm with small diameter $d = 200\ \mu$m in relation to its length ($\frac{d}{l} \approx 2 \cdot 10^{-4}$). Such cases approximately comply with the assumption of the one-dimensional model (of infinite cylinders). No significant differences in the rise and propagation of an arising action should be visible.

We use proMesh (Reiter, 2014) to construct the three dimensional cylindric soma with a length 9.8 mm and a diameter 200 μm (**Figure 3**).

This test domain we now use in order to first verify the the correct implementation of our discretization schema and second in order to see that we indeed obtain almost identical solutions in comparison with those produced by NEURON.

First is obtained, if the computed solution converges as the computational grid fineness is increased. In order to assess the second point, we have to compare the one dimensional solution of NEURON with the three-dimensional solution of our model. By construction of the one dimensional cable equation, each quantity, although computed on every point of a line, actually represents a volumetric quantity. Thus, the one-dimensional model assumes for all quantities to be radial symmetric and iso-potential on cross-sections of a three-dimensional cylinder. Considering this particularity, we can blow up the solution of NEURON to a three-dimensional solution and compare it with the solution of our model or we compare NEURON's solution with our solution recorded on the cylinder axis. For the sake of simplicity, we use the second way considering that its difference with the volumetric comparison is just the factor of the cross-section area.

Because for three dimensional numeric computations, domains have to be discretized, even simple cylinders never correspond to ideal cylinders, which, however, are the basis of the one-dimensional model. Thus, we will always expect small quantitative differences in such a comparison and, therefore, we are already satisfied to evaluate the differences with NEURON with the aid of an Euclidean integral norm

$$||f||_{L^2([a,b])} = \sqrt{\int_a^b |f|^2 dx}, \tag{36}$$

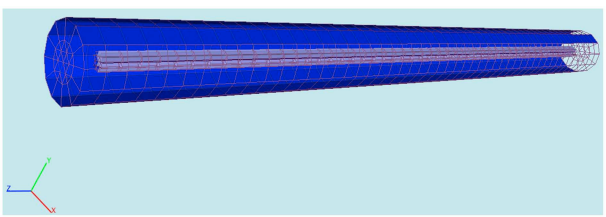


FIGURE 3 | Computational domain constructed with proMesh (Reiter, 2014) for the comparison of the 3D-model's results with NEURON. The cylinder represents a soma having a length of 9.8 mm and a diameter of 200 μm. The purple is the intracellular and the blue the extracellular space. The domain is discretized with tetrahedrals.

where the interval $[a, b]$ corresponds to the time interval of the simulation. Furthermore, in order to get this measure dimensionless, we will consider the relative error between the solution of neuron $V_{m_{\text{NEURON}}}$ and the solution computed at refinement level x, denoted with $V_{m_{\text{Level }x}}$, over the interval $[0, T]$

$$\frac{||V_{m_{\text{NEURON}}} - V_{m_{\text{Level }x}}||_{L^2([0,T])}}{||V_{m_{\text{Level }x}}||_{L^2([0,T])}}. \tag{37}$$

Yet, qualitative measures like propagation speed and signal width should be identical.

Concerning the numeric convergence at grid refinement, we computed the solution on our cylinder, composed by a tetrahedral grid, at two levels of refinement and observed the desired convergence (**Figure 4**). This behavior should serve as benchmark for the right implementation of the finite volume discretization schema.

The solution between the standard cable equation and the three dimensional model are qualitatively undistinguishable (**Figure 3**). The small numerical differences (**Table 1**) are due to the aforementioned reasons: the cylinder in the computation is a disretization of an ideal one, the cylinder's length is finite (the standard cable equation assumes infinite cylinders). Moreover, since the three-dimensional model additionally considers the coupling of the extracellular potential on the membrane, so that there are always to be expected some subtile differences in the solutions, which are reflected in **Table 1**.

However as regards the emerging of the action potential (**Table 1**, **Figure 4**), the propagation speed of $5\frac{m}{s}$, and the signal width (**Table 1**, **Figure 4**) we receive identical results.

Simulation on a Small Network of Four Idealized Neurons

With a computationally quite demanding simulation, we also solve the Equations (17–19) on a more complicated geometry representing four idealized neurons with chemical synapses (**Figure 5**).

The simulation is demanding, because we have a non-linear time-dependent domain problem in three dimensions. It means we solve several a huge linear systems in each time step within Newton's method. Thereby, the time step to be chosen is constrained by the fast dynamics of the active membrane's gating variables, which in our case is chosen with 10 μs, while we aim to simulate the time period of 14 ms. This means we need to compute the solution for 1400 time steps, which is time demanding despite parallel procedures due to the geometry's complexity.

We constructed the computational domain given by a small network of four neurons with the help of an algorithm developed in Niklas Antes' master thesis (Antes, 2009). Each cell consists of a myelinated axon (diameter $d \approx 5\ \mu$ m), a soma ($d \approx 20\ \mu$m) and dendrites ($d \approx 10\ \mu$m). The cells are several hundred micrometer separated among each other.

As regards the transmembrane current $j_{\text{all}}(x, V_m)$ (Equation 12) for the different cell parts, we just considered passive properties on the dendrites while an active membrane reflecting Hodgkin–Huxley dynamics for the soma as well as for the

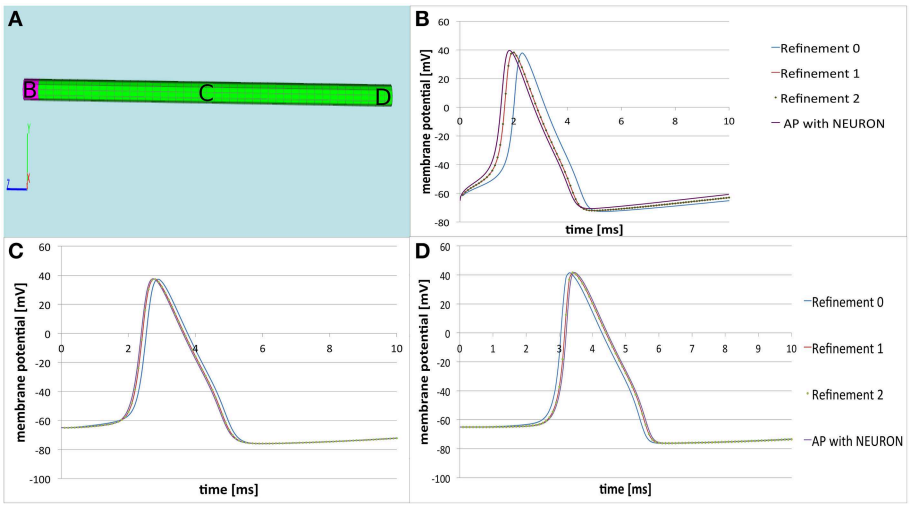


FIGURE 4 | Comparison of the three-dimensional model with NEURON. (A) Computational domain with the marked areas **(B–D)** where the membrane potential is recorded. **(B–D)** Time courses of the membrane potential at the corresponding areas. The solution of the three dimensional model at refinement level 0 is the blue line. After two refinements the solution converges—the red line representing the solution at refinement level 1 coincides with the solution represented at refinement level 2 (dotted green line). This implies the correct implementation of the applied finite volume discretization schema. We see that the solution produced by the three-dimensional model (dotted green line) is almost the same as the solution produced by NEURON (purple line). The small differences are due to the nature of the three-dimensional modeling procedure (see text).

TABLE 1 | Relative error of the computed solution in comparison with NEURON.

Solution on refinement level x	$\frac{\|\|Vm_{\text{NEURON}} - Vm_{\text{Level } x}\|\|_{L^2([0,T])}}{\|\|Vm_{\text{Level } x}\|\|_{L^2([0,T])}}$
$Vm_{\text{Level 0}}$	0.2002
$Vm_{\text{Level 1}}$	0.1174
$Vm_{\text{Level 2}}$	0.1174

The relative error between the solution computed with NEURON Vm_{NEURON} and the solution computed on refinement level x, denoted with $Vm_{Level\ x}$ is very small. This implies that qualitative characteristics like propagation speed, signal width as well are very similar. The small differences measured here can be explained with the nature of the three-dimensional model which automatically considers the extracellular potential in the signal processing and which works with discretized and finite domains (in this case: cylinders are supposed to be ideal and infinite for the standard cable equation).

nodes of Ranvier. On the myelinated sheaths, the transmembrane current $j_{\text{all}}(x, V_m)$ is composed of the first term in Equation (12) only, the capacitive current. Furthermore, two of the cells (cell 1 and cell 4, see **Figure 5**) own external input areas by which the network can be stimulated.

Because we simulate the relatively small time period of 14 ms, we let the synapses work as pre-defined strong post-synaptic current pulses of some nA, which are triggered as soon as the membrane potential at the pre-synapse indicates that an action potential has arrived. This is assumed to happen when the membrane potential at the pre-synapse exceeds the value of 5 mV.

For the sake of simplicity, we choose a constant intra- and extracellular conductivity $\sigma_{\text{in}} = 2\frac{mS}{cm}$, $\sigma_{\text{out}} = 20\frac{mS}{cm}$.

We activate the network by stimulating cell number one (see **Figure 5**) with approximately 30 pA at each of its input areas over the whole simulation period of 14 ms. At the moment of $8ms$, we then stimulate cell number four with a current pulse of approximately 0.5 nA over 20μs. Although this stimulation of the fourth cell is not enough to generate an action potential alone, within the regime of this network and with the ephaptic current activated (Equation 20), an action potential arises (see **Figure 6**). This demonstrates that ephaptic interactions can have a decisive effect as to whether a neuron fires.

The model integrates the impact of the extracellular potential into the signal processing. Though its impact is rather small, it still can have a significant effect when combined with the right stimulation at the right time. Action potentials can arise, which otherwise would not show up (**Figure 6**).

Discussion

The three-dimensional passive model of Voßen et al. (2007) has been extended to a model with active membrane dynamics and has been reformulated mathematically with the aid of an extension of the membrane potential into the intracellular space. This reformulation, for the first time, facilitated numeric simulations of neuronal activity on three-dimensionally resolved idealized neurons generalizing the one dimensional cable equation by fully incorporating the three-dimensional extension of the neurons' geometry and by automatically considering the extracellular potential's influence on the membrane. As shown, the latter influence -though it is quite small- in combination with additional stimulation at the right timing can lead to an action potential which otherwise would not have arisen.

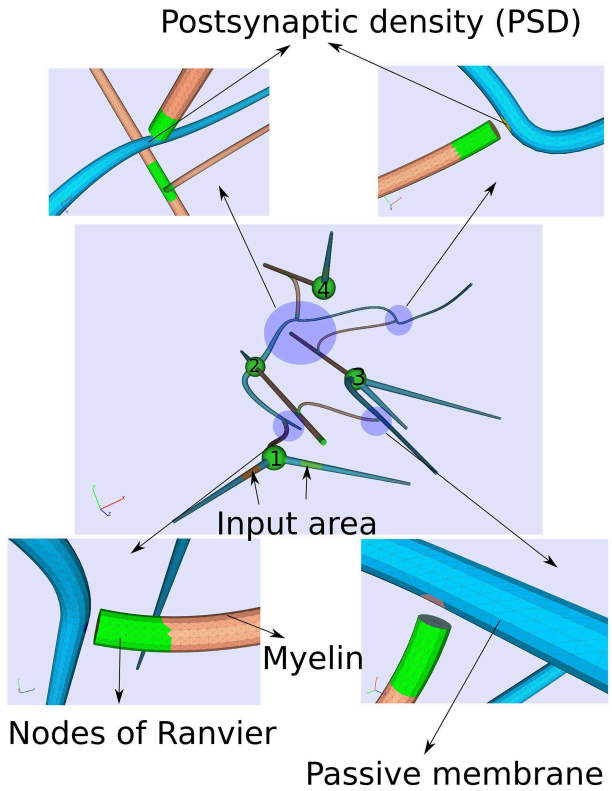


FIGURE 5 | Computational domain of four idealized neurons consisting of a soma, an axon, dendrites and chemical synapses. Computationally, the synapses are modeled by a postsynaptic current at the postsynaptic site as soon the membrane potential exceeds some threshold indicating an action potential at the pre-synapse. Due to the complexity of the computation, a small time period of around 14 ms is covered.

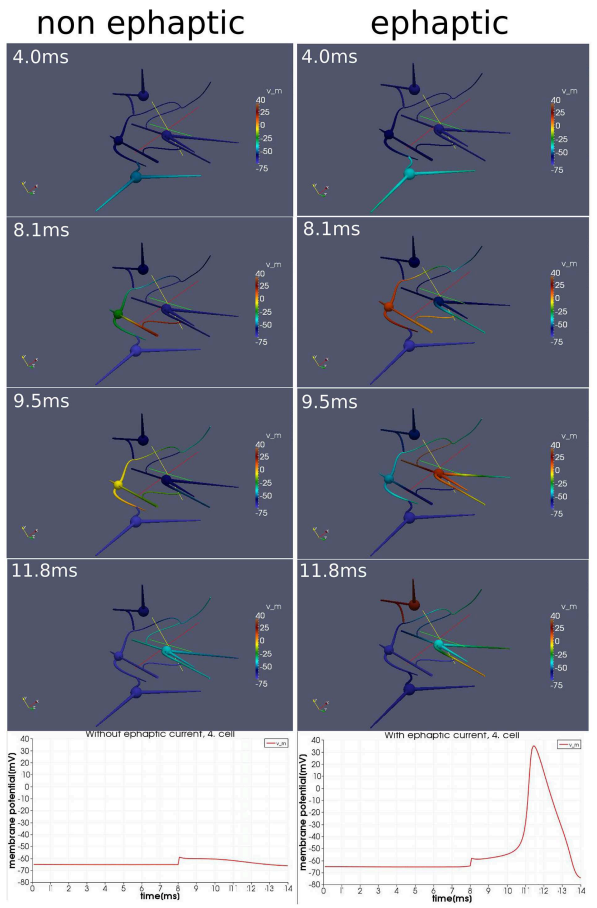


FIGURE 6 | Effect of ephaptic interactions. Two simulations on an idealized small network of four cells and idealized paradigm. One simulation (left column), in which the ephaptic current (Equation 20) is neglected and one simulation (right column) in which it is included. Both simulations are carried out with the same stimulation paradigm. An initial signal spreads through the network. Additionally around the moment of 8 ms, the forth cell (the upper cell of this network) is activated slightly with a current pulse so that it depolarizes just below the threshold for and action potential. Although the effects of ephaptic interactions are very small, we see that they can determine whether a neuron activates in particular circumstances.

For the sake of verifying the correct implementation of this model and because it should deliver similar results as the one-dimensional cable equation for the limit case of long and thin cylinders, we carried out a comparison with NEURON and obtained very good agreement between the two models.

Based on the assumption of charge-free non-membrane spaces -an assumption also used for the derivation of the standard cable equation-, we could provide strong theoretical evidence (to our knowledge for the first time) with the aid of the three-dimensional model that there aren't any current monopoles as the overall out-flux across the membrane balances out. A significant consequence of this behavior is that the leading term of the extracellular potential's multipole expansion vanishes so that it falls in space with higher powers of its distance to the transmembrane current source. In the work of Lindén et al. (2011), this very assumption has been applied for the extracellular potential in order to arrive at converging LFPs. The authors in Lindén et al. (2011) showed that a monopole behavior would lead to a diverging LFP.

We consider the ability to carry out realistic simulations with the cable equation on three-dimensionally resolved ideal neurons as important step and milestone on the way of refining and generalizing existing models for neuronal activity. This three dimensional model facilitates gaining a better understanding of all the processes involved in the signal processing, especially the influence of the extracellular potential activity on the membrane and the impact of the precise three-dimensional shape of the neuron's geometry. Concerning the ephaptic communication, it would be interesting to further investigate its influence on synchronous firing within networks. The latter point also seems to be very promising since lots of precise experimental geometric data are produced. Questions connecting function with geometry can be directly tackled with this model.

However, there is still a long way to go on this path, as the biggest challenge at the moment for our model is its computational demand. Further algorithmic and computational analysis needs to be invested in order to make applicable cutting edge solvers of linear systems arising from partial differential equations -like algebraic multi grid methods- on highly parallel

machines, even on graphic card clusters. As next steps, we want to focus on these improvements.

On the other hand, the computational efficiency is a big advantage for standard one dimensional cable equation. Once we accomplished this efficiency for the three-dimensional model, there are still lots of interesting applications which we wish to address- especially concerning backward modeling with questions like which are the underlying network properties in order to reproduce a given a extracellular potential activity wave.

Furthermore, we see the need of a deeper theoretical analysis of this model with the purpose to provide a mathematical proof that it converges to the standard cable equation for the limit case of infinite cylinders and vanishing extracellular resistivity.

Our long-range purpose is to generalize this model with homogenization and multi-scale techniques so that to be able to simulate the activity of bigger clusters of neuronal networks while also considering the detail in processing on the small scale.

Realized steps on this path will be hopefully items of future publications.

Acknowledgments

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007–2013) under grant agreement number 650003 (Human Brain Project).

Supplementary Material

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fncom.2015.00094

References

Anastassiou, C. A., Perin, R., Markram, H., and Koch, C. (2011). Ephaptic coupling of cortical neurons. *Nat. Neurosci.* 14, 217–223. doi: 10.1038/nn.2727

Antes, N. (2009). *Ein Werkzeug zur Erzeugung von Oberächengeometrien von Neuronen*. Master's Thesis, University of Heidelberg.

Barrett, R., Berry, M., Chan, T. F., Demmel, J., Donato, J., Dongarra, J., et al. (1987). *Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods*, Vol. 43. Philadelphia, PA: Society for Industrial and Applied Mathematics.

Buzsáki, G., Anastassiou, C. A., and Koch, C. (2012). The origin of extracellular fields and currents EEG, ECOG, LFP and spikes. *Nat. Rev. Neurosci.* 13, 407–420. doi: 10.1038/nrn3241

Gold, C., Henze, D. A., Koch, C., and Buzsáki, G. (2006). On the origin of the extracellular action potential waveform: a modeling study. *J. Neurophysiol.* 95, 3113–3128. doi: 10.1152/jn.00979.2005

Hines, M. L., and Carnevale, N. T. (1997). The neuron simulation environment. *Neural Comput.* 9, 1179–1209. doi: 10.1162/neco.1997.9.6.1179

Hodgkin, A. L., and Huxley, A. F. (1952). Propagation of electrical signals along giant nerve fibres. *Proc. R. Soc. Lond. B* 140, 177–183. doi: 10.1098/rspb.1952.0054

Holt, G. R. (1997). *A Critical Reexamination of Some Assumptions and Implications of Cable Theory in Neurobiology*. Ph.D. Thesis, California Institute of Technology.

Holt, G. R., and Koch, C. (1999). Electrical interactions via the extracellular potential near cell bodies. *J. Comput. Neurosci.* 6, 169–184. doi: 10.1023/A:1008832702585

Lindén, H., Tetzlaff, T., Potjans, T. C., Pettersen, K. H., Grün, S., Diesmann, M., et al. (2011). Modeling the spatial reach of the LFP. *Neuron* 72, 859–872. doi: 10.1016/j.neuron.2011.11.006

López-Aguado, L., Ibarz, J., and Herreras, O. (2001). Activity-dependent changes of tissue resistivity in the ca1 region *in vivo* are layer-specific: modulation of evoked potentials. *Neuroscience* 108, 249. doi: 10.1016/S0306-4522(01)00417-1

Rall, W. (1962). Theory of physiological properties of dendrites. *Ann. N. Y. Acad. Sci.* 96, 1071–1092. doi: 10.1111/j.1749-6632.1962.tb54120.x

Rall, W. (1964). "Theoretical significance of dendritic trees for neuronal input-output relations," in *Neural Theory and Modeling*, ed R. F. Reiss (Stanford, CA: Stanford University Press), 73–97.

Reiter, S. (2014). *Effiziente Algorithmen und Datenstrukturen für die Realisierung von Adaptiven, Hierarchischen Gittern auf Massiv Parallelen Systemen*. Ph.D. dissertation, Universität Frankfurt.

Scott, A. C. (1975). The electrophysics of a nerve fiber. *Rev. Mod. Phys.* 47, 487–533. doi: 10.1103/RevModPhys.47.487

Versteeg, H. K., and Malalasekera, W. (2007). *An Introduction to Computational Fluid Dynamics: The Finite Volume Method*. Harlow: Prentice Hall.

Vogel, A., Reiter, S., Rupp, M., Nägel, A., and Wittum, G. (2012). Ug 4: A novel flexible software system for simulating pde based models on high performance computers. *Comput. Vis. Sci.* 16, 165–179. doi: 10.1007/s00791-014-0232-9

Voßen, C., Eberhard, J. P., and Wittum, G. (2007). Modeling and simulation for three-dimensional signal propagation in passive dendrites. *Comput. Vis. Sci.* 10, 107–121. doi: 10.1007/s00791-006-0036-7

Xylouris, K., Queisser, G., and Wittum, G. (2010). A three-dimensional mathematical model of active signal processing in axons. *Comput. Vis. Sci.* 13, 409–418. doi: 10.1007/s00791-011-0155-7

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.

frontiers in
COMPUTATIONAL NEUROSCIENCE

METHODS ARTICLE
published: 26 September 2014
doi: 10.3389/fncom.2014.00120

Membrane current series monitoring: essential reduction of data points to finite number of stable parameters

Raoul R. Nigmatullin[1], Rashid A. Giniatullin[2,3] and Andrei I. Skorinkin[4,5,6]*

[1] *Theoretical Physics Department, Institute of Physics, Kazan Federal University, Kazan, Russia*
[2] *Department of Neurobiology, A.I. Virtanen Institute, University of Eastern Finland, Kuopio, Finland*
[3] *Laboratory of Neurobiology, Department of Physiology, Kazan Federal University, Kazan, Russia*
[4] *Department of Radioelectronics, Institute of Physics, Kazan Federal University, Kazan, Russia*
[5] *Department of Biophysics of Synaptic Processes, Kazan Institute of Biochemistry and Biophysics Russian Academy of Sciences, Kazan, Russia*
[6] *Department of Bioinformatics, Institute of Informatics, Kazan, Russia*

Edited by:
Tobias Alecio Mattei, Ohio State University, USA

Reviewed by:
Jianbo Gao, Wright State University, USA
Abdelmalik Moujahid, University of the Basque Country UPV/EHU, Spain

***Correspondence:**
Andrei I. Skorinkin, Radioelectronics Department, Institute of Physics, Kazan Federal University, 16 Kremliovskaja Str., Room 102, Kazan 420008, Russia
e-mail: askorink@yandex.ru

In traditional studies of changes in cell membrane potential or trans-membrane currents a large part of the recorded data presents "a pure noise." This noise results mainly from the random openings of membrane ionic channels. Different types of stationary or non-stationary noise analysis have been used in electrophysiological experiments for identification of channels kinetic states. But these methods have a limited power and often cannot answer to the main question of the experimental study: do external factors induce a significant change of channels kinetics? A new method suggested in the current study is based on the scaling properties of the beta-distribution function that allows reducing the series containing 200,000 and more data points to analysis of only 10–20 stable parameters. The following clusterization using the generalized Pearson correlation function allows taking into account the influence of an external factor and combine/separate different parameters of interest into a statistical cluster considering the influential parameter. This method which we call BRC (Beta distribution-Reduction-Clusterization) opens new possibilities in creation of a largely reduced database while extracting specific fingerprints of the long-term series. The BRC method was validated using patch clamp current recordings containing 250,000 data points obtained from the living cells and from open tip electrode. The numerical distinction between these two series in terms of the reduced parameters was obtained.

Keywords: noise analysis, detrended fluctuation analysis, fluctuation spectroscopy based on beta-distribution, sequence of the ranged amplitudes, membrane currents of neurons

INTRODUCTION

During electrophysiological studies it is common to record rather long tracks of signals. These signals are registered as temporal variations of cell membrane potential or trans-membrane currents induced by the opening of some ligand- or voltage-gated or even chaotic ionic channels. Usually the principal aim of such a study is the registration of some macroscopic signals—evoked or spontaneous—and the change of parameters of these signals characterizes the total effect of some actions that are located in the experimental object. But a large part of the record forms a so-called "empty track" containing a "pure noise" only. It is well known that this noise reflects mainly the result of random openings of transmembrane ionic channels. Different types of stationary or non-stationary noise analysis have been used for identification of these channels' states (Neher and Sakmann, 1976; Sigworth, 1980, 1985, 1986; Läuger, 1985; Traynelisa and Jaramilloa, 1998; Alvarez et al., 2002; Venkataramanan and Sigworth, 2002).

Unfortunately, these methods have not come into widespread use among physiologists since they often cannot answer the main question of the study: If this drug or this change of environment state induces the reliable change of channels condition or not?

Thus, there is an urgent task to develop a special language that can be compact and reliable in order to describe accurately very long current streams (long-time series) with hidden signals and noise in terms of a finite and statistically understandable set of reduced parameters. In this paper we want to show *how* to develop this special language based on an example of the analysis of signals recorded in rat's spinal cord slices. Besides this problem we want to show how to detect the presence of the biological object inside the experimental set. For this purpose we also recorded data representing the dependence of the current vs. time when the biological object is absent. Examples of currents recorded in a living cell and with empty electrodes are shown in **Figure 1**. It is well noticeable that these two signals are apparently very similar. Even though generally distinguishable by an experienced observer the reliability of these differences cannot be numerically evaluated without some special analytic methods.

To the authors' best knowledge one method is basically suitable for quantitative analysis of the different long-time series. This method was introduced by Peng et al. (1994) and nowadays it

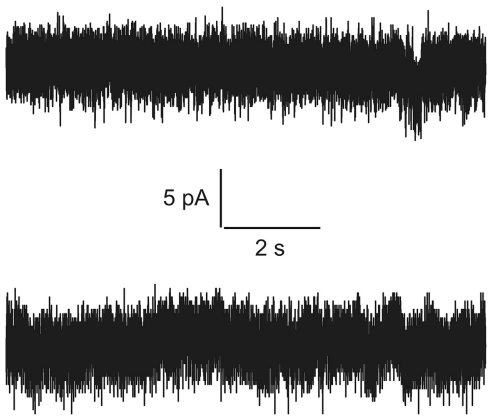


FIGURE 1 | The examples of currents recorded for empty electrode (top) and in a living cell (bottom). Note the high similarity of the tracks.

is known as detrended fluctuation analysis (DFA). It was well described in literature by their creators (Ossadnik et al., 1994; Peng et al., 1995) and found its application in analysis of biomedical (Penzel et al., 2003; Jospin et al., 2007; Burr et al., 2008) and other (Hausdorff et al., 1995, 1996) data. But it is necessary to note that the DFA algorithm works well only for certain types of non-stationary time series (especially having slowly varying trends), it is not designed to handle all possible non-stationarities in real-world data. This algorithm was *not* free also from uncontrollable errors that are associated with approximate fitting of detrended fluctuations by the segments of straight lines or by the parabolic or high order polynomials (Kantelhardt et al., 2001). The final straight line with power-law exponent α_{DFA} is obtained as a slope in a double-log scale as a result of the fitting procedure and contains the fitting error that depends also on the type of segmentation of the initial series considered. These uncontrollable errors (usually they are not properly analyzed in the literature) can lead to different results in calculation of the desired value of the α_{DFA} and other associated fitting parameters in analysis of the *same* long-time series.

A technique, called scale-dependent Lyapunov exponent (SDLE, see Gao et al., 2006, 2012b, 2013; Hu et al., 2010), provides a more comprehensive characterization of complex time series. Some of DFA's limitations have been overcome recently as well by using a new method called adaptive fractal analysis (AFA, see Gao et al., 2010, 2011, 2012a; Riley et al., 2012; Kuznetsov et al., 2013). AFA has been shown to be able to determine global trends, remove noise, perform fractal analysis and multiscale decomposition and present data as a curve. However, new tools could be developed specifically designed to show and estimate even mild differences between two long time series.

Thus, it would be desirable to have a new method with "high resolution" (10–20 significant parameters) to distinguish more accurately the experimental data and effect of treatments. In this paper we demonstrate such method based on some invariant properties of the beta-distribution function; furthermore this method admits a procedure that controls the error in each stage of its application. From our point of view the effectiveness of new approach is based on the monotone behavior of the primary fitting parameters that admit the secondary fit. This peculiarity allows compressing initial fitting parameters with the help of the secondary fit and present initial data set in more compact form.

The four fitting parameters (A, B, α, β) of beta-distribution can be interpreted and used for quantitative *reading* of fluctuations arising on different scales of the long-time series considered. In previous papers (Nigmatullin, 2010; Nigmatullin et al., 2012) based on the principle of the strong correlation of random sequences it was shown that the cumulative (integral) curve obtained from the sequence of the ranged amplitudes (SRA) can be described with high accuracy by means of the beta-distribution function. In other words, any *detrended* random sequence being transformed to the SRA (when all amplitudes of the initial sequence are sorted out and located in the descending order $y_1 > y_2 > \ldots > y_N$) after elimination of its mean value and subsequent integration, forms a bell-like curve $J(x)$ that can be fit (with controllable relative error) by the function:

$$J(x) \cong Jb(x) = A\,(x - x_0)^{\alpha}\,(x_N - x)^{\beta} + B. \qquad (1)$$

Here the limiting values $x_0 < x_N$ define the ends of the location interval of the random sequence considered. In many cases the parameters x_0, x_N are known. Other quantitative parameters (A, B, α, β) should be found from the fitting procedure of the function $J(x)$ to the curve $Jb(x)$. The power-law exponents (α, β) reflect the *fractal* properties of the random sequence considered and the presence of the memory that is expressed in the behavior of the corresponding SRAs. The criterion for the verification of the presence of memory in two random sequences which are compared is as follows. If one SRA being plotted with respect to another one forms a curve close to a straight line then these two random curves are defined as a having a relative memory and can be considered as being *strongly correlated*. This important property allows transforming any segment of a random sequence to a beta-distribution function and "read" this segment in terms of four unknown fitting parameters (A, B, α, β). Such transformation from 30 to 50 or even more initial points belonging to a random sequence can be read in terms of these four parameters only. This allows us to suggest a new type of spectroscopy based on some scaling properties of the beta-distribution. This transformation is called Fluctuation Spectroscopy based on Beta-Distribution (FSBD). In general we suggest a method which we call BRC (Beta distribution-Reduction-Clusterization). The basic problem that is solved in this paper by using the BRC method can be formulated as follows: *Is it possible to suggest a reliable method with controllable error that has a wide range of applicability and which has a flexible small set (10–20) of statistically understandable parameters for quantitative characterization of the differences between long-time series?*

MATERIALS AND METHODS

PREPARATION OF SPINAL CORD SLICES

Ten- to Twenty-days-old Wistar rats were deeply anesthetized with diethyl ether and killed by decapitation. After laminectomy, the spinal cord was excised, and immediately immersed in cold ($0 \div 4$°C) artificial cerebrospinal fluid containing (in mM): 126

NaCl, 26 $NaHCO_3$, 2.5 KCl, 1.25 NaH_2PO_4, 2 $CaCl_2$, 2 $MgCl_2$, and 10 glucose (bubbled with 95% O_2 and 5% CO_2; pH 7.3; 310 mOsm measured). Several transverse slices (250-μm thick) were prepared from the lumbosacral enlargement (L4-6) with a vibratome (VT1000S, Leica, Nussloch, Germany).

WHOLE-CELL RECORDINGS

Slices were transferred to a recording chamber ($300 \div 400\,\mu$l volume) and continuously superfused with oxygenated artificial cerebrospinal fluid at 3 ml/min and $22 \div 24$°C. Interneurons were visualized with an upright interference contrast microscope and a × 40 water immersion objective (Axioscope FS, Carl Zeiss, Oberkochen, Germany). Patch-pipettes (tip resistance, $5 \div 7\,M\Omega$) were prepared by a puller (Flaming-Brown P97; Sutter, Novato, CA, USA) from borosilicate capillaries and were filled with intracellular solution consisting of (in mM: potassium gluconate 140, NaCl 10, $MgCl_2$ 3, HEPES 10, EGTA 11; pH 7.3 adjusted with KOH; 300 mOsm measured).

Interneurons were voltage-clamped at −65 mV in the whole-cell configuration after obtaining GV seals (usually not less than 2 GV) by means of a patch-clamp amplifier (Axopatch 200B; Molecular Devices, Sunnyvale, CA, USA). Compensation of capacitance (Cm) and series resistance (Rs) was achieved with the inbuilt circuitry of the amplifier. Series resistance was compensated by $40 \div 70\%$ and did not change appreciably from the beginning to the end of the experiments, indicating stable recording conditions. The tracks used for comparison were recorded by the immersion of filled patch-pipettes in artificial cerebrospinal fluid; the patch-pipettes were voltage-clamped at −65 mV too.

Then all data were sampled at 10 kHz and stored on-line with a PC using the pClamp 10.0/Clampex 10.0 software package (Molecular Devices).

SCALING PROPERTIES OF THE BETA-DISTRIBUTION AND DESCRIPTION OF THE TREATMENT PROCEDURE

In this section we want to demonstrate the scaling properties of Expression (1). We subject x, x_0 and x_N in Expression (1) to the following scaling transformations, keeping the power-law exponents α and β invariable: $x = \xi \cdot x' + b$. $x_0 = \xi \cdot x'_0 + b$, $x_N = \xi \cdot x'_N + b$, which gives the following beta transformation:

$$Jb(x) \to Jb(x') = A'\left(x' - x'_0\right)^{\alpha} \cdot \left(x'_N - x'\right)^{\beta}, \tag{2}$$

where $A' = A \cdot \xi^{(\alpha+\beta)}$. This is the accurate mathematical result that follows from the scaling transformation of the initial coordinates.

In order to have a simple criterion for comparison of the two beta-distributions let us calculate the values of two extreme points $\bar{x}, \bar{x}'$ belonging to the functions $Jb(x)$ *and* $Jb(x')$ respectively.

$$\bar{x} = w_1 x_0 + w_2 x_N,\ \bar{x}' = w_1 x'_0 + w_2 x'_N,$$

$$w_1 = \frac{\beta}{\alpha+\beta} = \frac{x_N - \bar{x}}{\Delta},\ w_2 = 1 - w_1,\ \Delta = x_N - x_0,\ \Delta' = \frac{1}{\xi}\Delta,$$

$$\bar{H} = Jb\left(\bar{x}\right) = A w_1^{\beta} w_2^{\alpha} \Delta^{\alpha+\beta} + B,$$

$$\bar{H}' = Jb\left(\bar{x}'\right) = A' w_1^{\beta} w_2^{\alpha} \left(\Delta'\right)^{\alpha+\beta} + B,$$

$$\bar{H}' = Jb(\bar{x}') = A w_1^{\beta} w_2^{\alpha} \xi^{\alpha+\beta} \left(\frac{1}{\xi}\right)^{\alpha+\beta} + B \equiv \overline{H}. \tag{3}$$

From Expressions (3) it follows that for the scaling transformation (2) the heights $\bar{H}, \bar{H}'$ of the extreme points of the two bell-like distributions at the fixed values of the power-law exponents α and β and parameter B should *coincide* with each other.

Besides this criterion it is necessary to take into account the scaling relationship between the heights $\bar{H}$, $\bar{H}'$. If two power-law exponents α and β are subjected to the scaling transformation at the fixed value of the length $\Delta = x_N - x_0$:

$$\alpha' = \theta\alpha,\ \beta' = \theta\beta, \tag{4}$$

then simple manipulations lead to the second scaling relationship:

$$\frac{\bar{H}'}{A'} = \left(\frac{\bar{H}}{A}\right)^{\theta}. \tag{5}$$

Here the amplitudes A and A' are defined by relationships (1) and (2), respectively. The consideration of the scaling properties of the beta-distribution allows one to suggest the following two steps.

Step 1. This step includes the formation of the sequence of the range amplitudes (SRA) when all amplitudes located on the fixed length $\Delta = x_N - x_0$ are ordered in descending order $y_1(x_0) > y_2 > \ldots > y(x_N)$.

Step 2. Numerical integration of the SRA with respect to its mean value and subsequent fit to the function (1).

Figure 2 illustrates this transformation which is realized after application of these two steps.

Each sub-segment having equal length Δ is transformed to its SRA (**Figure 2A**) in Step 1, and the integration of the SRAs with respect to its subtracted mean value gives finally the desired bell-like curve that can be fit to Expression (1) in Step 2. Mathematically these two steps correspondingly are expressed as:

$$\begin{aligned} SRA(y(x_j)) &= sort(y(x_j)) \to \Delta SRA(y(x_j)) \\ &= SRA(y(x_j)) - \frac{1}{\Delta}\sum_{j=1}^{\Delta} SRA(y(x_j)) \\ &\equiv SRA(y(x_j)) - \langle \ldots \rangle. \end{aligned} \tag{6a}$$

Here the integer index j ($j = 1, 2, \ldots, N$) numerates the number of data points in the fixed segment $\Delta = x_N - x_0$ containing initially 30–50 data points.

$$\begin{aligned} J(x_j) = J(x_j) + \frac{1}{2}\left(x_j - x_{j-1}\right) \cdot \left(\Delta SRA(y(x_j))\right. \\ \left. + \Delta SRA(y(x_j))\right),\ J_0 = 0. \end{aligned} \tag{6b}$$

Figure 2 demonstrate the realization of these two steps [with the usage of Expression (6)] on a short segment belonging to the membrane current initial time segment (containing 250,000 data points). We should notice that the mean value $<\ldots>$ of the chosen segment should be subtracted and the integration procedure [the last row in (6)] should be realized with the help of the

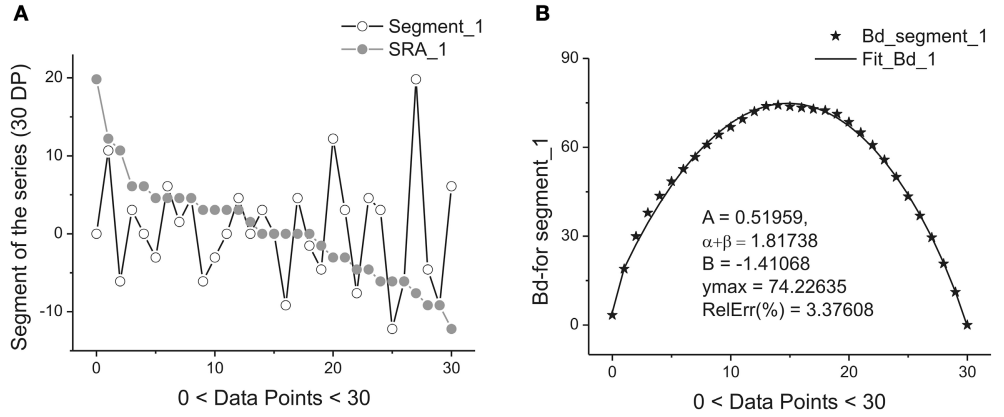


FIGURE 2 | Example of one segment (marked as Segment_1) containing 30 points. (A) The sequence of the ranged amplitudes (SRA) given in descending order and marked by gray stars. On vertical axes the values of the current are given in picoampers. **(B)** The bell-like curve (marked by crossed stars) obtained by integration from SRA_1 (shown on the previous **A** by gray stars) and its fit marked by the bold solid line. The fitting parameters of this curve are given inside of this figure. As it follows from this figure 30 data points are sufficient for providing the acceptable fit with the value of the relative error close to 3.5%.

trapezoid method. As a result of calculation of Expression (6) we obtain the desired bell-like curve $J(x_j)$.

Figure 2B shows the quality of the fitting of the bell-like curve obtained to the beta-distribution. In order to have the value of the relative error:

$$\mathrm{Re}lErr = \left(\frac{stdev(J(x) - Jb(x))}{mean(J(x))}\right) \cdot 100\%,$$

$$where\ stdev(f(x)) = \left[\frac{1}{N_\Delta}\sum_{j=1}^{N_\Delta}\left(f\left(x_j\right) - mean\left(f(x)\right)\right)^2\right]^{1/2},$$

$$mean\left(f(x)\right) = \frac{1}{N_\Delta}\sum_{j=1}^{N_\Delta} f(x_j), \tag{7}$$

to be limited to a few percentages (2–5)% we should choose the length of the minimal segment $\Delta_{\min}$ of the initial series containing initially 30–50 data points. In Expression (7) the value N_Δ defines the number of data points that enters in the segment of the length Δ. Thus, the first reduction criterion should be written as:

$$\Delta_{\min} \cdot \xi^k = N_{\mathrm{total}} \tag{8}$$

Here the scaling parameter ξ has the same meaning as in Expression (2).

This requirement allows one to consider the long-time series containing the total number of data points ($j = 1, 2, \ldots, N_{\mathrm{total}}$) in terms of the reduced parameters of the beta-distribution (A, B, α, β) depending on parameter k. Further it is convenient to rewrite condition (8) in the following form changing the numeration of the current parameter k:

$$\Delta_k = \frac{N_{\mathrm{total}}}{\xi^{K+1-k}}, \quad k = 1, 2, \ldots, K+1, \tag{9}$$

where [in comparison with (8)] the value Δ_1 should coincide with the minimal value $30 < \Delta_{\min} < 50$ giving the condition for finding the limiting value of K (the total number of segments is equaled to K + 1). In the opposite case, the value Δ_{K+1} should give the maximal length coinciding with the value N_{total}. As a result of this reduction procedure one can transform N_{total} data points to $4.(K+1)$ parameters. But this step is *not* sufficient. If the functions A_k, B_k, $(\alpha+\beta)_k$ have monotonic behavior one can realize further reduction to the *primary* set of the fitting parameters describing these functions.

Now it is necessary to explain why the sum of the parameters $(\alpha+\beta)$ is selected instead of considering each-power law exponent separately. This selection is based on the comparison of these exponents with the single power-law exponent α_{DFA} figuring as the basic parameter in the DFA. It is easy to see that relationship $\alpha + \beta = 1$ with $\alpha \approx \beta \approx 0.5$ (for this case beta-distribution looks like a semicircle) corresponds to a distribution with the absence of power-law correlations in the time series. From another side it gives for $\alpha_{DFA} = 0.5$. Comparison with these two power-law exponents leads us to the following approximate expression:

$$\alpha_{DFA} \cong \frac{1}{2}(\alpha + \beta). \tag{10}$$

One can notice also that Expression (10) does not contradict other well-known power-law exponents (Hausdorff et al., 1995; Burr et al., 2008) $\beta_f = 2\alpha_{DFA} - 1$ that is used for description of the power-law spectrum $S(f) \sim f_f^{-p}$ and decay of autocorrelation function $C(t) = \langle x_i x_{i+1}\rangle \sim t^{-1}$ with $\gamma = 2 - 2\alpha_{DFA}$. From the requirements (β_f, $\gamma > 0$) it follows that:

$$1 \le (\alpha+\beta) = 2\alpha_{DFA} \le 2. \tag{11}$$

We want to stress here that this requirement is *approximate* and can serve as an indication for division of long-time series with fractal structure (because it does not contradict with well-known inequalities) known before from series with self-similar structure.

The left-hand inequality follows from the requirement $\beta_f > 0$ and does not contradict with numerical results obtained in other papers (Penzel et al., 2003; Jospin et al., 2007; Burr et al., 2008). We should also note that the equality $(\alpha + \beta) = 2$ corresponds to a *uniform* amplitude distribution. The uniform distribution leads to the degeneration of the corresponding SRA to a straight line (Nigmatullin, 2010). The beta-distribution in this case is described by a parabolic curve. If one of the power-law exponent (say $\alpha \to 0$) then the position of extreme point $\bar{x} \to x_0$. Because of normalization $w_1 + w_2 = 1\,\beta \to 1$. This statement is valid also in the opposite case when $\alpha \to 1$, $\beta \to 0$. So, the last relationship (11) can be considered as a *specific fractal* test in our further calculations. Here we should also note that in practical applications the existence of the interval $0 < \alpha + \beta < 1$ and inequality $\alpha + \beta > 2$ also are *possible*. For the first case, for small values of α and β the beta-distribution degenerates to a rectangle-like curve. In the second case the values of the derivatives on the ends (x_0, x_N) of the beta-distribution have zero values. These two cases correspond to degeneration of the fractal properties of the time-series analyzed. The verification of relationship (11) on the Weierstrass-Mandelbrot function that represents itself the self-affine function (see its definition in Feder, 1988) confirms the relationship (11). So, for practical purposes it is useful to work with the combination of $(\alpha + \beta)$.

The statistical and geometrical meaning of other parameters entering to (1) can be explained as follows. The value of the amplitude A together with the height H of the beta-distribution is associated with intensity of the fluctuations analyzed. As one can see from **Figure 3A** the angle of the SRA slope counted off from zero point (after elimination of its mean value) is proportional to the height of the corresponding fluctuation that is expressed in the form of a beta-distribution in **Figure 3B**. If this angle approaches the vertical axis, the height of the distribution becomes large. In the opposite case when this angle tends to zero the height of the distribution is small. See **Figure 3B** where the first 14 beta-distributions are shown. The measure of asymmetry can be connected with parameters B and the values of weight factors $w_{1,2}$ that are defined by Expression (3). The value $w_1 = 0.5$ corresponds to the complete symmetry of the distribution in the horizontal direction. Any shift of this parameter to the left-($w_1 < 0.5$) or to the right-hand side ($w_1 > 0.5$) reflects the *horizontal asymmetry* of the distribution. A small asymmetry of this distribution in vertical direction is controlled by the parameter B.

Step 3. After selection of the scaling parameter ξ and the limiting value K from Expression (9) one can obtain a family of bell-like curves that can be fitted to Expression (1). The calculated fitting parameters $A_k, \alpha_k, \beta_k, B_k,\ k = 1, 2, \ldots, K+1$ from Expression (1) are obtained. The set of these bell-like curves and the corresponding fitting parameters forms the total fluctuation spectrum based on the beta-distribution (FSBD). Each part of this FSBD contains the corresponding beta-distribution:

$$Jb_k(x_j) = A_k \left(x_j - x_{0,k}\right)^{\alpha_k} \left(x_{N,k} - x_j\right)^{\beta_k} + B_k. \tag{12}$$

Step 4. In order to subject them to the scale-invariant properties described above it is necessary to average this family of distributions and consider only one weighted distribution:

$$\left\langle Jb_k(x_j)\right\rangle = \frac{1}{NBd_k}\sum_{j=1}^{NBd_k} Jb_k(x_j), \quad j = 1, 2, \ldots, NBd_k\,,$$

$$NBd_k = \frac{N_{total}}{\Delta_k}, \tag{13}$$

located in the given interval Δ_k. Here the parameter NBd_k coincides with number of beta-distributions calculated for the given k. **Figure 4** shows the averaged beta-distribution obtained for the cell number 3. If $N_{\text{total}} = 250{,}000$ then from condition (9) at the given $\Delta_1 = 32$ and $\xi = 2$ we obtain that $K = 13$. So, the total number of beta-distributions $NBd_1 = N_{\text{total}}/\Delta_1 = 8333$. The first 14 distributions belonging to this family is shown in **Figure 3B**.

Step 5. Further calculations are reduced to the analysis of the functional dependencies $A_k, \alpha_k, \beta_k, B_k,\ k = 1, 2, \ldots, K+1$ with

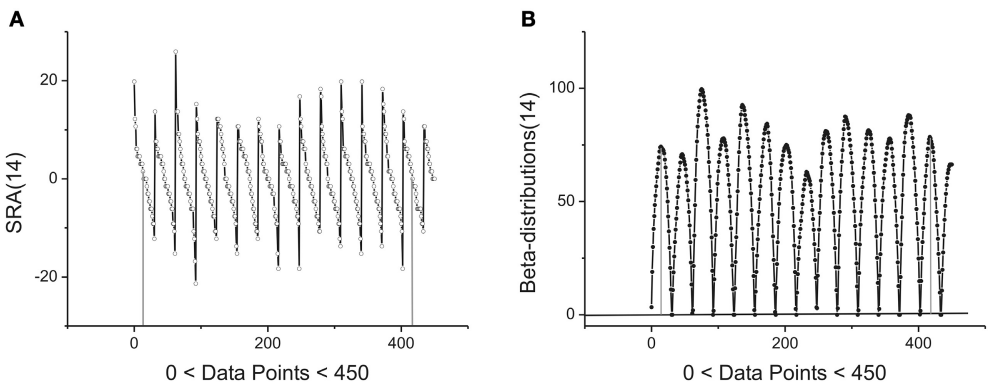


FIGURE 3 | Example of first 14 segments, each segment contains 30 points. (A) The first 14 SRAs calculated for the large-time membrane current sequence containing in total 250,000 data points. After elimination of their mean values (two limiting of them are shown by vertical gray lines) and subsequent integration one can obtain a family of the bell-like curves. They are shown below. **(B)** The first 14 beta-distributions obtained by numerical integration from the SRAs given of the previous panel. For the total sequence having 250,000 data points we have in general 8333 distributions of such kind. Two limiting heights are marked by solid gray lines.

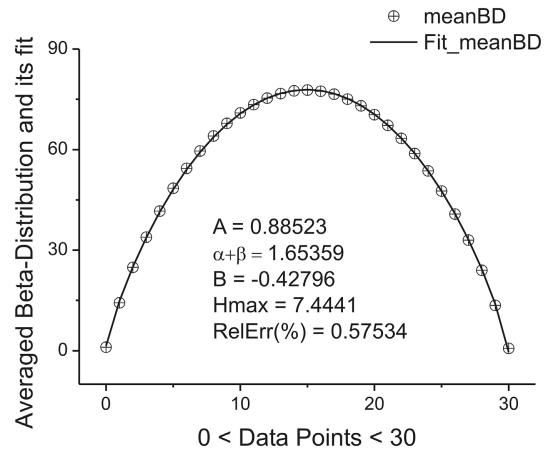


FIGURE 4 | The averaged beta-distribution (averaged with the usage of Expression (13) for the given $NBd_1 = N_{total}/\Delta_1 = 8333$). The fitting parameters are shown inside of this figure.

respect to the variable k. We define them as the *primary* fitting parameters characterizing the averaged distribution (13). Further analysis shows that the amplitude A_k has monotonic behavior and can be described by a simple exponential behavior:

$$\langle A_k \rangle = A_1 \cdot \exp(\lambda_a \cdot k) + A_0. \quad (14)$$

Preliminary calculations show that this monotonic behavior is conserved for the long-time series without any trend. The presence of trend distorts this behavior.

This dependence follows after substitution of Expression (9) in relationship (2) for the amplitudes. The perfect fit of this monotone curve is shown in **Figure 5A**. Other dependencies are not so simple but nevertheless they can be identified from simple power-law and exponential hypothesis with the help of the eigencoordinates (ECs) method (Baleanu et al., 2011; Ciurea et al., 2011). The dependences $<(\alpha+\beta)_k> \equiv S_k(\alpha\beta)$ and $<B_k>$ have also monotonic character and can be fitted by means of two simple functions:

$$S_k(\alpha\beta) \cdot k^{\nu} = A_{pl} \cdot k + B_{pl},$$
$$\langle B_k \rangle = B_1 \cdot \exp(\lambda_B \cdot k) + B_0 \quad (15)$$

These functions are shown, respectively, in **Figures 5B,C**. So, finally we obtain 10 fitting parameters that can be combined with 9 parameters figuring in Expressions (14) and (15) [λ_a, A_1, A_0], [ν, A_{pl}, B_{pl}], [λ_B, B_1, B_0] and the limiting value of parameter $w_{1,K+1}$. The behavior of this weight factor is shown in **Figure 5D**.

These ten parameters can be used as the *primary* set of the fitting parameters for creation of a specific "fingerprint" of the long-time series considered. The idea of clusterization of these parameters is discussed in Results Section. Further analysis shows that the distribution of the heights and mean values of the SRAs obtained for the family of distributions at Δ_1 also forms two other different beta-distributions. These distributions are *important* also for clusterization purposes because initially the information about the secondary distribution of the heights of the initially formed beta-distributions family and mean values of the corresponding SRA were *not* taken into account. The distributions of the heights and mean values together with their beta-distributions are shown in **Figures 6**, **7**, correspondingly. After fitting of these two distributions one can obtain in addition 5 significant parameters characterizing each beta-distribution separately.

$$\left[A_H, (\alpha+\beta)_H, w_{1,H}, \max(Bd_H), mean(SRA_H)\right],$$
$$\left[A_{mn}, (\alpha+\beta)_{mn}, w_{1,mn}, \max(Bd_{mn}), mean(SRA_{mn})\right]. \quad (16)$$

These ten additional parameters we define as the *secondary* fitting parameters. The statistical meaning of these parameters are the following. The parameters $A_{H,\,mn}$ characterize the amplitudes of beta-distributions referring, correspondingly, to the heights (H) and mean values (mn). The sum $(\alpha+\beta)_{H,\,mn}$ contains the information about their power-law exponents, $w_{1,\,H,\,mn}$ gives the information about their asymmetry, $\max(Bd_H, Bd_{mn})$ signifies their heights, and the fifth parameter $\mathrm{SRA}_{H,\,mn}$ contains information about the mean values of these two additional distributions.

From our point of view, these 20 (10 primary and 10 secondary) significant parameters [figuring in Expressions (14)–(16)] combined together can completely characterize the behavior of fluctuations associated with the long-time series analyzed and containing $N_{total} = 2.5.10^5 \div 10^6$ and even more data points.

CLUSTERIZATION OF FINAL PARAMETERS BASED ON THE GENERALIZED PEARSON CORRELATION FUNCTION

For clusterization purposes one can suggest more accurate selection of similar sequences based on *internal* correlations. For this aim we introduce the generalized Pearson correlation function (GPCF) (Nigmatullin, 2010; Nigmatullin et al., 2012).

$$GPCF_p = \frac{GMV_p(s_1, s_2)}{\sqrt{GMV_p(s_1, s_1) \cdot GMV_p(s_2, s_2)}}, \quad (17)$$

where expression:

$$GMV_p(s_1, s_2, \ldots, s_K) =$$
$$\left(\frac{1}{N}\sum_{j=1}^{N}\left|nrm_j(s_1) \cdot nrm_j(s_2) \cdot \ldots \cdot nrm_j(s_K)\right|^{mom_p}\right)^{1/mom_p}, \quad (18)$$

determines the generalized mean value (GMV)-function of the K-th order. Here the generalized mean value (GMV) function determines the mean value for all range of the moments (see Expression (19) below). The set of parameters (s_1,s_2,. . .,s_K) determines the type of the random sequence compared. The $GPCF_p$ determined by Expression (17) coincides with the conventional definition of the Pearson correlation coefficient at $mom_p = 1$. The set of moments are determined by the following expression:

$$mom_p = \exp(Ln_p), \; Ln_p = mn + \left(\frac{p}{P}\right) \cdot (mx - mn),$$
$$p = 0, 1, \ldots, P. \quad (19)$$

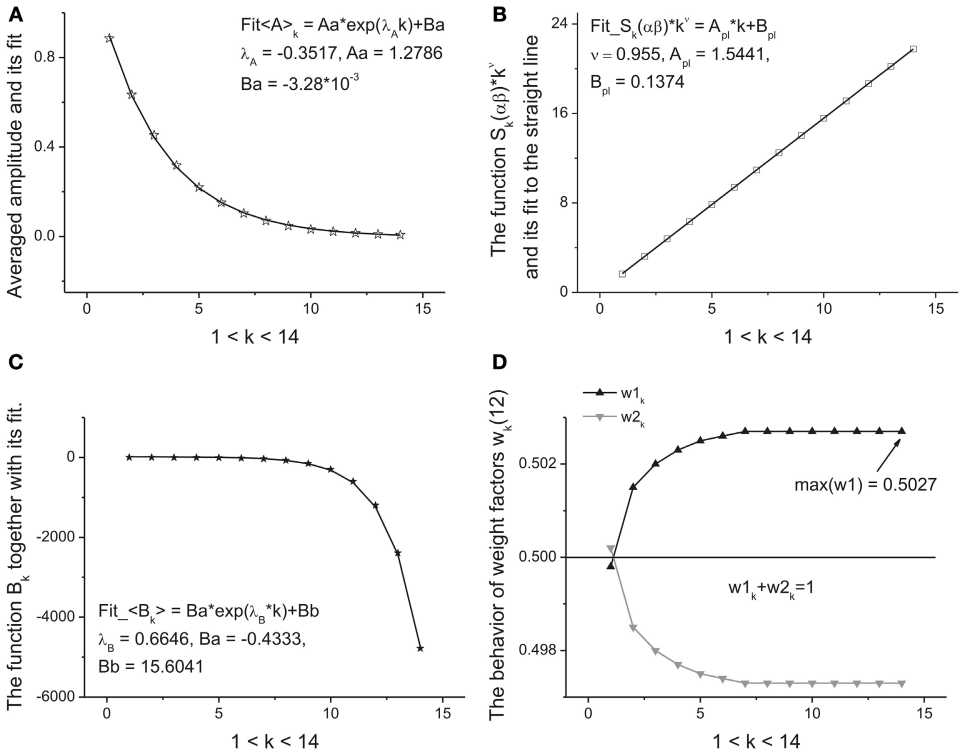


FIGURE 5 | The fitting curves of four parameters. (A) The fit of the amplitude obtained for the averaged beta-distributions for different values of k. See Expression (14) for details. The fitting parameters of the exponential function are given above of this figure. **(B)** The fit of the function $S_k(\alpha\beta) = (\alpha+\beta)_k$ defined by Expression (15). Being separated by the power-law exponent with $\nu = 0.955$ it represents the perfect straight line. The slope and intercept of this line are given above of this figure. **(C)** The fit of the monotonic decreasing function $<B_k>$ defined by Expression (15). The three fitting parameters of this function can be added to the previous ones for characterization of the given long-time series. **(D)** The behavior of the weight factors with respect to the parameter k. As the significant factor characterizing the behavior of the long-time sequence we use the maximal value $\max(w_1) = 0.5027$. So, from analysis of the **Figure 4** and in this figure we can extract 10 *primary* fitting parameters.

The value mom_p in (19) corresponds to the current moment from the interval $[0, P]$. The value P determines the final value of the linear function Ln_p located in the interval $[mn, mx]$. The values mn and mx define correspondingly the limits of the moments in the uniform logarithmic scale. In many practical cases these values are chosen as $mn = -15$, $mx = 15$ and P is chosen as an integer value located in the interval $[50 \div 100]$. This empirical choice is related to the fact that the transition region of the random sequences considered and expressed in the form of the *GMV*-functions is concentrated usually in the interval $Ln_p \in [-5, 5]$. The extended interval $[-15, 15]$ is taken usually for calculation of the limiting values of this function in the space of the fractional moments. The initial sequences are chosen in that way: the minimum of the GMV-function coincides with zero value while the upper value of this function coincides with the maximal value of the random sequence considered. In formula (18) the random sequence is normalized to the unit value in accordance with Expressions (A) and (B):

$$(A)\ nrm_j(y) = \frac{y_j^{(+)}}{\max\left(y_j^{(+)}\right)} - \frac{y_j^{(-)}}{\min\left(y_j^{(-)}\right)},$$

$$y_j^{(\pm)} = \frac{1}{2}\left(y_j \pm |y_j|\right), \tag{20a}$$

$$(B)\ nrm_j(y) = \frac{\Delta y_j}{\max\left(\Delta y_j\right)},\ \Delta y_j = y_j - \min\left(y_j\right). \tag{20b}$$

$$j = 1, 2, \ldots, N,\ 0 < nrm(y) < 1.$$

Here, as it was done above, the set y_j defines an initial random sequence that can contain a trend or can be compared with another trendless sequence. The symbol $|\ldots|$ and index j ($j = 1.2,\ldots,N$) determine the absolute value and number of the measured points, correspondingly. The second case (B) in [20(b)] corresponds to the case when the initial sequence is positive. If the limits mn and mx in (20) have opposite signs and accept sufficiently large values, then the GPCF function has two plateaus (equaled unit at small numbers of mn (i.e., $GPCF_{mn} = 1$) and another limiting value $GPCF_{mx}$ depends on the degree of internal correlation between two random sequences compared. This right-hand limit (defined as Lm) is located between two values:

$$M \equiv \min\left(GPCF_p\right) \le Lm \equiv GPCF_{mx} \le 1. \tag{21}$$

The appearance of two plateaus implies that all information about possible correlations is complete and further increasing of the limiting numbers (mx, mn) figuring in (19) is *useless*. Numerous

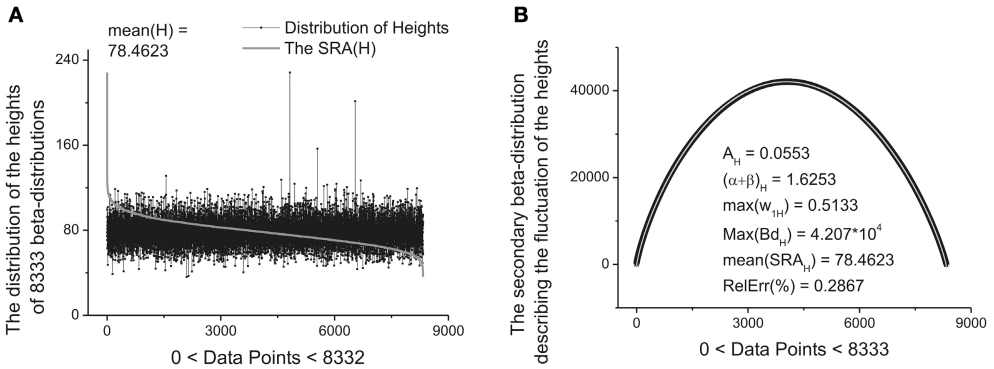


FIGURE 6 | The distribution of the heights of 8333 beta-distributions (when each distribution occupies only 30 data points). (A) Subtracting the mean value of this distribution [mean(H) = 78.4623] one can obtain the bell-like curve again. This curve can be fitted it to the secondary beta-distribution corresponding to the distribution of fluctuations of the heights. **(B)** The fit to beta-distribution function corresponding to fluctuations of the heights. The five fitting parameters of this distribution (shown inside of this figure) can be used as the statistically significant parameters for characterizing of the long-time series considered.

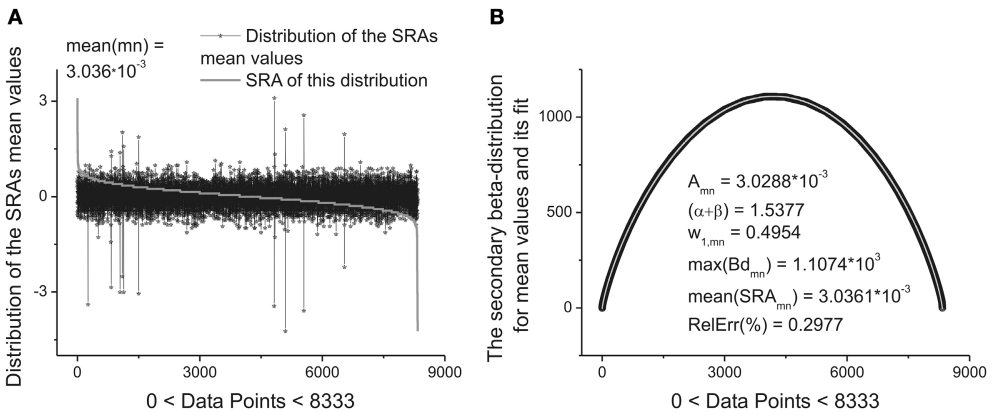


FIGURE 7 | The distribution of the mean values of 8333 beta-distributions (when each distribution occupies only 30 data points.) that were calculated in the initial analysis. (A) Subtracting the mean value of this distribution (mean(mn) = $3.036.10^{-3}$) one can obtain again the bell-like curve. This curve can be fitted it to the secondary beta-distribution corresponding to the distribution of mean values. **(B)** The fit to beta-distribution function corresponding to the fluctuations of the mean values. This information was lost at the preliminary analysis. The five fitting parameters of this distribution (shown inside of this figure) can be used as the statistically significant parameters for characterizing of the long-time series considered. So, in the results of this complete analysis one can obtain 20 statistically significant parameters that can be used for the detailed classification of the long-time series containing $2.5.10^5 \div 10^6$ data points.

tests showed that the high degree of correlations between two random sequences is achieved when $Lm = 1$, while the lowest correlations are observed when $Lm = M$. This empirical observation, having a general character for all random sequences, allows us to introduce new correlation parameter CC (complete correlation)—factor, which is determined as:

$$CC = M \cdot \left(\frac{Lm - M}{1 - M} \right). \tag{22}$$

We would like to stress here that this factor is determined on the *total* set of the fractional moments located between $\exp(mn)$ and $\exp(mx)$. As it was mentioned above, in practical calculations for many cases it is sufficient to put $mn = -15$ and $mx = +15$. The CC-factor accepts the unit values when the degree of correlation is high while the case $Lm = M$ corresponds to the lowest (remnant) degree of correlations that can be observed between the compared random sequences. In addition, we want to stress also the following fact. This CC-factor does *not* depend on the amplitudes of the random sequences. The pair random sequences compared should be normalized to the interval: $0 \leq |y_j| \leq 1$. It reflects the *internal* structure of correlations of the compared random sequences based presumably on the similarity of their probability distribution functions that are *not* known in many cases. Recent example related to application of the statistics of the fractional moments was considered in paper (Nigmatullin et al., 2012). So, the CC-factor (22) can be used for clusterization of the significant parameters based on the following idea. For a set of significant parameters referring to one qualitative factor one can calculate the limits of CC-factor:

$$cf_{\min} \leq CC \leq 1. \quad (23)$$

Here the low correlation limit cf_{min} is determined by the sampling volume and conditions of experiment that should be almost the same for two qualitative factors compared (control/influence of another qualitative factor).

RESULTS

PROCESSING OF THE LONG-TIME MEMBRANE CURRENT SERIES

In previous Section we described in details (**S1**–**S5**) basic steps of treatment of an arbitrary long-time series. Here we want to make some general remarks related to this procedure. If the long-time series considered contains the clearly expressed but random trend then its random behavior can disturb the monotonic behavior of the primary 9 parameters figuring in the fitting functions (15) and (16). In this cases one can recommend to apply the POLS (procedure of the optimal linear smoothing) described in papers (Baleanu et al., 2011; Ciurea et al., 2011; Nigmatullin et al., 2012) or simple numeric differentiation. These two procedures help to suppress the hidden random trend and obtain the monotonic behavior for the 9 parameters figuring in (15) and (16). In the shown figures we used the scaling factor $\xi = 2$. For the rational values of ξ from the interval (1, 2) Expression (9) can be modified as:

$$\Delta_k = \frac{N_{tot}}{\exp\left[(K+1-k)\ln(2)\cdot\mu\right]},\ \mu = \frac{\ln(\xi)}{\ln(2)}. \quad (24)$$

So, numerical calculations realized at $\xi = 1.5$ show that results are *not* changed essentially, only the integer variable k in Expressions (15) and (16) is replaced as $k \to \mu k$. We think that this method has a wide range of its applicability and these two modifications can be taken into account in order to express the long-range time series in terms of 20 significant parameters. In similar manner as it was treated the membrane currents for the randomly taken interneuron-3 one can treat other long-time series related to other (1, 2, 4, 5, 6, 7) interneurons. Besides, in order to differentiate these random sequences recorded *without* presence of a biological object we treated in the same manner 6 random sequences corresponding to empty electrode.

The next problem is associated with the finding of criterion of clusterization that helps to combine these "control" membrane currents to one strongly-correlated cluster based on the values of the significant parameters. For each cell these parameters are collected in **Table 1**. For 6 files corresponding to pure solute (without presence of the cell) the results are collected in **Table 2**. How to differentiate these 20 quantitative parameters (in this case a qualitative factor is associated with the presence/absence of a biological cell) from each other? The simplest classification can be related to calculation of the mean value and standard deviation of the calculated significant parameter in each row. But more effective scheme of clusterization based on the statistics of the fractional moments and the usage of the complete correlation factor is considered in the next section.

For the clusterization of final parameters we have used new correlation parameter *CC* described in "Materials and Methods" section Expression (22). The calculation of the CC-factor (in our case it is based on a set of membrane currents associated with 3 "control" measurements for each chosen cell from the total set of currents representing other 7 biologic cells) which is considered as the complex correlation matrix (see **Table 3**) having minimal dimension (7×7) leads to the minimal value $cf_{min} = 0.9238$. The result is not changed essentially if one calculates numerically the corresponding integrals with respect to their normalized significant parameters and then considers their CC-factors. The tendency of the strong correlations between columns of **Table 1** is conserved, only the boundary of the correlation interval is slightly increased achieving the value $Jcf_{min} = 0.9736$. So, using the method of clusterization based on the statistics of the fractional moments and Expression (22) one can say that all "control" currents measured for the sampling $7 \times 7 = 49$ form the strongly-correlated cluster with limits [0.9238, 1] for the initial set of significant parameters (20 parameters for each sampling) and [0.9736, 1] (for the corresponding integrals that are obtained by direct trapezoid method from the normalized significant parameters). In accordance with this method of clusterization one can make the following conclusion: if any another series having 20 significant parameters will give the CC-factor located in the interval [0.9238, 1] then it can be considered as the "friend" file belonging to this cluster, in the opposite case it can be considered as a "strange" file. For more reliable identification the saying above can be referred to the integrated columns formed from 20 normalized significant parameters. In the same manner we treated the files corresponding to the electrode currents recorded in normal saline solution without presence of biological object. The 20 desired parameters for 6 files are collected in **Table 2**. Their correlation matrix presented by **Table 4** form another cluster. But attempt to combine the currents corresponding to the living o cells with currents corresponding to empty electrodes located in saline solution is *unsuccessful*. If we compare the correlation matrix of **Table 5** with the previous ones (**Tables 3**, **4**) then one can notice that the last matrix is *uncorrelated* (all elements are close to zero). It means that the presence of the biologic cell completely changes the statistical structure of the current and from qualitative point of view the long-time random sequences of currents recorded for both cases (presence/absence of biological cell) are *different*.

So, new clusterization method helps to express quantitatively the internal factor as the presence/absence of the living cell (compare this statement with series shown on **Figure 1** where the corresponding currents look similar to each other). Definitely, more accurate measurements are needed in order to differentiate from many mixed factors that form a time-series for biological and non-biological objects a specific *predominant* factor that plays an essential role in this differentiation. But this problem merits a separate research.

DISCUSSION

It is well known that cellular membrane is the element which largely provides cell functioning. Cell membrane has so many functions that it is difficult even to list—anyone can find them all in each textbook on cell biology. In general the membrane provides all interaction of the cell with the external environment including the perception of the effect of active substances. Withal

Table 1 | The collection of 20 significant parameters calculated for 7 cells based on calculation of registered membrane currents.

Parameter	Cell-1	Cell-2	Cell-3	Cell-4	Cell-5	Cell-6	Cell-7
$\max(w_1)$	**0.4997**	0.5006	**0.5027**	0.5027	0.5019	0.5006	0.5003
λ_a	−0.3605	**−0.3613**	**−0.3517**	−0.3571	−0.3586	−0.3581	−0.3604
A_1	0.5719	0.5943	**1.2790**	**0.5683**	0.6528	0.8815	0.8768
A_0	−0.00149	−0.00149	**−0.00328**	**−0.00147**	−0.00167	−0.00224	−0.00219
ν	0.995	0.995	0.995	0.995	0.995	0.995	0.995
A_{pl}	**1.558**	1.558	**1.544**	1.552	1.554	1.553	1.556
B_{pl}	0.1308	0.1295	**0.1374**	**0.1277**	0.1306	0.1315	0.129
λ_B	**0.6666**	0.6650	0.6646	0.6654	0.6641	**0.6640**	0.6656
B_1	**−0.2111**	−0.2418	**−0.4333**	−0.2266	−0.2438	−0.3318	−0.3959
B_0	**7.807**	8.801	**15.60**	8.247	8.717	11.89	14.56
A_H	0.0292	0.0314	**0.0708**	**0.0265**	0.0379	0.0653	0.0440
$(\alpha + \beta)_H$	1.604	1.598	1.596	**1.613**	1.589	**1.560**	1.606
$\max(w_{1\,H})$	0.5208	**0.5136**	0.5158	0.5233	0.5160	**0.5257**	0.5174
$\max(Bd_H)$	18580	18980	**42070**	**18110**	21110	29110	28300
$\mathrm{mn}(\mathrm{SRA}_H)$	35.68	37.01	**78.46**	**34.87**	40.42	54.57	54.30
A_{mn}	0.00212	**0.00107**	**0.00303**	0.00123	0.00139	0.00242	0.00134
$(\alpha + \beta)_{mn}$	**1.517**	1.57	1.538	1.576	1.552	1.53	**1.592**
$\max(w_{1\,mn})$	0.5003	0.4997	**0.4954**	0.5011	**0.5013**	0.4991	0.5
$\max(Bd_{mn})$	650.4	**506.9**	**1107.0**	614.5	570.7	837.3	766.1
$\mathrm{mn}(SRA_{mn})$	$4.986\cdot10^{-4}$	$2.277\cdot10^{-4}$	**0.00304**	$2.205\cdot10^{-4}$	**−0.00157**	$6.655\cdot10^{-4}$	0.00258

Each column describing the chosen cell is obtained in the result of the averaging of three membrane currents with the length 250,000 data points. The first 10 primary parameters are marked by a double line. The minimal and maximal values of each significant parameter in each row are bolded.

Table 2 | The collection of 20 significant parameters calculated for 6 files corresponding to currents recorded with the empty electrode placed inside an artificial cerebrospinal fluid (the biological material is absent).

Parameter	File-1	File-2	File-3	File-4	File-5	File-6
$\max(w_1)$	0.5007	0.5029	0.4994	0.501	**0.4993**	**0.5032**
λ_a	**−0.3567**	**−0.3703**	−0.3576	−0.3623	−0.3674	−0.3606
A_1	$5.455\cdot10^{-9}$	$5.488\cdot10^{-9}$	$5.457\cdot10^{-9}$	$\mathbf{5.502\cdot10^{-9}}$	$\mathbf{5.417\cdot10^{-9}}$	$5.441\cdot10^{-9}$
A_0	$\mathbf{-2.555\cdot10^{-11}}$	$\mathbf{-1.282\cdot10^{-11}}$	$-1.47\cdot10^{-11}$	$-1.413\cdot10^{-11}$	$-1.358\cdot10^{-11}$	$-1.423\cdot10^{-11}$
ν	0.995	0.995	0.995	0.995	0.995	0.995
A_{pl}	1.562	**1.569**	1.557	1.562	**1.569**	**1.560**
B_{pl}	0.1242	**0.1278**	0.1315	0.1290	0.1294	**0.1320**
λ_B	**0.6395**	0.6506	0.6859	0.6398	0.6506	**0.6944**
B_1	$\mathbf{-3.349\cdot10^{-9}}$	$-2.479\cdot10^{-9}$	$-1.986\cdot10^{-9}$	$-3.336\cdot10^{-9}$	$-2.479\cdot10^{-9}$	$\mathbf{-1.649\cdot10^{-9}}$
B_0	$\mathbf{1.023\cdot10^{-7}}$	$\mathbf{7.672\cdot10^{-8}}$	$9.194\cdot10^{-8}$	$9.272\cdot10^{-8}$	$7.675\cdot10^{-8}$	$8.584\cdot10^{-8}$
A_H	$3.202\cdot10^{-10}$	$2.211\cdot10^{-10}$	$\mathbf{6.638\cdot10^{-12}}$	$3.232\cdot10^{-10}$	$2.231\cdot10^{-10}$	$\mathbf{3.238\cdot10^{-10}}$
$(\alpha + \beta)_H$	1.591	1.637	**1.585**	**1.721**	1.631	1.592
$\max(w_{1\,H})$	0.5104	0.5136	**0.4986**	**0.5704**	0.5436	0.5131
$\max(Bd_H)$	$1.820\cdot10^{-4}$	$1.847\cdot10^{-4}$	$\mathbf{3.583\cdot10^{-6}}$	$\mathbf{2.920\cdot10^{-4}}$	$1.907\cdot10^{-4}$	$1.853\cdot10^{-4}$
$\mathrm{mn}(\mathrm{SRA}_H)$	$\mathbf{3.471\cdot10^{-7}}$	$3.469\cdot10^{-7}$	$\mathbf{2.775\cdot10^{-13}}$	$3.171\cdot10^{-7}$	$3.369\cdot10^{-7}$	$3.448\cdot10^{-7}$
A_{mn}	$\mathbf{6.094\cdot10^{-12}}$	$\mathbf{8.315\cdot10^{-12}}$	$6.638\cdot10^{-12}$	$6.014\cdot10^{-12}$	$8.227\cdot10^{-12}$	$6.731\cdot10^{-12}$
$(\alpha + \beta)_{mn}$	**1.600**	1.563	1.585	1.556	1.569	**1.524**
$\max(w_{1\,mn})$	0.5009	**0.4964**	0.4986	0.5109	0.5064	**0.5169**
$\max(Bd_{mn})$	$\mathbf{3.733\cdot10^{-6}}$	$3.709\cdot10^{-6}$	$3.583\cdot10^{-6}$	$\mathbf{3.233\cdot10^{-6}}$	$3.711\cdot10^{-6}$	$3.385\cdot10^{-6}$
$\mathrm{mn}(SRA_{mn})$	$\mathbf{1.064\cdot10^{-11}}$	$7.818\cdot10^{-12}$	$2.775\cdot10^{-13}$	$1.004\cdot10^{-11}$	$7.821\cdot10^{-12}$	$\mathbf{2.557\cdot10^{-13}}$

The first 10 primary parameters are marked by a double line. The minimal and maximal values of each significant parameter in each row are bolded.

Table 3 | The correlation matrix of the calculated CC-factors [Expression (22)] for 20 parameters characterizing 7 neurons collected in the Table 1.

Cells	Cell-1	Cell-2	Cell-3	Cell-4	Cell-5	Cell-6	Cell-7
Cell-1	1	0.99876	0.92838	**0.99957**	0.99841	0.9767	**0.94824**
Cell-2	0.99876	1	**0.93615**	0.99954	**0.99981**	0.98465	0.95698
Cell-3	0.92838	0.93615	1	**0.93354**	0.93708	0.97193	**0.99714**
Cell-4	0.99957	0.99954	0.93354	1	**0.99933**	0.98166	**0.95451**
Cell-5	0.99841	0.99981	0.93708	0.99933	1	**0.98558**	**0.95776**
Cell-6	0.9767	0.98465	0.97193	0.98166	0.98558	1	0.9804
Cell-7	0.94824	0.95698	0.99714	0.95451	0.95776	0.9804	1

The maximal and minimal values of correlations in each row are bolded.

Table 4 | The correlation matrix of the calculated CC-factors for 20 parameters characterizing 6 empty electrode records collected in the Table 2.

Files	F-1	F-2	F-3	F-4	F-5	F-6
F-1	1	0.99581	**0.97238**	0.99076	**0.99642**	0.99624
F-2	0.99581	1	**0.97241**	**0.99767**	0.99995	0.99951
F-3	0.97238	0.97241	1	**0.97008**	**0.97244**	0.97237
F-4	0.99076	0.99767	0.97008	1	**0.99706**	**0.99588**
F-5	0.99642	0.99995	0.97244	0.99706	1	0.9996
F-6	0.99624	0.9995	0.97237	0.99588	0.9996	1

The maximal and minimal values of correlations in each row are bolded.

Table 5 | The correlation matrix of the CC-factors calculated for 7 cells and 6 empty electrodes.

Cells\Files	F-1	F-2	F-3	F-4	F-5	F-6
Cell-1	0.01809	0.01807	0.01387	0.01883	0.01805	0.01781
Cell-2	0.01772	0.01769	0.01357	0.01844	0.01767	0.01743
Cell-3	0.00889	0.00887	0.00666	0.00929	0.00886	0.00873
Cell-4	0.01807	0.01805	0.01386	0.0188	0.01802	0.01778
Cell-5	0.01679	0.01676	0.01282	0.01748	0.01674	0.01652
Cell-6	0.01343	0.01341	0.01018	0.014	0.01339	0.0132
Cell-7	0.01212	0.0121	0.00918	0.01264	0.01208	0.01191

the membrane comprises a lot of elements which produce so-called "membrane noise"—rather small variations of membrane potential or trans-membrane current; mainly they are different types of ion channels, transporters and pumps. There are many active substances affecting the operation of these elements so the action of these substances actually can be detected by analyzing the membrane noise. But even if some substance does not affect channels, transporters or pumps directly its action often can be detected by noise analysis too. For example if the substance affects G protein-coupled receptors or state of membrane lipids—in many cases it leads to the changes in the functioning of ion channels (Tillman and Cascio, 2003; Inanobe and Kurachi, 2014) and, accordingly, to the noise changes. So the analysis of the long-time series of noise can help to detect the action of many substances when we cannot detect this action differently.

For analysis of the long-time series we applied new BRC method based on the beta-distribution function. Four parameters of the beta-distribution function can be used for description of the local fluctuations and the averaged beta-distributions can be applied for *quantitative* reading of series containing large number of data points. The fluctuation spectroscopy based on beta distribution allows realizing the essential reduction $(2.5–10).10^5$ data points to 20 quantitative parameters *only* [see Expressions (14)–(16)] that contain the basic information calculated from three basic beta-distributions: (1) distribution over different segments (scales), (2) the secondary beta-distributions over their heights and (3) distributions over mean values. This reduction becomes possible thanks to the invariant properties that are expressed by formulae (3) and (5). We suppose that this approach can be applied successfully for the unified additional analysis of fluctuations of different long-time series that present the results of monitoring of biological, medical and other data reflecting the results of response of the complex system considered with respect to some external factor. In particular, this BRC method is applicable to testing the action of antagonist of receptor and ion channels when the modification based on different type of interaction (with binding site or with the open channel with different kinetics). In such experiments in order to understand the

mechanism of action of some new substances we only need to compare the FSBD parameter changes caused by this substance with typical changes stored in the database.

FUNDING

This work was partially supported (Andrei I. Skorinkin) by RF grant "Leading Scientific School" and RFBR grant.

ACKNOWLEDGMENTS

We are grateful to professor Peter Illes (Leipzig University, Germany) for the possibility to receive the used experimental data in his laboratory, we also thank professor Sverre Holm (University of Oslo, Norway) for useful discussions. The work is performed according to the Russian Government Program of Competitive Growth of Kazan Federal University.

REFERENCES

Alvarez, O., Gonzalez, C., and Latorre, R. (2002). Counting channels: a tutorial guide on ion channel fluctuation analysis. *Adv. Physiol. Educ.* 26, 327–341. doi: 10.1152/advan.00006.2002

Baleanu, C. M., Nigmatullin, R. R., Cetin, S. S., Baleanu, D., and Ozcelik, S. (2011). New method and treatment technique applied to interband transition in $GaAs_{1-x}$ P_x ternary alloys. *Cent. Eur. J. Phys.* 9, 729–739. doi: 10.2478/s11534-010-0068-y

Burr, R. L., Kirkness, C. J., and Mitchell, P. H. (2008). Detrended fluctuation analysis of the ICP predicts outcome following traumatic brain injury. *IEEE Trans. Biomed. Eng.* 55, 2509–2518. doi: 10.1109/TBME.2008.2001286

Ciurea, M. L., Lazanu, S., Stavaracher, I., Lepadatu, A.-M., Iancu, V., Mitroi, M. R., et al. (2011). Stress-induced traps in multilayered structures. *J. Appl. Phys.* 109, 013717. doi: 10.1063/1.3525582

Feder, J. (1988). *Fractals.* New York, NY: Plenum Press.

Gao, J. B., Hu, J., Tung, W. W., and Cao, Y. H. (2006). Distinguishing chaos from noise by scale-dependent Lyapunov exponent. *Phys. Rev. E* 74, 066204. doi: 10.1103/PhysRevE.74.066204

Gao, J. B., Sultan, H., Hu, J., and Tung, W. W. (2010). Denoising nonlinear time series by adaptive filtering and wavelet shrinkage: a comparison. *IEEE Signal Proc. Let.* 17, 237–240. doi: 10.1109/LSP.2009.2037773

Gao, J. B., Hu, J., and Tung, W. W. (2011). Facilitating joint chaos and fractal analysis of biosignals through nonlinear adaptive filtering. *PLoS ONE* 6:e24331. doi: 10.1371/journal.pone.0024331

Gao, J. B., Hu, J., Mao, X., and Perc, M. (2012a). Culturomics meets random fractal theory: insights into long-range correlations of social and natural phenomena over the past two centuries. *J. R. Soc. Int.* 9, 1956–1964. doi: 10.1098/rsif.2011.0846

Gao, J. B., Hu, J., Tung, W. W., and Blasch, E. (2012b). Multiscale analysis of physiological data by scale-dependent Lyapunov exponent. *Front. Physiol.* 2:110. doi: 10.3389/fphys.2011.00110

Gao, J. B., Gurbaxani, B. M., Hu, J., Heilman, K. J., Emanuele, V. A., Lewis, G. F., et al. (2013). Multiscale analysis of heart rate variability in nonstationary environments. *Front. Physiol.* 4:119. doi: 10.3389/fphys.2013.00119

Hausdorff, J. M., Peng, C.-K., Ladin, Z., Wei, J. Y., and Goldberger, A. L. (1995). Is walking a random walk? Evidence for long-range correlations in the stride interval of human gait. *J. Appl. Physiol.* 78, 349–358.

Hausdorff, J. M., Purdon, P., Peng, C.-K., Ladin, Z., Wei, J. Y., and Goldberger, A. L. (1996). Fractal dynamics of human gait: stability of long-range correlations in stride interval fluctuations. *J. Appl. Physiol.* 80, 1448–1457.

Hu, J., Gao, J. B., Tung, W. W., and Cao, Y. H. (2010). Multiscale analysis of heart rate variability: a comparison of different complexity measures. *Ann. Biomed. Eng.* 38, 854–864. doi: 10.1007/s10439-009-9863-2

Inanobe, A., and Kurachi, Y. (2014). Membrane channels as integrators of G-protein-mediated signaling. *Biochim. Biophys. Acta* 1838, 521–531. doi: 10.1016/j.bbamem.2013.08.018

Jospin, M., Caminal, P., Jensen, E. W., Litvan, H., Vallverdu, M., Struys, R. F., et al. (2007). Detrended fluctuation analysis of EEG as a measure of depth of anesthesia. *IEEE Trans. Biomed. Eng.* 54, 840–846. doi: 10.1109/TBME.2007.893453

Kantelhardt, J. W., Koscielny-Bunde, E., Rego, H. H. A., Havlin, S., and Bunde, A. (2001). Detecting long-range correlations with detrended fluctuation analysis. *Physica A* 295, 441–454. doi: 10.1016/S0378-4371(01)00144-3

Kuznetsov, N., Bonnette, S., Gao, J. B., and Riley, M. A. (2013). Adaptive fractal analysis reveals limits to fractal scaling in center of pressure trajectories. *Ann. Biomed. Eng.* 41, 1646–1660. doi: 10.1007/s10439-012-0646-9

Läuger, P. (1985). Structural fluctuations and current noise of ionic channels. *Biophys. J.* 48, 369–373. doi: 10.1016/S0006-3495(85)83793-0

Neher, E., and Sakmann, B. (1976). Noise analysis of drug induced voltage clamp currents in denervated frog muscle fibres. *J. Physiol.* 258, 705–729.

Nigmatullin, R. R. (2010). Universal distribution function for the strongly-correlated fluctuations: general way for description of random sequences. *Commun. Nonlinear Sci. Numer. Simulat.* 15, 637–647. doi: 10.1016/j.cnsns.2009.05.019

Nigmatullin, R. R., Ionescu, C., and Baleanu, D. (2012). NIMRAD: novel technique for respiratory data treatment. *J. Signal Image Video Process*. doi: 10.1007/s11760-012-0386-1

Ossadnik, S. M., Buldyrev, S. V., Goldberger, A. L., Havlin, S., Mantegna, R. N., Peng, C.-K., et al. (1994). Correlation approach to identify coding regions in DNA sequences. *Biophys. J.* 67, 64–70. doi: 10.1016/S0006-3495(94)80455-2

Peng, C.-K., Buldyrev, S. V., Havlin, S., Simons, M., Stanley, H. E., and Goldberger, A. L. (1994). Mosaic organization of DNA nucleotides. *Phys. Rev. E* 49, 1685–1689. doi: 10.1103/PhysRevE.49.1685

Peng, C.-K., Havlin, S., Stanley, H. E., and Goldberger, A. L. (1995). Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series. *Chaos* 5, 82–87. doi: 10.1063/1.166141

Penzel, T., Kantelhardt, J. W., Grote, L., Peter, J.-H., and Bunder, A. (2003). Comparison of deterended fluctuation analysis and spectral analysis for heart rate variability in sleep and sleep apnea. *IEEE Trans. Biomed. Eng.* 50, 1143–1151. doi: 10.1109/TBME.2003.817636

Riley, M. A., Kuznetsov, N., Bonnette, S., Wallot, S., and Gao, J. B. (2012). A tutorial introduction to adaptive fractal analysis. *Front. Physiol.* 3, 1–10. doi: 10.3389/fphys.2012.00371

Sigworth, F. J. (1980). The variance of sodium current fluctuations at the node of Ranvier. *J. Gen. Physiol.* 307, 97–129.

Sigworth, F. J. (1985). Open channel noise. I. Noise in acetylcholine receptor currents suggests conformational fluctuations. *Biophys. J.* 47, 709–720. doi: 10.1016/S0006-3495(85)83968-0

Sigworth, F. J. (1986). Open channel noise. II. A test for coupling between current fluctuations and conformational transitions in the acetylcholine receptor. *Biophys. J.* 49, 1041–1046. doi: 10.1016/S0006-3495(86)83732-8

Tillman, T. S., and Cascio, M. (2003). Effects of membrane lipids on ion channel structure and function. *Cell Biochem. Biophys.* 38, 161–190. doi: 10.1385/CBB:38:2:161

Traynelisa, S. F., and Jaramilloa, F. (1998). Getting the most out of noise in the central nervous system. *Trends Neurosci.* 21, 137–145. doi: 10.1016/S0166-2236(98)01238-7

Venkataramanan, L., and Sigworth, F. J. (2002). Applying hidden Markov models to the analysis of single ion channel activity. *Biophys. J.* 82, 1930–1942. doi: 10.1016/S0006-3495(02)75542-2

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.

Received: 04 June 2014; paper pending published: 11 July 2014; accepted: 08 September 2014; published online: 26 September 2014.

Citation: Nigmatullin RR, Giniatullin RA and Skorinkin AI (2014) Membrane current series monitoring: essential reduction of data points to finite number of stable parameters. Front. Comput. Neurosci. ***8****:120. doi: 10.3389/fncom.2014.00120*

This article was submitted to the journal Frontiers in Computational Neuroscience.

frontiers in
COMPUTATIONAL NEUROSCIENCE

METHODS ARTICLE
published: 01 October 2013
doi: 10.3389/fncom.2013.00122

Fast monitoring of epileptic seizures using recurrence time statistics of electroencephalography

***Jianbo Gao*[1,2]* and *Jing Hu*[1,2]**

[1] *Institute of Complexity Science and Big Data Technology, Guangxi University, Nanning, China*
[2] *PMB Intelligence LLC, Sunnyvale, CA, USA*

Edited by:
Tobias A. Mattei, Ohio State University, USA

Reviewed by:
Tobias A. Mattei, Ohio State University, USA
Paul Rapp, Uniformed Services University of the Health Sciences, USA

***Correspondence:**
Jianbo Gao, Institute of Complexity Science and Big Data Technology, Guangxi University, 100 Daxue Road, Nanning 530005, China
e-mail: jbgao.pmb@gmail.com

Epilepsy is a relatively common brain disorder which may be very debilitating. Currently, determination of epileptic seizures often involves tedious, time-consuming visual inspection of electroencephalography (EEG) data by medical experts. To better monitor seizures and make medications more effective, we propose a recurrence time based approach to characterize brain electrical activity. Recurrence times have a number of distinguished properties that make it very effective for forewarning epileptic seizures as well as studying propagation of seizures: (1) recurrence times amount to periods of periodic signals, (2) recurrence times are closely related to information dimension, Lyapunov exponent, and Kolmogorov entropy of chaotic signals, (3) recurrence times embody Shannon and Renyi entropies of random fields, and (4) recurrence times can readily detect bifurcation-like transitions in dynamical systems. In particular, property (4) dictates that unlike many other non-linear methods, recurrence time method does not require the EEG data be chaotic and/or stationary. Moreover, the method only contains a few parameters that are largely signal-independent, and hence, is very easy to use. The method is also very fast—it is fast enough to on-line process multi-channel EEG data with a typical PC. Therefore, it has the potential to be an excellent candidate for real-time monitoring of epileptic seizures in a clinical setting.

Keywords: EEG, recurrence time, seizure detection, seizure propagation, brain complexity

1. INTRODUCTION

Epilepsy is a relatively common brain disorder which may be very debilitating. It affects approximately 1% of the world population (Jallon, 1997) and three million people in the United States alone. It is characterized by intermittent seizures. During a seizure, the normal activity of the central nervous system is disrupted. The concrete symptoms include abnormal running/bouncing fits, clonus of face and forelimbs, or tonic rearing movement as well as simultaneous occurrence of transient EEG signals such as spikes, spike and slow wave complexes or rhythmic slow wave bursts. Clinical effects may include motor, sensory, affective, cognitive, automatic and physical symptomatology. Although epilepsy can be treated effectively in many instances, severe side effects may result from constant medication. Even worse, some patients may become drug-resistant not long after treatment. To make medications more effective, timely detection of seizure is very important.

In the past several decades, considerable efforts have been made to detect/predict seizures through non-linear analysis of EEGs (Kanz and Schreiber, 1997; Gao et al., 2007). Representative non-linear methods proposed for seizure prediction/detection include approaches based on correlation dimension (Lehnertz and Elger, 1995, 1997; Martinerie et al., 1998; Aschenbrenner-Scheibe et al., 2003), Kolmogorov entropy (van Drongelen et al., 2003), permutation entropy (Cao et al., 2004), short time largest Lyapunov exponent (STLmax) (Iasemidis et al., 1990; Lai et al., 2003), dissimilarity measures (Protopopescu et al., 2001; Quyen et al., 2001), long-range-correlation (Hwa and Ferree, 2002; Gao et al., 2006b, 2007, 2011b; Valencia et al., 2008), power-law sensitivity to initial conditions (Gao et al., 2005b), scale-dependent Lyapunov exponent (SDLE) (Gao et al., 2006a, 2012a,b), and synthesis of linear/non-linear methods by using neural networks (Adeli et al., 2007). Readers interested in "what is epilepsy, where, when, and why (how) do seizures occur?" are referred to the April, 2007 issue of *Journal of Clinical Neurophysiology*.

Note that most of the proposed methods assume that EEG signals are chaotic and stationary. As a result, they tend to have performances that are signal- and patient-dependent due to the noisy and non-stationary nature of the EEG within and across patients. In addition, they are computationally expensive. Consequentially, studies of epilepsy still heavily involve visual inspection of multi-channel EEG signals by medical experts. Visual inspection of long (e.g., tens of hours or days) EEG data is, however, tedious, time-consuming, and in-efficient. Therefore, it is important to develop new non-linear seizure monitoring approaches.

In this paper, we explore recurrence time based analysis of EEG (Gao, 1999, 2001; Gao and Cai, 2000; Gao et al., 2003), with the goal of potentially on-line monitoring the occurrence and propagation of seizures. The method does not assume that the underlying dynamics of EEGs be chaotic or stationary. More importantly, it has been tested to be able to readily detect very subtle changes in signals (Gao, 2001; Gao et al., 2003).

When developing a new method, it is important to compare its performance with that of existing methods. For seizure

detection, such a task has been greatly simplified by our recent studies (Gao et al., 2011a, 2012a). By comparing seizure detection using a variety of complexity measures from deterministic chaos theory, random fractal theory, and information theory, we have found that the variations of those complexity measures with time have two patterns—either similar or reciprocal (Gao et al., 2011a). More importantly, we have gained fundamental understanding about the connections among different complexity measures through a new multiscale complexity measure, the SDLE. These results are recapitulated in **Figure 1**. While we leave the details to our prior works (Gao et al., 2006a, 2007, 2012a,b), these results suggest that it would be sufficient for us to compare the performance of the recurrence time based method for seizure detection with the performance of any of the existing complexity measures. Since some of the EEG data examined here had also been analyzed by the STLmax method and documented results exist, we shall compare our recurrence time method with the STLmax method. We shall show that the recurrence time method is both more accurate and faster than the STLmax method in detecting seizures from EEG.

The remainder of the paper is organized as follows. In section 2, we describe the data used here and the recurrence time method and the STLmax method for seizure detection. In section 3, we compare the performance of the recurrence time and STLmax method for seizure detection, as well as study seizure propagation. In section 4, we make a few concluding remarks.

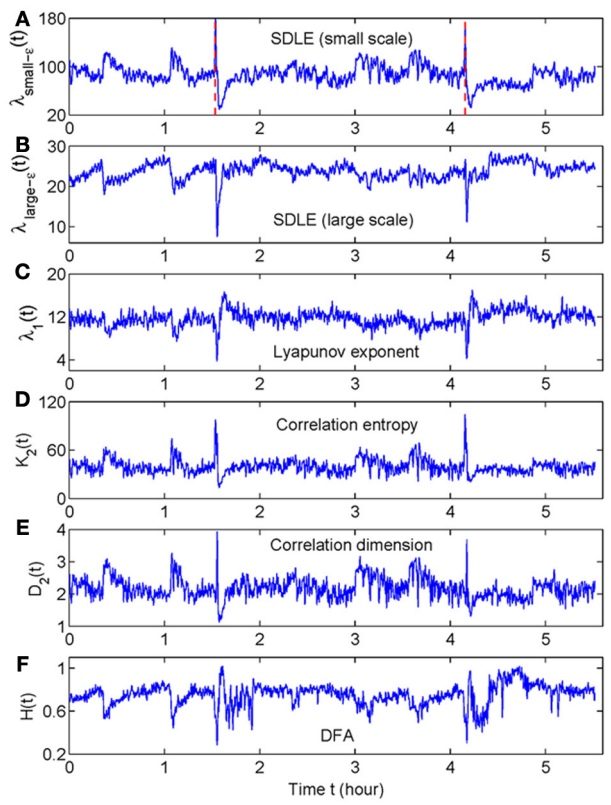


FIGURE 1 | The variation with time of (A) $\lambda_{small-\varepsilon}$, (B) $\lambda_{large-\varepsilon}$, (C) Lyapunov exponent, (D) correlation entropy, (E) correlation dimension, and (F) the Hurst parameter obtained using DFA. Adapted from Gao et al. (2011a).

2. MATERIALS AND METHODS

In this section, we first describe EEG data used here, then describe the recurrence time method and the short-time Lyapunov exponent (STLmax) method.

2.1. DATA

The EEG signals analyzed here are human EEG. They were recorded intracranially with approved clinical equipment by the Shands hospital at the University of Florida, with a sampling frequency of 200 Hz. **Figure 2** shows our typical 28 electrode montage used for subdural and depth recordings.

Intracranial EEG is also called depth EEG, and is considered less contaminated by noise or motion artifacts. However, the clinical equipment used to measure the data has a pre-set, unadjustable maximal amplitude, which is around 5300 μV. This causes clipping of the signals when the signal amplitude is higher than this threshold. This is often the case during seizure episodes, especially for certain electrodes. To a certain extent, clipping complicates seizure detection, since certain seizure signatures may not be captured by the measuring equipment. However, we did not apply any filtering or conditioning methods to preprocess the raw EEG signals when we use our recurrence time method. The good results presented below thus suggest that the method is very reliable.

Altogether we have data of seven patients. The total duration of the measurement for each patient was up to about 3 days, as shown in the 2nd column of **Table 1**. There were only one or a few

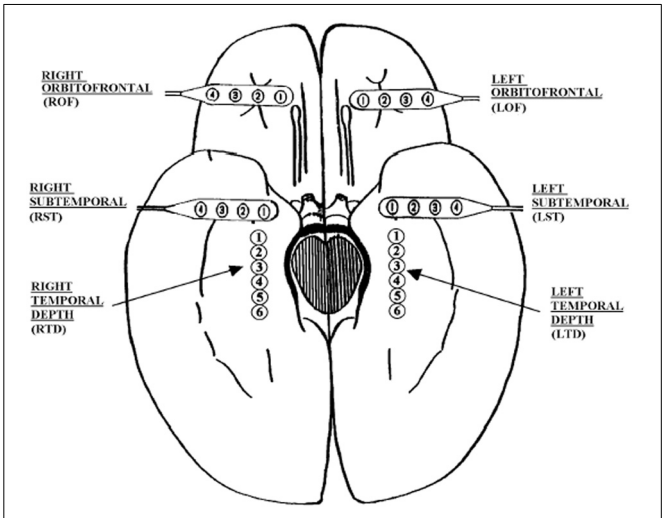


FIGURE 2 | Schematic diagram of the depth and subdural electrode placement. This view from the inferior aspect of the brain shows the approximate location of depth electrodes, oriented along the anterior–posterior plane in the hippocampi (RTD, right temporal depth; LTD, left temporal depth), and subdural electrodes located beneath the orbitofrontal and subtemporal cortical surfaces (ROF, right orbitoftrontal; LOF, left orbitofrontal; RST, right subtemporal; LST, left subtemporal).

Table 1 | Performance of the T_2 and the STLmax method for seven patients' data.

Data set	Length (hours)	Total number of seizures	STLmax performance		T2 performance	
			Sensitivity (%) Overall: 74%	False alarm per hour Mean: 0.05	Sensitivity (%) Overall: 83%	False alarm per hour Mean: < 0.01
P92	35	7	100	0.09	100	0.00
P93	64	23	78	0.02	78	0.02
P148	76	17	58	0.07	76	0.00
P185	47	19	73	0.02	89	0.04
P40	5.3	1	100	0.00	100	0.00
P256	4.5	1	100	0.00	100	0.00
P130	5.7	2	50	0.18	100	0.00

The total number of seizures was determined by examining clinical symptons and all 28 channel video-EEG data by medical experts. Note the five missed seizures for patient P93 are all subclinical seizures, whose information does not appear to be reflected by the EEG dynamics.

seizures for some patients while there were several tens of seizures for some other patients, as shown in the 3rd column of **Table 1**. Some of the seizures were considered subclinical, i.e., not manifested in the EEG signals. Sometimes the EEG signals may contain signatures distinctly different from background non-seizure signals, due to, for example, the fact that the patient may be eating food, drinking, etc. These non-seizure signatures typically may also be picked up by a seizure monitoring method. In this study, we shall focus on the behavior of the recurrence time and STLmax method in detecting seizures using only three channels EEG data without any preprocessing. As we shall see later, reliable decisions can be made based on single channel EEG data. There appears to be no need to combine multiple channels data.

2.2. RECURRENCE TIME BASED METHOD FOR SEIZURE DETECTION

The method involves first partitioning a long EEG signal into (overlapping or non-overlapping) blocks of data sets of short length k, and compute the so-called mean recurrence time of the 2nd type, $\overline{T}_2(r)$, for each data subset. For non-stationary and transient time series, it has been found (Gao, 1999, 2001; Gao and Cai, 2000; Gao et al., 2003) that $\overline{T}_2(r)$ will be different for different blocks of data subsets.

Let us first define the recurrence time of the 2nd type. Suppose we are given a scalar time series $\{x(i), i = 1, 2, \ldots\}$. We first construct vectors of the form: $X_i = [x(i), x(i + L), \ldots, x(i + (m - 1)L)]$, with m being the embedding dimension and L the delay time (Packard et al., 1980; Takens, 1981; Sauer et al., 1991). $\{X_i, i = 1, 2, \ldots, N\}$ then represents certain trajectory in a m-dimensional space. Next, we arbitrarily choose a reference point X_0 on the reconstructed trajectory, and consider recurrences to its neighborhood of radius r: $B_r(X_0) = \{X : \|X - X_0\| \leq r\}$. The recurrence points of the 2nd type are defined as the set of points comprised of the first trajectory point getting inside the neighborhood from outside. These are denoted as the dark solid circles in **Figure 3**. The trajectory may stay inside the neighborhood for a while, thus generating a sequence of points, as designated by open circles in **Figure 3**. These are called sojourn points (Gao, 1999). It is clear that there will be more such points when the size of the neighborhood gets larger as well as when the trajectory is sampled more densely. The summation of the recurrence points of the second kind and the sojourn points is called the recurrence points of the first kind. These are often called nearest neighbors of the reference point X_0, and have been used by all other chaos theory-based non-linear methods.

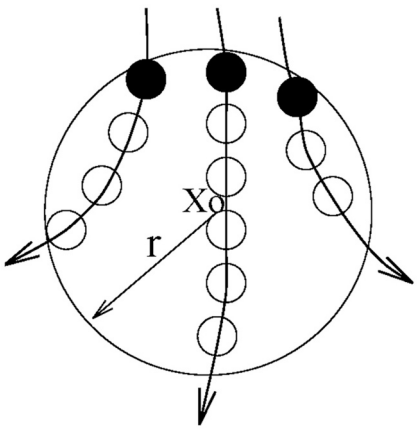


FIGURE 3 | A schematic showing the recurrence points of the second type (solid circles) and the sojourn points (open circles) in $B_r(X_0)$.

Let us be more precise mathematically. We denote the recurrence points of the 1st type by $S_1 = \{X_{t_1}, X_{t_2}, \ldots, X_{t_i} \ldots\}$, and the corresponding Poincare recurrence time of the 1st type by $\{T_1(i) = t_{i+1} - t_i, i = 1, 2, \ldots\}$. Note the time is computed based on successive returns, not based on the returning points and the reference point. Also note $T_1(i)$ may be 1 (for continuous time systems, this means one unit of the sampling time), for some i. This occurs when there are at least one sojourn point. Existence of such points makes further quantitative analysis difficult. Thus, we remove the sojourn points from the set S_1 (which can be easily achieved by monitoring whether the recurrence times of the first type are one or not). Let us denote the remaining set by $S_2 = \{X_{t'_1}, X_{t'_2}, \ldots, X_{t'_i} \ldots\}$. S_2 then defines a time sequence $\{T_2(i) = t'_{i+1} - t'_i, i = 1, 2, \ldots\}$. These are called the recurrence times of the 2nd type.

$T_2(i)$ has a number of interesting properties: (1) For periodic motions, so long as the size of the neighborhood is not too large, $T_2(i)$ accurately estimates the period of the motion. (2) For discrete sequences, the entire Renyi entropy spectrum can be computed from the moments of T_2 (Gao et al., 2005a). (3) For chaotic motions, $T_2(i)$ is closely related to the Lyapunov exponent, and hence, Kolmogorov entropy (Gao and Cai, 2000). (4) For chaotic motions, $T_2(i)$ is related to the information dimension d_1 by a simple scaling law (Gao, 1999; Gao et al., 2003),

$$T_2(r) \sim r^{d_1-\alpha}, \tag{1}$$

where α takes on value 0 or 1, depending on whether the sojourn points form very few isolated points inside the neighborhood $B_r(X_0)$, thus contribute dimension 0, or form a smooth curve inside $B_r(X_0)$, thus contribute dimension 1. These properties make the recurrence time based method very versatile and powerful in detecting signal transitions.

We now explain how the mean recurrence time of the 2nd type can be computed. We simply evaluate this quantity for every reference point in a window, then take the mean of those times. Such calculation is carried out for all the data subsets, resulting in a curve which describes how $\overline{T}_2(r)$ varies with time. It has been observed (Gao, 1999, 2001; Gao and Cai, 2000; Gao et al., 2003) that the variations of $\overline{T}_2(r)$ coincide very well with sudden changes in the signal dynamics, such as bifurcations or transitions from regular motions to chaotic motions in non-stationary data, and vise versa. An example is shown in **Figure 4** using the transient logistic map described by

$$x(n+1) = a(n)x(n)[1-x(n)], \quad a(n) = a(n-1) + 10^{-5} \tag{2}$$

We observe from **Figure 4** that the method not only detects all the bifurcations in the signal, but also gives the exact periods of periodic signals. Note that some changes in a signal may be difficult to detect visually (Gao, 2001).

Since there are altogether four parameters involved, namely, the embedding dimension m and delay time L, the window length k for the data subsets, and the neighborhood size r, how shall we select them properly? To better illustrate the ideas, we postpone the discussion to section 3.1.1.

2.3. STLmax METHOD FOR SEIZURE DETECTION

The basic idea is to compute the largest positive Lyapunov exponent for each window's EEG signal using the Wolf et al.'s algorithm (Wolf et al., 1985) or its simple variants. Therefore, it is sufficient to describe the Wolf et al.'s algorithm (Wolf et al., 1985) and point out how it can be modified.

To apply the Wolf et al.'s algorithm (Wolf et al., 1985), one selects a reference trajectory and follows the divergence of its neighboring trajectory from it. Denote the reference and the neighboring trajectories by $X_i = [x(i), x(i+L), \ldots, x(i+(m-1)L)]$, $X_j = [x(j), x(j+L), \ldots, x(j+(m-1)L)]$, $i = 1, 2, \ldots,$ $j = K, K+1, \ldots,$ respectively. At the start of the time (which corresponds to $i = 1$), X_K is usually taken as the nearest neighbor of X_1. That is, $j = K$ minimizes the distance between X_j and X_1. When time evolves, the distance between X_i and X_j also changes. Let the spacing between the two trajectories at time t_i and t_{i+1} be d'_i and d_{i+1}, respectively. Assuming $d_{i+1} \sim d'_i e^{\lambda_1(t_{i+1}-t_i)}$, the rate of divergence of the trajectory, λ_1, over a time interval of $t_{i+1} - t_i$ is then

$$\frac{\ln(d_{i+1}/d'_i)}{t_{i+1}-t_i}.$$

To ensure that the separation between the two trajectories is always small, when d_{i+1} exceeds certain threshold value, it has to be renormalized: a new point in the direction of the vector of d_{i+1} is picked up so that d'_{i+1} is very small compared to the size of the attractor. After n repetitions of stretching and renormalizing the spacing, one obtains the following formula:

$$\lambda_1 = \sum_{i=1}^{n-1}\left[\frac{t_{i+1}-t_i}{\sum_{i=1}^{n-1}(t_{i+1}-t_i)}\right]\left[\frac{\ln(d_{i+1}/d'_i)}{t_{i+1}-t_i}\right] = \frac{\sum_{i=1}^{n-1}\ln(d_{i+1}/d'_i)}{t_n - t_1}. \tag{3}$$

Note that this algorithm assumes but does not verify exponential divergence. In fact, the algorithm can yield a positive value of λ_1 for any type of noisy process so long as all the distances involved are small. The reason for this is that when d'_i is small, evolution would move d'_i to the most probable spacing, which is typically much larger than d'_i. Then, d_{i+1}, being in the middle step of this evolution, will also be larger than d'_i; therefore, a quantity calculated based on Equation (3) will be positive. This argument makes it clear that the algorithm cannot distinguish chaos from noise. In

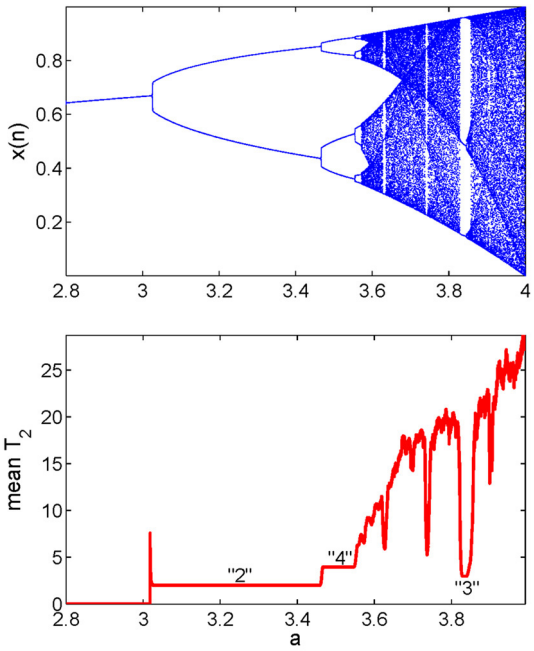


FIGURE 4 | State transitions in the transient logistic map.

other words, even if the algorithm returns a positive λ_1 from EEG data, one cannot conclude that the data are chaotic.

It is worth noting that in practice, to simplify implementation of the algorithm, one may replace the renormalization procedure described above by requiring that d'_{i+1} is constructed whenever $t_{i+1} = t_i + T$, where T is a small time interval. Such a procedure may be called periodic renormalization. In contrast, the original version of the algorithm is an aperiodic renormalization.

3. RESULTS

3.1. SEIZURE DETECTION USING RECURRENCE TIME METHOD

As we pointed out earlier, the method contains four parameters: the embedding dimension m and delay time L, the window length k for the data subsets, and the neighborhood size r. In this subsection, we first discuss how to choose these four parameters properly. Then we evaluate the effectiveness of the method for detecting epileptic seizures. For ease of presentation, we assume that the data have been normalized to the unit interval [0, 1] before further analysis.

3.1.1. Parameter selection

First, we consider the window length k for data subsets. Since our purpose is to find transitions in the signal dynamics, the data subset has to be small. In order to estimate the interesting statistics reliably, a rule of thumb is that so long as a data subset contains several periods of "oscillations", it would be fine (assuming the motion defines certain periodicity-like time scales). For our EEG sampled with a frequency of 200 Hz, we have found that K in the range of 500–2000 are all fine. **Figures 5A,B** show two examples, for $k = 1000$ and 2000, respectively. Clearly, in both cases, the two seizures have been detected correctly.

Next, we consider the size r of the neighborhood. It can be readily appreciated that when r is large, there will be a lot of recurrences, while when r is small, recurrences will be rather rare. This means $\overline{T}_2(r)$ will be large for small r but small for large r. Such expectations have been extensively observed in practice. For EEG signals, we have found that although the values of $\overline{T}_2(r)$ may vary with r, the pattern of the variation basically remains the same for a wide range of r. Two examples are shown in **Figures 5B,C**, where r differs by a factor of 2. Our experience is that choice of this parameter is not very critical, in so far as seizure monitoring is concerned.

Finally, we consider the embedding parameters. As is well known, the embedding parameters critically control the geometrical structure formed by the constructed vectors. Because of this feature, optimal embedding is a critical issue, especially when geometrical or dynamical quantities of the dynamics are concerned, such as the fractal dimension, Lyapunov exponents, and Kolmogorov entropy. For an in-depth discussion of this issue, we refer to Gao et al. (2007). Here, we wish to point out that the time scales associated with the motion are typically much less sensitive to the embedding parameters than the quantities such as the fractal dimension, Lyapunov exponents, and Kolmogorov entropy. To appreciate this feature, we have schematically shown in **Figure 6** two different sets of embeddings. It is clear that the reconstructed trajectory shown in **Figure 6A** is fairly uniform, while that in **Figure 6B** is less so. One can readily conceive that when **Figure 6B** is further squeezed, the embedding quality is even worse. Judged by most optimal embedding criteria, the embedding shown in **Figure 6A** is considered a much better one than that shown in **Figure 6B**. However, it can be readily seen that $\overline{T}_2(r)$ for both **Figures 6A,B** are more or less the same. This means that the selection of m and L for computing $\overline{T}_2(r)$ is much less critical than that for computing other dynamical quantities. One good rule of thumb is that as long as the geometrical structure formed by the vectors are reasonably

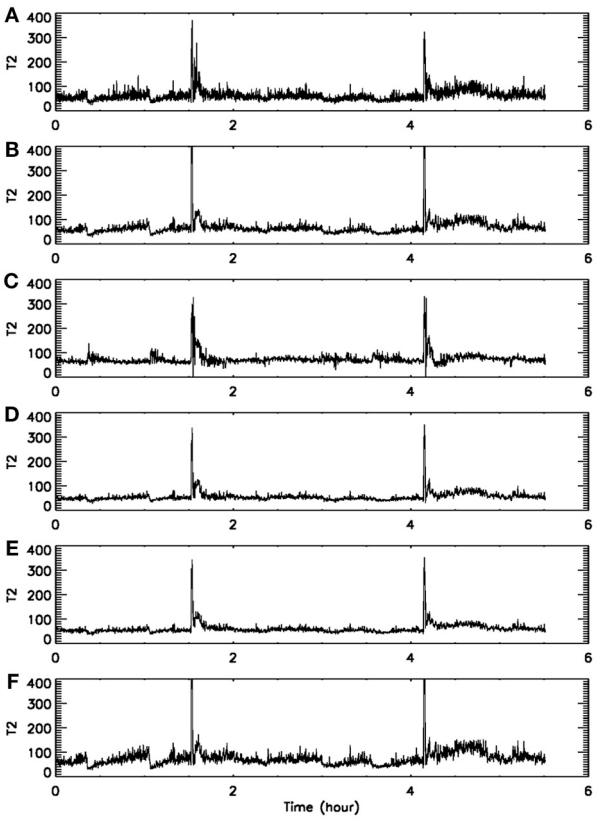


FIGURE 5 | Dependence of $\overline{T}_2$ on the parameters of the algorithm. **(A–F)** Correspond to $(k, m, L, r) = (1000, 4, 4, 2^{-4})$, $(2000, 4, 4, 2^{-4})$, $(2000, 4, 4, 2^{-3})$, $(2000, 3, 4, 2^{-4})$, $(2000, 4, 2, 2^{-4})$, and $(2000, 4, 6, 2^{-4})$, respectively.

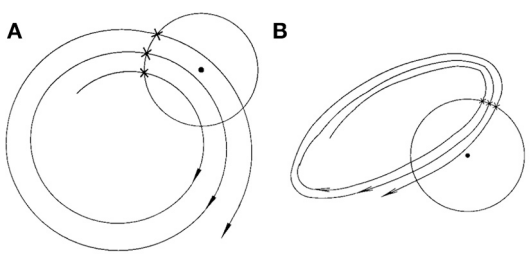


FIGURE 6 | A schematic showing the effect of embedding on the recurrence times of the second type. **(A)** and **(B)** show examples of the reconstructed trajectory that is fairly uniform or less uniform, respectively.

space-filling, the embedding is considered fine. Our experience with computing $\overline{T}_2(r)$ from EEG is that $3 \leq m \leq 6$ are all fine, and with a sampling frequency of 200 Hz, L may be chosen 2–6. This discussion may be better appreciated by comparing **Figures 5B,D–F**, where four sets of (m, L) are illustrated. Clearly, all the parameter combinations have detected the two seizures accurately.

To summarize, the recurrence method is much less sensitive to the parameters when compared with other nonlinear methods, where embedding and other parameters have to be chosen carefully, and have to be specifically adapted to each patient for good results. For our recurrence time method, however, we have used the same parameter combination $(k, m, L, r) = (2000, 4, 4, 2^{-4})$ for all seven patients' data.

3.1.2. Performance evaluation of the method

To illustrate the idea, we shall arbitrarily pick up three channels of EEG data,[1] from one patient, and compare the patterns of variation of $\overline{T}_2(r)$ with that of STLmax. One typical result is shown in **Figure 7**. Vertical dotted lines indicate the seizure occurrence time determined by medical experts by viewing videotapes as well as the EEG signals. There are three seizures in **Figure 7** during the period of time plotted. We observe that $\overline{T}_2(r)$ curves very cleanly and accurately detect all the seizures occurred. In fact, if one ignores the propagation-related slight timing difference (on the order of a few seconds up to 1 min; this will be further discussed later) among different electrodes, then most of the channels can be considered equivalent. In other words, decision can be based on single channel EEG data. This feature makes automatic detection of seizure by thresholding almost trivial. In contrast, the STLmax curves are much noisier than the $\overline{T}_2(r)$ curves. Although STLmax curves can be further post-processed to better reveal seizure information (Iasemidis et al., 2003), those features are still much weaker than those revealed by the recurrence time method.

To more systematically compare the performance of the two methods in detecting seizures, we have computed positive detection (or equivalently, sensitivity) and false alarm per hour for the two methods. Positive detection is defined as the ratio between the number of seizures correctly detected and the total number of seizures. The false alarm per hour is simply the number of falsely detected seizures divided by the total time period. **Table 1** summarizes the results. Clearly, the recurrence time method is more accurate than the STLmax method. This accuracy becomes even more attractive if one notices that the recurrence time method only involves simple thresholding, while the STLmax method involves a lot of further analysis (Iasemidis et al., 2003).

[1] In fact, the three chosen channels EEG data may not correspond to where a seizure was localized. This further indicates the robustness of our method.

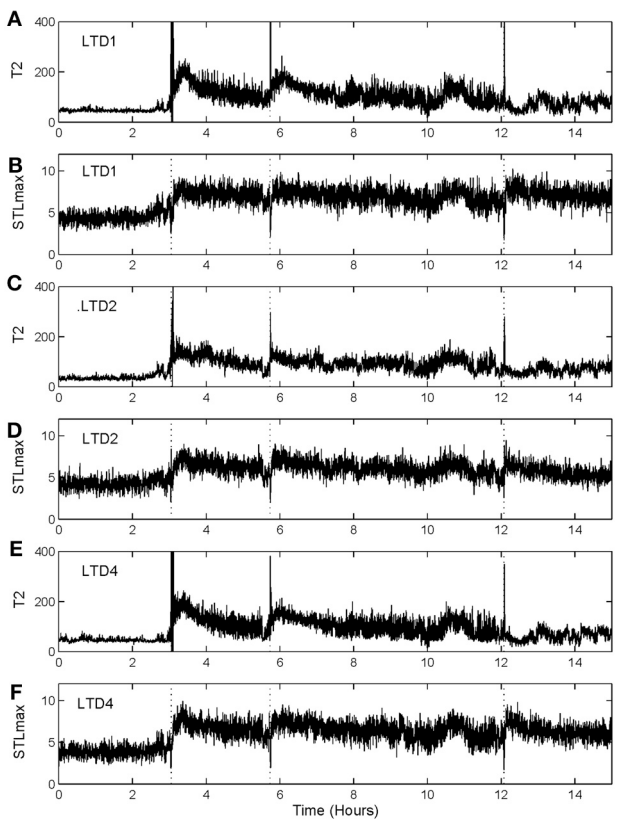


FIGURE 7 | $\overline{T}_2(r)$ and STLmax vs. time curves for the EEG signals measured from three electrodes for a human patient. (A) and **(B)** are for electrode LTD1, **(C)** and **(D)** for electrode LTD2, and **(E)** and **(F)** for electrode LTD4.

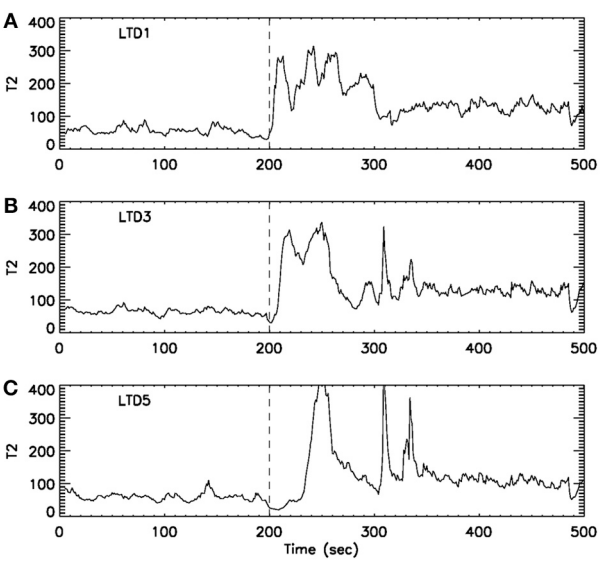


FIGURE 8 | $\overline{T}_2(r)$ curves for EEG signals measured by three electrodes. The dashed vertical line indicate the seizure starting position around 200 s. The seizure lasted for about 2 min. Note that from **(A)** LTD1 to **(B)** LTD3, the seizure activity is delayed about 10 s, while from **(A)** LTD1 to **(C)** LTD5, the seizure activity is delayed about 30–40 s.

3.1.3. Computational cost

The recurrence time method is very fast. With an ordinary PC (CPU speed less than 2 GHz), computation of $\overline{T}_2(r)$ from one channel EEG data of duration 1 h with sampling frequency of 200 Hz takes about 1 min CPU time. Computation of STLmax, on the other hand, takes more than 10 min. Hence, the recurrence time based method is much faster than the STLmax method. In fact, even with an ordinary PC, one is able to process all 28 channels of 1-h EEG data in about half an hour, therefore, faster than the data being continuously collected. With a more powerful PC, of course, the speed becomes even faster. Such a speed implies that the method can be used to real-time on-line process continuously collected all channels of EEG data. From an engineering perspective, the fast computation of recurrence time statistics can be considered overwhelming.

3.2. PROPAGATION OF EPILEPTIC SEIZURES IN THE BRAIN

Formation and propagation of epileptic seizures in the brain is an outstanding example of complex spatial-temporal pattern formations. One of the most desirable ways of studying these problems is to understand how and when information flows from one region of the system to other regions. To resolve this issue, it is critical to accurately providing timing information for interesting events occurring in the system. With the exact timing information, one can then use concepts such as cross correlation and cross spectrum, mutual information, or measures from chaos theory, such as related to cross recurrence plots, to more fully characterize the spatial–temporal patterns. Recurrence time method can effectively provide such a timing information. To illustrate this point, we have shown in **Figure 8** an example of analysis of multi-channel EEG signals using the recurrence time method. For the specific seizure studied, it was known that the seizure occurred around 200 s, and lasted about 2 min. While the recurrence time method has accurately detected the seizure, we note that the seizure activity recorded by electrode LTD3 and LTD5 was about 10 and 40 s later than that indicated by electrode LTD1, respectively. Hence, the recurrence time method not only accurately detects the seizure, but also provides invaluable timing information for the development of the seizure.

4. CONCLUSIONS

Motivated by developing a non-linear method without the limitations of assuming that EEG signals are chaotic and stationary, we have proposed a recurrence time based approach to characterize brain electrical activity. The method is very easy to use, as it only contains a few parameters that are largely signal-independent. It very accurately detects epileptic seizures from EEG signals. Most critically, the method is very fast—it is fast enough to real-time on-line process multi-channel EEG data with a typical PC. Therefore, it has the potential to be an excellent candidate for real-time monitoring of epileptic seizures in a clinical setting.

The recurrence time method is also able to accurately give the timing information critical for understanding seizure propagation. Therefore, it may help characterize epilepsy type, lateralization and seizure classification (Holmes, 2008; Napolitano and Orriols, 2008; Plummer et al., 2008). To more thoroughly understand the capabilities of recurrence time method in characterizing seizure propagation, it would be desirable to combine recurrence time analysis of EEG with studies based on MEG and MRI exams.

REFERENCES

Adeli, H., Ghosh-Dastidar, S., and Dadmehr, N. (2007). A wavelet-chaos methodology for analysis of eegs and eeg subbands to detect seizure and epilepsy. *IEEE Trans. Biomed. Eng.* 54, 205–211. doi: 10.1109/TBME.2006.886855

Aschenbrenner-Scheibe, R., Maiwald, T., Winterhalder, M., Voss, H., Timmer, J., and Schulze-Bonhage, A. (2003). How well can epileptic seizures be predicted? an evaluation of a nonlinear method. *Brain* 126, 2616. doi: 10.1093/brain/awg265

Cao, Y., Tung, W., Gao, J., Protopopescu, V., and Hively, L. (2004). Detecting dynamical changes in time series using the permutation entropy. *Phys. Rev. E* 70:046217. doi: 10.1103/PhysRevE.70.046217

Gao, J. (1999). Recurrence time statistics for chaotic systems and their applicaitons. *Phys. Rev. Lett.* 83, 3178–3181. doi: 10.1103/PhysRevLett.83.3178

Gao, J. (2001). Detecting nonstationarity and state transitions in a time series. *Phys. Rev. E* 63:066202. doi: 10.1103/PhysRevE.63.066202

Gao, J., and Cai, H. (2000). On the structures and quantification of recurrence plots. *Phys. Lett. A* 270, 75–87. doi: 10.1016/S0375-9601(00)00304-2

Gao, J., Cao, Y., Gu, L., Harris, J., and Principe, J. (2003). Detection of weak transitions in signal dynamics using recurrence time statistics. *Phys. Lett. A* 317, 64–72. doi: 10.1016/j.physleta.2003.08.018

Gao, J., Cao, Y., Qi, Y., and Hu, J. (2005a). Building innovative representations of dna sequences to facilitate gene finding. *IEEE Intell. Syst.* 20, 34–39. doi: 10.1109/MIS.2005.100

Gao, J., Tung, W., Cao, Y., Hu, J., and Qi, Y. (2005b). Power-law sensitivity to initial conditions in a time series with applications to epileptic seizure detection. *Physica A* 353, 613–624. doi: 10.1016/j.physa.2005.01.027

Gao, J., Cao, Y., Tung, W., and Hu, J. (2007). *Multiscale Analysis of Complex Time Series – Integration of Chaos and Random Fractal Theory, and Beyond*. Hoboken, NJ: Wiley. doi: 10.1002/9780470191651

Gao, J., Hu, J., and Tung, W. (2011a). Complexity measures of brain wave dynamics. *Cogn. Neurodyn.* 5, 171–182. doi: 10.1007/s11571-011-9151-3

Gao, J., Hu, J., and Tung, W. (2011b). Facilitating joint chaos and fractal analysis of biosignals through nonlinear adaptive filtering. *PLoS ONE* 6:e24331. doi: 10.1371/journal.pone.0024331

Gao, J., Hu, J., and Tung, W. (2012a). Entropy measures for biological signal analysis. *Nonlin. Dyn.* 68, 431–444. doi: 10.1007/s11071-011-0281-2

Gao, J., Hu, J., Tung, W., and Blasch, E. (2012b). Multiscale analysis of physiological data by scale-dependent lyapunov exponent. *Front. Fractal Physiol.* 2:110. doi: 10.3389/fphys.2011.00110

Gao, J., Hu, J., Tung, W., and Cao, Y. (2006a). Distinguishing chaos from noise by scale-dependent lyapunov exponent. *Phys. Rev. E* 74:066204. doi: 10.1103/PhysRevE.74.066204

Gao, J., Hu, J., Tung, W., Cao, Y., Sarshar, N., and Roychowdhury, V. (2006b). Assessment of long range correlation in time series: how to avoid pitfalls. *Phys. Rev. E* 73:016117. doi: 10.1103/PhysRevE.73.016117

Holmes, M. (2008). Dense array eeg: methodology and new hypothesis on epilepsy syndromes. *Epilepsia* 49(Suppl. 3), 3–14. doi: 10.1111/j.1528-1167.2008.01505.x

Hwa, R., and Ferree, T. (2002). Scaling properties of fluctuations in the human electroencephalogram. *Phys. Rev. E* 66:021901. doi: 10.1103/PhysRevE.66.021901

Iasemidis, L., Pardalos, P., Shiau, D., Chaovalitwongse, W., Narayanan, K., Kumar, S., et al. (2003). Prediction of human epileptic seizures based on optimization and phase changes of brain electrical activity. *Optim. Method Softw.* 18, 81–104. doi: 10.1080/1055678021000054998

Iasemidis, L., Sackellares, J., Zaveri, H., and Williams, W. (1990). hase space topography and the lyapunov exponent of electrocorticograms in partial seizures. *Brain Topogr.* 2, 187–201. doi: 10.1007/BF01140588

Jallon, P. (1997). Epilepsy in developing countries. *Epilepsia* 38, 1143–1151. doi: 10.1111/j.1528-1157.1997.tb01205.x

Kanz, H., and Schreiber, T. (1997). *Nonlinear Time Series Analysis*. Cambridge, NY: Cambridge University Press.

Lai, Y., Harrison, M., Frei, M., and Osorio, I. (2003). Inability of lyapunov exponents to predict epileptic seizures. *Phys. Rev. Lett.* 91:068102. doi: 10.1103/PhysRevLett.91.068102

Lehnertz, K., and Elger, C. (1995). Spatiotemporal dynamics of the primary epileptogenic area in temporal-lobe epilepsy characterized by neuronal complexity loss. *Electroencepphalogr. Clin. Neurophysiol.* 95, 108–117. doi: 10.1016/0013-4694(95)00071-6

Lehnertz, K., and Elger, C. (1997). Neuronal complexity loss in temporal lobe epilepsy: effects of carbamazepine on the dynamics of the epileptogenic focus. *Electroencepphalogr. Clin. Neurophysiol.* 103, 376–380. doi: 10.1016/S0013-4694(97)00027-1

Martinerie, J., Adam, C., Quyen, M. L. V., Baulac, M., Clemenceau, S., Renault, B., et al. (1998). Epileptic seizures can be anticipated by non-linear analysis. *Nat. Med.* 4, 1173–1176. doi: 10.1038/2667

Napolitano, C., and Orriols, M. (2008). Two types of remote propagation in mesial temporal epilepsy: analysis with scalp ictal EEG. *J. Clin. Neurophysiol.* 25, 69–76. doi: 10.1097/WNP.0b013e31816a8f09

Packard, N., Crutchfield, J., Farmer, J., and Shaw, R. (1980). Gemomtry from time-series. *Phys. Rev. Lett.* 45, 712–716. doi: 10.1103/PhysRevLett.45.712

Plummer, C., Harvey, S., and Cook, M. (2008). Eeg source localization in focal epilepsy: where are we now? *Epilepsia* 49, 201–218. doi: 10.1111/j.1528-1167.2007.01381.x

Protopopescu, V., Hively, L., and Gailey, P. (2001). Epileptic event forewarning from scalp EEG. *J. Clin. Neurophysiol.* 18, 223–245. doi: 10.1097/00004691-200105000-00003

Quyen, M. L. V., Martinerie, J., Navarro, V., Boon, P., D'Have, M., Adam, C., et al. (2001). Anticipation of epileptic seizures from standard EEG recordings. *Lancet* 357, 183. doi: 10.1016/S0140-6736(00)03591-1

Sauer, T., Yorke, J., and Casdagli, M. (1991). Embedology. *J. Stat. Phys.* 65, 579–616. doi: 10.1007/BF01053745

Takens, F. (1981). "Detecting strange attractors in turbulence," in *Dynamical Systems and Turbulence*. Lecture Notes in Mathematics, Vol. 898, eds D. A. Rand and L. S. Young (Berlin: Springer-Verlag), 366.

Valencia, M., Artieda, J., Alegre, M., and Maza, D. (2008). Influence of filters in the detrended fluctuation analysis of digital electroencephalographic data. *J. Neurosci. Methods* 170, 310–316. doi: 10.1016/j.jneumeth.2008.01.010

van Drongelen, W., Nayak, S., Frim, D., Kohrman, M., Towle, V., Lee, H., et al. (2003). Seizure anticipation in pediatric epilepsy: use of kolmogorov entropy. *Pediatr. Neurol.* 29, 207–213. doi: 10.1016/S0887-8994(03)00145-0

Wolf, A., Swift, J., Swinney, H., and Vastano, J. (1985). Determining lyapunov exponents from a time series. *Physica D* 16, 285. doi: 10.1016/0167-2789(85)90011-9

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.

Received: 28 July 2013; accepted: 04 September 2013; published online: 01 October 2013.

Citation: Gao J and Hu J (2013) Fast monitoring of epileptic seizures using recurrence time statistics of electroencephalography. Front. Comput. Neurosci. 7:122. doi: 10.3389/fncom.2013.00122

This article was submitted to the journal Frontiers in Computational Neuroscience.

frontiers in
COMPUTATIONAL NEUROSCIENCE

OPINION ARTICLE
published: 25 February 2014
doi: 10.3389/fncom.2014.00016

Astronomical apology for fractal analysis: spectroscopy's place in the cognitive neurosciences

Damian G. Kelty-Stephen*

Department of Psychology, Grinnell College, Grinnell, IA, USA
**Correspondence: keltysda@grinnell.edu*

Edited and reviewed by:
Tobias A. Mattei, Ohio State University, USA

Keywords: spectroscopy, fractal, power law, time series, molecular cloud, stars, perception, perception-action

Fractal structure offers new leverage for understanding cognition (Dixon et al., 2012; Kelty-Stephen and Dixon, 2012, 2013). A minority in neuroscience feels very strongly about this point, finding it either crucial (Friston et al., 2012; Van Orden et al., 2012) or patently absurd (e.g., Wagenmakers et al., 2012). The majority remain understandably mystified or bored by opaque math and ponderous debate. I propose to re-present the point through analogy to a field far removed from neuroscience, namely, astronomy, in the hopes of making the common threads clearer and less threatening. One field gazes deep into the brain; the other gazes up and away from anything on Earth. However, both kinds of scientists seek physicochemical accounts of comparably high-dimensional systems (Mesulam, 2008). They must take imperfect measurements and use elegant strategies to probe these measurements for what is not plainly obvious to the naked eye.

Fractal structure (or its absence) and its implication in cognition grows rather inoffensively out of spectral methods (i.e., "spectroscopy") that elevated astronomy from guesswork to extremely sophisticated inquiry. The comparison of 20 years of neuroscience exploring fractal structure in cognition (e.g., Gilden et al., 1995) to 200 years of spectroscopy in astronomy is humblingly instructive (see Hearnshaw, 2010). Far from undermining physicochemical accounts of the heavens, since its recognition in astronomy (e.g., de Vaucouleurs, 1970; Mandelbrot, 1977), fractal structure has supported physicochemical accounts of star formation in ways non-fractal models could not (e.g., Larson, 2005). Comparing our 20 years with astronomy's 200, I am prepared not to live to see the fruition of similar attempts in neuroscience. I hope only to illustrate that neuroscience might learn a lot from astronomy's cosmopolitan views of spectroscopy.

We forget easily that modern astronomy was not always the scientific success we know today. Despite unresolved questions, we are awash in precise physical and chemical information about 10^{11} stars living for billions of years in each of 20^{11} galaxies (Geach, 2011; Tolstoy, 2011). Roughly 180 years ago, Comte (1835) predicted that we would never know the physicochemical details of the heavens. Astronomy was only as good as telescopes with the strongest magnification, and astronomy would never be more than guesswork projected into kinematics of these magnified dots and smears. Comte's words reflected an ignorance of the initial evidence from a new method called "spectroscopy." And it was the subsequent development of spectroscopy that allowed astronomers to bury Comte's disparaging assessment.

What Comte didn't know about spectroscopy was that astronomical measures of celestial dots and smears carry richly patterned optical information (e.g., Fraunhofer, 1817). The full spectrum of electromagnetic radiation reached Earth only incompletely. Between star and telescope lay rich molecular clouds of dust and gas. Decomposing this radiation into a spectrum of oscillations at different scales revealed the composition of the molecular clouds because specific configurations of electrons absorbed and emitted light from specific ranges of the electromagnetic spectrum. For instance, Lockyer (1869) and Janssen (1869) identified the element later known as helium based on its absorbing and emitting light waves of length 587.6 nanometers—or, equivalently, light waves oscillating at a frequency of 5.1×10^{14} Hz. Specific elements composing the universe absorbed energy at specific scales of space and time. Here was the key to the universe's composition and to quashing Comte's prophecy of ignorance.

Spectroscopy denotes the broad class of analyses depicting how an observable's distribution over a wide range of measurement scales. Different kinds of spectra entail different sorts of axis labels. "Power" spectra plot oscillatory power (i.e., amplitude squared) against oscillatory wavelength or, inversely, frequency. "Energy" and "mass" spectra plot quantity across spatial scales. Scientists care about spectroscopy because, as with light through celestial molecular clouds, the distribution of observables varies with scale, and this relationship usually provides insights into the processes underlying phenomena we care about. Sometimes these processes exhibit selective response to characteristic scales, as in helium's emission spectra. Other measurements exhibit response over a continuous range of scales, and this response can increase or decrease with scale. Fractal structure is nothing but an extremely specific example of this latter case, namely, a spectrum exhibiting power-law (and thus scale-invariant) growth or decay across scales. Here we encounter a rather large fact that often goes unmentioned in the debates: There are truly no "fractal analyses"—only fractal or non-fractal patterns revealed by spectroscopic methods.

Neuroscience has a fondness for characteristic scales. For instance, evoked response potential (ERP) data suggests

that cortical activity exhibits different voltage profiles across time depending on the engagement of separate neural/cognitive mechanisms. A peak of negative voltage at 400 ms (i.e., the "N400") after visual presentation of a letter string indicates recognition that the letter string is pronounceable (e.g., Rossi et al., 2011). Whereas absorption/emission of light at 5.1×10^{14} Hz was the astronomers' first glimpse of helium, perhaps N400s at ($400\ ms^{-1}=$) 2.5 Hz is a glimpse of a similarly elemental mechanism in cognitive processes. However, neuroscience focuses its spectroscopic strategies on molecular details of blood flow and metabolites (Minati et al., 2007; Murkin and Arango, 2009). However, these molecular details alone don't address flexible, task-sensitive operation of cognitive processes of language comprehension (White et al., 2012). So long as these mechanisms are known by their characteristic time scales, why hasn't neuroscience situated the N400 on a spectrum too?

One obstacle is that spectroscopy needs long, densely sampled time series. Any single stream of ERP data is so noisy that observing N400s in single-participant data requires averaging over at least 45 trials (e.g., Niedeggen et al., 1999). Otherwise, we might collect prolonged series of ERP data of a participant viewing continuous text of pronounceable letter strings. Reading pace is ∼250 ms/word (Rayner and Clifton, 2009). Let us imagine the resulting ERP signal: N400 peaks for each string, spaced 250 ms apart over time. The emission line in power-spectral analysis of this ERP signal would appear at ($250\ ms^{-1}=$) 4 Hz. Dyslexic readers take 500 ms/word longer (Russeler et al., 2007), and their N400 peaks might be spaced by ($250 + 500=$) 750 ms, producing a peak in a spectrum of ERP data at ($750\ ms^{-1}=$) 1.33 Hz. Just as a peak voltage at 400 ms might signify a phonotactic mechanism's characteristic scale, the gap between 1.33 and 4 Hz should indicate the difference in reading mechanisms between dyslexic and typical readers. After all, wasn't it a similar spectral difference that helped astronomers distinguish helium from sodium?

Results from reading reaction times tell a different story. Over the course of reading a 14000-word story, reading time per word decrease according to Newell and Rosenbloom's (1981) ubiquitous power-law of learning (Wallot et al., 2013). Also, rather than looking at the power spectrum of ERP signals, we might examine the power spectrum of trial-by-trial reading times. Whereas our above ERP series are imaginary, the latter power spectra have been empirically recorded and presented many times over (e.g., Van Orden et al., 2003; Holden et al., 2009; Wallot and Van Orden, 2011). These spectra show that fluctuations in reading-time series resemble 1/f noise, an inverse power-law relationship between oscillatory power and frequency. Rather than having cleanly individuated peaks like emission spectra, the power spectra from these reading-time series show a continuous slow decrease in oscillatory power with greater frequencies. Rather than individuated peaks (i.e., characteristic time scales), these spectra show similar decreases in power across all scales. Often hotly contested as statistical artifacts of "simpler" behavior of cognitive processes at characteristic scales, these patterns have survived statistical rigors (Delignières and Marmelat, 2012).

Statistical rigor notwithstanding, origins and relevance of fractal patterns in neuroscience remain as hotly contested. My own view aligns with one expressed in astronomical literature: fractal patterns reflect cascade dynamics both supported by and giving rise to structures at many scales (Larson, 2005). Astronomy and neuroscience alike have grappled with the realization that structures must somehow embody stability but also flexibility. Stars are not static, homogeneous objects distinct from their contexts—no matter the convenience of this notion for brief measurement and modeling. Stars condense out of clouds, undergo developmental phases, and collapse or explode, and so on. Structures exhibiting characteristic scales demand reconciliation with the fractal patterns inherited from the Big Bang (Mohaved et al., 2011). Similarly, independent mechanisms underpinning cognition are no more static or distinct. Brain structures and cognitive structures reflect relatively stable configurations of neural dynamics within contexts structured at multiple scales (Buzaki, 2006). They exhibit relatively stable short-range functions, but this stability is relative to longer-term variation across the time scales of learning, the life span, and species evolution. The hierarchical nesting of these multiple scales engenders cascades giving rise to structure, and these cascades are no less valid a factor in a physicochemical account than electron configurations. In this light, fractal results that can be (rigorously!) demonstrated to reflect cascade dynamics support a physicochemical account of structure, in astronomy and neuroscience alike.

Spectroscopic work relating fractal patterns to changes in the organization of observed structures supports the foregoing proposals. Fractal modeling of cloud dispersion predicts galactic emission spectra (Bottorff and Ferland, 2001) as well as temperature changes associated with star formation (Pan and Padoan, 2009). In cognitive tasks, bodily movements (e.g., of eye-gaze, hand, foot, or posture) incident to exploring task environments exhibit fractal power spectra. These power-law exponents describing these spectra serve to predict the flexibility of cognitive performance in the same tasks. That is, fractal fluctuations in the human body support the ability of cognitive systems to fine-tune their perceptual judgments (Stephen and Hajnal, 2011; Palatinus et al., 2013) or to discover new representations of problem-solving tasks (Stephen and Dixon, 2009; Stephen et al., 2009). Moreover, these effects of fractal patterning in exploratory behaviors may predict individual-trial performance above and beyond average differences in reaction times due to traditional cognitive processes (Stephen and Anastas, 2011).

The central appeal of fractal results in cognition and neuroscience, to my view, is that they may offer us a framework for aligning physicochemical accounts of neural, cognitive phenomena with physicochemical accounts pursued in different domains. Reaching for a relatively more generic physicochemical framework in which insights from different domains might be mutually relevant and compatible interests me. Not only that, it strikes me as an ideal way of grounding our tests of physicochemical guesses for neuroscience upon stronger physicochemical foundations. Evidence of fractality in domains beyond cognition and neuroscience is a reason that neuroscientists cite

for being unimpressed: for instance, the fact that many more systems are found to exhibit fractal fluctuations than are agreed upon to be "cognitive" is taken to entail that fractality is not important to cognition (Botvinick, 2012). This logic seems to presume that welcome causal players in cognitive theory include only those that maintain the (pre-theoretical) distinction between cognitive systems and non-cognitive ones. Cognitive neuroscience sometimes takes great comfort in asserting the fundamental difference of cognitive systems from all others (Wagenmakers et al., 2012).

Perhaps similarity between cognitive neuroscience and other physicochemically-oriented fields is unwelcome. I find declaring one's own scientific field to require special and different explanation from other scientific fields no more compelling than Comte (1835) found pre-spectroscopic astronomy's guesswork at dots and smears in telescope images. We already have one Big Bang from which to weave cosmological history, and the simple assertion that cognitive systems are fundamentally different from everything else post-Big Bang will require another. Any such cognitive Big Bang (e.g., "when something might have had the first thought") seems less like compelling explanation and more like reluctance to face what may be humbling physicochemical realities. I remain cautiously confident that spectroscopy should be as valuable to cognitive neuroscience as it has been to astronomy in discerning common explanatory ground with other physicochemical disciplines.

Fractal and non-fractal results from spectroscopy appear important to me because they make falsifiable the interesting physicochemical hypothesis that development of structure in nervous systems depends on cascades. When this hypothesis fails to be interesting, I will oblige my critics and stop worrying about fractals.

ACKNOWLEDGMENTS

I would like to thank Zsolt Palatinus and Emma Kelty-Stephen for their kind, patient feedback, and I would also like to thank Tobias Mattei for inviting this submission.

REFERENCES

Bottorff, M., and Ferland, G. (2001). Fractal quasar clouds. *Astrophys. J.* 549, 118–132. doi: 10.1086/319083

Botvinick, M. (2012). Commentary: why I am not a dynamicist. *Top. Cogn. Sci.* 4, 78–83. doi: 10.1111/j.1756-8765.2011.01170.x

Buzaki, G. (2006). *Rhythms of the Brain*. New York, NY: Oxford University Press. doi: 10.1093/acprof:oso/9780195301069.001.0001

Comte, A. (1835). *Cours de Philosophie Positive, II, 19th Lesson*. Paris: Bachelier.

Delignières, D., and Marmelat, V. (2012). Fractal fluctuations and complexity: current debates and future challenges. *Crit. Rev. Biomed. Eng.* 40, 485–500. doi: 10.1615/CritRevBiomedEng.2013006727

de Vaucouleurs, G. H. (1970). The case for a hierarchical cosmology. *Science* 167, 1203–1213. doi: 10.1126/science.167.3922.1203

Dixon, J. A., Holden, J. G., Mirman, D., and Stephen, D. G. (2012). Multifractal dynamics in the emergence of cognitive structure. *Top. Cogn. Sci.* 4, 51–62. doi: 10.1111/j.1756-8765.2011.01162.x

Fraunhofer, J. (1817). Bestimmung des Brechungs- und des Fabenzerstreungs-Vermögens verschiedener glasarten, in Bezug auf die Vervollkommnung achromatischer Fernröhre. *Gilberts Ann. Phys.* 56, 264–313. doi: 10.1002/andp.18170560706

Friston, K., Breakspear, M., and Deco, G. (2012). Perception and self-organized instability. *Front. Comput. Neurosci.* 6:44. doi: 10.3389/fncom.2012.00044

Geach, J. E. (2011). The lost galaxies. *Sci. Am.* 304, 46–53. doi: 10.1038/scientificamerican0511-46

Gilden, D. L., Thornton, T., and Mallon, M. W. (1995). 1/*f* noise in human cognition. *Science* 267, 1837–1839. doi: 10.1126/science.7892611

Hearnshaw, J. (2010). Auguste Comte's blunder: an account of the first century of stellar spectroscopy and how it took one hundred years to prove that Comte was wrong! *J. Astron. Hist. Herit.* 13, 90–104.

Holden, J. G., Van Orden, G., and Turvey, M. T. (2009). Dispersion of response times reveals cognitive dynamics. *Psychol. Rev.* 116, 318–342. doi: 10.1037/a0014849

Janssen, P. J. (1869). Observations spectrales prises pendant l'éclipse du 18 août 1868, et méthode d'observation des protuberances en dehors des éclipses. *Compt. Rend. Hebd. Acad. Sci.* 68, 367–375.

Kelty-Stephen, D. G., and Dixon, J. A. (2012). When physics isn't "just physics": complexity science invites new measurement frames for exploring the physics of cognitive and biological development. *Crit. Rev. Biomed. Eng.* 40, 471–483. doi: 10.1615/CritRevBiomedEng.2013006693

Kelty-Stephen, D. G., and Dixon, J. A. (2013). Notes on a journey from symbols to multifractals: a tribute to Guy Van Orden. *Ecol. Psychol.* 25, 204–211. doi: 10.1080/10407413.2013.810469

Larson, R. B. (2005). Thermal physics, cloud geometry and the stellar initial mass function. *Mon. Not. R. Astron. Soc.* 359, 211–222. doi: 10.1111/j.1365-2966.2005.08881.x

Lockyer, J. N. (1869). Spectroscopic observations of the Sun. No. II. *Phil. Trans. R. Soc. Lond.* 159, 425–444. doi: 10.1098/rstl.1869.0015

Mandelbrot, B. B. (1977). *Fractals: Form, Chance, and Dimension*. New York, NY: Freeman.

Mesulam, M. (2008). Representation, inference, and transcendent encoding in neurocognitive networks of the human brain. *Ann. Neurol.* 64, 367–378. doi: 10.1002/ana.21534

Minati, L., Grisoli, M., and Bruzzone, M. G. (2007). MR spectroscopy, functional MRI, and diffusion-tensor imaging in the aging brain: a conceptual review. *J. Geriatr. Psychiatry Neurol.* 20, 3–21. doi: 10.1177/0891988706297089

Mohaved, M. S., Ghasemi, F., Rahvar, S., and Tabar, M. R. (2011). Long-range correlation in cosmic microwave background radiation. *Phys. Rev. E* 84:021103. doi: 10.1103/PhysRevE.84.021103

Murkin, J. M., and Arango, M. (2009). Near-infrared spectroscopy as an index of brain and tissue oxygenation. *Br. J. Anaesth.* 103, i3–i13. doi: 10.1093/bja/aep299

Newell, A., and Rosenbloom, P. S. (1981). "Mechanisms of skill acquisition and the law of practice," in *Cognitive Skills and their Acquisition*, ed J. R. Anderson (Hillsdale, NJ: Erlbaum), 1–55.

Niedeggen, M., Rossler, F., and Jost, K. (1999). Processing of incongruous mental calculation problems. evidence for an arithmetic N400 effect. *Psychophysiology* 36, 307–324. doi: 10.1017/S0048577299980149

Palatinus, Z. S., Dixon, J. A., and Kelty-Stephen, D. G. (2013). Fractal fluctuations in quiet standing predict the use of mechanical information for haptic perception. *Ann. Biomed. Eng.* 41, 1625–1634. doi: 10.1007/s10439-012-0706-1

Pan, L., and Padoan, P. (2009). The temperature of interstellar clouds from turbulent heating. *Astrophys. J.* 692, 594. doi: 10.1088/0004-637X/692/1/594

Rayner, K., and Clifton, C. Jr. (2009). Language processing in reading and speech perception is fast and incremental: implications for event potential related research. *Biol. Psychol.* 80, 4–9. doi: 10.1016/j.biopsycho.2008.05.002

Rossi, S., Jürgenson, I. B., Hanulíková, A., Telkemeyer, S., Wartenburger, I., and Obrig, H. (2011). Implicit processing of phototactic cues: evidence from electrophysiological and vascular responses. *J. Cogn. Neurosci.* 23, 1752–1764. doi: 10.1162/jocn.2010.21547

Russeler, J., Becker, P., Johannes, S., and Munte T. F. (2007). Semantic, syntactic, and phonological processing of written language in adult developmental dyslexic readers: an event-related brain potential study. *BMC Neurosci.* 8:52. doi: 10.1186/1471-2202-8-52

Stephen, D. G., and Anastas, J. (2011). Fractal fluctuations in gaze speed visual search. *Atten. Percept. Psychophys.* 73, 666–677. doi: 10.3758/s13414-010-0069-3

Stephen, D. G., Boncoddo, R. A., Magnuson, J. S., and Dixon, J. A. (2009). The dynamics of insight: mathematical discovery as a phase transition. *Mem. Cogn.* 37, 1132–1149. doi: 10.3758/MC.37.8.1132

Stephen, D. G., and Dixon, J. A. (2009). The self-organization of insight: entropy and power laws in problem solving. *J. Prob. Solving* 2, 72–101. doi: 10.7771/1932-6246.1043

Stephen, D. G., and Hajnal, A. (2011). Transfer of calibration between hand and foot: functional equivalence and fractal fluctuations. *Atten. Percept. Psychophys.* 73, 1302–1328. doi: 10.3758/s13414-011-0142-6

Tolstoy, E. (2011). Galactic paleontology. *Science* 333, 176–178. doi: 10.1126/science.1207392

Van Orden, G., Holden, J. G., and Turvey, M. T. (2003). Self-organization of cognitive performance. *J. Exp. Psychol. Gen.* 132, 331–350. doi: 10.1037/0096-3445.132.3.331

Van Orden, G., Hollis, G., and Wallot, S. (2012). The blue-collar brain. *Front. Physiol.* 3:207. doi: 10.3389/fphys.2012.00207

Wagenmakers, E.-J., van der Maas, H. L. J., and Farrell, S. (2012). Abstract concepts require concrete models: why cognitive scientists have not yet embraced nonlinearly coupled, dynamical, self-organized critical, synergistic, scale-free, exquisitely context-sensitive, interaction-dominant, multifractal, interdependent brain-body-niche systems. *Top. Cogn. Sci.* 4, 87–93. doi: 10.1111/j.1756-8765.2011.01164.x

Wallot, S., Hollis, G., and van Rooij, M. (2013). Connected text reading and differences in text reading fluency in adult readers. *PLoS ONE* 8:e71914. doi: 10.1371/journal.pone.0071914

Wallot, S., and Van Orden, G. (2011). Grounding language performance in the anticipatory dynamics of the body. *Ecol. Psychol.* 23, 157–184. doi: 10.1080/10407413.2011.591262

White, L., Mattys, S. L., and Wiget L. (2012). Segmentation cues in conversational speech: robust semantics and fragile phonotactics. *Front. Psychol.* 3:375. doi: 10.3389/fpsyg.2012.00375

Received: 30 January 2014; accepted: 03 February 2014; published online: 25 February 2014.

Citation: Kelty-Stephen DG (2014) Astronomical apology for fractal analysis: spectroscopy's place in the cognitive neurosciences. Front. Comput. Neurosci. ***8****:16. doi: 10.3389/fncom.2014.00016*

This article was submitted to the journal Frontiers in Computational Neuroscience.

frontiers in
COMPUTATIONAL NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 14 March 2014
doi: 10.3389/fncom.2014.00022

Chunking dynamics: heteroclinics in mind

Mikhail I. Rabinovich[1], Pablo Varona[2]*, Irma Tristan[1] and Valentin S. Afraimovich[3]

[1] *BioCircuits Institute, University of California, San Diego, La Jolla, CA, USA*
[2] *Grupo de Neurocomputación Biológica, Departamento de Ingeniería Informática, Escuela Politécnica Superior, Universidad Autónoma de Madrid, Madrid, Spain*
[3] *Instituto de Investigación en Comunicación Óptica, Universidad Autónoma de San Luis Potosí, San Luis Potosí, México*

Edited by:
Tobias Alecio Mattei, Ohio State University, USA

Reviewed by:
Maurizio Mattia, Istituto Superiore di Sanità, Italy
Hiroshi Okamoto, RIKEN Brain Science Institute, Japan

Correspondence:
Pablo Varona, Grupo de Neurocomputación Biológica, Departamento de Ingeniería Informática, Escuela Politécnica Superior, Universidad Autónoma de Madrid, C/Francisco Tomás y Valiente, 11, 28049 Madrid, Spain
e-mail: pablo.varona@uam.es

Recent results of imaging technologies and non-linear dynamics make possible to relate the structure and dynamics of functional brain networks to different mental tasks and to build theoretical models for the description and prediction of cognitive activity. Such models are non-linear dynamical descriptions of the interaction of the core components—brain modes—participating in a specific mental function. The dynamical images of different mental processes depend on their temporal features. The dynamics of many cognitive functions are transient. They are often observed as a chain of sequentially changing metastable states. A stable heteroclinic channel (SHC) consisting of a chain of saddles—metastable states—connected by unstable separatrices is a mathematical image for robust transients. In this paper we focus on hierarchical chunking dynamics that can represent several forms of transient cognitive activity. Chunking is a dynamical phenomenon that nature uses to perform information processing of long sequences by dividing them in shorter information items. Chunking, for example, makes more efficient the use of short-term memory by breaking up long strings of information (like in language where one can see the separation of a novel on chapters, paragraphs, sentences, and finally words). Chunking is important in many processes of perception, learning, and cognition in humans and animals. Based on anatomical information about the hierarchical organization of functional brain networks, we propose a cognitive network architecture that hierarchically chunks and super-chunks switching sequences of metastable states produced by winnerless competitive heteroclinic dynamics.

Keywords: cognitive dynamics, stable heteroclinic channel, transient dynamics, low dimensionality of brain activity, hierarchical sequences, chunking and superchunking, cognition modeling principles

INTRODUCTION

Chunking is a dynamical phenomenon that the brain uses for processing long informational sequences. The concept of chunk was introduced by Miller (1956). His key notion is that short-term storage is not rigid but amenable to strategies such as chunking that can expand its capacity. Miller's work drew plenty of attention to the concept of short-term memory and its functional characteristics. Chunking involves two processes: concatenation of units in a block and segmentation of the blocks. In general, chunking is related to the hierarchical organization of perceptual, cognitive, or behavioral sequential activity. In particular, in motor control (see Rosenbaum et al., 1983) sequences can consist of subsequences and these can in turn consist of sub-sub-sequences, etc. The natural hierarchical organization of long sequences is a result of the activity of specific brain functional networks. Such networks include many different brain areas and some of them are also organized in a hierarchical manner. A well-known example is Broca's area that has been suggested to act as a "supramodal syntactic processor," able to process any type of hierarchically organized sequences (Grossman, 1980; Tettamanti and Weniger, 2006), a hypothesis based on the findings that this region is not only involved in processing language syntax (Musso et al., 2003), but also in syntax like aspects of non-linguistic tasks, for example, the performance of specific movements and music (Fadiga et al., 2009) as several fMRI studies (Bahlmann et al., 2008, 2009) seem to confirm. Clerget et al. hypothesize that motor behavior shares some similarities with language (Clerget et al., 2013), namely that a complex action can be viewed as a chain of subordinate movements, which need to be combined according to certain rules in order to reach a given goal (Dehaene and Changeux, 1997; Dominey et al., 2003; Botvinick, 2008).

What are the mechanisms that transform the extremely complex, noisy, and many-dimensional brain activity into a rather regular, low-dimensional, and even predictable cognitive behavior, e.g., what are the mechanisms underlying the dynamics of the mind, including chunking? This is one of the most challenging questions in today's neuro- and cognitive science. Recent continuous advances in non-invasive brain imaging allow assessing the structural connectivity of the brain and the corresponding evolution of the spatio-temporal activity in detail.

In our view, metastability is a key element of transient cognitive dynamics participating in chunking processes. The idea of the spatiotemporal organization of brain dynamic activity through transient, metastable states emerged more than 15 years ago (Kelso, 1995; Friston, 1997). According to this scenario, such dynamics can be represented as a sequential switching between different metastable states (for a description of the mathematical basis of this scenario see Rabinovich et al., 2008a,b). Metastable

transient dynamics represent a balance between the segregation of focused cognitive processing and the flexible integration of distributed brain areas. Such integration is necessary for the performance of a specific cognitive function (Bressler and Kelso, 2001; Meehan and Bressler, 2012). The existence of connections that are prevalent over long periods of time supports the well-regarded concept of a hierarchical organization of neural processing (Engel et al., 2001), which is the basis for the understanding of the origin of the chunking dynamics. Because the dimensionality of cognition depends on the number of activated (in contrast to the potentially observable) metastable states, it is important to remember that the brain chooses the necessary metastable states and suppresses those which are irrelevant to the goal of the cognitive process, resulting in a reduced dimensionality. The low-dimensionality of brain cognitive dynamics is based on two important issues: first, the manner of the cognitive task encoding—an external or internal stimulus determining a specific cognitive task excites a set of elements of the community networks which are responsible for the performance of such cognitive activities; and second, the existence of a specific hierarchical organization of the global brain networks that operate for the performance of a specific cognitive task by a moderate number of brain modes.

Based on experimental data suggesting that the processing of sequential cognitive activity on computational grounds is implemented in the brain by spatiotemporally pattern dynamics (see also Sahin et al., 2009), we build here a general dynamical model that produces hierarchical chunking of sequences, which suggests a plausible neural mechanism of chunking dynamics in the brain. This model is reasonably low-dimensional, which allows a detailed dynamical analysis.

MATERIALS AND METHODS

A top-down approach to model transient cognitive dynamics taking into account the experimental observations described in the introduction is to use kinetic equations for the description of spatiotemporal mental modes that contain the discussed metastable states as equilibrium points. The set of brain patterns that sequentially change in the process of the cognitive task performance determine the spatial structure of the modes and the associated connection matrix among them. Using such type of models we can integrate our knowledge about the description of brain activity based on these new ideas related to heteroclinic sequences and their interactions, i.e., heteroclinic networks.

As a top-down departing point, we need a mathematical object that can describe robust transient dynamics and their associated information processing. Once we have this object, we can implement it through a set of canonic equations that can be used to study transient activity at different brain description levels, and in particular to address chunking dynamics. A mathematical image of robust transient sequential dynamics must have two principal features. First, it must be resistant to noise and reliable even in the context of small variations in initial conditions, so that the succession of states visited by the system (its trajectory, or transient) is stable. Second, the transients must be input-specific to contain information about what caused them. These are two fundamental contradictions regarding the use of transient dynamics for the description of brain activity. Transient dynamics are inherently unstable: any transient depends on initial conditions and cannot be reproduced from arbitrary initial conditions. On the other hand, dynamical robustness in principle prevents sensitivity to informative perturbations. These contradictions can be solved through the concept of metastability, which was introduced to cognitive science at the end of the last century (Kelso, 1995; Friston, 1997, 2000; Fingelkurts and Fingelkurts, 2006; Oullier and Kelso, 2006; Gros, 2007; Ito et al., 2007).

A stable heteroclinic channel (SHC) is a mathematical object that meets the above discussed requirements, which can implement such stable transients. A SHC is defined by a sequence of successive metastable "saddle" states that are connected by separatrices. Under proper conditions, all the trajectories in the neighborhood of these saddle metastable states that form the chain remain in the channel, ensuring robustness and reproducibility over a wide range of control parameters (Rabinovich et al., 2008b). The stability of a channel means that trajectories in the channel do not leave it until the end of the channel is reached.

A simple model to implement SHCs is a generalized Lotka–Volterra equation with N interactive elements:

$$\frac{dA_i(t)}{dt} = A_i(t)F\left(\sigma_i\left(S_k\right) - \sum_{j=1}^{N}\rho_{ij}A_i(t)\right) + A_i(t)\eta_i(t)$$
$$i = 1, \ldots, N \qquad (1)$$

where $A_i(t) \geq 0$ is the activity rate of element i, σ_i is the gain function that controls the impact of the stimulus, S_k is an environmental stimulus, ρ_{ij} determines the interaction between the variables, η_i represents the noise level, and F is a function, in the simplest case a linear function. The state portrait of the system often contains a heteroclinic sequence linking saddle points. These saddles can be interpreted as successive and temporary winners in a never-ending competitive game, i.e., winnerless competition (WLC) dynamics (Rabinovich et al., 2001, 2006). In neural systems, because a representative model must produce sequences of connected neuronal population states (the saddle points), the neural connectivity ρ_{ij} must be asymmetric, as determined by the theoretical examination of this model (Huerta and Rabinovich, 2004). Although many connection statistics probably work for stable heteroclinic-type dynamics, it is likely that connectivity within biological networks is, to some extent at least, the result of optimization by evolution and synaptic plasticity. It is important to emphasize that Equation (1) is just an elementary building block for different levels of the chunking hierarchy that we will describe below.

Models like the generalized Lotka–Volterra equations allow establishing the conditions necessary for transient stability, and display stable, sequential, and cyclic activation of its components, the simplest variant of WLC. A network with several degrees of freedom and asymmetric connections can generate structurally stable sequences—transients, each shaped by one input. Asymmetric inhibitory connectivity helps to solve the apparent paradox that sensitivity and reliability can coexist in a network (Huerta and Rabinovich, 2004; Nowotny and Rabinovich, 2007; Rabinovich et al., 2008b; Rabinovich and Varona, 2011). The

neurons or modes participating in a SHC are assigned by the stimulus, by virtue of their direct and/or indirect input from the neurons activated by that stimulus. The joint action of the external input and a stimulus-dependent connectivity matrix defines the stimulus-specific heteroclinic channel. In addition, asymmetric inhibition coordinates the sequential activity and keeps a heteroclinic channel stable.

The WLC concept is directly related to the sequential dynamics of metastable states that are activated by inputs that do not destroy the origin of a competitive process. This paradigm can explain and predict many dynamical phenomena in neural networks with excitatory and inhibitory synaptic connections. Based on the requirement of the stability, this formalism has been used (i) to assess the dynamical origin of finite working memory (WM) capacity based upon WLC amongst available informational items (Bick and Rabinovich, 2009; Rabinovich et al., 2012); (ii) to build a dynamical model of information binding for transients that can describe the interaction of different sensory information flows that are generated concurrently (Rabinovich et al., 2010a); (iii) to model the sequential interaction between emotion and cognition (Rabinovich et al., 2010b); (iv) to represent attention dynamics (Rabinovich et al., 2013); and (v) to assess the dynamics of pathological states in mental disorders (Bystritsky et al., 2012; Rabinovich et al., 2013). Here we focus on a model of hierarchical chunking dynamics that can represent several forms of cognitive activity such as WM and speech construction.

As we discussed in the Introduction, chunking is grouping or categorizing related issues or information into smaller, most meaningful and compact units. Think about how hard it would be to read a long review paper without chapters, subchapters, paragraphs, and separated sentences. Chunking is a naturally occurring process that can be actively used to break down problems in order to think, understand, and make improvisation more efficiently. This is because it is easier to process chunked tasks or perceptional data. In particular, it is much easier to learn and recall such data. Mathematically, the "chunking principle" can be viewed as the transformation of a chain of metastable states along a transient process to the chain of groups of such states. It is a key dynamical idea that nature may use to make cognitive information processing more effective in the context of a complex environment.

Chunking processes in human perception, learning, and performance of a cognitive task can be both automatic and directly linked to the environmental stimuli, and controllable by a goal-oriented intrinsic signal (Gobet et al., 2001). It is important to note that chunking is a strategy that supports increasing speed and accuracy through the formation of hierarchical memory structures and complex task-dependent behavioral sequences. Two competitive processes form temporal chunking sequences—one separates long sequences into shorter groups of information items to be easily performed, and the second connects them to express a long sequence as a unified thought or behavioral action (Friederici et al., 2011; Chekaf and Matha, 2012).

Hierarchical chunking dynamics can be implemented in a model of cognitive networks whose information processing relies on SHCs. **Figure 1** illustrates a chunking heteroclinic cognitive network for two hierarchical informational groups—elementary items and chunking (integrated) informational items including many elementary units interacting through dynamical connections. It is reasonable to hypothesize that functionally there are two different cognitive networks from at least two different hierarchical levels that are responsible for the: (i) organization of the sequence of items inside chunks, and (ii) the formation of the chunk sequence. In particular, this hypothesis is supported by an experiment with chunking during visuomotor sequence learning (Sakai et al., 2003). It has been shown that each motor cluster is processed as a single memory unit—a chunk. A learned visuomotor sequence is a sequence of chunks that contains several elementary movements. The authors of this work have shown that a key role in the process of chunking formation is played by a brain network including the dominant parietal area, the basal ganglia, and the presupplementary motor area (see also Ribas-Fernandes et al., 2011 and Bor and Seth, 2012, where authors discuss the chunking structure of conscious processes).

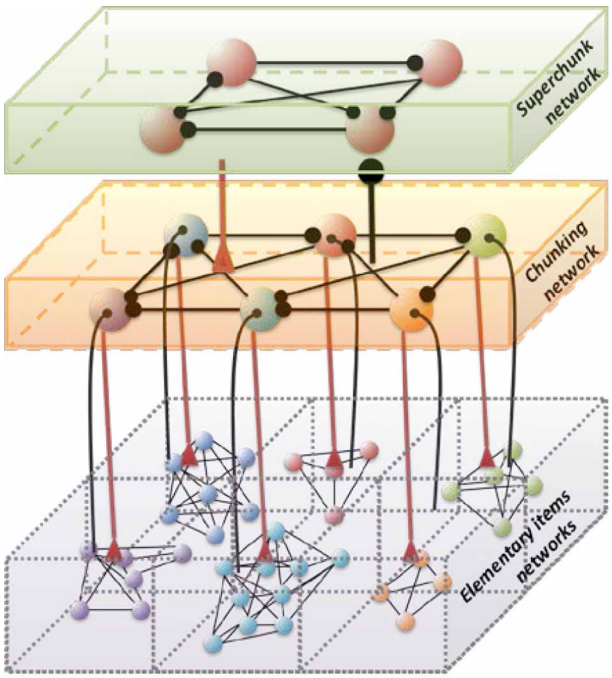


FIGURE 1 | Architecture of the three level cognitive network responsible for the grouping of informational items. Each level of hierarchy is described by its own Lotka–Volterra type Equations (see 2–6) with connection matrices ρ, ξ and ς. Black circles represent inhibitory connections; triangles represent excitatory connections responsible for the choosing of the informational items. Spheres represent the informational items or units (metastable stables). Different colors indicate different chunks. All connections inside the elementary items are inhibitory.

Below we suggest a three level hierarchical model for the description of the chunking dynamics. Inhibition plays a key role in this model as is responsible for the execution of three functions: (i) competition between elementary informational items in order to produce stable sequences of metastable states, (ii) generation of the chunking sequence, and (iii) control of the performance of the sequential task. In recent years, the investigation of the hierarchical control between different levels of representation and

information processing has become one of the hot subjects in cognitive science. This issue is important for understanding how the mind controls behavior and itself. In particular, the relationship between chunking (a sequence-level process) and task-set inhibition (a task-level process) in the performance of task sequences was investigated in (Koch et al., 2006; Schneider, 2007; Li et al., 2010), for a description of "chunks of chunks"—"superchunks" see Rosenberg and Feigenson (2013).

To understand the emergence of hierarchical chunking dynamics in a model we need to depart from Equation (1) in the following direction, c.f. **Figure 1**):

$$\dot{X}_i^{lk} = X_i^{lk}\left(\sigma_i^{lk}(S,C)\cdot Y^{lk} - \sum_j^{N^{lk}} \rho_{ij}^{lk}(S,C)X_j^{lk}\right) \quad (2)$$

$$\tau\dot{Y}^{lk} = Y^{lk}\left(\left(V^l - \beta(C)\sum_i^{N^{lk}} X_i^{lk}\right) - Z^{lk}\right) \quad (3)$$

$$\theta(C)\dot{Z}^{lk} = \sum_m^M \xi_l^{km}(S,C)Y^{lm} - Z^{lk} \quad (4)$$

$$T\,\dot{V}^l = V^l\left(\left(1 - \delta(C)\sum_j^{M^l} Y^{lj}\right) - W^l\right) \quad (5)$$

$$\Theta(C)\dot{W}^l = \sum_q^P \varsigma^{lq}(S,C)V^q - W^l \quad (6)$$

Here X_i^{lk} characterizes the -th informational item associated with the k-th chunk and l-th superchunk, $\sigma_i^{lk}(S,C)$ is the growth rate for each informational item determined by the stimulus S and the cognitive task C, and $\rho_{ij}^{lk}(S,C)$ is the matrix of inhibitory connections among basic informational items. In this model Y^{lk} characterizes the k-th chunk associated to the l-th superchunk V^l, with corresponding characteristic times τ and T, respectively, and $\beta(C)$ represents the strength of the inhibition between the informational items and the chunk, and $\delta(C)$ between the chunks and the superchunk. Also, Z^{lk} describes the synaptic dynamics for the k-th chunk associated to the l-th superchunk with $\xi_l^{km}(S,C)$, the matrix of inhibitory connections between chunks (black circles in **Figure 1**); and W^l describes the synaptic dynamics for the l-th superchunk with $\varsigma^{lq}(S,C)$, the matrix of inhibitory connections between superchunks, the corresponding characteristic times are $\theta(C)$ and $\Theta(C)$. In this model, $\beta(C)$ and $\delta(C)$ are adaptation parameters that determine the timing relationship between a basic informational chain and the chunking and superchunking modulation. The chunking variables also satisfy the generalized Lotka–Volterra—canonic equations which allows them to form a stable sequence. Because of this, in fact, chunking variables play the role of cognitive controllers. The parameters for Equations (3)–(5) in the simulations below were chosen with this scope. Since chunking dynamics has to take into account of the characteristic time of the chunk formation, the competition between different chunks has to be delayed—we used for this an inhibition described by a first order kinetic model. At the same time, the competition among elementary informational items is implemented by fixed weight ρ_{ij} instantaneous synapses. The same logic has been applied for the description of the highest level of the hierarchy—the superchunks.

RESULTS: HIERARCHICAL SEQUENCES—CHUNKING AND SUPER-CHUNKING

Let us first represent the phase portrait of a simple two-level chunking dynamics. We carried out numerical simulations of the model for the dynamics within chunks of informational items for the following parameters $N^k = 3$, $M = 3$ (number of "chunks" or "episodes"), $\sigma^1 = [7.24, 5.85, 8.30]$, $\sigma^2 = [9.93, 6.00, 5.18]$, $\sigma^3 = [8.29, 7.86, 9.16]$, and given these values, $\rho_{ii}^k = 1.0$, $\rho_{i_{n-i}i_n}^k = \frac{\sigma_{i_{n-1}}^k}{\sigma_{i_n}^k} + 0.51$, and $\rho_{i_{n+i}i_n}^k = \frac{\sigma_{i_{n+1}}^k}{\sigma_{i_n}^k} - 0.5$, $i = 1,\ldots,N^{lk}$, $k = 1,\ldots,M$ as well as the parameters considered for the synaptic dynamics described by Equations (3) and (4): $\tau = 0.7$, $\theta = 2.0$, $\xi^{kk} = 1.0$, $\xi^{k_n k_{n+1}} = 1.4$ and $\xi^{k_n k_{n-1}} = 0.5$, $k = 1,\ldots,M$ and $\beta = 0.01$. The results of these simulations are shown in **Figures 2**, **3**.

Figure 2 shows the phase portrait of the chunking dynamics when the superchunk formation is absent: the system is described by Equations (2)–(4), $V = 1$. This example illustrates a closed chunking sequence (green) that consists of several heteroclinic cycles that represent the elementary chunks (blue). In general, the number of elementary items in each chunk are different and the chunking sequence can be open.

Figure 3 illustrates the timing between chunks along the sequence. The emergence of the chunking sequence shown in **Figure 2** is the result of a modulational instability in the two-level hierarchical network whose dynamics is described by Equations (2)–(4). This instability is oscillatory. The characteristic period of the oscillation is ΔT. The analytical investigation of the dependence of ΔT on the control parameters τ, θ, β and connection matrices $\boldsymbol{\rho}$, $\boldsymbol{\xi}$ is a non-realistic problem because of the non-linear feedback between the dynamical variables X and Y. However, it is reasonable to think that the key parameter in this problem is

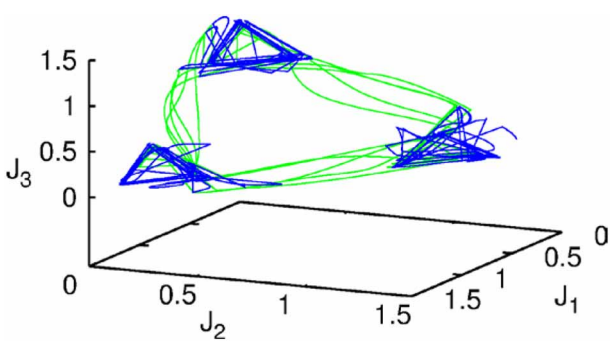


FIGURE 2 | The projection of a nine-dimensional phase portrait of a two-level chunking hierarchical dynamics in the space of the three-dimensional auxiliary variables [see the Equations (2)–(4)] $\boldsymbol{J_1 = Y^1 + 0.04\cdot(X_1^1 + X_1^2 + X_1^3)}$, $\boldsymbol{J_2 = Y^2 + 0.04\cdot(X_2^1 + X_2^2 + X_2^3)}$, $\boldsymbol{J_3 = Y^3 + 0.04\cdot(X_3^1 + X_3^2 + X_3^3)}$. Blue represents the elementary informational item activity—individual chunk. Green represents the chunking sequence.

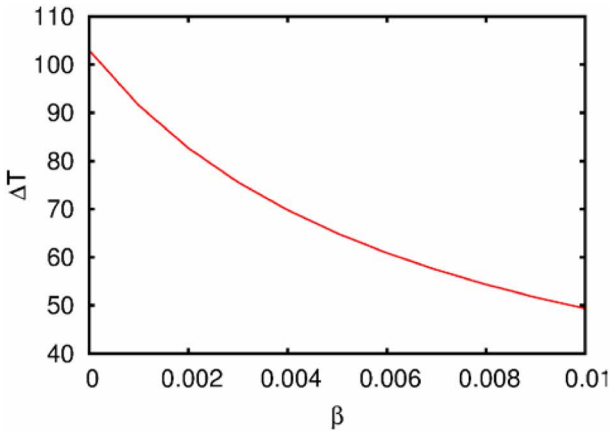


FIGURE 3 | The dependence of the chunking interval timing [see Equation (1)] on the control parameter β. *One* can see that the chunking interval strongly decreases together with the increasing of the adaptation parameter β. When β increases the effective excitation of variable *Y* decreases.

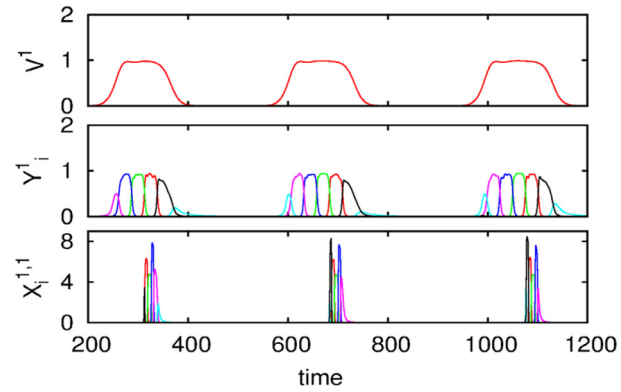


FIGURE 4 | Time series of the sequences of the three-level hierarchy—108 items groupped in 18 chunks of 6 items; these chunks form 3 superchunks of 6 elements each displaying reproducible dynamics according to the model (2)–(6). Different colors correspond to different items inside each group (switching the color means moving from the previous item to the next one).

β which determines the level of excitability of variable *Y* and, according to the feedback, also controls the excitability of *X* (term $\sigma_i^{lk}(S, C) \cdot X_i^{lk} \cdot Y^{lk}$) in the right hand side of Equation (2). In **Figure 3** we represent the numerical analysis of the dependence of ΔT on the parameter β—increasing β, i.e., decreasing the excitability leads to the decreasing of the timing interval ΔT.

We also carried out numerical simulations of a high-dimensional model that describes the dynamics of chunk and super-chunk formation with the following parameters: $N^{lk} = 6$, $M^l = 6$ (number of chunks), $P = 3$ (number of superchunks), $\sigma^{l1} = [6.94, 5.11, 8.94, 5.86, 8.33, 9.62]$, $\sigma^{l2} = [5.48, 5.66, 5.39, 9.89, 9.99, 5.82]$, $\sigma^{l3} = [7.65, 8.98, 9.21, 6.02, 5.71, 5.12]$, $\sigma^{l4} = [7.61, 7.73, 5.62, 7.93, 5.80, 5.39]$, $\sigma^{l5} = [5.11, 9.99, 5.52, 5.66, 5.50, 8.21]$, $\sigma^{l6} = [5.84, 9.39, 7.08, 5.16, 8.37, 6.87]$, and given these values, $\rho_{ii}^{lk} = 1.0$, $\rho_{i_{n-i}i_n}^{lk} = \frac{\sigma_{i_{n-1}}^{lk}}{\sigma_{i_n}^{lk}} + 0.5$ 1, $\rho_{i_{n+i}i_n}^{lk} = \frac{\sigma_{i_{n+1}}^{lk}}{\sigma_{i_n}^{lk}} - 0.5, i = 1, \ldots, N^{lk}, k = 1, \ldots, M^l, l = 1, \ldots$,P, and $\rho_{ii_n}^{lk} = \rho_{i_{n-1}i_n}^{lk} + \frac{\sigma_i^{lk} - \sigma_{i_{n-1}}^{lk}}{\sigma_{i_n}^{lk}} + 2$, $i \neq \{i_{n-1}, i_n, i_{n+1}\}$, as well as the parameters considered for the synaptic dynamics between chunks described by the equations $\tau = 0.8$, $\theta = 2.0$, $\xi_l^{kk} = 1.0$, $\xi_l^{k_n k_{n-1}} = 0.5$, $\xi_1^{k_n k_{n+1}} = 1.4$, $\xi_2^{k_n k_{n+1}} = 1.3$, $\xi_3^{k_n k_{n+1}} = 1.5$, $k = 1, \ldots, M^l$, $l = 1, \ldots, P$, $\xi_l^{kk_n} = \xi_l^{k_{n-1}k_n} + 2$, $k \neq \{k_{n-1}, k_n, k_{n+1}\}$, and $\beta = 0.01$. Finally, the parameters for the synaptic dynamics between superchunks were $T = 5$, $\Theta = 10$, $\varsigma^{ll} = 1.0$, $\varsigma^{l_n l_{n-1}} = 0.5$, $\varsigma^{l_n l_{n+1}} = 1.4$, $l = 1, \ldots, P$, and $\delta = 0.01$. The result of these simulations are displayed in **Figure 4**, which shows three levels of information hierarchy: original informational chain (lower panel), chunked chain (middle panel), and superchunking chain (upper panel).

As illustrated in **Figure 2**, the sequence of chunks can be considered as a heteroclinic cycle of metastable states where each metastable state itself is a heteroclinic cycle of elementary informational items. Based on this self-similarity, we can expect that the chunking chain as a result of a second heteroclinic instability generates the next level of modulation—the superchunk sequence. Our expectation is confirmed in **Figure 4** that shows the time series of the three level network (2)–(6) (c.f. **Figure 1**) dynamics. In this figure, one can see the generation of sequences of superchunks. All together, the sequences informational items, chunks and superchunks can be interpreted as "words," "sentences," and "paragraphs."

For the sake of simplicity we have illustrated here the phenomenon of stability just for a closed-loop clustered chunking-superchunking sequence. In the general case of open sequence, it is possible to formulate the sufficient conditions for the existence and stability of the non-closed channel based on the estimation of the saddle values of the metastable states (elementary items)—the channel is stable in the case that all of them are larger than one in absolute value (Afraimovich et al., 2004; Bick and Rabinovich, 2010). The formulation of the necessary conditions is a more complex problem and is still under consideration. The imposed stability conditions determine the behavior of the trajectories inside the neighborhood of the heteroclinic network independently of the initial conditions as computer experiments have confirmed (Afraimovich et al., 2004; Bick and Rabinovich, 2010).

The above described numerical results can be justified by an analytical study of the system

$$\begin{cases} \dot{X}_i^k = X_i^k \left(\sigma_i^k \cdot Y^k - \sum_{j=1}^{N^k} \rho_{ij}^k X_j^k \right), \\ \tau \dot{Y}^k = Y^k \left(1 - \beta \sum_{i=1}^{N^k} X_i^k - Z^k \right), \\ \theta \dot{Z}^k = \sum_{m=1}^{M} \xi^{km} Y^m - Z^k \end{cases} \tag{7}$$

$i = 1, \ldots, N^k$, $k = 1, \ldots, M$. For the sake of simplicity, let us assume that $\tau = \theta << 1$, so one can apply geometric singular perturbation theory (see, for instance, Jones, 1995; Hek, 2010 and references therein). In order to avoid confusion, it is important to say that the assumption $\tau = \theta << 1$ implies that, in contrast to the dynamics of X, the chunking dynamics is a composition of fast and slow motions. The fast motions lead variables Y-th and Z-th to a neighborhood of the slow manifold in the phase space. The evolution of the chunk variables on this manifold in the vicinity of the metastable states is much slower than the X variables. This corresponds to the intuitively clear fact that the "enveloping" variables mimic the averaging dynamics of X. Computer experiments confirm this explanation (see **Figure 4**).

The limit slow manifold has the equations $Y^k\left(1 - \beta\sum_{i=1}^{N^k} X_i^k - Z^k\right) = 0$, $\sum_{m=1}^{M} \xi^{km} Y^m - Z^k = 0$, thus, $\sum_{m=1}^{M} \xi^{km} Y^m = 1 - \beta\sum_{i=1}^{N^k} X_i^k$. Denote by ξ the $m \times m$-matrix ξ^{km}. If $\det\xi \neq 0$, we find

$$Y^k = \frac{1}{\det\xi}\left(\sum_{m=1}^{M} \eta^{mk} - \beta\sum_{m=1}^{M} \eta^{mk}\sum_{i=1}^{N^m} X_i^m\right) \tag{8}$$

Table 1 | Sequential dynamics in neural and cognitive systems.

Phenomenon/image	Model	References	Comments
Voting paradox / Structurally stable heteroclinic cycle	Kinetic (rate) equation, Lotka–Volterra model	Krupa, 1997; Stone and Armbruster, 1999; Ashwin et al., 2003; Postlethwaite and Dawes, 2005	J. C. Borda and the Marquis de Condorcet (De Borda, 1781; Saari, 1995) analyzed the process of plurality elections at the French Royal Academy of Sciences. They predicted the absence of a winner in a 3 step voting process (Condorcet's triangle)
Learning sequences	Hopfield type non-symmetric networks with time delay including spiking neuron models	Amari, 1972; Kleinfeld, 1986; Sompolinsky and Kanter, 1986; Minai and Levy, 1993; Deco and Rolls, 2005	Networks proposed to explain the generation of rhythmic motor patterns and the recognition and recall of sequences
Latching dynamics	Potts network is able to hop from one discrete attractor to another under random perturbation to make a sequence	Treves, 2005; Russo et al., 2008; Russo and Treves, 2011; Linkerhand and Gros, 2013	The dynamics can involve sequences of continuously latching transient states
Sequential memory with synaptic dynamics / Chaotic itinerancy sequences of Milnor attractors or attractor ruins	Spike-frequency-adaptation mechanism Noisy dynamical systems. Cantor coding	Tsuda, 2009	Proposed to be involved in episodic memory and itinerant process of cognition
Winnerless sequential switchings along metastable states/Stable heteroclinic channel	Generalized coupled Lotka–Volterra equations	Afraimovich et al., 2004; Rabinovich et al., 2008a,b	Information processing with transient dynamics at many different description levels from simple networks to cognitive processes
Winnerless competitive dynamics in spiking brain networks	Random inhibitory networks of spiking neurons in the striatum	Ponzi and Wickens, 2010	Neurons form assemblies that fire in sequential coherent episodes and display complex identity–temporal spiking patterns even when cortical excitation is constant or fluctuating noisily
Sequences of sequences / Hierarchical transient sequences	Recognition of sequence of sequences based on a continuous dynamical model	Kiebel et al., 2009	Speech can be considered as a sequence of sequences and can be implemented robustly by a dynamical model based on Bayesian inference. recognition dynamics disclose inference at multiple time scales

where η^{km} is the cofactor of the entry $\xi^{mk}$of the matrix ξ. Substituting this expression into the first equation of the system (7) we obtain the system

$$\dot{X}_i^k = X_i^k\left(\sigma_i^k \frac{1}{\det\xi}\sum_{m=1}^{M}\eta^{mk} - \sum_{j=1}^{N^k}\rho_{ij}^k X_j^k - \frac{\beta}{\det\xi}\sum_{m=1}^{M}\eta^{mk}\sum_{i=1}^{N^k}X_i^m\right) \tag{9}$$

$i = 1, \ldots, N^k,\ k = 1, \ldots, M$, which is similar to the binding model described in Rabinovich et al. (2010a). In particular, the "in-chunk" dynamics in (9) corresponds to the dynamics in the modality subspace in Rabinovich et al. (2010a). The main peculiarity of the system (9) is that the rates of coupling coefficients between different chunks have the common factor β, so if $\beta = 0$ then the interaction between different chunks is absent. Similarly to the study in Rabinovich et al. (2010a), one can impose conditions under which there exists a heteroclinic cycle for each chunk and successive heteroclinic connections between saddle points in different cycles. The last claim has the form $\beta > \beta_{cr}$ where β_{cr} depends on the parameters of the system (9). If τ is small then because of the geometric singular perturbation theory, the imposed conditions shall guarantee the existence of a heteroclinic network in the original system (7) corresponding to the "in-chunk" and "inter-chunk" dynamics.

Observations on the temporal chunk signal have focused on the use of pauses in behavior to probe chunk structures in WM. On the basis of some of these studies, a hierarchical process model has been proposed, which consists of four hierarchical levels describing different kind of pauses. The lowest level consists of pauses between strokes within letters. On higher levels, there are pauses between letters, words, and phrases. Each level is associated with a larger amount of processing when retrieving these chunks from memory (Cheng and Rojas-Anaya, 2006). Writing may be an effective approach to the study of cognitive phenomena that involves the processing of chunks. In Cheng and Rojas-Anaya (2003), it was demonstrated that in the writing of simple number sequences the duration of pauses between written elements (digits) that are within a chunk are shorter than the pauses between elements across the boundary of chunks. This temporal signal is apparent in un-aggregated data for individual participants in single trials. Mathematically the time intervals between chunks and super-chunks are controlled by parameter β (see Equation 3).

DISCUSSION

In this paper we have shown how the architecture of hierarchical mental model networks affected their associated functions. The discussed examples illustrate that networks with metastable states having several unstable separatrices exhibit very diverse cognitive functions (behavior). Complex heteroclinic networks allow completely new dynamical phenomena, and one of the primary challenges is the assessment of the existence and stability of hierarchical—chunking processes that can represent cognitive activity.

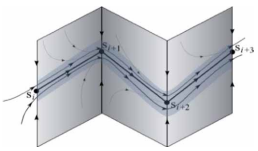
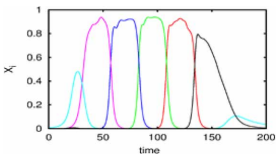
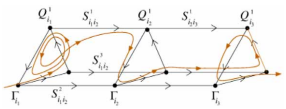
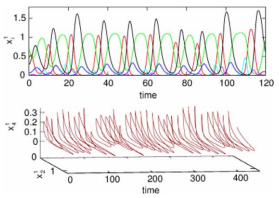
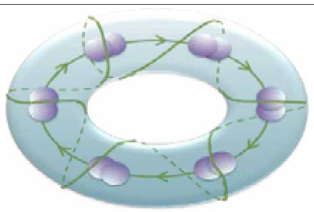
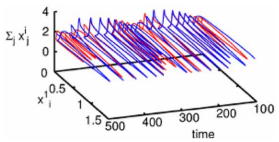
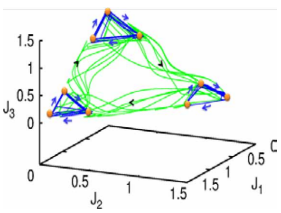
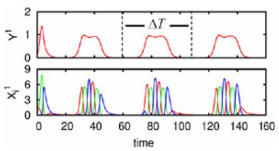
It is important to remind that the modeling of cycling and sequential dynamics in behavior and cognition has a long history (see several representative efforts in **Table 1**). Most of these models are based on Hopfield type networks. The main problem there is to keep the stability of the recall sequences against noise.

The results of chunking dynamics reported in this paper can be viewed as relevant in the description of different cognitive tasks. For example, in WM, humans encode items and synthesize them. With that, we give meaning to ideas and find a relevant place for them in our cognitive world. In these actions the interaction between WM and chunking are reciprocal—first of all WM is the "engine" of chunking, and on the other hand, the chunking makes WM capacity higher.

The model of chunking dynamics discussed in this paper relies on heteroclinic dynamics. It is important to emphasize that the main features of the SHC do not depend on the specific model used. The conditions of existence and the dynamical features of SHCs can be implemented in a wide variety of models: from simple Lotka–Volterra descriptions to complex Hodgkin–Huxley models, and from small networks to large ensembles of many elements (Varona et al., 2002; Venaille et al., 2005; Nowotny and Rabinovich, 2007; Rabinovich et al., 2012). The intrinsic hierarchical nature of the SHC at different temporal and spatial scales allows implementing many types of cognitive dynamics. Within this framework, brain networks can be viewed as non-equilibrium systems and their associated computations as unique patterns of transient activity, controlled by incoming input. The results of these computations can be reproducible, robust against noise, and easily decoded. Using asymmetric inhibition appropriately, the space of possible states of large neural systems can be restricted to connected saddle points, forming SHCs. These channels can be thought of as underlying reliable transient brain dynamics. **Table 2** summarizes four types of heteroclinic networks that can describe different aspects of sequential dynamics in cognitive processes: (i) A canonic heteroclinic network that produces reproducible sequential switching from one metastable state to another inside one modality (like in a simple WM task); (ii) A network displaying inhibitory-based heteroclinic binding dynamics that is responsible for the stable perception of a subject based on three different modalities; (iii) Two different modalities dynamically coordinated by excitatory connections; (iv) A chunking heteroclinic network that controls the grouping of elements of sequential behavior.

Mathy and Feldman have recently suggested to use the Kolmogorov complexity and compressibility (Mathy and Feldman, 2012) for the definition of a "chunk": a chunk is a unit in a maximally compressed code. The authors presented a series of experiments in which they manipulated the compressibility of stimulus sequences by introducing sequential patterns of variable length. To explore the influence of chunking on the capacity limits of WM, and departing from Bick and Rabinovich (2009), authors in Li et al. (2013) have suggested a model for chunking in sequential WM. This model also uses hierarchical bidirectional inhibition-connected neural networks with WLC. Assuming no interaction between a basic sequence and a chunked sequence, and the existence of an upper bound to the inhibitory weights the network, authors show that chunking increases the number of memorized items in WM from the "magical number" 7–16 items. The optimal number of chunks and the number of the memorized items in each chunk correspond to the "magical number 4."

Table 2 | Heteroclinics in mind.

Phenomenon	Network formalism*	Phase portrait	Time series
Sequential heteroclinic switching	$\dot{X}_i = X_i\left(\sigma_i - \sum_{j=1}^{N} \rho_{ij} X_j\right)$		
Sequential heteroclinic binding and information flow	$\dot{X}_i^l = X_i^l\left(\sigma_i^l - \sum_{j=1}^{N} \rho_{ij}^l X_j^l - \sum_{m=1}^{L}\sum_{j=1}^{N} \xi_{ij}^{lm} X_j^m\right)$		
Heteroclinic cooperation	$\tau_i^m \dot{X}_i^m = X_i^m \cdot \left[\sigma_i^m - \sum_{j=1}^{K^m} \rho_{ij}^m X_j^m + \sum_{k=1}^{M}\sum_{j=1}^{K^m} \xi_{ij}^{mk} X_j^k\right]$		
Hierarchical chunking memory and learning	$\dot{X}_i^k = X_i^k\left(\sigma_i^k \cdot Y^k - \sum_{j}^{N^k} \rho_{ij}^k X_j^k\right)$ $\tau \dot{Y}^k = Y^k\left(\left(1 - \beta \sum_{i}^{N^k} X_i^k\right) - Z^k\right)$ $\theta \dot{Z}^k = \sum_{m=1}^{M} \xi^{km} Y^m - Z^k$		

*See the definition of the variables and parameters in the text.

Recent experiments have confirmed the existence of three levels of cognitive hierarchy—see Rosenberg and Feigenson (2013). In this paper authors reported that infants can unify the representation of chunks into "*super-chunks.*"

The chunking models discussed above can be generalized on more complex cases. In particular, by adding attention control in the network hierarchy, it is possible to analyze the binding of sequences of chunks. The brain could use such binding to perform many cognitive functions like the coordination of visual perception with speech comprehension, or the coordination of music chunks and word chunks in singing processes. It is well-known that viewing a speaker's articulatory movements substantially improves a listener's ability to understand spoken words, especially under noisy environmental conditions like in a crowded cocktail party. Ross and coauthors claimed that this effect is most pronounced when the auditory input is weakest. As a result of attentional binding—multisensory integration—, substantial gain in multisensory speech enhancement is achieved at even the lowest signal-to noise ratios (Ross et al., 2007).

The dynamics of hierarchical heteroclinic networks is also able to explain and predict the coordination of behavioral elements with different time scales (for a study about the coordination of sensorimotor dynamics see Jantzen and Kelso, 2007). Functionally, such kind of synchronization can be the result of learning—the changing of the strength of inhibitory connections between agents at the different levels of the hierarchy in order to coordinate the dynamics with different time scales (see **Figure 3**). Additionally, it is important to note that the winnerless competitive learning process itself can be chaotic (Komarov et al., 2010), which provides wider possibilities for adaptability.

ACKNOWLEDGMENTS

Mikhail I. Rabinovich acknowledges support from ONR grant N00014310205. Pablo Varona was supported by MINECO TIN2012-30883. Irma Tristan acknowledges support from the UC-MEXUS-CONACYT Fellowship. Valentin S. Afraimovich was partially supported by Ohio University Glidden Professorship program.

REFERENCES

Afraimovich, V. S., Zhigulin, V. P., and Rabinovich, M. I. (2004). On the origin of reproducible sequential activity in neural circuits. *Chaos* 14, 1123. doi: 10.1063/1.1819625

Amari, S.-I. (1972). Learning patterns and pattern sequences by self-organizing nets of threshold elements. *IEEE Trans. Comput.* C-21, 1197–1206. doi: 10.1109/T-C.1972.223477

Ashwin, P., Field, M., Rucklidge, A. M., and Sturman, R. (2003). Phase resetting effects for robust cycles between chaotic sets. *Chaos* 13, 973–981. doi: 10.1063/1.1586531

Bahlmann, J., Schubotz, R. I., and Friederici, A. D. (2008). Hierarchical artificial grammar processing engages Broca's area. *Neuroimage* 42, 525–534. doi: 10.1016/j.neuroimage.2008.04.249

Bahlmann, J., Schubotz, R. I., Mueller, J. L., Koester, D., and Friederici, A. D. (2009). Neural circuits of hierarchical visuo-spatial sequence processing. *Brain Res.* 1298, 161–170. doi: 10.1016/j.brainres.2009.08.017

Bick, C., and Rabinovich, M. I. (2009). Dynamical origin of the effective storage capacity in the brain's working memory. *Phys. Rev. Lett.* 103:218101. doi: 10.1103/PhysRevLett.103.218101

Bick, C., and Rabinovich, M. I. (2010). On the occurrence of stable heteroclinic channels in Lotka–Volterra models. *Dyn. Syst.* 25, 1–14. doi: 10.1080/14689360903322227

Bor, D., and Seth, A. K. (2012). Consciousness and the prefrontal parietal network: insights from attention, working memory, and chunking. *Front. Psychol.* 3:1–14. doi: 10.3389/fpsyg.2012.00063

Botvinick, M. M. (2008). Hierarchical models of behavior and prefrontal function. *Trends Cogn. Sci.* 12, 201–208. doi: 10.1016/j.tics.2008.02.009

Bressler, S. L., and Kelso, J. A. S. (2001). Cortical coordination dynamics and cognition. *Trends Cogn. Sci.* 5, 26–36. doi: 10.1016/S1364-6613(00)01564-3

Bystritsky, A., Nierenberg, A. A., Feusner, J. D., and Rabinovich, M. (2012). Computational non-linear dynamical psychiatry: a new methodological paradigm for diagnosis and course of illness. *J. Psychiatr. Res.* 46, 428–435. doi: 10.1016/j.jpsychires.2011.10.013

Chekaf, M., and Matha, F. (2012). "Chunking memory span of categorizable objects," in *53rd Annual Meeting of the Psychonomic Society* (Minneapolis, MN: Psychonomic Society Publication).

Cheng, P. C. H., and Rojas-Anaya, H. (2003). "Writing out a temporal signal of chunks: patterns of pauses reflect the induced structure of written number sequences," in *Proceedings of the 27th Annual Conference of the Cognitive Science Society* (Mahwah, NJ: Lawrence Erlbaum), 424–429.

Cheng, P. C.-H., and Rojas-Anaya, H. (2006). "A temporal signal reveals chunk structure in the writing of word phrases," in *Proceedings of the Twenty-Eighth Annual Conference of the Cognitive Science Society* (Mahwah, NJ: Lawrence Erlbaum).

Clerget, E., Andres, M., and Olivier, E. (2013). Deficit in complex sequence processing after a virtual lesion of left BA45. *PLoS ONE* 8:e63722. doi: 10.1371/journal.pone.0063722

De Borda, J. C. (1781). *Memoire sur les Elections au Scrutin*. Paris: Academie Royale des Sciences.

Deco, G., and Rolls, E. T. (2005). Sequential memory: a putative neural and synaptic dynamical mechanism. *J. Cogn. Neurosci.* 17, 294–307. doi: 10.1162/0898929053124875

Dehaene, S., and Changeux, J. P. (1997). A hierarchical neuronal network for planning behavior. *Proc. Natl. Acad. Sci. U.S.A.* 94, 13293–13298. doi: 10.1073/pnas.94.24.13293

Dominey, P. F., Hoen, M., Blanc, J.-M., and Lelekov-Boissard, T. (2003). Neurological basis of language and sequential cognition: evidence from simulation, aphasia, and ERP studies. *Brain Lang.* 86, 207–225. doi: 10.1016/S0093-934X(02)00529-1

Engel, A. K., Fries, P., and Singer, W. (2001). Dynamic predictions: oscillations and synchrony in top-down processing. *Nat. Rev. Neurosci.* 2, 704–716. doi: 10.1038/35094565

Fadiga, L., Craighero, L., and D'Ausilio, A. (2009). Broca's area in language, action, and music. *Ann. N.Y. Acad. Sci.* 1169, 448–458. doi: 10.1111/j.1749-6632.2009.04582.x

Fingelkurts, A. A., and Fingelkurts, A. A. (2006). Timing in cognition and EEG brain dynamics: discreteness versus continuity. *Cogn. Process.* 7, 135–162. doi: 10.1007/s10339-006-0035-0

Friederici, A. D., Bahlmann, J., Friedrich, R., and Makuuchi, M. (2011). The neural basis of recursion and complex syntactic hierarchy. *Biolinguistics* 5, 87–104. Available online at: http://www.biolinguistics.eu/index.php/biolinguistics/article/view/170

Friston, K. J. (1997). Transients, metastability, and neuronal dynamics. *Neuroimage* 5, 164–171. doi: 10.1006/nimg.1997.0259

Friston, K. J. (2000). The labile brain. II. Transients, complexity and selection. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 355, 237–252. doi: 10.1098/rstb.2000.0561

Gobet, F., Lane, P. C. R., Croker, S., Cheng, P. C.-H., Jones, G., Oliver, I., et al. (2001). Chunking mechanisms in human learning. *Trends Cogn. Sci.* 5, 236–243. doi: 10.1016/S1364-6613(00)01662-4

Gros, C. (2007). Neural networks with transient state dynamics. *New J. Phys.* 9, 109. doi: 10.1088/1367-2630/9/4/109

Grossman, M. (1980). A central processor for hierarchically-structured material: evidence from broca's aphasia. *Neuropsychologia* 18, 299–308. doi: 10.1016/0028-3932(80)90125-6

Hek, G. (2010). Geometric singular perturbation theory in biological practice. *J. Math. Biol.* 60, 347–386. doi: 10.1007/s00285-009-0266-7

Huerta, R., and Rabinovich, M. (2004). Reproducible sequence generation in random neural ensembles. *Phys. Rev. Lett.* 93:238104. doi: 10.1103/PhysRevLett.93.238104

Ito, J., Nikolaev, A. R., and van Leeuwen, C. (2007). Dynamics of spontaneous transitions between global brain states. *Hum. Brain Mapp.* 28, 904–913. doi: 10.1002/hbm.20316

Jantzen, K. J., and Kelso, J. S. (2007). "Neural coordination dynamics of human sensorimotor behavior: a review," in *Handbook of Brain Connectivity*, eds V. K. Jirsa and A. McIntosh (Berlin; Heidelberg: Springer), 421–461.

Jones, C. K. R. T. (1995). Geometric singular perturbation theory. *Lect. Notes Math.* 1609, 44–118. doi: 10.1007/BFb0095239

Kelso, J. A. S. (1995). *Dynamic Patterns: The Self-Organization of Brain and Behavior*. Cambridge, MA: The MIT Press.

Kiebel, S. J., von Kriegstein, K., Daunizeau, J., and Friston, K. J. (2009). Recognizing sequences of sequences. *PLoS Comput. Biol.* 5:e1000464. doi: 10.1371/journal.pcbi.1000464

Kleinfeld, D. (1986). Sequential state generation by model neural networks. *Proc. Natl. Acad. Sci. U.S.A.* 83, 9469–9473. doi: 10.1073/pnas.83.24.9469

Koch, I., Philipp, A. M., and Gade, M. (2006). Chunking in task sequences modulates task inhibition. *Psychol. Sci.* 17, 346–350. doi: 10.1111/j.1467-9280.2006.01709.x

Komarov, M. A., Osipov, G. V., and Burtsev, M. S. (2010). Adaptive functional systems: learning with chaos. *Chaos* 20, 045119. doi: 10.1063/1.3521250

Krupa, M. (1997). Robust heteroclinic cycles. *J. Nonlinear Sci.* 7, 129–176. doi: 10.1007/BF02677976

Li, G., Ning, N., Ramanathan, K., He, W., Pan, L., and Shi, L. (2013). Behind the magical numbers: hierarchical chunking and the human working memory capacity. *Int. J. Neural Syst.* 23:1350019. doi: 10.1142/S0129065713500196

Li, K. Z. H., Blair, M., and Chow, V. S. M. (2010). Sequential performance in young and older adults: evidence of chunking and inhibition. *Neuropsychol. Dev. Cogn. Sect. B Aging Neuropsychol. Cogn.* 17, 270–295. doi: 10.1080/13825580903165428

Linkerhand, M., and Gros, C. (2013). Generating functionals for autonomous latching dynamics in attractor relict networks. *Sci. Rep.* 3:2042. doi: 10.1038/srep02042

Mathy, F., and Feldman, J. (2012). What's magic about magic numbers? Chunking and data compression in short-term memory. *Cognition* 122, 346–362. doi: 10.1016/j.cognition.2011.11.003

Meehan, T. P., and Bressler, S. L. (2012). Neurocognitive networks: findings, models, and theory. *Neurosci. Biobehav. Rev.* 36, 2232–2247. doi: 10.1016/j.neubiorev.2012.08.002

Miller, G. A. (1956). The magical number seven plus or minus two: some limits on our capacity for processing information. *Psychol. Rev.* 63, 81–97. doi: 10.1037/h0043158

Minai, A. A., Levy, W. B. (1993). "Sequence learning in a single trial," in *INNS World Congress on Neural Networks* (Hillsdale, NJ: Erlbaum), 505–508.

Musso, M., Moro, A., Glauche, V., Rijntjes, M., Reichenbach, J., Büchel, C., et al. (2003). Broca's area and the language instinct. *Nat. Neurosci.* 6, 774–781. doi: 10.1038/nn1077

Nowotny, T., and Rabinovich, M. I. (2007). Dynamical origin of independent spiking and bursting activity in neural microcircuits. *Phys. Rev. Lett.* 98:128106. doi: 10.1103/PhysRevLett.98.128106

Oullier, O., and Kelso, J. A. S. (2006). Neuroeconomics and the metastable brain. *Trends Cogn. Sci.* 10, 353–354. doi: 10.1016/j.tics.2006.06.009

Ponzi, A., and Wickens, J. (2010). Sequentially switching cell assemblies in random inhibitory networks of spiking neurons in the striatum. *J. Neurosci.* 30, 5894–5911. doi: 10.1523/JNEUROSCI.5540-09.2010

Postlethwaite, C. M., and Dawes, J. H. P. (2005). Regular and irregular cycling near a heteroclinic network. *Nonlinearity* 18, 1477–1509. doi: 10.1088/0951-7715/18/4/004

Rabinovich, M., Huerta, R., and Laurent, G. (2008a). Neuroscience. Transient dynamics for neural processing. *Science* 321, 48–50. doi: 10.1126/science.1155564

Rabinovich, M., Tristan, I., and Varona, P. (2013). Neural dynamics of attentional cross-modality control. *PLoS ONE* 8:e64406. doi: 10.1371/journal.pone.0064406

Rabinovich, M., Volkovskii, A., Lecanda, P., Huerta, R., Abarbanel, H. D., and Laurent, G. (2001). Dynamical encoding by networks of competing neuron groups: winnerless competition. *Phys. Rev. Lett.* 87:68102. doi: 10.1103/PhysRevLett.87.068102

Rabinovich, M. I., Afraimovich, V. S., Bick, C., and Varona, P. (2012). Information flow dynamics in the brain. *Phys. Life Rev.* 9, 51–73. doi: 10.1016/j.plrev.2011.11.002

Rabinovich, M. I., Afraimovich, V. S., and Varona, P. (2010a). Heteroclinic binding. *Dyn. Syst. An Int. J.* 25, 433–442. doi: 10.1080/14689367.2010.515396

Rabinovich, M. I., Huerta, R., Varona, P., and Afraimovich, V. S. (2006). Generation and reshaping of sequences in neural systems. *Biol. Cybern.* 95, 519–536. doi: 10.1007/s00422-006-0121-5

Rabinovich, M. I., Huerta, R., Varona, P., and Afraimovich, V. S. (2008b). Transient cognitive dynamics, metastability, and decision making. *PLoS Comput Biol* 4:e1000072. doi: 10.1371/journal.pcbi.1000072

Rabinovich, M. I., Muezzinoglu, M. K., Strigo, I., and Bystritsky, A. (2010b). Dynamical principles of emotion-cognition interaction: mathematical images of mental disorders. *PLoS ONE* 5:e12547. doi: 10.1371/journal.pone.0012547

Rabinovich, M. I., and Varona, P. (2011). Robust transient dynamics and brain functions. *Front. Comput. Neurosci.* 5:24. doi: 10.3389/fncom.2011.00024

Ribas-Fernandes, J. J. F., Solway, A., Diuk, C., McGuire, J. T., Barto, A. G., Niv, Y., et al. (2011). A neural signature of hierarchical reinforcement learning. *Neuron* 71, 370–379. doi: 10.1016/j.neuron.2011.05.042

Rosenbaum, D. A., Kenny, S. B., and Derr, M. A. (1983). Hierarchical control of rapid movement sequences. *J. Exp. Psychol. Hum. Percept. Perform.* 9, 86–102. doi: 10.1037/0096-1523.9.1.86

Rosenberg, R. D., and Feigenson, L. (2013). Infants hierarchically organize memory representations. *Dev. Sci.* 16, 610–621. doi: 10.1111/desc.12055

Ross, L. A., Saint-Amour, D., Leavitt, V. M., Javitt, D. C., and Foxe, J. J. (2007). Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments. *Cereb. Cortex* 17, 1147–1153. doi: 10.1093/cercor/bhl024

Russo, E., Namboodiri, V. M. K., Treves, A., and Kropff, E. (2008). Free association transitions in models of cortical latching dynamics. *New J. Phys.* 10:015008. doi: 10.1088/1367-2630/10/1/015008

Russo, E., and Treves, A. (2011). An uncouth approach to language recursivity. *Biolinguistics* 5, 133–150. Available online at: http://www.biolinguistics.eu/index.php/biolinguistics/article/view/186

Saari, D. G. (1995). *Basic Geometry of Voting*. Berlin: Springer. doi: 10.1007/978-3-642-57748-2

Sahin, N. T., Pinker, S., Cash, S. S., Schomer, D., and Halgren, E. (2009). Sequential processing of lexical, grammatical, and phonological information within Broca's area. *Science* 326, 445–449. doi: 10.1126/science.1174481

Sakai, K., Kitaguchi, K., and Hikosaka, O. (2003). Chunking during human visuomotor sequence learning. *Exp. Brain Res.* 152, 229–242. doi: 10.1007/s00221-003-1548-8

Schneider, D. W. (2007). Task-set inhibition in chunked task sequences. *Psychon. Bull. Rev.* 14, 970–976. doi: 10.3758/BF03194130

Sompolinsky, H., and Kanter, I. I. (1986). Temporal association in asymmetric neural networks. *Phys. Rev. Lett.* 57, 2861–2864. doi: 10.1103/PhysRevLett.57.2861

Stone, E., and Armbruster, D. (1999). Noise and O(1) amplitude effects on heteroclinic cycles. *Chaos* 9, 499–506. doi: 10.1063/1.166423

Tettamanti, M., and Weniger, D. (2006). Broca's area: a supramodal hierarchical processor? *Cortex* 42, 491–494. doi: 10.1016/S0010-9452(08)70384-8

Treves, A. (2005). Frontal latching networks: a possible neural basis for infinite recursion. *Cogn. Neuropsychol.* 22, 276–291. doi: 10.1080/02643290442000329

Tsuda, I. (2009). Hypotheses on the functional roles of chaotic transitory dynamics. *Chaos* 19, 015113. doi: 10.1063/1.3076393

Varona, P., Rabinovich, M. I., Selverston, A. I., and Arshavsky, Y. I. (2002). Winnerless competition between sensory neurons generates chaos: a possible mechanism for molluscan hunting behavior. *Chaos* 12, 672–677. doi: 10.1063/1.1498155

Venaille, A., Varona, P., and Rabinovich, M. I. (2005). Synchronization and coordination of sequences in two neural ensembles. *Phys. Rev. E Stat. Nonlin. Soft. Matter Phys.* 71:61909. doi: 10.1103/PhysRevE.71.061909

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.

Received: 18 October 2013; accepted: 10 February 2014; published online: 14 March 2014.

Citation: Rabinovich MI, Varona P, Tristan I and Afraimovich VS (2014) Chunking dynamics: heteroclinics in mind. Front. Comput. Neurosci. ***8****:22. doi: 10.3389/fncom.2014.00022*

This article was submitted to the journal Frontiers in Computational Neuroscience.

frontiers in
COMPUTATIONAL NEUROSCIENCE

HYPOTHESIS AND THEORY ARTICLE
published: 15 May 2014
doi: 10.3389/fncom.2014.00055

A non-linear dynamical approach to belief revision in cognitive behavioral therapy

David Kronemyer* and Alexander Bystritsky

Anxiety and Related Disorders Program, David Geffen School of Medicine, Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, CA, USA

Edited by:
Tobias A. Mattei, Ohio State University, USA

Reviewed by:
Tobias A. Mattei, Ohio State University, USA
Fatemeh Bakouie, Amirkabir University of Technology, Iran

****Correspondence:***
David Kronemyer, Anxiety and Related Disorders Program, David Geffen School of Medicine, Semel Institute for Neuroscience and Human Behavior, University of California, 300 UCLA Medical Plaza, Room 2330, Los Angeles, CA 90095-6968, USA
e-mail: dkronemyer@mednet.ucla.edu

Belief revision is the key change mechanism underlying the psychological intervention known as cognitive behavioral therapy (CBT). It both motivates and reinforces new behavior. In this review we analyze and apply a novel approach to this process based on AGM theory of belief revision, named after its proponents, Carlos Alchourrón, Peter Gärdenfors and David Makinson. AGM is a set-theoretical model. We reconceptualize it as describing a non-linear, dynamical system that occurs within a semantic space, which can be represented as a phase plane comprising all of the brain's attentional, cognitive, affective and physiological resources. Triggering events, such as anxiety-producing or depressing situations in the real world, or their imaginal equivalents, mobilize these assets so they converge on an equilibrium point. A preference function then evaluates and integrates evidentiary data associated with individual beliefs, selecting some of them and comprising them into a belief set, which is a metastable state. Belief sets evolve in time from one metastable state to another. In the phase space, this evolution creates a heteroclinic channel. AGM regulates this process and characterizes the outcome at each equilibrium point. Its objective is to define the necessary and sufficient conditions for belief revision by simultaneously minimizing the set of new beliefs that have to be adopted, and the set of old beliefs that have to be discarded or reformulated. Using AGM, belief revision can be modeled using three (and only three) fundamental syntactical operations performed on belief sets, which are expansion; revision; and contraction. Expansion is like adding a new belief without changing any old ones. Revision is like adding a new belief and changing old, inconsistent ones. Contraction is like changing an old belief without adding any new ones. We provide operationalized examples of this process in action.

Keywords: AGM theory, belief revision, cognitive behavioral therapy, cognitive restructuring, exposure/response prevention, non-linear dynamical psychiatry, systematic desensitization

Non-linear dynamical psychiatry recently has taken two different directions. The first is the granular description of neurological systems from a bottom-up, micro level, in order to characterize a cognitive phenotype such as emotion or attention (illustrative is Rabinovich et al., 2010a). The second is the functional description of psychopathology and corollary intervention strategies from a top-down, macro level, in order to characterize the course and progression of psychiatric disorders (illustrative is Bystritsky et al., 2012). Drawing on both, in this review we set forth a theory of belief revision for the intervention strategy known as cognitive behavioral therapy (CBT). CBT postulates that psychiatric disorders such as anxiety and depression are not caused by acts, transactions, events or circumstances in the real world, or by one's imaginal reconstruction of them. Rather, they result from one's attitude, orientation or outlook toward them. Persons who are anxious or depressed hold dysfunctional beliefs about themselves, others, their environment and the future. Dysfunctional beliefs are caused by an invalidating environment, deficient information-gathering practices and breakdowns in one's belief formation system (Warman et al., 2007). They often are accompanied by dysregulated emotions (Linehan, 1993). As a result, persons holding them engage in problematic or undesired behavior that is personally distressful or socially maladaptive, for example, anger, impulsivity, self-harm, self-isolation or substance abuse ("target behavior").

Belief revision is the primary therapeutic technology underlying CBT. As we will explain, it comes in two types. The first, called "cognitive restructuring," reformulates old beliefs and changes them into new ones. As a result, one is able to reregulate one's emotions and modify or abandon target behavior. The second results from behavioral change through a process called "systematic desensitization" or "exposure/response prevention." It extinguishes old, conditioned target behavior and introduces new more flexible, adaptive behavior. This in turn reformulates or discards old beliefs and reregulates emotions, reinforcing the newly-learned behavior. In both cases, the new behavior then stabilizes, consolidates and strengthens the new beliefs. Both are forms of belief revision: the former, more cognitively-based than behavioral; and the latter, more behaviorally-based than cognitive. Belief revision also reduces the intensity of interoceptive alarms activated by the sympathetic nervous system when stressed, such as those characteristic of panic (Khalsa et al., 2009; Domschke et al., 2010). CBT widely is regarded as the paradigm of an empirically-supported therapy (EST) (Butler et al., 2006), which

should make it particularly amenable to a cognitive science-based approach.

Our central premise is that belief revision in CBT is an integral component of a non-linear dynamical process of psychological change as conceptualized, for example, by Bystritsky et al. (2013). Anxiety and mood disorders have three essential components, which are alarms, beliefs and coping strategies (A-B-C). Alarms can be evaluated using conventional metrics such as their frequency, intensity, duration and onset. Coping strategies–a form of behavior–can be evaluated by whether they are distressful, maladaptive, or effective in down-regulating the incidence of target behavior and the intensity of correlative alarms. Beliefs are more difficult to integrate into a theory of non-linear dynamical systems. They have several unique characteristics as cognitive phenotypes, which prevent them from fitting well into the canonical model. One might not even notice one has beliefs to begin with, unless and until they are activated by environmental triggers, interoceptive sensations or undesired behavioral consequences.

Alternatively, we propose and demonstrate a set-theoretical, semantically-based approach to belief revision known as AGM theory, and show how it is the most plausible candidate to perform belief revision within a non-linear, dynamical framework. AGM is an acronym of the last names of its inventors, Alchourrón et al. (1985). It sets forth the requirements for non-delusional belief change in light of new evidence, and that one's resulting updated knowledge base must meet, in order to remain intuitively appealing (Carnota and Rodríguez, 2011, p. 2). As we discuss at §3, AGM operationalizes the cognitive component of CBT. Its objective is to define the necessary and sufficient conditions for belief revision by simultaneously minimizing the set of new beliefs that have to be adopted, and the set of old beliefs that have to be discarded or reformulated. Using AGM, belief revision can be modeled using three (and only three) fundamental syntactical operations performed on belief sets, which are expansion; revision; and contraction. Expansion is like adding a new belief without changing any old ones. Revision is like adding a new belief and changing old, inconsistent ones. Contraction is like changing an old belief without adding any new ones.

SOME RELEVANT CONSIDERATIONS ABOUT BELIEF

The nature of belief and what it is to believe in something (a doxastic state) both long have been central pre-occupations of psychology and epistemology (Schwitzgebel, 2010). It is beyond the scope of this review to discuss exhaustively the voluminous literature on belief, which has accumulated relentlessly since antiquity. We will, however, briefly develop several characteristics of belief pertinent to its integration into a theory of non-linear dynamical systems, which any theory of belief revision must take into account[1].

[1]Some of the other issues affecting beliefs that are beyond the scope of this review include (for starters): the subjective, phenomenological experience of belief; taxonomies of different types of beliefs; the relationship between beliefs and emotions; the role of memory; subjective probability theory; Bayesian epistemology; Dempster-Shafer theory; theories of reasoning; and rationality. In addition we do not here address objections such as logical omniscience, monotonicity and whether language (and beliefs) can be analyzed using a logical structure, to begin with.

A consensus definition is that beliefs are "states of mind that have the property of being about things–things in the world, as well as abstract things, events in the past and things only imagined" (Churchland and Churchland, 2013, p. 1). Russell (1921/2005) and colleagues famously developed a theory of propositions and propositional attitudes. What beliefs are about is their substantive propositional content, i.e., (that "x"). Belief is an attitude, orientation or outlook toward that propositional content, i.e., BEL("x"). The set of all of one's beliefs at time t_1 is one's knowledge base k_1. Beliefs are different than simple reference to people, places or things; informal or colloquial uses (Grice, 1975); as well as other modes of discourse such as performatives (Austin, 1962)[2]. While all of its individual elements are controversial in various respects, for our purposes, **Figure 1** depicts the standard model of belief, with components including perceptual, cognitive, emotional, linguistic and behavioral processing.

BELIEFS ARE BASED ON EVIDENCE

Evidence is a set of epistemological claims adduced to support a belief set. Relevant evidence enables one to devise and then test various hypotheses the belief set generates (Glymour, 1975; Hartmann and Sprenger, 2010). One is justified in believing that "x" to the extent one has good evidence for "x" (Feldman and Conee, 1985; Joyce, 2011). In the case of psychiatric disorders such as anxiety or depression, evidentiary data are things one might cite or rely on to support a contention that what one is *afraid* will occur, actually *will* occur. The feared outcome or consequence does not *actually* have to occur, rather, the evidence gives credence to the belief or prediction that it will.

From a clinical standpoint, the client is not responding to an object of fear; instead, to an internal symbolic representation of it, which (among other properties) has a compelling sense of reality. The client's behavioral expressions and coping strategies in turn are not a reaction to the feared object, but rather to the set of beliefs surrounding it, comprising the client's vision of what the feared object is, or might be. Under these circumstances, evidence is nothing more than the way things seem. One is "right to believe everything he believes as strongly as he believes it until it is rendered improbable by something else he believes" (Swinburne, 2011, p. 202). This support function often is conditional (Joyce, 2003). A conditional belief is one with the form

[2]In linguistics the study of how language actually is used is known as deixis (Brisard, 2011). Deixis is an example of how one's environment pragmatically imposes itself on one's beliefs. Although a word's semantic meaning may be fixed, what it actually means can vary with a number of factors, such as person, place and time. All of these are susceptible to ambiguous reference if viewed in isolation. It may not be clear, for example, who is designated by a pronoun. Spatial locutions such as "here" or "there" may designate more than one location, and temporal ones such as "now" and "then" might apply to different times (Corazza, 2011; Hanks, 2011). By constraining the limits of potential communication systems, ambiguity in natural languages actually may be adaptive (Piantadosi et al., 2012; Solé and Seoane, 2014). Deictic reference is a sub-category of indexical reference, which expands these principles to any context-sensitive use. Example: a vague expression with a hidden or latent variable, or one that has a particular meaning unique to a local community (such as "urban slang"), which often is uninterpretable absent specialized knowledge (Braun, 2007).

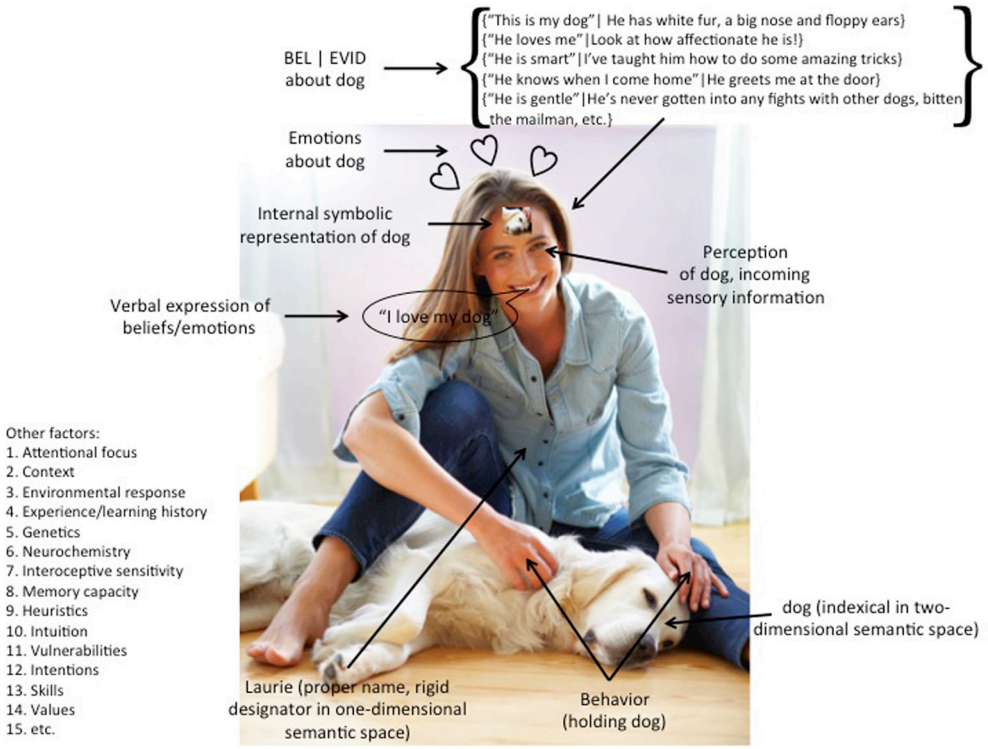


FIGURE 1 | Depiction of canonical belief model. Photograph licensed from Getty Images

BEL(x)|{EVID$_1$, EVID$_2$,. . . EVID$_n$}, which reads "BEL(that "x") assuming {EVID$_1$, EVID$_2$,. . . EVID$_n$}" (Arlo-Costa, 2007).

In psychiatry, evidence often is clinical observations of patient behavior or patient reports of symptoms set forth in the Diagnostic and Statistical Manual (DSM-5) (American Psychiatric Association, 2013). An example of the former: BEL("This person is depressed") | EVID("She has insomnia or hypersomnia nearly every day and significant weight loss when not dieting or weight gain, or decrease or increase in appetite nearly every day"). An example of the latter: BEL("I'm depressed") | EVID("I have markedly diminished interest or pleasure in all, or almost all, activities most of the day, nearly every day; and I have feelings of worthlessness or excessive or inappropriate guilt nearly every day"). Evidence also can be third-person observations or patient reports of them. Example: EVID("She always is fighting with her friends") or EVID("My parents always told me so"). Persons also may have corollary beliefs about their beliefs (Paulus and Stein, 2010). For example, one might BEL("Therapy/pharmacology doesn't help") or BEL("I'm going to have this for the rest of my life"). They also might be reflexive, as in BEL("I'm afraid of experiencing the symptoms of panic disorder").

REFERENTIAL OPACITY

A sentence's reference is what it designates. Sentences about beliefs are referentially "opaque" in that co-designating terms are not intersubstitutable (Quine, 1953/1980). To use a famous example, Oedipus married Jocasta; Oedipus believed Jocasta was his girlfriend; Oedipus didn't know Jocasta was his mother. This reads as follows: there was a time (t_1) when Oedipus believed "Jocasta was his girlfriend" (BEL$_1$) given the supply of evidentiary data {EVID$_1$, EVID$_2$,. . . EVID$_n$} then available to him. Even though true, Oedipus didn't believe at t_1 "Jocasta was his mother" (BEL$_2$), i.e., BEL$_2 \notin k_1$. He discovered this only at t_2, when (to his consternation) his knowledge base was k_2.

It follows that sentences about beliefs are informative in a way that "the sum of the angles of a triangle is 180°" is not. Another famous example from Gottlob Frege: one believes the morning star rises in the east; one also believes the evening star sets in the west; one doesn't know both are the planet Venus. Even though both sentences refer to the same thing, their meanings or "senses" are different (Zalta, 2012). Failures of reference do not require one to postulate intentional conduct. They may be due to something as simple as accident or mistake (Austin, 1956/1970)[3]. The main

[3]A related concept is intensionality, later developed by Rudolf Carnap (1947/1988). Intension roughly is the same thing as meaning or sense. It contrasts with extension, which roughly is the same thing as reference. For Carnap, two phrases or sentences have the same extension if they designate the same thing, i.e. they both are true or false with regards to it, so that one can be substituted for the other. Intensional ones fail this test, at least for our actual world. There is, however, a possible world or state-description with different conditions, in which there is substitutability of identity. That possible world could be our actual world at a different point in time, or even the knowledge base of different persons. Beliefs, according to Carnap, are neither extensional or intensional, because one can believe x but not y or z without realizing they

problem with belief reports is that they rely on a client's interpretation of her subjective phenomenological experience (Dattilio et al., 2010).

BELIEFS ARE SUBJECTIVE

Referential opacity is a set-theoretical way of saying that beliefs are inherently subjective. As *homo credens*, people are infinitely capable of believing any number of different things (Shermer, 2012). One might believe in unicorns, global warming, conspiracy theories, that the sun revolves around the earth, or that they are the present King of France. It is not our intention to restrict the content of different beliefs, or the types of evidence that may be adduced to support them.

Psychiatrists and psychologists have devised numerous ways to find out *what* people believe, including observing them, testing them and asking them. In this sense, beliefs are "epistemically objective." Implausible as it may seem, in the near future, it might even be possible to read a person's mind using neurotechnologies such as fMRI (Harris et al., 2008; Poldrack et al., 2011); neuropsychiatric phenomics (Bilder et al., 2009a,b); connectionist-type principles (Sporns et al., 2005); or interactionist-type principles (Stumpf et al., 2008)[4].

One of the perennial issues in cognitive science is whether these methods ever will be sufficient to account for belief's phenomenological texture. There is something unsatisfying about the neuromaterialistic/neurodeterministic program of extracting the substantive propositional content of a belief from neurological events. The reason why is because beliefs are underdetermined neurophysiologically; a single neurological state potentially could give rise to any number of different beliefs (they are "multiply realizable," (Levine, 1983, 1999); there is an "explanatory gap" between the two, Davidson, 1970, 1974). Further, they only can be held by the person who believes them. In this sense they are "ontologically subjective," as features or ascriptive predicates attributable only to that person (Dehaene, 2014, p. 9; Searle, 1995, pp. 7–9)[5]. From a clinical standpoint, there is no such thing as a standardized set of beliefs. Any approach to psychometric assessment that attempts to construct a taxonomy of typical beliefs, whether normative or pathological, most likely will not be successful, because beliefs fundamentally are distinctive, unique and personal. The clinician and the client must become co-investigators to identify them and the evidence ostensibly supporting them.

BELIEFS ARE MEDIATED AND MODERATED

Beliefs are mediated and moderated by any number of different factors such as background, upbringing, life experiences, information processing strategies, temperament, attributional style, other beliefs, context, culture, motivation, and the presence of environmental cues and situational primes (Hope et al., 2010). They may be teleological or subject to confirmation bias. People deploy a variety of heuristic reasoning strategies to arrive at the beliefs they hold, including hypothesis formation, generalization and anomaly resolution. Reasoning has a rational basis rooted in probabilistic approaches to problem-solving (Kahneman and Tversky, 1979; Tversky and Kahneman, 1983; Oaksford and Chater, 2007). These strategies have evolved over time to facilitate our ability to make decisions in situations with incomplete information as to potential outcomes (Kahneman et al., 1982; Shafer and Tversky, 1985; Kahneman, 2003; Michalewicz and Fogel, 2004). They include everything from educated guesses to intuitive judgments and common sense. Induction is an important aspect of human reasoning (Heit and Rotello, 2010; Johnson-Laird, 2010), as are techniques to evaluate the evidence in support of individual beliefs such as Bayesian reasoning and Dempster-Shafer theory (Curley, 2007; Zhao and Osherson, 2010; Zhao et al., 2012). There also is a complex relationship between cognition and emotion (§2.1.4, below; Pessoa, 2008, 2014). Beliefs are thought; emotions are felt. Just as one can have beliefs about one's emotions, so does one's emotional state affects one's belief-generating system. As with the subjective nature of beliefs (§1.3, above), while all of these are controversial in various respects, it is not our intention to restrict the nature, scope and extent of potential belief influencers.

CONDITIONS OF SATISFACTION

A proposition has the property that it is true or false in the real world (McGrath, 2012). Beliefs, on the other hand, have conditions of satisfaction–what happens when things are the way one believes them to be. BEL("It's raining") is satisfied if in fact it is raining. Under those circumstances, we say the belief is "true." Beliefs have a "mind-to-world" direction of fit, in that the belief corresponds, to some extent, with reality (Searle, 1983).

PSYCHOPATHOLOGY DISRUPTS THE ENTIRE BELIEF TEMPLATE

One of the best ways to consider belief as a psychological construct is to examine counterfactual cases (Langdon and Connaughton, 2013). Persons who are anxious or depressed have beliefs that are dysfunctional and experienced as negative and invalidating (Bernstein et al., 2010, 2013). Example: BEL("If I try to do this, I'm going to fail").

The main problem with dysfunctional beliefs is they cannot be assigned a truth value, as in BEL ("The cat is on the mat" | There is a creature of the genus and species *felis catus* lying prone upon a rectangle of flooring material). Rather, one *thinks* conditions of satisfaction have been met, or thinks *others* think they have, when in fact they have not. Example: BEL("I'm a terrible person") does not imply one in fact is a terrible person (under some plausible

all refer to the same thing. Phrases or sentences are "intensionally isomorphic" if in fact this intersubstitutability relationship nonetheless exists.

[4]The Human Connectome Project was established in September, 2010 by the U.S. National Institutes of Health (Vance, 2010). In April, 2013, the U.S. announced its BRAIN Initiative, a $1 billion connectionist-type project. It joined a similar €1 billion venture, the Human Brain Project, announced in January, 2013 by the E.U. (Abbott, 2013; Reardon, 2014). Internet companies such as the Allen Institute for Brain Science (Carey, 2012); Google (Markoff, 2012); and Vicarious (Albergotti, 2014) have similar objectives. Because connectionism results in something akin to a static, point-in-time wiring diagram, it is the opposite of non-linear dynamical psychiatry, see §4. Connectionism has obvious applications to artificial intelligence (AI), beyond the scope of this review to investigate further.

[5]Eliminative materialists such as Churchland et al. necessarily are committed to a theory that psychological disorders are a result of brain malfunction, for example, defective or impaired neurochemistry (Matthews, 2013).

consensus definition of what that means), or that others think so. Initially, negatively-valenced beliefs arise from misinterpretation of exteroceptive and interoceptive evidence and from information processing deficits (Paulus and Stein, 2010; Boden et al., 2012). Misevaluation of conditions of satisfaction then causes one to misjudge the evidence supporting the feared outcomes ("cost biases") (Nelson et al., 2010a,b).

Normatively, we are inclined to impose certain minimum requirements on a set of beliefs in order to maximize the likelihood there will be a match between beliefs and conditions of satisfaction. These include conformity, conditioning and coherence (Howson, 2009).

CONFORMITY

Conformity disregards the substantive propositional content ("x") of BEL("x") and requires only that one not endorse ("-x") simultaneously. Actual human reasoning might not be quite that simple. Research shows that people deal with inconsistencies not by attempting to refute one of the premises, but rather by trying to explain their origins, which has the side effect of revising their beliefs (Khemlani and Johnson-Laird, 2011).

CONDITIONING

Conditioning means that one should hold BEL("x") only for so long as {EVID$_1$, EVID$_2$,... EVID$_n$} support (x) and that one must update (x) in light of new, incoming EVID. Such an update may involve modifications to the belief's conditions of satisfaction. Acquiring, maintaining and using new evidence in order to revise and update beliefs is a crucial human survival strategy (Patterson and Barbey, 2013). When incorrect or obsolete, conceptual knowledge must be repaired by integrating and explaining new material (Friedman and Forbus, 2011).

COHERENCE

Coherence means that only tautological falsehoods qualify for a probability assignment of $p(x = 0)$ and only tautological truths qualify for $p(x = 1)$. Thus one should not assign $p(\text{BEL}) = 0$ to (BEL = "the sum of the angles of a triangle is 180°"), §1.2, above. Rather, one should assign it $p(\text{BEL}) = 1$.

Although they seem sensible, these axioms often do not apply to psychopathological states, because cognitive processing systems are impaired and emotion processing systems are dysregulated. Persons holding dysfunctional beliefs also may not be able to reason normatively. For example, they may disbelieve a set of propositions (e.g., evolution, global warming), which (most) everybody else believes (Perring, 2010). They may be indifferent to antecedent beliefs and stored knowledge; misunderstand inferential relationships; prioritize anomalous perceptual experiences; and lack a coherent theory of mind (Davies and Coltheart, 2000). It also makes sense to think of sentences expressing the ideations of persons with psychiatric disorders (§1.2, above) as ultra-opaque, thus even less amenable to substitutability of identity.

Their ability to evaluate evidence also may be impaired. Normatively, one relies on evidence to support a belief that what one *thinks* will occur, actually *does* occur. The evidence does not contradict, and in fact supports, the belief. In problematic cases, though, one does not have to believe a feared outcome or consequence actually *will* occur. Rather, all one has to believe is that the evidence supports the *belief* that it will, regardless of whether it happens or not (Joyce, 2011; §1.1, above). In such cases, the evidence supporting the belief is misaligned with reality (Warman et al., 2007; Möller, 2012). Clearly this is a slippery slope. If people can believe whatever they want, then what's to stop them, particularly if they have a mental disorder?

SUBJECTIVE PROBABILITY THEORY

There are two modern epistemic interpretations of probability, which are logicism and subjectivism (Galavotti, 2011). Logicism contends that probability is a person-independent, normative relationship between real-world facts or events. Subjectivism is the theory that probability is one's degrees of belief (Hájek, 2011). Under the logicist interpretation, a tautological statement (such as $A \rightarrow B$; A; $\therefore B$) is certain regardless of what people may think about it. Its probability p within a sample space Ω is 1 and in principle a large number of other beliefs can be incorporated within Ω so long as they are complementary (§1.6.3, above). Under the subjectivist interpretation, different persons can believe whatever they want and assign their beliefs different p-values, even given the same evidence, permitting wide intersubjective belief variation.

Subjectivism almost certainly is true when considering a person's individual beliefs (§1.3, above). It breaks down, however, when considering a set comprising different beliefs, all held by the same person. This surely is normative. It would be odd for a person only to have one belief. Most people probably hold tens of thousands, perhaps hundreds of thousands, of beliefs, and their knowledge base most likely expands over time (Ohlsson, 2011, p. 293). The problem is not about subjectivism. Rather, it is about probability. Probability assessments do not occur on an interval scale, making it impossible to combine them or determine something analogous to their "mean" probability function using a linear pooling methodology (Wallsten et al., 1997)[6]. Beliefs comprising belief sets are interdependent, not independent. As a result, they cannot be evaluated using a differential equation or structural equation modeling approach. A differential equation approach will not work, because one cannot parameterize the values of the variables in order to create a belief change trajectory or phase portrait within a vector field. A structural equation modeling approach will not work, because one needs dimensionality reduction. For example, if one holds 13 separate beliefs, the binominal coefficient is 715. Their interaction effects are 13! (13 factorial), or 6,227,020,800. Beliefs simply cannot be converted into numbers. They are not variables with values. Consequently, there must be some other way to fit beliefs into a non-linear dynamical model.

BELIEFS HAVE SEMANTIC, PROPOSITIONAL CONTENT

The solution is that beliefs have semantic, propositional content. Semantic content need not be expressed in complete sentences or

[6]Primarily for this reason, it is not clear that a comprehensive Bayesian approach to belief formulation and revision (for a summary, see Davies and Egan, 2013) is viable.

even phrases. It can be concepts that either are the semantic content or that combine to form it (Laurence and Margolis, 2012). Beliefs are just such a conceptual state. Unlike variables populated by values, they must be elicited using a natural language and then comprised into sets at various stages of the belief generating process (t_1, t_2,. . . t_n). One selects beliefs and includes them as members of belief sets by promoting or prioritizing them ahead of others, based on one's credences in the evidence supporting them, or levels of confidence in their conditions of satisfaction (§1.5, above; Makinson, 2009; Dietrich and List, 2013). Credences are situated along a continuum ranging from complete certainty of falsehood (does not meet perceived conditions of satisfaction) to complete certainty of truth (meets perceived conditions of satisfaction), depending on the evidence (Joyce, 2009).

Preference functions

Individual beliefs are organized into sets by preference or ranking functions (γ), which assess the occurrence or persistence of the belief (Spohn, 2009). In order to assign a preference function, one must adopt a theory of utility to determine what counts as a desirable (utility-maximizing) action; establish degrees of belief; rank preferences; and determine what evidence counts as confirming what beliefs (Johnson-Laird, 2010, 2013; Meacham and Weisberg, 2011). The higher a belief's preference function, the more likely it is to provide a basis for behavior (Segerberg et al., 2009)[7]. Following this compilation process, different belief sets then can be evaluated in order to determine the nature, scope and extent of belief revision, most likely by a human skilled in use of the language in which the beliefs are expressed[8]. It is likely that different beliefs impose contrasting and disparate semantic burdens, based on factors such as prevalence, complexity, and the number of inferences involved.

Semantic encoding

An example of a technique that has been devised to elicit beliefs is the articulated thoughts in simulated situations (ATSS) think-aloud paradigm, initially developed by Davison et al. (Zanov and Davison, 2010). Computational semantics attempts to model key features of natural language processes such as word meaning, sentence meaning, pragmatic usage and background knowledge (Stone, 2014). Recent initiatives include WordNet (Princeton University, 2010); latent semantic analysis (LSA) (University of Colorado Boulder, 1998); and SNePS (SNePS. Research Group., 2013). WordNet is a lexical database that groups words into sets of distinct cognitive concepts. LSA evaluates word similarity by similarity of context of use. SNePS is a natural language knowledge representation and reasoning system. A SNePS sub-routine models belief revision to maintain conformity, conditioning and coherence (§1.6.1, §1.6.2, §1.6.3, above). It too requires both individual beliefs and their relationships to be semantically encoded. One of the research priorities of several of today's most prominent internet companies is to develop algorithms for natural language recognition. Apple acquired Siri in April 2010 (Wortham, 2010); Facebook announced Graph Search in January 2013 (Sengupta, 2013); Google announced Hummingbird in September 2013 (Miller, 2013); Yahoo announced SkyPhrase in December 2013 (Goel, 2013); and in February 2014, Wolfram released software intended to answer natural language queries with real-world information as a kind of "computational knowledge-engine" potentially demonstrating a form of "machine intelligence" (Lecher, 2014). One of the main challenges of these initiatives will be to capture the numerous shades and nuances of meanings used by fluent language speakers–the senses of words, in Fregean terms (§1.2, above).

Semantic entailment

Closely related are problems of semantic entailment, that is, when a phrase or sentence commits one to other associated concepts. A classic example: "Socrates lived in Greece" should be inferred from "Socrates lived in Athens." Words are organized into "semantic/associative neighborhoods within a larger network of words and links that bind the network together" (Nelson et al., 2013, p. 797); Schroeter (2012) characterizes it as a two-dimensional semantic space comprising rules for assigning values to words and sentences. Specifying exactly what these neighborhoods and networks are is challenging, because (as with semantic encoding, §1.8.2, above) it depends on acquiring paraphrases, lexical semantic relationships, and inferences in contexts such as question answering, information extraction and summarization–similar to the usages employed by a natural language speaker (Dagan et al., 2009).

BELIEFS DO NOT EXIST IN ISOLATION

As semantic entailment illustrates, beliefs are components of complex domains, knowledge sets and networks (Davidson, 1994/2005). The limits of certitude on the one hand and psychopathology on the other allow for a wide variety of different {BEL | EVID} (Huber, 2009). One has an extensive set of unspecific background beliefs, which are culturally sensitive and context-dependent. They are "encoded in our linguistic formulation of the problem" (Weisberg, 2011, p. 507). Activities such as data selection, acquisition and learning require constant revision to one's knowledge base. Belief formation is subject to the overwhelming intervention of human experience, chance events and real-world constraints (Oaksford and Chater, 2007).

Quine and Ullian (1978) refer to this as a "web of belief"–"The totality of our so-called knowledge or beliefs, from the most casual matters of geography and history to the profoundest laws of atomic physics or even of pure mathematics and logic, is a man-made fabric which impinges on experience only along the edges" (Quine, 1953/1980, p. 42). Another way to look at beliefs is how they fit into what Searle (1995) calls the "background"–"all of those abilities, capacities, dispositions, ways of doing things and general know-how that enable us to carry out our intentions and apply our intentional states generally" (Searle, 2010, p. 31); or, the "foundational, non-representational non-rule-governed, dispositional structure of everyday understanding that underpins both our perception and our reasoning" (Rhodes and Gipps, 2008, p. 295).

[7]Other than noting its important function, it is beyond the scope of this review to assess γ's mechanism of action.

[8]Obviously this may be any type of language capable of performing this function.

DYNAMICS OF NATURAL LANGUAGE FORMATION

Another important factor involved in belief semantics is the dynamics of natural language formation. Any language must have certain minimal constructs and features. These include generativity (one can create an indefinite number of new sentences from its component elements); discreteness (semantic elements, such as words, retain their identity, even in different syntactical contexts); compositionality (smaller language units, such as words, can be combined to form more complex ones, such as sentences); predictability; and recursion (phrases can be embedded within phrases to create new sentences) (Hauser et al., 2002; Studdert-Kennedy, 2005; Searle, 2007). Noam Chomsky famously theorized there was a universal human linguistic structure, which he called "generative grammar" (Chomsky, 1955, 1965). For Chomsky, syntax was the essential component of language, as opposed to semantics (meaning and reference) and pragmatics (how language actually is used) (Chomsky, 1977)[9].

LANGUAGE AND MIND

It is beyond the scope of this review to investigate the complex relationships between language and mind (for a current overview, see Gleitman and Papafragou, 2012, 2013). Issues include criticism of Chomsky's views; whether logical variables represent the propositional contents of mental states and that cognition consists in manipulating them, a view most closely associated with Jerry Fodor (1975); criticism of Fodor's views; the linguistic relativity hypothesis (Swoyer, 2003); whether one can observe thoughts or emotions without labeling them (Linehan, 1993); or whether simply changing the way one labels them is effective to initiate cognitive/affective/behavioral change (Lieberman et al., 2007; Hayes et al., 2012). Our concern is not just a matter of choosing new words to describe beliefs, but rather reformulating beliefs, which then are expressed using words. At a minimum, we are in accord with Davidson (1975), who holds that belief is central to thought and that to have a belief requires the ability to express it using words[10].

The substantive propositional content of an individual belief is interesting and important, particularly for determining just which dysfunctional beliefs typically align with different types of psychopathology. We are more interested, though, in the relationship of an individual belief to the other constituents of the belief set of which the individual belief is a member, and how that set's membership changes or is reformulated between t_1 and t_n. Belief revision does not involve alteration or replacement of that which the belief is about, i.e., the "x" in BEL(that "x"). It is not a form of reality modification. Rather, the focus of change is belief considered as a propositional attitude (§1, above). The nature, scope and extent of belief revision only can be evaluated by inspecting modifications to the semantics of sets of {BEL | EVID} at k_1 and k_n.

INTEGRATING BELIEF INTO A NON-LINEAR DYNAMICAL SYSTEM

Given these complex conditions, how can belief revision using CBT be integrated into a theory of non-linear, dynamical systems? As set forth at our Introduction, above, belief revision essentially involves two separate pathways: one through cognition, the other through behavior. CBT straightforwardly uses interventions directed toward both. The first, cognitive restructuring, requires belief revision in order to initiate behavioral change. The second, exposure/response prevention, requires behavioral change in order to initiate belief revision. Both cognitive restructuring and exposure/response prevention are mechanisms of belief revision from k_1 to k_2 ($k_1 \Delta k_2$). **Figure 2** illustrates their respective critical paths for a client presenting with borderline personality disorder, DSM-5 §301.83.

COGNITIVE RESTRUCTURING

Cognitive restructuring is the therapeutic technology underlying the "cognitive" component of CBT (Spiegler and Guevremont, 2009). It contends that belief revision is the active ingredient motivating behavioral change: if belief set k_1 at time t_1 is modified to belief set k_n at time t_n, then more adaptive behavior will follow (Leahy, 2001, p. 23). Cognitive restructuring erodes dysfunctional beliefs through several steps: (1) identify them; (2) marshal disconfirming evidence against them; (3) deconstruct them by challenging and refuting them; (4) replace them with alternative, more functional beliefs; and then (5) conduct behavioral experiments to see how the environment responds (Huppert, 2009; McMillan and Lee, 2010; Morina et al., 2011). Examples of cognitive-oriented interventions include decatastrophizing, disputing the evidence, detecting logical errors, chain analysis, situational analysis, etc. (Leahy and Rego, 2012).

Clinical interventions look something like these: If one is afraid of snakes, that belief can be challenged through a series of counter-examples. A herpetologist might be concerned with the snake's various anatomical features. A veterinarian might be concerned with its health. A herpetoculturist might be concerned with its taxonomy. Some people have them as pets, or pose with them for photographs, or perform with them in theatrical productions. Each of these persons has a different, proactive mental stance toward things that are (or that appear to be) snakes, none of which are threatening. Or, if a person with lived experience concedes suicidal ideations or reports parasuicidal target behavior, then one way to interrupt her might be to evaluate the evidence and establish the active ingredients of a life worth living: "We have no reliable information that persons who are dead have a better quality of life than persons who are alive. If you're dead, then therapy won't work and you won't be able to get better."

It follows that in order to recalibrate one's belief-generating system, one must modify one's credences in the evidence supporting the pathological belief. The first step in cognitive restructuring is to elicit BEL(x). Then, for example, BEL$^{\ominus}$("I'm afraid of x") at t_1 might get cognitively restructured into something

[9]The logical underpinnings of natural languages is an involved subject, beyond the scope of this review; for recent discussions, see Carruthers (2012) and Scholz (2011). Culbertson and Adger (2014) recently concluded that some grammatical rules (such as placing adjectives closest to the noun they modify) are innate.

[10]Davidson also contends that one must be aware one has a belief in order to hold it to begin with, because if one didn't, then one wouldn't be able to change it, because one wouldn't be able to recognize that the underlying belief was false. This type of metacognitive awareness might be helpful for eliciting beliefs, §1.8.2, above. However, we concur with Laurence and Margolis (2012) that such a requirement overstates the case.

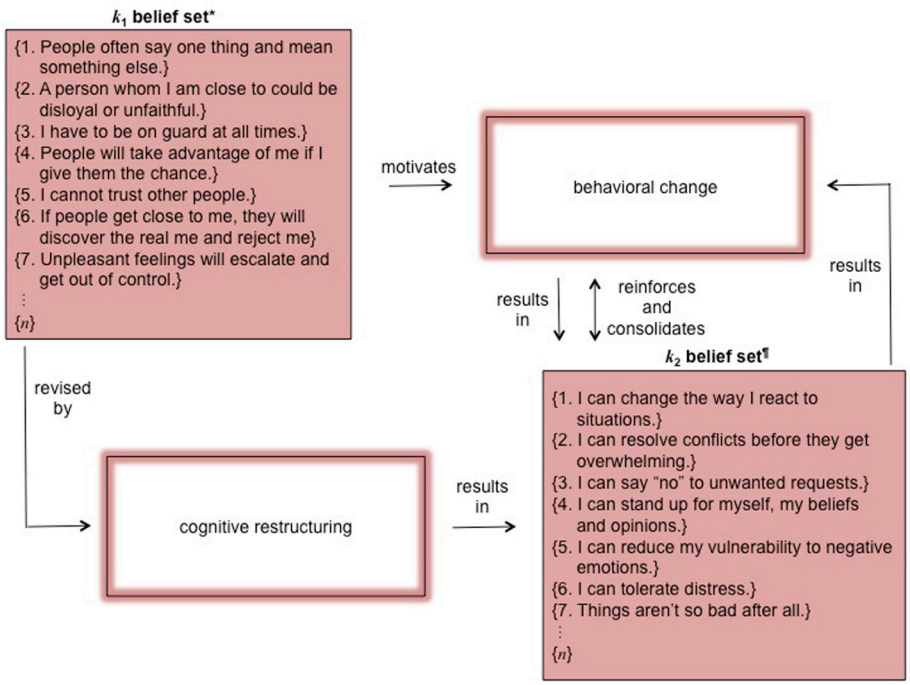


FIGURE 2 | Schematic of non-linear dynamical belief revision processes in CBT. *Adapted from Bahr et al., 2008. ¶Adapted from Linehan, 1993. k_1 is one's knowledge base at time t_1; k_2, at t_2; this example uses beliefs characteristic for a person presenting with symptoms consistent with a diagnosis of borderline personality disorder.

like BEL$^{\oplus}$("There've been times when I've encountered x and it wasn't so bad") at t_2. Positive belief attributions (BEL$^{\oplus}$) supplant negative ones (BEL$^{\ominus}$). Following cognitive restructuring, one then searches for discrepant evidence to confirm BEL$^{\oplus}$ and disconfirm BEL$^{\ominus}$, giving one a good reason to reformulate one's behavioral repertoire (Garland et al., 2010; Morina et al., 2011; Lightsey et al., 2012). Like belief, fear simply is another propositional attitude, i.e., {fear(x) | EVID}. Once one has accumulated enough relevant evidence, the choice clearly is framed: spend a significant portion of one's time entrained to the feared outcome, vs. the likelihood it actually will occur (i.e., conditions of satisfaction will be met, §1.5, above). From an assessment standpoint, this likely would require one to have good metacognitive awareness, that is, the ability to reflect upon, understand and control their learning (Schraw and Dennison, 1994) in order to be able to identify and articulate their beliefs. A related concept from attachment theory is that of reflective functioning, that is, the ability to observe and describe one's own mental state (Fonagy et al., 1991).

Cognitive restructuring presents several issues:

1. It is difficult to challenge entrenched beliefs, even when they result in target behavior. Although maladaptive, to some extent they relieve immediate personal distress. Over time they are reinforced and become a conditioned response to the circumstances triggering them, which consolidate around their utility and effectiveness (Hartley and Phelps, 2012).
   Example: aerophobia (fear of flying). In effect one has become fear-conditioned: the unconditioned stimulus (flying) initially provokes anxiety (unconditioned response), then becomes paired or associated with other typically-innocuous contexts or situations extrapolated from or analogized to the original one (such as acrophobia, fear of heights, the conditioned stimulus) (Samanez-Larkin et al., 2008). The resulting thought-pathways become ingrained with experience as they are reinforced by sufficient confirming evidence that maintains the associated beliefs until they become conditioned, learned responses (Tryon and McKay, 2009). One keeps doing the same thing over and over again because one is afraid of the perceived consequences of doing anything else.
2. Cognitive restructuring readily can morph into a form of escape/avoidance, if misapplied, because it feeds into intellectualization rather than the emotional, felt experience of a genuinely feared outcome. From a clinical perspective, too much thinking can become therapy-interfering, because one might approach the feared outcome as a puzzle to be solved. If this happens, then cognitive restructuring might backfire and one's tolerance of the feared outcome deteriorates even further. Feelings and thoughts both are in continuous competition for the same cognitive resources.
3. Because it involves a series of complex mental events, cognitive restructuring may be too complicated for many persons, especially those presenting with delusional features or severely dysregulated emotions (§4, below). They barely may be able to tolerate their dysfunctional beliefs, much less generate new ones. Persons with body dysmorphic disorder (BDD), for example, have a granular information processing style so they recall selective details of their appearance, rather than larger organizational design features (Feusner et al., 2007).

This makes it difficult for them to generalize from a specific exposure addressing a particular feared outcome to more global cognitive change. While one might become inoculated or desensitized to a particular trigger, establishing it also applies in other contexts requires deducing there is a more pervasive relationship between them–which is a cognitive process. In effect one must blunt the impulse toward fractalization.

If one adopts the wrong cognitive hypothesis, then it will be ineffective to revise the associated belief set. In order to be successful, cognitive restructuring must correctly identify the ultimate fear: "I'll lose control," "I'll be judged," "I'll be embarrassed and humiliated," "I'm going to die," etc. If one is afraid of physiological symptoms such as those characteristic of panic, then the question should be, what happens next? For example, if a client presents with symptoms consistent with a diagnosis of social anxiety disorder (SAD), such as vasodilation (blushing), then the consequence might be that "people think I'm an idiot." If people think one's an idiot, then the next consequence might be "I'll be rejected and abandoned." If one's rejected and abandoned, then the next consequence might be "I'll lose my job and my relationships," etc. If the terminal fear is not adequately specified, then target behavior actually might increase over baseline, because rather than contending with dysfunctional beliefs, one just has animated or enlivened them. The reason why is because one *thinks* one has handled the problem, but one really hasn't (§1.6, above). One just has deferred dealing with it. As a result, further triggers will continue to recruit and redeploy cognitive, affective and physiological assets to support it (Smits et al., 2008; Olthuis et al., 2012).

4. Cognitive restructuring essentially is a process of "out with the old, in with the new" using interventions such as those described at §2.1, above (Leahy and Rego, 2012). Because CBT regards dysfunctional beliefs as distortions or errors in thinking, such a challenge might be experienced as emotionally invalidating (Leahy, 2001, p. 58; Linehan, 1993, p. 92). Familiar (and to some extent serviceable) beliefs may be revealed as unrealistic, mistaken, distorted, or even irrational. As a result, subsequent behavior might just exchange one cognitive/affective state (e.g., anxiety) for another (e.g., "I'm deficient" or "I'm defective"). In this respect, dialectical behavior therapy (DBT) augments CBT case conceptualization. It emphasizes emotional validation in addition to cognitive restructuring. It is not enough to focus only on beliefs and behavior, because emotions (and their associated interoceptive sensations) also are an integral component of the same equation. In fact, if anything, in a contest between emotions and cognitions, emotions most likely will win out, because they are more fundamental and, in a sense, primordial (LeDoux, 1996; Damasio, 1999; Afraimovich et al., 2011; Frazzetto, 2013). A recent study by Moser et al. (2014) concluded that positively reinterpreting negative emotional experiences (such as those associated with fearful outcomes) is one of belief revision's key mechanisms, with well-defined neurological correlates. The equation *should* read: {dysfunctional beliefs} + {emotional dysregulation} = {target behavior}[11].
5. CBT uses phrases such as "downward arrow technique" (Persons et al., 2006) and "chain analysis" (Lynch et al., 2006) as metaphors for complex cognitive processes, without considering their component elements. This leaves beliefs in a kind of mysterious "black box"–something everyone knows must be addressed, but without unpacking their underlying logic and structure. What CBT lacks (and what we offer) is a theory of belief revision–which beliefs get changed, why those instead of others, and what the constraints are.
6. Cognitive therapy is a means to behavioral change, not an end in and of itself. During cognitive restructuring, one develops hypotheses that exposure/response prevention either will falsify or prove. For example, if a person with SAD undergoes cognitive therapy and concludes, "Well, I guess it's not so bad if I speak up at meetings," but then never does so, cognitive restructuring will not have been effective.

EXPOSURE/RESPONSE PREVENTION

CBT's second critical path is behavioral intervention based around the concept of progressive desensitization-exposure/response prevention to a feared outcome, rather than escape/avoidance of it. It proposes that the main driver for therapeutic change is behavior, not cognition. It assumes that it is difficult for cognition alone to motivate new behavior; that one of the main reasons why persons engage in target behavior is to attempt to induce their environment to respond; that when reinforcement contingencies are altered, behavioral modification follows; and that psychological change occurs as a result. Instead of being the driving force motivating behavioral change, cognition brings up the rear. This dichotomy is similar to that between thought and action, or thinking vs. doing.

Using this approach, the first question always must be "how did the behavior get to be the way that it is." Often this can be explained using classical and operant conditioning paradigms. Sometimes people enact coping strategies to prevent something bad from happening; occasionally, it may even be pleasurable. If, however, actions have *not* had effects, then it is necessary to supply them in order to consequate that behavior. The next step is to unpair or decouple a conditioned stimulus from an unconditioned one, or to extinguish target behavior that previously has been reinforced (and the entire cycle giving rise to it), by establishing prospective environmental contingencies; acquiring skills; enacting new behavior; and then evaluating evidence as to how the environment responds (Spiegler and Guevremont, 2009). At each stage, behavioral markers demonstrate that the feared outcome did not occur.

Target behavior typically is a form of escape/avoidance. It may be accommodating and protective in the short term, because it reduces the threat posed by dysfunctional beliefs (§2.1.1, above;

[11]While we spend considerable time analyzing pathways between cognition and behavior (§4), it is beyond the scope of this review to expand our analysis to include emotions and affect. For speculation on this point, see (Afraimovich et al., 2011; Huntsinger and Schnall, 2013); and (Rabinovich et al., 2010a).

Hofer, 2010). However, it is ineffective over the long term, as novel and even more threatening stimuli arise in the world and present for interpretation and action (Roemer et al., 2002; Carter et al., 2008; Lee et al., 2010). It does not affect one's pre-existing vulnerabilities and the environmental affordances that trigger or activate them. It does not down-regulate dysfunctional beliefs or dysregulated emotions. Instead, by impeding assimilation of accurate information, it maintains judgmental biases, emotional vulnerability and alarm sensitivity–a kind of "contrast avoidance" (Taylor and Alden, 2010; Newman and Llera, 2011, p. 226).

Adaptive new behavior, on the other hand, is generated by stepwise exposure followed by systematic desensitization or response prevention. Initially this is a "fragile behavioral state" and can be recovered "spontaneously or subsequent to environment influences, such as context changes or stress" (Herry et al., 2010, p. 599). As one confronts the feared stimulus, the fear becomes extinguished through a reverse inhibitory learning process, allowing for more flexible control of conditioned response by forming a consolidated extinction memory. With continued or reinitiated exposure, post-behavior cognitions consolidate and become further refined, dampening responsiveness in the brain's fear-sensitive network (Hauner et al., 2012; Trouche et al., 2013). Similar to cognitive restructuring (§2.1.3, above), in order to be an effective intervention, exposure/response prevention must be autogenic, i.e., personalized more or less exactly to falsifying or validating a specific feared outcome–the one that matters the most.

Example: if one is afraid of heights and things that move quickly, then an escape/avoidance strategy would be not to engage with them. An exposure/response prevention strategy, on the other hand, would be to take opposite action by (say) going on a series of roller-coaster rides at an amusement park, starting with those that are small and innocuous but then building up over the course of a day to those that are taller and faster. At each step one take's stock of one's mental condition, notices that one still is alive and breathing, thereby habituating or acclimating oneself to more challenging stimuli, resulting in cognitive change. Example: if one is afraid of driving on the freeway, then an escape/avoidance strategy would be to take surface streets. What happens, though, if the surface streets all are blocked and the only way to get to one's destination is by taking the freeway? The escape/avoidance strategy no longer works. A more adaptive exposure/response prevention strategy would be to progressively expose oneself to driving on the freeway by (say) traveling from one on-ramp to one off-ramp at a time, then gradually building this up to two, then three, etc. Example: rather than engaging in a difficult and potentially futile process of weighing pros and cons in order to motivate herself not to drink alcohol, a person with substance over-use issues alters her behavioral regimen not to drive by liquor stores and restructures her social network to exclude those persons maintaining it.

Behavior modification is powerful. Some theorists contend that in a contest between beliefs and behavior (i.e., cognitive restructuring versus exposure/response prevention followed by belief consolidation), behavior always will win; see e.g., Gipps (2013) and Longmore and Worrell (2007). Historically, committed behaviorists denied one has beliefs to begin with; rather, one only is disposed to respond to stimuli (Pavlov, 1927/2003; Skinner, 1947; Ryle, 1949/2009). Today, along similar lines, eliminative materialists such as Churchland and Churchland (1998) and Dennett (1992) deny beliefs are anything more than folk-psychological explanations (this phrase is intended to be mildly derisive) of complex neurological events (Bickle et al., 2010). The weakness of this formulation is what originally lead to the cognitive revolution, as exemplified, for example, by Chomsky's (1959) critique of Skinner's (1957/1991) *Verbal Behavior*. Behavior does not, however, occur in a vacuum. There must be some threshold level of belief revision in order to stimulate it, most likely based on the salience of an initial belief or belief set, its relevance to current goals, or its resonance with a particular feature of the environment. In principle this should be similar to the way that intention redirects attention from the default mode network to some other neural construct or constructs (Buckner et al., 2008; Rabinovich et al., 2012a). Attention focuses intentional orientedness, causing heightened self-monitoring, resulting in greater interoceptive sensitivity (Simmons et al., 2006; Woody and Nosen, 2009), one of the main precursors to belief change.

Thereafter, the role of cognition primarily is to consolidate revised beliefs and build behavioral insight. Beliefs are conjectures or predictions about conditions of satisfaction and the evidence supporting them. The only way to accumulate evidence is by enacting behavioral experiments and seeing what happens. From a clinical standpoint, the client can assume the role of an anthropologist, investigating the behavior of a strange tribe, of which she also happens to be a member. If there is insufficient evidence to support a belief, or the evidence disconfirms it, then there is no particular reason why it should be retained as a component element of a belief set. Discrepant evidence creates "expectation violations" (disconfirms pathogenic beliefs), modifying behavioral vectors previously directed toward averting feared outcomes, thereby raising the cognitive accessibility of alternative and more flexible belief formulations. In many instances, the cognitive objective is not to eradicate fear, but rather to tolerate ambiguity. Using a variation of the Rescorla and Wagner (1972) model, Craske et al. (2012) recently advocated that while it may become semi-perturbed, the pairing or coupling between the conditioned stimulus and the unconditioned stimulus never really is eradicated. Instead, it is inhibited or attenuated. It follows that variability in fear level, or reintroducing elements of the unconditioned stimulus concurrently with the conditioned stimulus during exposure, is more likely to create a durable learning experience. Doing so *maximally* violates expectations, eliciting more improvisational and extemporaneous behavior, thereby promoting belief revision (Kircanski et al., 2012). The goal is not so much extinction (from a behavioral standpoint) as it is acceptance (from a cognitive standpoint)—which is a completely different skill. As the Viennese novelist (and, in retrospect, proto-ACT theorist) Robert Musil (1930-43) declared: "one must live with uncertainty, yet not be caught in hesitation."

Cognition also extrapolates or pluralizes revised beliefs to analogous contexts. When one masters a skill in a certain domain, that mastery experience carries over to others. Only the target behavior will be affected without generalization effects. While this may be acceptable insofar as it goes, especially in refractory cases,

exposure/response prevention will have limited success unless it also addresses adjacent beliefs (Arntz, 2002; Bryant et al., 2003). To continue with the example from §2.1.6, above, if a person with SAD starts mindlessly speaking up at meetings, that will not in and of itself change cognition. It simply is a form of unregulated exposure/response prevention. It may even become a form of escape/avoidance if she engages in it unthinkingly in order to avoid cognitive dissonance, a necessary precursor to extinction. The more that target behavior is effective as a form of escape/avoidance, the more difficult it will be to create a counteracting exposure/response prevention, precipitating belief revision. Reciprocally, some persons who hold severely dysfunctional beliefs or who are considerably emotionally dysregulated may lack the cognitive capacity to perform generalization operations (§4, below). In such cases, target behavior must be specified even more precisely, otherwise it will not be extinguished, or some other undesired behavior will be reinforced instead.

AUTOMATIC NEGATIVE THOUGHTS, INTERMEDIATE BELIEFS, CORE BELIEFS

How do cognitive restructuring and exposure/response prevention integrate with the epistemology of CBT? Received Beck-Ellis theory (Ellis, 1994; Beck, 2011) holds that doxastic agents have a hierarchy of automatic thoughts, intermediate beliefs and core beliefs. There now are several dozen recognized schools of CBT, all of which trace their provenance back to Beck and Ellis (Emmelkamp et al., 2010).

Automatic thoughts

For Beck (2011), automatic thoughts are an undercurrent of cognitions and self-talk, subject to articulation on query or in response to an analogous simulation (Zanov and Davison, 2010). They rarely are conscious in the sense of a state one is aware of, however they typically are accessible and available to other cognitive processes (van Gulick, 2004).

Intermediate beliefs

Automatic thoughts are linked to core beliefs by intermediate beliefs. Beck (2011) assumes the role played by intermediate beliefs is unproblematic (p. 205), however they can be difficult to formulate and it is not clear anybody ever has held an intermediate belief. In principle they should be rules or assumptions in the form of conditional if-then statements such as: "If I (engage in rigid behavioral coping pattern), then (I'll be insulated from a core belief I'll experience as aversive)" or "Unless I (engage in rigid behavioral coping pattern), then (I'll be exposed to a core belief I'll experience as aversive)." For example, if one unexpectedly is running late for work because the bus is running late, intermediate beliefs might be: "If I'm always on time for meetings, then I'm not inadequate" (or, "Unless I'm always on time for meetings, then I'm inadequate"). They should not, however, be idiographic. Thus, "If I'm on time for meetings, then I'll do well at work" is not a proper formulation of an intermediate belief. Rather, it is more of an expression of a particular coping style, connecting to an individual instance of behavior, not a pattern of behavior. Nor should intermediate beliefs be depersonalized. Thus, "People who frequently are late for meetings typically end up losing their jobs" also is not a proper formulation of an intermediate belief, because the outcome does not tie to a more generalizable core belief.

Core beliefs

A core belief is not an actual thought in an epistemological sense. E.g., if the automatic thought is "I'm running out of money," then the associated core belief might be, "One needs a lot of money in order to be safe," even though one never actually thinks that particular core belief. Uncovering it is cognitive restructuring's *raison d'être*. It is tempting to think of a core belief as an implicit conclusion derived from the application of a rule (an intermediate belief) to a premise (an automatic thought). All three are components of an information processing system (Beck, 2011, p. 33) or a way for people to "organize their experience in a coherent way in order to function adaptively" (Beck, 2011, p. 35).

Still, it is not clear what comprises a set of core beliefs. Is it just a single belief, or a set of multiple, interdependent beliefs? Although they acknowledge the possibility that there are many of them, all of the Beck-Ellis examples treat beliefs as singletons rather than as elements of belief sets. It seems implausible that individual beliefs, regardless of how entrenched, proximately cause (or explain) a complex phenomenon such as human behavior. It seems more likely that human behavior is the outcome of a dynamic, interactive network of beliefs (and that it reciprocally influences them).

It also is unclear just what causes what. Does a trigger–a real-world or imaginal event–activate core beliefs or automatic thoughts? Once set in motion, which causes which? Beck (2011) has little to say about the relationships between automatic thoughts, intermediate beliefs and core beliefs other than core beliefs "activate" automatic thoughts (p. 32) and "underlie" (p. 36) both them and intermediate beliefs. Intermediate beliefs "influence" one's view of the situation or event (p. 35), which "trigger" automatic thoughts (p. 38) (Beck apparently views these different verb formulations as synonymous).

BELIEF REVISION–THREE AND ONLY THREE FUNDAMENTAL SYNTACTICAL OPERATIONS

While CBT provides useful tools that can be used to induce or facilitate belief revision such as cognitive restructuring or exposure/response prevention, the problems with Beck's (2011) formulation (§2.3, above) make clear that it comes up short to explain just how they do so. At best, from a clinical standpoint, they just "soften" a set of dysfunctional beliefs, or point out why individual beliefs are implausible (Beck) or illogical (Ellis). We contend that the process of belief revision in CBT can be better characterized using AGM[12].

[12]Since their original (1985) paper, AGM theory has evolved and undergone significant further developments (Makinson, 2003; Costa and Pedersen, 2011; Gärdenfors, 2011). While there are other theories of belief revision (Fermé and Hansson, 2011), AGM is the one that has acquired the most traction in the literature. The concept of *k*, whether and how BEL represents or stands for a psychological state, all of the AGM postulates and all of the operations potentially performable on *k* have been discussed and challenged extensively. It is beyond the scope of this review to analyze these various permutations.

According to AGM, a person's knowledge base k comprises a number of individual beliefs, BEL_1, BEL_2,... BEL_n, which combine together to form belief sets. AGM provides a set of ecological rules for how beliefs dynamically evolve by examining the interaction effect of k_1's and k_2's respective belief sets at equilibrium points t_1 and t_2 during the process of belief revision. The problem AGM is trying to solve is to minimize the set of $BEL_{new} \in k_2$ and the set of $BEL_{old} \notin k_1$ *simultaneously*, so as to maximally preserve both k_1's and k_2's inductive cores. Unlike k_1, k_2 is less subjectively distressing and leads to more adaptive or normative behavior.

This is interesting and important because it defines the necessary and sufficient conditions for belief revision–what has to happen and that is all that has to happen. It therefore specifies the minimum requirements necessary for successful cognitive restructuring or belief modification following exposure/response prevention. From a clinical standpoint, maybe this is all one can expect, particularly with difficult cases. It can accommodate a diverse belief set, limited only by one's strategies to interpret beliefs, semantically encode them by assigning them substantive propositional content (that "x") and then identify the resulting doxastic commitments, which gives it explanatory power. It deemphasizes the distinction between automatic thoughts, intermediate beliefs and core beliefs. All beliefs are targets for revision at any equilibrium point. This better explains the subjective phenomenological experience of belief revision. It also recognizes there are different related beliefs at t_1, t_2, etc. Some motivate behavioral change, e.g., $k_1 =$ ("If I enact behavioral experiment y then z will happen"). Others reinforce it, e.g., k_2 following skills acquisition or exposure/response prevention $=$ ("This is how the environment responded"). It is a dynamical system because it changes and evolves in real time. It is nonlinear because the "x" of BEL(x) is idiographic, idiosyncratic and unpredictable.

During belief revision, elements of belief sets are modified or replaced using three (and only three) fundamental syntactical operations, which are expansion (EXP); revision (REV); and contraction (CON). Particular beliefs are the semantics this architecture supports (Fermé and Hansson, 2011).

EXPANSION (EXP)

EXP is like adding a new belief without deleting any old ones. EXP (expressed as $k_1 + BELx$) occurs when one accepts, acknowledges or incorporates a BEL_{new} into k_1. $k_2 = (k_1 + BEL_{new})$: BEL_{new} is added to k_1; no $\exists(BEL\ x \in k_1)$ is deleted or removed from k_1; and on conclusion of belief revision, $\{(BEL_1 \ldots BEL_n) \cup BEL_{new}\} \subseteq k_2$, with the caveat it also is the smallest possible set of $(k_2 \cup BEL_{new})$. Although it might be, BEL_{new} does not necessarily have to be consistent with k_1. Since AGM does not restrict the substantive propositional content "x" of BEL_{new} (§1.3, above), it can have either $\oplus$ or $\ominus$ valence. If it has $\oplus$ valence ($BELx^{\oplus}$), then it contributes to cognitive restructuring at t_2. If it has $\ominus$ valence ($BELx^{\ominus}$), then either it does not contribute to cognitive restructuring, or may even reinforce k_1.

For this reason, EXP might be confusing for an AGM agent. $BEL_{old\ominus}$ remain as elements of her belief set, even as they are joined by BEL_{new}, which can either be $BEL^{\oplus}$, $BEL^{\ominus}$ or ambiguous. To continue with our previous example, the trigger is running late for a meeting at work because one's bus is late. Under such circumstances, one's beliefs might be: $BEL_{1\ominus}$ ("My boss is going to get angry"), $BEL_{2\ominus}$ ("My colleagues will disrespect me") and $BEL_{3\ominus}$ ("My opinion doesn't count"). One then acquires a new belief $BEL_{4\ominus}$ ("I need this paycheck to support myself"). $BEL_{4\ominus}$ is not inconsistent with $\{BEL_{1\ominus}, BEL_{2\ominus}, BEL_{3\ominus}\}$. For these reasons, we hypothesize that it is unlikely EXP alone will result in successful cognitive restructuring or belief consolidation following exposure/response prevention. **Figure 3** depicts this outcome.

REVISION (REV)

REV is like adding a new belief and deleting old, inconsistent ones. As with EXP, REV (expressed as k_1^* $BELx$) occurs when one accepts a BEL_{new} or admits it to one's k_1 knowledge base. $k_2 = (k_1 + BEL_{new})$: BEL_{new} is added to k_1; on conclusion, $\{(BEL_1 \ldots BEL_n) \cup BEL_{new}\} \subseteq k_2$. The main difference between REV and EXP is that with REV, a BEL_{old} must be *deleted* from k_1 so that k_2 is consistent with k_1.

Pragmatic Closure

k is "logically closed" if it represents *all* of one's beliefs, even though they may be difficult or impossible to specify. Every BEL

$\{k_1\}$ BEL_{old}	Δ_{bel}	$\{k_2\}$ BEL_{new}
$BEL_1^{\ominus}$ = "My boss is going to get angry."		$BEL_1^{\ominus}$ = "My boss is going to get angry."
$BEL_2^{\ominus}$ = "My colleagues will disrespect me."		$BEL_2^{\ominus}$ = "My colleagues will disrespect me."
$BEL_3^{\ominus}$ = "My opinion doesn't count."		$BEL_3^{\ominus}$ = "My opinion doesn't count."
EXP	$BEL_4^{\ominus}$ = "I need this paycheck to support myself."	$BEL_4^{\ominus}$ = "I need this paycheck to support myself."

FIGURE 3 | EXP.

logically derivable from k already $\in k$, i.e., k includes not only BEL but also all BEL consequences. Stand-alone beliefs sometimes are referred to as "basic beliefs" and consequences as "derived beliefs"–those beliefs one is epistemically committed to hold, even though one might not actively do so (Gabbay et al., 2010). Since k_1 is logically closed in this sense, only *one* anomalous BEL(x) is sufficient to create inconsistency; an inconsistent $k(x)$ sometimes is notated as $k(x)\perp$. In this respect, REV incorporates the concept of conformity (§1.6.1, above)[13].

Frame of discernment

To some extent the problem of logical closure is solved by the concept of "frame of discernment." The domain of all possible beliefs must be truncated in order to engage in practical inference and reason from belief to action. One's frame of discernment is the set of all of the beliefs comprising k that are useful to answer, in a practical context, the question of what one believes. It is notated Θ where (BEL $\in \Theta \in k$); we might say one's Θ is "pragmatically closed" in order for one to be able to function effectively in the world. Example: when one adopts the set $\Theta_1 = \{\text{red, white, yellow}\}$ as the frame for the question "What color rose is Bill wearing today?" one formalizes the variable x with those possible values. The frame $\Theta_2 = \{\text{white, blue}\}$ might answer the question "What color shirt is Bill wearing today?" The frame for the conjoined question "What color rose and what color shirt is Bill wearing today?" is $\Theta_1 \times \Theta_2 = \{$(red, white), (red, blue), (white, white), (white, blue), (yellow, white), (yellow, blue)$\}$ (Liu et al., 1991). Frame of discernment narrows down a potentially unwieldy set of beliefs into something more pragmatically serviceable[14].

To continue with our earlier example, let's say that at k_2 one has acquired $\text{BEL}_{\text{new}\oplus}$ ("The last time I was late for work, my boss was understanding"). Because it is $\text{BEL}^{\oplus}$, it is inconsistent with $\{\text{BEL}_{1\ominus}, \text{BEL}_{2\ominus}, \text{BEL}_{3\ominus}\}$. The objective of cognitive restructuring or belief consolidation following exposure/response prevention is for k_1 to be *in*consistent with k_2. It follows that BEL_{old} should be $\text{BEL}^{\ominus}$ and BEL_{new} should be $\text{BEL}^{\oplus}$, otherwise, there would not be any therapeutic change. Cognitive restructuring is teleological in that it is undertaken with a specific objective in mind, which is belief change and resulting behavior modification. For these reasons, we hypothesize that REV is the paradigm case of successful cognitive restructuring (see **Figure 4**).

CONTRACTION (CON)

CON is like deleting an old belief without adding any new ones. CON (expressed as $k_1 \div \text{BEL}x$) is when one rejects a BEL_{old} or deletes it from her knowledge base. $k_2 = (k_1 - \text{BEL}_{\text{old}})$: $k_2$supersedes k_1; $k_2 \subseteq (k_1 \mid k_2 \nrightarrow \text{BEL}_{\text{old}})$; but from which no ($\text{BEL}x \in k_1$) has been unnecessarily deleted. Because a BEL has been deleted from one's k_1 belief set, CON is a process of

[13]There are several other possible operations one can perform using REV: "partial meet revision" and "transitively relational partial meet revision." We do not cover these, here. Logical closure may be unrealistic in a real-world environment, because one might not recognize derived beliefs, even if they are specified. One draws on numerous other beliefs, facts assumptions and knowledge about the world in order to function effectively within it. It is unlikely one ever is in command of all possibly relevant evidence pertaining to a belief or beliefs. It most likely would be impossible to specify fully all of the beliefs comprising one's knowledge base, a project that in effect would require axiomatizing all human knowledge (Dreyfus, 1992; Shanahan, 2009).

[14]A related concept is partition dependence, which is the psychological pattern of how one divides up a set of possible outcomes into particular events. Doing so influences the perceived likelihood those events will occur. Combining events into a common partition lowers their perceived probability. Conversely, unpacking events into separate partitions increases their perceived probability (Sonnemann et al., 2013). For example, apocryphally, Eskimos have numerous words for "snow," because that phenomenon allegedly is far more prevalent where they live than elsewhere (Martin, 1986). They need a vocabulary with greater subtlety and nuance to describe its various aspects. This in turn increases the probability an event will be interpreted as snow-like, because a set of phenomena (e.g. cold wet stuff falling from the sky) with its associated beliefs (e.g. if you stay out in it too long, you will freeze) has been parsed out into separate partitions. Rabinovich et al. (2014, p. 1) recently characterized this as "chunking"–a dynamical strategy agents use to "perform information processing of long sequences by dividing them in shorter information items" thereby making "more efficient use of short-term memory by breaking up long strings of information."

$\{k_1\}$ BEL_{old}	Δ_{bel}	$\{k_2\}$ BEL_{new}
$\text{BEL}_{1}{}^{\ominus}$ = "My boss is going to get angry."		$\text{BEL}_{1}{}^{\ominus}$ = deleted.
$\text{BEL}_{2}{}^{\ominus}$ = "My colleagues will disrespect me."		$\text{BEL}_{2}{}^{\ominus}$ = deleted.
$\text{BEL}_{3}{}^{\ominus}$ = "My opinion doesn't count."		$\text{BEL}_{3}{}^{\ominus}$ = deleted.
REV	$\text{BEL}_{4}{}^{\oplus}$ = "The last time I was late for work, my boss was understanding."	$\text{BEL}_{4}{}^{\oplus}$ = "The last time I was late for work, my boss was understanding."

FIGURE 4 | REV.

"epistemic entrenchment." In rejecting BEL_{old}, one also may have to disavow other BELx that imply or are implied by it. Which beliefs should be deleted? From the standpoint of CBT:

1. One should start with those beliefs that violate the requirements of conformity, conditioning and coherence (§1.6.1, §1.6.2, §1.6.3, above). Because of coherence, BEL$x \notin k_n$ trivially is non-entrenched and tautologies are fully entrenched.
2. Next, since an AGM agent strives for minimal change and maximum information value, she should relinquish those beliefs with the least-explanatory power and supporting evidence, because they are less entrenched. The more entrenched beliefs dominate ("$\leq$") the lesser entrenched beliefs when $\{(BEL_1 \rightarrow BEL_2) \rightarrow (BEL_1 \leq BEL_2)\}$ so that k_2 comprises the "inclusion maximal" set $(BEL_1, BEL_2, \ldots BEL_n) \mid (k_1 \nrightarrow BEL_{old})$ and there is minimal information loss. AGM refers to the beliefs that stay as "remainders." The remainders comprising k_2 are the maximally-large set of BEL following deletion of BEL_{old} that do not imply any BEL_{old}, or their derivatives, remaining in k_1.
3. The exact mix of BEL$x^{\oplus}$ and BEL$x^{\ominus}$ selected by CON is determined by the preference function γ (§1.8.1, above), which specifies the minimum set of (BEL$x \in k_1$) that ought to be retained in k_2. γ should select $k(x)$ in order of plausibility; $(k_2 \gamma k_1)$ represents k_2 as more likely than k_1, given BEL_{new}. In other words, γ should select those BELx most likely to result in a more functional (less dysfunctional) k_2. It follows that the most preferred candidates γ should select to delete from k_1 (after steps 3.3.1 and 3.3.2) are $BEL^{\ominus}$, such as automatic negative thoughts and their corollary intermediate beliefs and core beliefs, in order to maximize CON's effectiveness. The remainders then will be $BEL^{\oplus}$.
4. If γ selects a maximally-consistent set of k_1 that $\nrightarrow BEL_{old})$ to become k_2, then CON is a "partial meet contraction." If k_2 ends up being populated with only one BELx (unlikely), then CON is a "maxichoice contraction." If CON selects all of the BEL comprising k_1 (thus k_2 ends up being populated with all of the them), then CON is a "full meet contraction[15]."

   We hypothesize that CON is the most problematic maneuver for an AGM agent, because its contribution to cognitive restructuring depends on whether it operates on a $BEL^{\oplus}$ or a $BEL^{\ominus}$. If the BEL that are being deleted are $BEL^{\ominus}$, then the remainders will be $BEL^{\oplus}$. This corresponds with the intuitive requirement that successful cognitive restructuring should eliminate dysfunctional $BEL^{\ominus}$, while leaving $BEL^{\oplus}$ alone. On the other hand, it also illustrates a way in which cognitive restructuring might backfire, for example, if one is so committed to a $BEL^{\ominus}$ that a $BEL^{\oplus}$ is deleted as a consequence. If the belief that is being deleted is a $BEL^{\oplus}$, then the remainders all may end up being $BEL^{\ominus}$, because they are well-entrenched. An example might be recovery following extinction using a classical conditioning model, which occurs when $k_1 \subseteq \{(k_1 \div BEL_{new}) + BEL_{old}\}$. This means that if k_1 was EXP by BEL_{old}, but one somehow readopted or reincorporated BEL_{old} into her k_1 belief set, then the effect of cognitive restructuring would be reversed. Or, the BEL set $\in k_2$ could be an ambiguous mixture of both $BEL^{\ominus}$ and $BEL^{\oplus}$, in which case cognitive restructuring would only be partially successful. Building on our previous examples, **Figure 5** illustrates an instance of successful belief revision using CON.

[15]There are several other possible operations one can perform using CON, including "transitively relational partial meet contraction." We do not cover these, here.

INTEGRATING AGM INTO A THEORY OF NON-LINEAR DYNAMICAL BELIEF REVISION

We conceptualize belief revision using AGM as an emergent property of a complex, self-organizing system involving huge numbers of neurons broadly distributed throughout different brain regions, including the prefrontal cortex (PFC), Broca's area and Wernicke's area (Cogan et al., 2014). There now has been considerable research imaging regions of the brain activated by BEL(x), starting approximately with Greene et al. (2001), continuing through Harris et al. (2008) and d'Acremont et al. (2013). Other studies examine brain regions activated by semantic processing–the words in which beliefs are expressed. Huth et al. (2012) used WordNet (§1.8.2, above) to identify 1705 object and action categories from several hours of nature movies. When they projected them to research participants undergoing fMRI, they were able to map semantic selectivity into smooth gradients covering much of the cortex. Crangle et al. (2013) presented their research participants with 48 spoken-word and visual depictions of sentences about the geography of Europe, half of which were true and half of which were false. They used WordNet and LSA (§1.8.2, above) to extract and classify their propositional content–the x in BEL(x). The resulting semantic processing was associated with characteristic features of EEG recordings. Costanzo et al. (2013) presented research participants undergoing fMRI with 140 line drawings or pictures of objects (visual stimuli) together with corresponding nouns spoken aloud (auditory stimuli). They found that both converged and were processed in the same regions of the brain during superordinate semantic categorization.

Semantic memory long has been recognized as a fundamental component of human cognition (McRae and Jones, 2013). It is "general knowledge about the world, including concepts, facts and beliefs" and is acquired through experience, thereby "grounding knowledge in distributed representations across brain regions that are involved in perceiving or acting" (Yee et al., 2014, p. 353). Semantic network structure plays a key role in the formulation of ideas and the ways in which they are combined and conceptually associated (Goñi et al., 2011; Marupaka et al., 2012). It accommodates both abstract concepts and concrete ones, the former associated with the medial PFC and the superior temporal sulcus, the latter associated with the bilateral intraparietal sulcus (Wilson-Mendenhall et al., 2013). It represents cognitive information either as specific autobiographical episodes or more general semantic knowledge, each with different subjective experiences (Heisz et al., 2014). Rabinovich et al. (2012b, p. 81) characterize it as a "space of interconnected information items," where "each item [is a separate] dynamical element" and "the dynamics of thinking (or consciousness) is a flow in a semantic space."

$\{k_1\}$ BEL_{old}	Δ_{bel}	$\{k_2\}$ BEL_{new}
$BEL_1^{\ominus}$ = "My boss is going to get angry."		$BEL_1^{\ominus}$ = deleted.
$BEL_2^{\ominus}$ = "My colleagues will disrespect me."		$BEL_2^{\ominus}$ = deleted.
$BEL_3^{\ominus}$ = "My opinion doesn't count."		$BEL_3^{\ominus}$ = deleted.
$BEL_4^{\oplus}$ = "The last time I was late for work, my boss was understanding."		$BEL_4^{\oplus}$ = "The last time I was late for work, my boss was understanding."
CON	$BEL_1^{\ominus}$ = "My boss is going to get angry."	
CON	$BEL_2^{\ominus}$ = "My colleagues will disrespect me."	
CON	$BEL_3^{\ominus}$ = "My opinion doesn't count."	

FIGURE 5 | CON.

This body of work supports a conclusion that {BEL | EVID} is not a specific topological location or ontogenetic landscape within the brain. Rather, it is a type of neural activity or pattern of activation that occurs within a comprehensive neural system. When one believes something, one enters into a series of hybrid doxastic/semantic states, which can be functionally represented as a non-linear, dynamical process–a belief revision network occurring in a global workspace–such as that depicted at **Figure 6** (while **Figure 6** depicts a two-dimensional surface, it should be understood as a multi-dimensional space; **Figure 7** depicts an alternative perspective).

It also requires a reconceptualization of the relationship between beliefs and semantics. Unlike an fMRI or EEG recording depicting brain activity, a belief set cannot be described as a geometrical object or in statistical terms. Rather, it is an encoded set of semantic propositions, embodying emergent semantic properties in its very organization (Juarrero, 1999). A belief set creates an internal symbolic mental representation based on one's assessment of its conditions of satisfaction (§1.5, above); one can imagine the conditions of satisfaction being enacted or realized[16]. It interacts with other brain regions responsible for perception, cognition, emotion, language and behavior. They are embedded within a manifold or phase plane together with physiological assets such as blood flow and oxygen. The phase plane is in a constant state of flux, flexibly changing in response to environmental constraints and internal demands (Kelso, 1999). Belief revision is a dynamic pattern of activity occurring within the phase plane.

Some beliefs initially are stored in long-term memory. These most likely are enduring, persistent beliefs about self, others, world and future; background or network beliefs of the sort described at §1.9, above; and core beliefs of the sort described at §2.3.3, above. They are recalled into short-term memory in response to decision points, environmental affordances and outcomes, and other multiple attractors. The network's attractors constitute a "self-organized space with emergent properties that can only be characterized as semantic" because they "embody [word] meaning[s] or sense[s] in the organization

[16]Mental images are controversial (for a summary of recent work, see Doumas and Hummel, 2012; Markman, 2012; Reisberg, 2014; and Shea, 2013). We are not committed to a theory that one creates actual, static mental representations in the brain. They are not pictures, rather, "depictive representations interpreted by cognitive processes at play in other systems" (Borst, 2014, p. 84). They have "several levels of complexity, from sparse, atomic concepts to complex, knowledge intensive ones" (Rips et al., 2012, p. 177). An agent's behavior must be flexible in order to respond to her circumstances, and mental representations play an important role in enabling her to do so (Egan, 2012, p. 250). Perception, for example, may be more of a process whereby a perceiver skillfully interacts with her environment. The real world presents way too much information for the perceiver to encapsulate it in an isomorphic mental image. Rather it is like a gigantic external memory, supplying a series of cues, which the perceiver can access as necessary (Noë, 2004). We do not, of course, contend that one literally perceives the words comprising the semantic formulation of one's belief set (in a manner similar to the way the Arnold Schwarzenegger character in the movie Terminator III movie was able to scroll through different belief-action options before selecting a particular alternative).

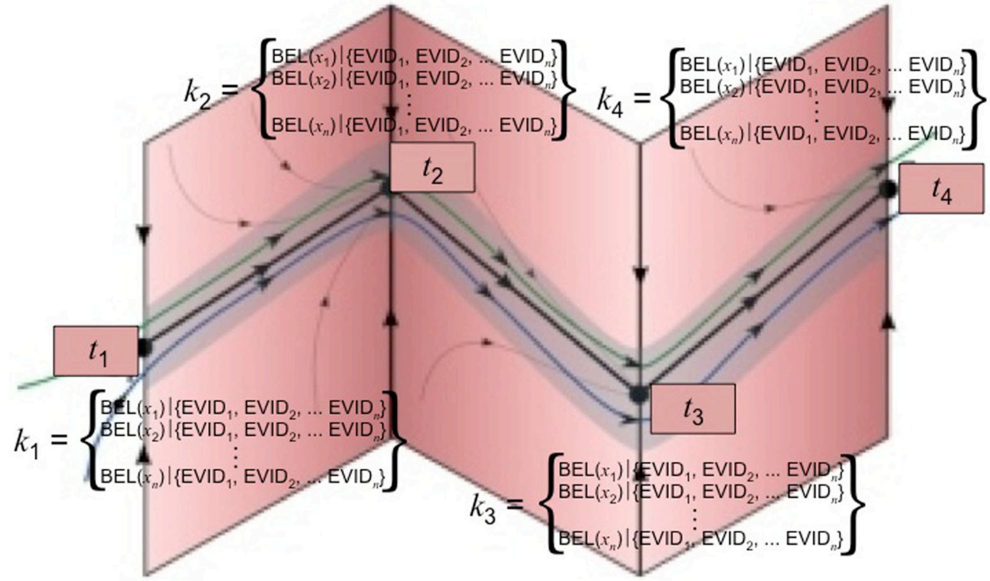


FIGURE 6 | Hypothesized pathways for belief revision–conceptualization 1. Adapted from (Rabinovich et al., 2010a). Used with permission.

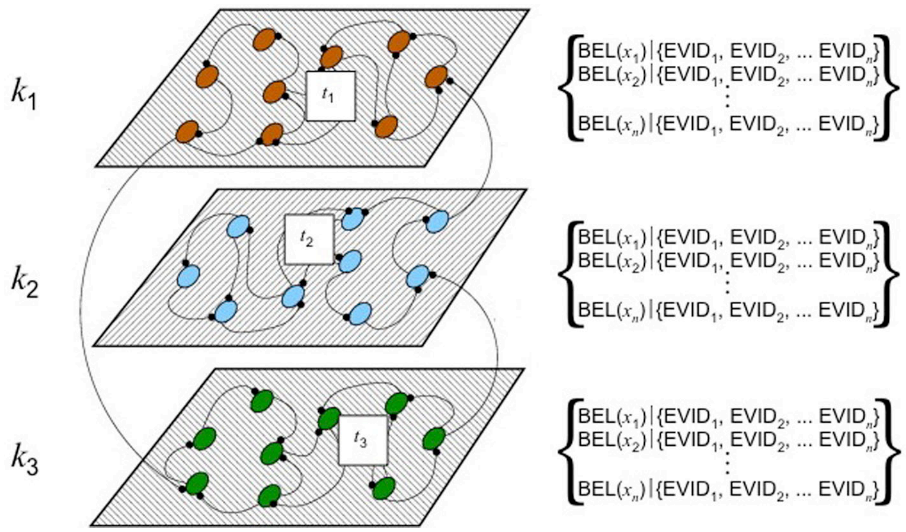


FIGURE 7 | Hypothesized pathways for belief revision–conceptualization 2. Adapted from (Rabinovich et al., 2010b). Used with permission.

of the relationships that constitute the higher-dimensional space" (Juarrero, 1999, p. 167). Initially, the phase plane represents all possible states of the belief-generating and belief-revision systems. It has a large number of degrees of freedom. It is unstable in that small changes to initial conditions–both perceived and imaginal–have the potential to become radically amplified, resulting in any number of different multi-stable belief sets. While the output belief set at k_n depends to some extent on the input belief set at k_1, k_n is asymmetrical and cannot be reliably predicted by k_1. Arguably, it exhibits chaotic dynamics because it would be difficult to specify the individual beliefs comprising the belief set as it evolves into novel and surprising states that are unexpectedly both deterministic and stochastic (non-deterministic) (Nicolis and Prigogine, 1989).

The belief revision system is transient. At t_1, all possible belief trajectories (starting with the system's initial conditions) intersect the phase plane in a structure similar to a Poincaré surface. As it evolves forward in time, it is bombarded with evidence–information derived from its interactions with the environment and subsequent interpretations. It becomes destabilized and undergoes non-equilibrium, dissipative phase transition. Individual beliefs transverse each attractor's basin of attraction and converge into specific belief sets, which consolidate at saddle equilibrium points $\{t_1, t_2 \ldots t_n\}$. They can be conceptualized

as a form of Mandelbrot fractal. Broader attractor basins capture or entrain a wider range of beliefs, depending on their strength. Because of the system's chaotic dynamics and each point's turbulent behavior, they resemble strange attractors. Convergence results in heteroclinic binding (Rabinovich et al., 2010b) of different evidentiary data to individual beliefs, which recruit resources and attempt to gain priority using the preference function γ as described at §1.8.1, above. The system bifurcates as new beliefs are formulated based on {BEL | EVID} (§1.1, above), revised conditions of satisfaction (§1.5, above), new evidence/information received as a result of interactions with the environment (§2.2, above), and associated evaluative processes.

Belief revision occurs as belief sets sequentially progress or are deflected from one metastable state to another, forming a heteroclinic channel. The separatices are ridges defining its boundaries. They constrain the flow of resources available to each belief set by modifying the phase plane or the possible trajectories of movements within it. As one belief set begins to dominate, it acquires and sustain coherence, crowding out the semantic space potentially accessible to other beliefs. At some point it reaches critical mass and overcomes an inertial threshold, compelling its migration from t_1 to t_n. During this process, the k_1 belief set competes with the k_2 belief set (then k_2 with k_3, etc.) to alter its composition using CON, EXP, or REV, either in response to cognitive restructuring or exposure/response prevention with associated environmental feedback, followed by belief revision.

Since the individual beliefs comprising each belief set displace each other (using CON, EXP, or REV), this is a zero-sum, inhibitory process. The sequence of equilibrium points in the heteroclinic channel form a heteroclinic belief revision network. This process typically remains non-conscious until at t_n, when elements of the belief set acquire salience or otherwise are extracted using typical CBT clinical techniques and protocols[17]. The combination of non-linearity and non-equilibrium, context-sensitive constraints initially permits multiple solutions, which have the potential to emerge from and be expressed within a diversified assortment of behaviors (Nicolis and Prigogine, 1989). Numerous beliefs compete in a kind of winnerless competition (Rabinovich et al., 2010a). As it stabilizes, though, the belief revision network appropriates a single behavioral output channel. The behavior semantically satisfies the intentions motivating it (the conditions of satisfaction of the associated belief sets, §1.5, above). Upon its conclusion at t_n, the reformulated beliefs comprising the k_n belief set are inserted (or reinserted) back into long-term memory. The behavioral stream transfers to an adjacent nonlinear dynamical system for action. Since emotion regulation also plays an important role in belief revision (Boden and Gross, 2013), associated emotions also are reregulated (§2.1.4, above)[18].

Cognition and behavior comprise a single autocatalytic unit and it is difficult to assess their respective influences at any t_n. Neurocognitive methods do not yet have sufficient precision to discriminate between the two (Morrison and Knowlton, 2012). There are no studies persuasively isolating the cognitive component from the behavioral one. Both require selective deployment of attentional, cognitive and affective resources. Unless belief revision was assessed immediately following cognitive intervention, before enactment of any behavior, it would not be possible to isolate the floor effect of cognitive change and control for reinforcement effects, because cognitive change already would be in the process being incrementally reinforced (for an early and unpersuasive attempt to do so based on the concept of "self-focused attention," see Wells, 2006). Any kind of change arguably results in a form of behavior. A recent study on the efficacy of mindfulness-based cognitive therapy (Kuyken et al., 2010)–seemingly, the paradigm case of a cognitive intervention–correctly noted that "these interactive mediation effects indicate that treatment changes the nature of the relationship between cognitive reactivity and outcome" (p. 1110).

What we can say is that together, they comprise a heterogeneous, self-organized, complex adaptive system (Juarrero, 1999) (in this sense, realizing Beck's concept of cognition as an information processing system, §2.3.3, above). Both are temporally and contextually embedded, exchanging information and energy with each other depending on the task at hand, the level of one's skills or expertise to accomplish it, and feedback from the environment. Structure and patterns emerge from repeated cycling involving the cooperation of many individual parts (Thelen and Smith, 2000). Although the system initially is out of equilibrium, with high entropy, it self-organizes by assuming a structure allowing it to operate more efficiently (Guastello and Liebovitch, 2009). Repeated behavioral stimulation and learning history facilitate signal transmission between neurons. Neural plasticity promotes Hebbian-type long-term potentiation, which in turn cascades into further hybrid cognitive-behavioral activation and reinforcement, strengthening attractors and facilitating the development of more predictable belief trajectories within the semantic phase plane. "Through repeated activation of a pattern the connections between units that are activated simultaneously become stronger and the whole pattern becomes an attractor." Thus, even if only partially activated, "the network can complete the pattern by a process of iterative spreading activation" so "the previously learned pattern is recovered in a number of updating cycles in which the activation level of each unit is adjusted according to the activation levels of the other units and the strength of the connections between the units" (Pecher, 2013, p. 359). As a result, conditions of satisfaction (§1.5, above) are revised, together with their corresponding internal symbolic mental representations (§1.1, above). These brain-environmental interactions comprise a negative feedback

[17] In this we are in accord with Dehaene (2014, p. 8) and Searle (1992, p. 152) to the effect that "The notion of an unconscious mental state implies accessibility to consciousness. We have no notion of the unconscious except as that which is potentially conscious." Metaphorically, beliefs are like objects within a multi-dimensional hologram; at any given time we are able to observe only a small portion of them within a potentially vast space-time continuum. Our characterization of the belief-generation and belief-modification process does not implicate any particular theory of action or agency, other than the basic principle that behavior is the action-expression of belief.

[18] Though we disagree with Boden and Gross' naive model of how this works (pp. 591-2), which appears to be the result of reading too much literature on acceptance and commitment therapy (ACT).

loop if they increase the incidence of target behavior; a positive one, if it decreases.

From a clinical standpoint, many cognitive interventions (such as mindfulness) are inherently mental and remain thoroughly solipsistic even as they reinforce and are reinforced by new behavior. Many principles of acceptance and commitment therapy (ACT) are cognitively front-loaded, for example, using metaphor as a means of identifying and developing a valued direction and defusing from one's private mental experiences (Hayes et al., 2012). Other examples are motivational interviewing for substance abuse (Miller and Rollnick, 2012); cognitive behavioral analysis system of psychotherapy (CBASP) for depression (McCullough, 2000); and cognitive processing therapy for PTSD (Resick et al., 2002). Behavioral factors, on the other hand, more clearly dominate interventions such as behavioral activation for depression; exposure/response prevention treatment for obsessive-compulsive disorder or attention deficit disorder; and prolonged exposure therapy for PTSD (Foa et al., 2007). With its dual emphases on learning (cognitive) then applying (behavioral) skills, DBT for borderline personality disorder (§2.1.4, above; Linehan, 1993) lies somewhere in the middle.

In some instances behavioral therapy is a more plausible intervention than cognitive therapy, and vice versa. Unquestionably it is possible to train up organisms with little cognitive processing capacity to demonstrate learned behavior. A 700-kg alligator, for example, has a brain that would fit comfortably inside of a teaspoon (Coulson and Herbert, 1981), yet still is capable of learning in the sense of (Squire and Kandel, 1998)[19]. In principle, it would be amenable to behavioral therapy. At some point, though, higher-order propositions must be expressed using natural language or a natural language equivalent[20]. Without it, propositions would neither be true nor false; the concept of truth builds upon veridical experience. Nor would beliefs have conditions of satisfaction (§1.5, above), nor would psychopathological beliefs have none (§1.6, above). Unlike behavior therapy, cognitive therapy depends on semantics. For this reason, as per §2.1.3, above, it is unclear whether persons with thought disorders can benefit from it (compare Grant et al., 2012 with Aggarwal and Basu, 2013; for a current overview, see Bachman and Cannon, 2012; and Jauhar et al., 2014). While of course outcomes lie on a continuum, arguably, it would be ineffective in principle for those toward the far end of the spectrum. If a person remains impervious to environmental feedback–she is unable to develop adaptive cognitions and activate belief revision–we are inclined to say that something is impeding the assimilation of new evidence, or that her information processing systems require recalibration. Functionally, she may be in a concrete operational stage, or otherwise incapable of abstract thought or metacognition. Having a theory of mind–being able to think about thoughts–may be a necessary component of psychological change (Saxe and Young, 2014). One solution from an operant conditioning perspective might be to increase positive reinforcement ($R^{\oplus}$) or to titrate down punishment using negative reinforcement ($R^{\ominus}$) in order to upregulate the desired behavior, with a view toward mobilizing additional cognitive resources.

Most likely cognition and behavior shuttle back and forth quickly depending on the client's perceptions, emotions, language capability, attentional focus, the context in which behavior occurs, the nature of the transaction the client is having with her/his environment, experience/learning history, genetics, neurochemistry, interoceptive sensitivity, memory capacity, heuristics, intuition, vulnerabilities, intentions, skills, values, and a variety of other factors. Their different trajectories oscillate (Schultz and Heimberg, 2008) in what Rabinovich et al. (2010b) would characterize as a heteroclinic channel between metastable states. Because the brain is a complex system with a variety of different inputs and outputs, neither cognition nor behavior can be controlled in isolation (Ruths and Ruths, 2014). From a clinical standpoint, target behavior should progressively and dynamically reduce. As depicted at **Figures 8**, **9**, their relationship is transactional. The exact mix of each depends not only on the type of therapy but also stages in the therapeutic process. For example, the manic phase of bipolar disorder (DSM-5 §296.xx) might be more amenable to cognitive therapy, whereas the depressive phase might be more amenable to behavioral therapy (Leahy, 2005). Daugherty et al. (2009) characterized this as a Liénard oscillator with autonomous forcing. From the standpoint

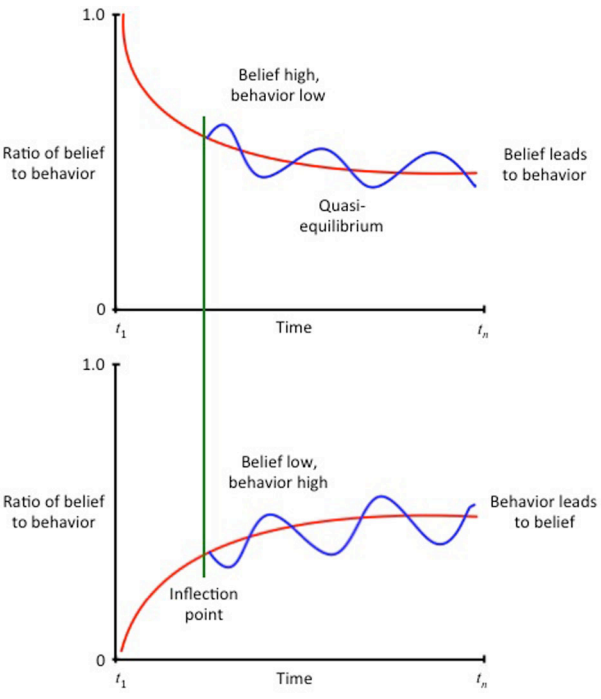


FIGURE 8 | Transactional relationships between beliefs and behavior–conceptualization 1.

[19]This is the double entendre behind the title of B.F. Skinner's famous paper "Superstition in the Pigeon" (1947). Superstition is a form of cognition, whereas pigeons only are capable of learned behavior.

[20]There is no bright-line test for this, either. The meaning of simple propositions can be enacted using language-like behavior, such as Quine's famous example of a speaker using ostension to point to a rabbit, while uttering the word "gavagi" to designate a rabbit-like stage or rabbit-like behavior (Quine, 1964).

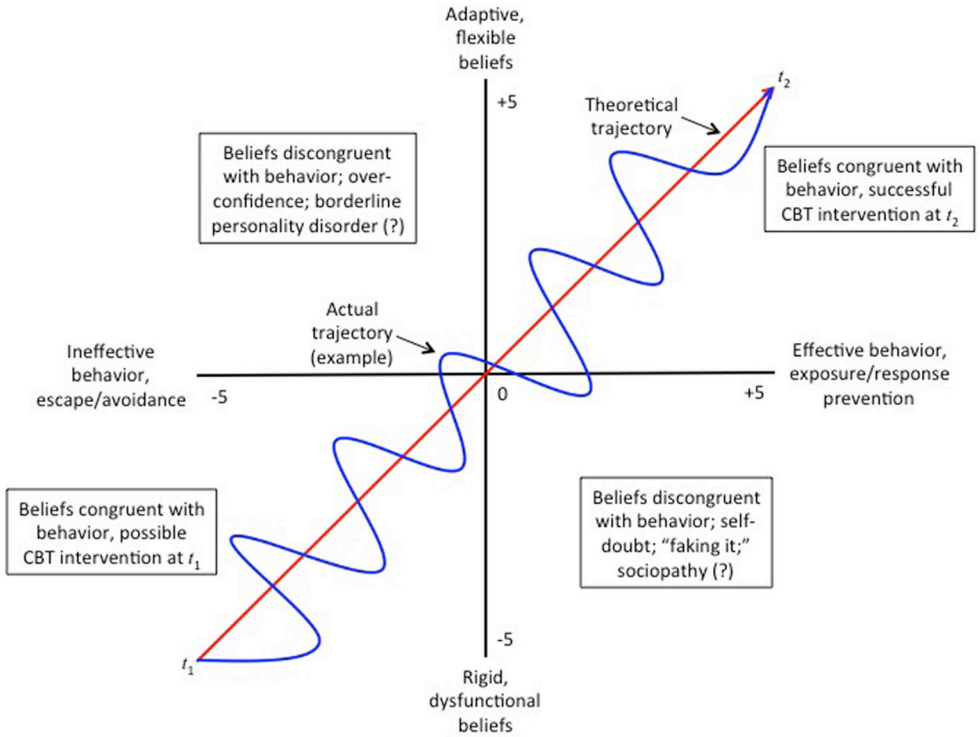


FIGURE 9 | Transactional relationships between beliefs and behavior–conceptualization 2.

of belief revision semantics, the theme of the substantive propositional content ("*x*") remains the same, even as the propositional attitude toward it changes, e.g., if the domain is "affection," then manic = "adorable" whereas depressed = "unlovable." Conceptually, behavioral reformulation and cognitive reconstruction serially propel it in a dynamic progression from t_1 through t_n as different inhibitory and stimulating paradigms take effect. At some point in this process–an extremely interesting one from the standpoint of cognitive science–their trajectories intersect and one transitions into the other. Both are active ingredients of therapeutic change.

CONCLUSION

The ultimate goal of cognitive restructuring or belief consolidation following exposure/response prevention should be thorough overhaul of a meaningful subset of one's entire belief system. Simply inducing doubt is not sufficient. An example of such a paradigm shift might be a prisoner on death row who is exonerated by new DNA evidence, resulting in radical reformation of her knowledge base, or Dostoyevsky's experience in front of a mock firing squad (Bloom, 2005). This is every bit as profound and disruptive as the transition from Ptolemaic astronomy to Copernican astronomy, or from Newtonian physics to Einstein physics, or through the so-called three waves of cognitive behavioral therapy (Hayes, 2004). Thomas Kuhn (1962/2012) labeled these "scientific revolutions"–on an individual level, they might be labeled "personal revolutions."

In addition to making a case for AGM, one of our main objectives in this review has been to illustrate a point of intersection between cognitive science and clinical psychology, two fields which long have enjoyed an uneasy *rapprochement* (Macleod, 2010). "The study of psychopathology has. . . become an important facet of the cognitive sciences, and the cognitive sciences have, in turn, exerted an important influence on many regions of psychiatry" (Cratsley and Samuels, 2013, p. 413). One of the characteristics of many cognitive science theories is that while each step of the argument makes sense, when viewed as a complete chain of inferential reasoning, the transition from premises to conclusion may be implausible, in a C.P. Snow (1959/2012)-type sense. Like a salmon swimming upstream, one ends up in a very small pond. Clinical psychology, in turn, depends operationally on protocols that first were devised over a quarter of a century ago. The prospects for *détente* are not as far-fetched as they initially might seem. For example, on April 1, 2014, the Max Planck Society announced a €5 million investment in a new center for computational psychiatry to be based in London and Berlin, with a view toward uncovering relationships between cognition and psychopathology of the sort we hypothesize (Siddique, 2014).

We submit that the best way to think of our initiative is that it is an exercise in translational research. It applies a form of nonlinear analysis to the study of complex systems in cognitive science and behavioral analysis. Even though it may not exactly mirror actual, common sense psychological activity, logical reasoning should "clarify, sharpen, systematize the purely semantic-level

characterization of the demands on any such implementation, biological or not" (Dennett, 1984/2006, p. 449); to "provide an account of our cognitive architecture–which specifies the basic operations, component parts, and organization of the mind" (Samuels, 2012). It also demonstrates how recent work in experimental cognitive science can be combined with clinical psychology to inform the process of psychological change.

REFERENCES

Abbott, A. (2013). Solving the brain. *Nature* 499, 272–274. doi: 10.1038/499272a

Afraimovich, V., Young, T., Muezzinoglu, M. K., and Rabinovich, M. I. (2011). Nonlinear dynamics of emotion-cognition interaction: when emotion does not destroy cognition? *Bull. Math. Biol.* 73, 266–284. doi: 10.1007/s11538-010-9572-x

Aggarwal, M., and Basu, D. (2013). Cognitive therapy in patients with schizophrenia. *JAMA Psychiat*. 70, 543–544. doi: 10.1001/jamapsychiatry.2013.284

Albergotti, R. (2014). Zuckerberg, musk invest in artificial intelligence company. *Wall Street J.* Available online at: http://blogs.wsj.com/digits/2014/03/21/zuckerberg-musk-invest-in-artificial-intelligence-company-vicarious/

Alchourrón, C. E., Gärdenfors, P., and Makinson, D. (1985). On the logic of theory change: partial meet contraction and revision functions. *J. Symbolic Logic* 50, 510–530. doi: 10.2307/2274239

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders, 5th Edn*. Washington, DC: Author.

Arlo-Costa, H. (2007). The logic of conditionals. *Stanford Encyclopedia Philos*. Available online at: http://plato.stanford.edu/entries/logic-conditionals/

Arntz, A. (2002). Cognitive therapy versus interoceptive exposure as treatment of panic disorder without agoraphobia. *Behav. Res. Ther.* 40, 325–341. doi: 10.1016/S0005-7967(01)00014-6

Austin, J. L. (1956/1970). A plea for excuses. *Proc. Aristotelian Soc.* 57, 1–30. Reprinted in (1970). *J. L. Austin–Philosophical Papers (175–204)*, eds J. O. Urmson and G. J. Warnock Oxford: Oxford University Press.

Austin, J. L. (1962). *How to Do Things with Words*. New York, NY: Oxford University Press.

Bachman, P., and Cannon, T. (2012). "The cognitive neuroscience of thought disorder in schizophrenia," in *The Oxford Handbook of Thinking and Reasoning*, eds K. J. Holyoak and R. G. Morrison (New York, NY: Oxford University Press), 673–700.

Bahr, S. S., Brown, G. K., and Beck, A. T. (2008). Dysfunctional beliefs and psychopathology in borderline personality disorder. *J. Pers. Disord.* 22, 165–177. doi: 10.1521/pedi.2008. 22.2.165

Beck, J. S. (2011). *Cognitive Behavior Therapy–Basics and Beyond, 2nd Edn*. New York, NY: Guilford Press.

Bernstein, A., Stickle, T. R., and Schmidt, N. B. (2013). Factor mixture model of anxiety sensitivity and anxiety psychopathology vulnerability. *J. Affect. Disord.* 149, 406–417. doi: 10.1016/j.jad.2012.11.024

Bernstein, A., Stickle, T. R., Zvolensky, M. J., Taylor, S., Abramowitz, J., and Stewart, S. (2010). Dimensional, categorical, or dimensional-categories: testing the latent structure of anxiety sensitivity among adults using factor-mixture modeling. *Behav. Ther.* 41, 515–529. doi: 10.1016/j.beth.2010.02.003

Bickle, J., Mandik, P., and Landreth, A. (2010). The philosophy of neuroscience. *Stanford Encyclopedia Philos*. Available online at: http://plato.stanford.edu/entries/neuroscience/

Bilder, R. M., Sabb, F. W., Cannon, T. D., London, E. D., Jentsch, J. D., Parker, D. S., et al. (2009a). Phenomics: the systematic study of phenotypes on a genome-wide scale. *Neuroscience* 164, 30–42. doi: 10.1016/j.neuroscience.2009.01.027

Bilder, R. M., Sabb, F. W., Parker, D. S., Kalar, D., Chu, W. W., Fox, J., et al. (2009b). Cognitive ontologies for neuropsychiatric phenomics research. *Cogn. Neuropsychiatry* 14, 419–450. doi: 10.1080/13546800902787180

Bloom, H. (2005). *Fyodor Dostoevsky*. Langhorne, PA: Chelsea House Publishers.

Boden, M. T., and Gross, J. J. (2013). "An emotion regulation perspective on belief change," in *The Oxford Handbook of Cognitive Psychology*, ed D. Reisberg (Oxford: Oxford University Press), 585–599.

Boden, M. T., John, O. P., Goldin, P. R., Werner, K., Heimberg, R. G., and Gross, J. J. (2012). The role of maladaptive beliefs in cognitive-behavioral therapy: evidence from social anxiety disorder. *Behav. Res. Ther.* 50, 287–291. doi: 10.1016/j.brat.2012.02.007

Borst, G. (2014). "Neural underpinning of object mental imagery, spatial imagery, and motor imagery," in *The Oxford Handbook of Cognitive Neuroscience*, vol. 1 eds K. N. Ochsner and S. M. Kosslyn (New York, NY: Oxford University Press), 74–87.

Braun, D. (2007). Indexicals. *Stanford Encyclopedia Philos.* Available online at: http://plato.stanford.edu/entries/indexicals/

Brisard, F. (2011). "H.P. Grice," in *Philosophical Perspectives for Pragmatics,* eds M. Sbisà, J.-O. Östman, and J. Verschueren (Philadelphia, PA: John Benjamins Publishing Co.), 104–124 doi: 10.1075/hoph.10.10bri

Bryant, R. A., Moulds, M. L., Guthrie, R. M., Dang, S. T., and Nixon, R. D. (2003). Imaginal exposure alone and imaginal exposure with cognitive restructuring in treatment of posttraumatic stress disorder. *J. Consult. Clin. Psychol.* 71, 706–712. doi: 10.1037/0022-006X.71.4.706

Buckner, R. L., Andrews-Hanna, J. R., and Schacter, D. L. (2008). The brain's default network: anatomy, function, and relevance to disease. *Ann. N.Y. Acad. Sci.* 1124, 1–38. doi: 10.1196/annals.1440.011

Butler, A. C., Chapman, J. E., Forman, E. M., and Beck, A. T. (2006). The empirical status of cognitive-behavioral therapy: a review of meta-analyses. *Clin. Psychol. Rev*. 26, 17–31. doi: 10.1016/j.cpr.2005.07.003

Bystritsky, A., Khalsa, S. S., Cameron, M. E., and Schiffman, J. (2013). Current diagnosis and treatment of anxiety disorders. *P T.* 38, 30–57.

Bystritsky, A., Nierenberg, A. A., Feusner, J. D., and Rabinovich, M. (2012). Computational non-linear dynamical psychiatry: a new methodological paradigm for diagnosis and course of illness. *J. Psychiatr. Res.* 46, 428–435. doi: 10.1016/j.psychires.2011.10.013

Carey, B. (2012). Paul Allen gives millions for brain research. *N.Y. Times*. Available online at: http://www.nytimes.com/2012/03/22/health/research/paul-allen-adding-300-million-for-brain-research.html

Carnap, R. (1947/1988). *Meaning and Necessity*. Chicago, IL: University of Chicago Press.

Carnota, R., and Rodríguez, R. (2011). "AGM Theory and artificial intelligence," in *Belief Revision Meets Philosophy of Science, Logic, Epistemology, and the Unity of Science*, eds E. Olsson and S. Enqvist (Dordrecht, NL: Springer Science+Business Media), 1–42. doi: 10.1007/978-90-481-9609-8_1

Carruthers, P. (2012). "Language in cognition," in *The Oxford Handbook of Cognitive Science*, eds E. Margolis, R. Samuels, and S.P. Stich (New York, NY: Oxford University Press), 382–401. doi: 10.1093/oxfordhb/9780195309799.003.0016

Carter, M. M., Forys, K. L., and Oswald, J. C. (2008). "The cognitive-behavioral model," in *Handbook of Clinical Psychology*–Vol. 1–*Adults*, eds A. Gross and M. Hersen (Hoboken, NJ: Wiley), 171–204.

Chomsky, N. (1955). Logical syntax and semantics: their linguistic relevance. *Language* 31, 36–45. doi: 10.2307/410891

Chomsky, N. (1959). Review of verbal behavior, by B. F. Skinner. *Language* 35, 26–58.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT University Press.

Chomsky, N. (1977). *Essays on Form and Interpretation*. New York, NY: North-Holland.

Churchland, P. M., and Churchland, P. S. (1998). *On the Contrary: Critical Essays 1987-1997*. Cambridge, MA: MIT Press.

Churchland, P. M., and Churchland, P. S. (2013). "What are beliefs?" in *The Neural Basis of Human Belief Systems*, eds F. Krueger and J. Grafman (New York, NY: Psychology Press), 1–18.

Cogan, G. B., Thesen, T., Carlson, C., Doyle, W., Devinsky, O., and Pesaran, B. (2014). Sensory-motor transformations for speech occur bilaterally. *Nature* 507, 94–98. doi: 10.1038/nature12935

Corazza, E. (2011). "Indexicals and demonstratives," in *Philosophical Perspectives for Pragmatics*, eds M. Sbisà, J.-O. Östman, and J. Verschueren (Philadelphia, PA: John Benjamins Publishing Co), 131–152. doi: 10.1075/hoph.10.12cor

Costa, H. A., and Pedersen, A. P. (2011). "Belief revision," in *The Continuum Companion to Philosophical Logic*, eds R. Pettigrew and L. Horsten (New York, NY: Continuum International Publishing Group), 450–502.

Costanzo, M. E., McArdle, J. J., Swett, B., Nechaev, V., Kemeny, S., Xu, J., et al. (2013). Spatial and temporal features of superordinate semantic processing studied with fMRI and EEG. *Front. Hum. Neurosci.* 7:293. doi: 10.3389/fnhum.2013.00293

Coulson, R. A., and Herbert, J. D. (1981). Relationship between metabolic rate and various physiological and biochemical parameters: a comparison of alligator,

man and shrew. *Comp. Biochem. Physiol.* 69A, 1–13. doi: 10.1016/0300-9629(81)90632-0

Crangle, C. E., Perreau-Guimaraes, M., and Suppes, P. (2013). Structural similarities between brain and linguistic data provide evidence of semantic relations in the brain. *PLoS ONE.* 8:e65366. doi: 10.1371/journal.pone.0065366

Craske, M. G., Liao, B., Brown, L., and Vervliet, B. (2012). Role of inhibition in exposure therapy. *J. Exp. Psychopathol.* 3, 322–345. doi: 10.5127/jep.026511

Cratsley, K., and Samuels, R. (2013). "Cognitive science and explanations of psychopathology," in *The Oxford Handbook of Philosophy and Psychiatry*, eds K. W. M. Fulford, M. Davies, R. G. T. Gipps, G. Graham, J. Z. Sadler, G. Stanghellini and T. Thornton (Oxford: Oxford University Press), 413–433. doi: 10.1093/oxfordhb/9780199579563.013.0027

Culbertson, J., and Adger, D. (2014). Language learners privilege structured meaning over surface frequency. *PNAS* 10, 2014. doi: 10.1073/pnas.1320525111

Curley, S. P. (2007). The application of Dempster-Shafer theory demonstrated with justification provided by legal evidence. *Judgm. Decis. Mak.* 2, 257–276.

d'Acremont, M., Schultz, W., and Bossaerts, P. (2013). The human brain encodes event frequencies while forming subjective beliefs. *J. Neurosci.* 33, 10887–10897. doi: 10.1523/jneurosci.5829-12.2013

Dagan, I., Dolan, B., Magnini, B., and Roth, D. (2009). Recognizing textual entailment: Rational, evaluation and approaches. *Nat. Lang. Eng.* 15, i–xvii. doi: 10.1017/S1351324909990209

Damasio, A. (1999). *The Feeling of What Happens*. New York, NY: Harcourt, Inc.

Dattilio, F. M., Edwards, D. J. A., and Fishman, D. B. (2010). Case studies within a mixed methods paradigm: toward a resolution of the alienation between researcher and practitioner in psychotherapy research. *Psychotherapy (Chic)*. 47, 427–441. doi: 10.1037/a0021181

Daugherty, D., Roque-Urrea, T., Urrea-Roque, J., Troyer, J., Wirkus, S., and Porter, M. A. (2009). Mathematical models of bipolar disorder. *Commun. Nonlinear Sci.* 14, 2897–2908. doi: 10.1016/j.cnsns.2008.10.027

Davidson, D. (1970). "Mental events," in *Experience and Theory,* eds L. Foster, and J. W. Swanson (Amherst, MA: University of Massachusetts Press), 79–101. Reprinted in (2001). *Essays on Actions and Events, 2nd Edn*. (Oxford: Oxford University Press), 207–244. Reprinted in N. J. Block (ed.). (1980). *Readings in Philosophy of Psychology,* Vol. 1, (Cambridge, MA: Harvard University Press), 107–119. Reprinted in (1991). *The Nature of Mind,* ed D. M. Rosenthal (Oxford: Oxford University Press), 247–256. Reprinted in (1992). *The Philosophy of Mind: Classical Problems/Contemporary Issues,* eds B. Beakley and P. Ludlow (Cambridge, MA: MIT Press), 137–150. Reprinted in (1999). *Mind and Cognition: An Anthology, 2nd Edn,* ed W.G. Lycan (Malden, MA: Blackwell), 35–46. Reprinted in (2002). *Philosophy of Mind–Classical and Contemporary Readings,* eds D. J. Chalmers (New York, NY: Oxford University Press), 116–125. doi: 10.1093/0199246270.001.0001

Davidson, D. (1974). "Psychology as philosophy," in *Philosophy of Psychology,* ed S. Brown (London: Macmillan Press), 41–52. Reprinted in (2001). *Essays on Actions and Events, 2nd Edn,* (Oxford: Oxford University Press), 229–245. Reprinted in (1976). *The Philosophy of Mind,* ed J. Glover (Oxford: Oxford University Press), 101–110. Reprinted in (2006). *Philosophy of Psychology,* ed J. L. Bermúdez (New York, NY: Routledge), 22–30. doi: 10.1093/0199246270.001.0001

Davidson, D. (1975). "Thought and talk," in *Inquires into Truth and Interpretation*, eds D. Davidson (Oxford: Oxford University Press), 155–170.

Davidson, D. (1994/2005). "The social aspect of language," in *Truth, Language and History*, ed M. Cavell (Oxford: Oxford University Press), 109–126.

Davies, M., and Coltheart, M. (2000). "Belief revision: biases and deficits," in *Pathologies of Belief*, eds M. Coltheart and M. Davies (Malden, MA: Blackwell Publishers, Inc), 22–27.

Davies, M., and Egan, A. (2013). "Delusion: cognitive approaches–Bayesian inference and compartmentalization," in *The Oxford Handbook of Philosophy and Psychiatry*, eds K. W. M. Fulford, M. Davies, R. G. T. Gipps, G. Graham, J. Z. Sadler, G. Stanghellini, and T. Thornton (Oxford: Oxford University Press), 688–727. doi: 10.1093/oxfordhb/9780199579563.013.0042

Dehaene, S. (2014). *Consciousness and the Brain–Deciphering How the Brain Codes Our Thoughts*. New York, NY: Viking.

Dennett, D. C. (1984/2006). "Cognitive wheels: the frame problem of AI," in. *Minds, Machines and Evolution: Philosophical Studies*, ed C. Hookway (Cambridge: Cambridge University Press), 128–151. Reprinted (1987). *The Robot's Dilemma: The Frame Problem in Artificial Intelligence*, ed Z. W. Pylyshyn (New York, NY: Ablex Publishing), 41–64). Reprinted in (2006). *Philosophy of Psychology: Contemporary Readings*, ed J. L. Bermúdez (New York, NY: Routledge), 433–454. Reprinted in (1998). *Brainchildren: Essays on Designing Minds*, D. C. Dennett, (Cambridge, MA: MIT University Press), 181–206.

Dennett, D. C. (1992). *Consciousness Explained*. New York, NY: Back Bay Books.

Dietrich, F., and List, C. (2013). Reasons for (prior) belief in Bayesian epistemology. *Synthese* 190, 787–808. doi: 10.1007/s11229-012-0185-9

Domschke, K., Stevens, S., Pfleiderer, B., and Gerlach, A. L. (2010). Interoceptive sensitivity in anxiety and anxiety disorders: an overview and integration of neurobiological findings. *Clin. Psychol. Rev.* 30, 1–11. doi: 10.1016/j.cpr.2009.08.008

Doumas, L. A. A., and Hummel, J. E. (2012). "Computational models of higher cognition," in *The Oxford Handbook of Thinking and Reasoning*, eds K. J. Holyoak and R. G. Morrison (New York, NY: Oxford University Press), 52–66.

Dreyfus, H. L. (1992). *What Computers Still can't Do.* Cambridge, MA: MIT Press.

Egan, F. (2012). "Representationalism," in *The Oxford Handbook of Cognitive Science*, eds E. Margolis, R. Samuels, and S. P. Stich (New York, NY: Oxford University Press), 250–272.

Ellis, A. (1994). "The essence of rational emotive behavior therapy: a comprehensive approach to treatment," in *Personality Theories: Critical Perspectives*. Available online at: http://www.sagepub.com/personalitytheoriesstudy/01/resources2.htm

Emmelkamp, P. M. G., Ehring, T., and Powers, M. B. (2010). "Philosophy, psychology, causes, and treatments of mental disorders," in *Cognitive and Behavioral Theories in Clinical Practice*, eds N. Kazantzis, M. A. Reinecke, and A. Freeman (New York, NY: Guilford Press), 1–27.

Feldman, R., and Conee, E. (1985). Evidentialism. *Philos. Stud.* 48, 15–34. doi: 10.1007/BF00372404

Fermé, E., and Hansson, S. O. (2011). AGM 25 years: twenty-five years of research in belief change. *J. Philos. Logic* 40, 295–331. doi: 10.1007/s10992-011-9171-9

Feusner, J. D., Townsend, J., Bystritsky, A., and Bookheimer, S. (2007). Visual information processing of faces in body dysmorphic disorder. *Arch. Gen. Psychiat.* 64, 1417–1425. doi: 10.1001/archpsyc.64.12.1417

Foa, E. B., Hembree, E. A., and Rothbaum, B. O. (2007). *Prolonged Exposure Therapy for PTSD–Emotional Processing of Traumatic Experiences*. New York, NY: Oxford University Press.

Fodor, J. (1975). *The Language of Thought*. Cambridge, MA: Harvard University Press.

Fonagy, P., Steele, M., Steele, H., Moran, G. S., and Higgitt, A. C. (1991). The capacity for understanding mental states: the reflective self in parent and child and its significance for security of attachment. *Infant. Ment. Health J.* 12, 201–218. doi: 10.1002/1097-0355(199123)12:3<201::AID-IMHJ2280120307>3.0.CO;2-7

Frazzetto, G. (2013). *Joy, Guilt, Anger, Love: What Neuroscience Can–and Can't–Tell Us About How We Feel*. New York, NY: Penguin Books.

Friedman, S. E., and Forbus, K. D. (2011). Repairing incorrect knowledge with model formulation and metareasoning. *IJCAI/AAII* 2011, 887–893. doi: 10.5591/978-1-57735-516-8/IJCAI11-154

Gabbay, D. M., Rodrigues, O. T., and Russo, A. (2010). "Introducing revision theory," in *Revision, Acceptability and Context – Theoretical and Algorithmic Aspects*, eds D. M. Gabbay, O. T. Rodrigues, and A. Russo (New York, NY: Springer Science+Business Media), 13–54. doi: 10.1007/978-3-642-14159-1_2

Galavotti, M. C. (2011). "The modern epistemic interpretation of probability: logicism and subjectivism," in *Handbook of the History of Logic–Vol. 10–Inductive Logic*, eds D. M. Gabbay, S. Hartmann, and J. Woods (Amsterdam, NL: Elsevier), 153–203.

Gärdenfors, P. (2011). Notes on the history of ideas behind AGM. *J. Philos. Logic.* 40, 115–120. doi: 10.1007/s10992-011-9174-6

Garland, E. L., Fredrickson, B., Kring, A. M., Johnson, D. P., Meyers, P. S., and Penn, D. L. (2010). Upward spirals of positive emotions counter downward spirals of negativity: insights from the broaden-and-build theory and affective neuroscience on the treatment of emotion dysfunctions and deficits in psychopathology. *Clin. Psychol. Rev*. 30, 849–864. doi: 10.1016/j.cpr.2010.03.002

Gipps, R. T. (2013). "Cognitive behavior therapy: a philosophical appraisal," in *The Oxford Handbook of Philosophy and Psychiatry*, eds K. W. M. Fulford, M. Davies, R. G. T. Gipps, G. Graham, J. Z. Sadler, G. Stanghellini, and T. Thornton (Oxford: Oxford University Press), 1245–1263. doi: 10.1093/oxfordhb/9780199579563.013.0072

Gleitman, L., and Papafragou, A. (2012). "New perspectives on language and thought," in *The Oxford Handbook of Thinking and Reasoning*, eds K. J. Holyoak and R. G. Morrison (New York, NY: Oxford University Press), 543–567.

Gleitman, L., and Papafragou, A. (2013). "Relations between language and thought," in *The Oxford Handbook of Cognitive Psychology*, ed D. Reisberg (Oxford: Oxford University Press), 504–523.

Glymour, C. (1975). Relevant evidence. *J. Philos.* 72, 403–426. doi: 10.2307/2025011

Goel, V. (2013). A new tool aims to help Facebook users dig deep. *N.Y. Times*. Available online at: http://www.nytimes.com/2013/07/08/technology/a-new-tool-aims-to-help-facebook-users-dig-deep.html

Goñi, J., Arrondo, G., Sepulcre, J., Martincorena, I., Mendizábal, N. V. D., Corominas-Murtra, B., et al. (2011). The semantic organization of the animal category: evidence from semantic verbal fluency and network theory. *Cogn. Process.* 12, 183–196. doi: 10.1007/s10339-010-0372-x

Grant, P. M., Huh, G. A., Perivoliotis, D., Stolar, N. M., and Beck, A. T. (2012). Randomized trial to evaluate the efficacy of cognitive therapy for low-functioning patients with schizophrenia. *Arch. Gen. Psychiatry* 69, 121–127. doi: 10.1001/archgenpsychiatry.2011.129

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., and Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science* 293, 2105–2108. doi: 10.1126/science.1062872

Grice, H. P. (1975). "Logic and conversation," in *Syntax and Semantics,* Vol. 3, *Speech Acts*, eds P. Cole and J. L. Morgan (New York, NY: Academic Press), 41–58.

Guastello, S. J., and Liebovitch, L. S. (2009). "Introduction to nonlinear dynamics and complexity," in *Chaos and Complexity in Psychology–The Theory of Nonlinear Dynamical Systems*, eds S. J. Guastello, M. Koopmans, and D. Pincus (Cambridge: Cambridge University Press), 1–40.

Hájek, A. (2011). Interpretations of probability. *Stanford Encyclopedia Philos*. Available online at: http://plato.stanford.edu/entries/probability-interpret/

Hanks, W. F. (2011). "Deixis and indexicality," in *Foundations of Pragmatics*, eds W. Bublitz and N. R. Norrick (Berlin, DE: De Gruyter Mouton), 315–346. doi: 10.1515/9783110214260.315

Harris, S., Sheth, S. A., and Cohen, M. S. (2008). Functional neuroimaging of belief, disbelief, and uncertainty. *Ann. Neurol.* 63, 141–147. doi: 10.1002/ana.21301

Hartley, C. A., and Phelps, E. A. (2012). Anxiety and decision-making. *Biol. Psychiatry* 72, 113 –118. doi: 10.1016/j.biopsych.2011.12.027

Hartmann, S., and Sprenger, I. (2010). "Bayesian epistemology," in *Routledge Companion to Epistemology*, eds S. Bernecker and D. Pritchard (London, UK: Routledge), 609–620.

Hauner, K. K., Mineka, S., Voss, J. L., and Paller, K. A. (2012). Exposure therapy triggers lasting reorganization of neural fear processing. *PNAS* 10, 2014. doi: 10.1073/pnas.1205242109

Hauser, M. D., Chomsky, N., and Fitch, T. (2002). The faculty of language: what is it, who has it, and how did it evolve? *Science* 298, 1569–1579. doi: 10.1126/science.298.5598.1569

Hayes, S. C. (2004). Acceptance and commitment therapy, relational frame theory, and the third wave of behavioral and cognitive therapies. *Behav. Ther.* 35, 639–665. doi: 10.1016/S0005-7894(04)80013-3

Hayes, S. C., Strosahl, K. D., and Wilson, K. G. (2012). *Acceptance and Commitment Therapy, 2nd Edn*. New York, NY: Guilford Press.

Heisz, J. J., Vakorin, V., Ross, B., Levine, B., and McIntosh, A. R. (2014). A trade-off between local and distributed information processing associated with remote episodic versus semantic memory. *J. Cogn. Neurosci.* 26, 41–53. doi: 10.1162/jocn_a_00466

Heit, E., and Rotello, C. M. (2010). Relations between inductive reasoning and deductive reasoning. *J. Exp. Psychol. Learn. Mem. Cogn.* 36, 805–812. doi: 10.1037/a0018784

Herry, C., Ferraguti, F., Singewald, N., Letzkus, J. J., Ehrlich, I., and Lüthi, A. (2010). Neuronal circuits of fear extinction. *Eur. J. Neurosci.* 31, 599–612. doi: 10.1111/j.1460-9568.2010.07101.x

Hofer, M. (2010). "New concepts in the evolution and development of anxiety," in *Anxiety Disorders–Theory, Research, and Clinical Perspectives*, eds H. B. Simpson, Y. Neria, R. Lewis-Fernández, and F. Schneier (Cambridge: Cambridge University Press), 59–68. doi: 10.1017/CBO9780511777578.008

Hope, D. A., Heimberg, R. G., and Turk, C. L. (2010). *Managing Social Anxiety–A Cognitive-Behavioral Therapy Approach, 2nd Edn*. Oxford: Oxford University Press.

Howson, C. (2009). "Epistemic probability and coherent degrees of belief," in *Degrees of Belief*, eds F. Huber and C. Schmidt-Petri (New York, NY: Springer Science+Business Media B.V.), 97–119. doi: 10.1007/978-1-4020-9198-8_5

Huber, F. (2009). "Belief and degrees of belief," in *Degrees of Belief*, eds F. Huber and C. Schmidt-Petri (New York, NY: Springer Science+Business Media), 1–33. doi: 10.1007/978-1-4020-9198-8_1

Huntsinger, J. R., and Schnall, S. (2013). "Emotion-cognition interactions," in *The Oxford Handbook of Cognitive Psychology*, ed D. Reisberg (Oxford: Oxford University Press), 571–584.

Huppert, J. D. (2009). The building blocks of treatment in cognitive-behavioral therapy. *Isr. J. Psychiat. Relat. Sci.* 46, 245–250.

Huth, A. G., Nishimoto, S., Vu, A. T., and Gallant, J. L. (2012). A continuous semantic space describes the representation of thousands of object and action categories across the human brain. *Neuron* 76, 1210–1224. doi: 10.1016/j.neuron.2012.10.014

Jauhar, S., McKenna, P. J., Radua, J., Fung, E., Salvador, R., and Laws, K. R. (2014). Cognitive-behavioural therapy for the symptoms of schizophrenia: systematic review and meta-analysis with examination of potential bias. *Br. J. Psychiatry*. 204, 20–29. doi: 10.1192/bjp.bp.112.116285

Johnson-Laird, P. N. (2010). Mental models and human reasoning. *Proc. Natl. Acad. Sci. U.S.A.* 107, 18243–18250. doi: 10.1073/pnas.1012933107

Johnson-Laird, P. N. (2013). Mental models and cognitive change. *J. Cog. Psychol.* 25, 131–138. doi: 10.1080/20445911.2012.759935

Joyce, J. M. (2003). Bayes theorem. *Stanford Encyclopedia Philos*. Available online at: http://plato.stanford.edu/entries/bayes-theorem/

Joyce, J. M. (2009). "Accuracy and coherence: prospects for an alethic epistemology of partial belief," in *Degrees of Belief*, eds F. Huber and C. Schmidt-Petri (New York, NY: Springer Science+Business Media), 263–297. doi: 10.1007/978-1-4020-9198-8_11

Joyce, J. M. (2011). "The development of subjective Bayesianism," in *Handbook of the History of Logic*–Vol. 10–*Inductive Logic*, eds D. M. Gabbay, S. Hartmann, and J. Woods (Amsterdam, NL: Elsevier),. 415–745.

Juarrero, A. (1999). *Dynamics in Action–Intentional Behavior as a Complex System*. Cambridge, MA: MIT Press.

Kahneman, D. (2003). A perspective on judgment and choice–mapping bounded rationality. *Am. Psychol.* 58, 697–720. doi: 10.1037/0003-066X.58.9.697?

Kahneman, D., Slovic, P., and Tversky, A. (eds.). (1982). *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press.

Kahneman, D., and Tversky, A. (1979). Prospect theory: an analysis of decision under risk. *Econometrica* 47, 263–292. doi: 10.2307/1914185

Kelso, J. A. S. (1999). *Dynamic Patterns–The Self-Organization of Brain and Behavior*. Cambridge, MA: MIT Press.

Khalsa, S. S., Rudrauf, D., Feinstein, J. S., and Tranel, D. (2009). The pathways of interoceptive awareness. *Nat. Neurosci.* 12, 1494–1496. doi: 10.1038/nn.2411

Khemlani, S. S., and Johnson-Laird, P. N. (2011). The need to explain. *Q. J. Exper. Psychol.* 64, 2276–2288. doi: 10.1080/17470218.2011.592593

Kircanski, K., Mortazavi, A., Castriotta, N., Baker, A. S., Mystkowski, J. L., Yi, R., et al. (2012). Challenges to the traditional exposure paradigm: variability in exposure therapy for contamination fears. *J. Behav. Ther. Exp. Psychiatry* 43, 745–751. doi: 10.1016/j.jbtep.2011.10.010

Kuhn, T. S. (1962/2012). *The Structure of Scientific Revolutions*. Chicago, IL: University of Chicago Press.

Kuyken, W., Watkins, E., Holden, E., White, K., Taylor, R. S., Byford, S., et al. (2010). How does mindfulness-based cognitive therapy work? *Behav. Res. Ther.* 48, 1105–1112. doi: 10.1016/j.brat.2010.08.003

Langdon, R., and Connaughton, E. (2013). "The neuropsychology of belief formation," in *The Neural Basis of Human Belief Systems*, eds F. Krueger and J. Grafman (New York, NY: Psychology Press), 19–42.

Laurence, S., and Margolis, E. (2012). "The scope of the conceptual," in *The Oxford Handbook of Philosophy of Cognitive Science*, eds E. Margolis, R. Samuels, and S. P. Stich (New York, NY: Oxford University Press), 291–317.

Leahy, R. L. (2001). *Overcoming Resistance in Cognitive Therapy*. New York, NY: Guilford Press.

Leahy, R. L. (2005). "Cognitive therapy," in *Psychological Treatment of Bipolar Disorder*, eds S. L. Johnson and R. L. Leahy (New York, NY: Guilford Press), 139–161.

Leahy, R. L., and Rego, S. A. (2012). "What is cognitive restructuring?" in *Cognitive Behavior Therapy: Core Principles for Practice*, eds W. T. O'Donohue and J. E. Fisher (Hoboken, NJ: John Wiley and Sons, Inc.), 133–158. doi: 10.1002/9781118470886.ch6

Lecher, C. (2014). Stephen Wolfram wants to make computer language more human. *Popular Science*. Available online at: http://www.popsci.com/article/technology/stephen-wolfram-wants-make-computer-language-more-human

LeDoux, J. (1996). *The Emotional Brain*. New York, NY: Touchstone.

Lee, J. K., Orsillo, S. M., Roemer, L., and Allen, L. B. (2010). Distress and avoidance in generalized anxiety disorder: exploring the relationships with intolerance of uncertainty and worry. *Cogn. Behav. Ther.* 39, 126–136. doi: 10.1080/16506070902966918

Levine, J. (1983). Materialism and qualia: the explanatory gap. *Pac. Philos. Quart.* 64, 354–361.

Levine, J. (1999). "Conceivability, identity, and the explanatory gap," in *Toward a science of consciousness III: The third Tucson discussions and debates (complex adaptive systems)*, eds S. R. Hameroff, A. W. Kaszniak, and D. Chalmers (Cambridge, MA: MIT University Press), 3–12.

Lieberman, M. D., Eisenberger, N. I., Crockett, M. J., Tom, S. M., Pfeifer, J. H., and Way, B. M. (2007). Putting feelings into words. *Psychol. Sci.* 18, 421–428. doi: 10.1111/j.1467-9280.2007.01916.x

Lightsey, O. R., Johnson, E., and Freeman, P. (2012). Can positive thinking reduce negative affect? A test of potential mediating mechanisms. *J. Cogn. Psychol.* 26, 71–88. doi: 10.1891/0889-8391.26.1.71

Linehan, M. M. (1993). *Cognitive-Behavioral Treatment of Borderline Personality Disorder*. New York, NY: Guilford Press.

Liu, W., McTear, M. F., and Hong, J. (1991). "Propagating beliefs among frames of discernment in Dempster-Shafer theory," in *AI and Cognitive Science'90*, eds M. F. McTear and N. Creaney (London, UK: Springer-Verlag), 367–377.

Longmore, R. J., and Worrell, M. (2007). Do we need to challenge thoughts in cognitive behavior therapy? *Clin. Psychol. Rev.* 27, 173–187. doi: 10.1016/j.cpr.2006.08.001

Lynch, T. R., Chapman, A. L., Rosenthal, M. Z., Kuo, J. R., and Linehan, M. M. (2006). Mechanisms of change in dialectical behavior therapy: theoretical and empirical observations. *J. Clin. Psychol.* 62, 459–480. doi: 10.1002/jclp.20243

Macleod, C. (2010). Current directions at the juncture of clinical and cognitive science: a commentary on the special issue. *Appl. Cognit. Psychol.* 24, 450–463. doi: 10.1002/acp.1697

Makinson, D. (2003). Ways of doing logic: what was new about AGM 1985. *J. Logic. Comput.* 13, 5–15.

Makinson, D. (2009). "Levels of belief in nonmonotonic reasoning," in *Degrees of Belief*, eds F. Huber and C. Schmidt-Petri (New York, NY: Springer Science+Business Media), 341–354. doi: 10.1007/978-1-4020-9198-8_13

Markman, A. B. (2012). "Knowledge representation," in *The Oxford Handbook of Thinking and Reasoning*, eds K. J. Holyoak and R. G. Morrison (New York, NY: Oxford University Press), 36–51.

Markoff, J. (2012). How many computers to identify a cat? 16,000. *N.Y. Times*. Available online at: http://www.nytimes.com/2012/06/26/technology/in-a-big-network-of-computers-evidence-of-machine-learning.html?pagewanted=all

Martin, L. (1986). Eskimo words for snow: a case study in the genesis and decay of an anthropological example. *Am. Anthropol.* 88, 418–423. doi: 10.1525/aa.1986.88.2.02a00080

Marupaka, N., Iyer, L. R., and Minai, A. A. (2012). Connectivity and thought: the influence of semantic network structure in a neurodynamical model of thinking. *Neural. Netw.* 32, 147–158. doi: 10.1016/j.neunet.2012.02.004

Matthews, E. (2013). "Mental disorder: can Merleau-Ponty take us beyond the mind-brain problem," in *The Oxford Handbook of Philosophy and Psychiatry*, eds K. W. M. Fulford, M. Davies, R. G. T. Gipps, G. Graham, J. Z. Sadler, G. Stanghellini, and T. Thornton (Oxford: Oxford University Press), 531–544. doi: 10.1093/oxfordhb/9780199579563.013.0033

McCullough, J. P. (2000). *Treatment for Chronic Depression: Cognitive Behavioral Analysis System of Psychotherapy (CBASP)*. New York, NY: Guilford Press.

McGrath, M. (2012). Propositions. *Stanford Encyclopedia Philos.* Available online at: http://plato.stanford.edu/entries/propositions/

McMillan, D., and Lee, R. (2010). A systematic review of behavioral experiments vs. exposure alone in the treatment of anxiety disorders: a case of exposure while wearing the emperor's new clothes? *Clin. Psychol. Rev*. 30, 467–478. doi: 10.1016/j.cpr.2010.01.003

McRae, K., and Jones, M. (2013). "Semantic memory," in *The Oxford Handbook of Cognitive Psychology*, D. Reisberg (Oxford: Oxford University Press), 206–219.

Meacham, C. J. G., and Weisberg, J. (2011). Representation theorems and the foundations of decision theory. *Australas. J. Philos.* 89, 641–663. doi: 10.1080/00048402.2010.510529

Michalewicz, Z., and Fogel, D. (2004). *How to Solve it: Modern Heuristics, 2nd Edn*. New York, NY: Springer Verlag. doi: 10.1007/978-3-662-07807-5

Miller, C. C. (2013). Google alters search to handle more complex queries. *N.Y. Times*. Available online at: http://bits.blogs.nytimes.com/2013/09/26/google-changes-search-to-handle-more-complex-queries/?_r=0

Miller, W. R., and Rollnick, S. (2012). *Motivational Interviewing: Helping People Change, 3rd Edn*. New York, NY: Guilford Press.

Möller, H.-J. (2012). How close is evidence to truth in evidence-based treatment of mental disorders? *Eur. Arch. Psychiat. Clin. Neurosci.* 262, 277–289. doi: 10.1007/s00406-011-0273-8

Morina, N., Deeprose, C., Pusowski, C., Schmid, M., and Holmes, E. A. (2011). Prospective mental imagery in patients with major depressive disorder or anxiety disorders. *J. Anxiety Disord.* 25, 1032–1037. doi: 10.1016/j.janxdis.2011.06.012

Morrison, R. G., and Knowlton, B. J. (2012). "Neurocognitive methods in higher cognition," in *The Oxford Handbook of Thinking and Reasoning*, eds K. J. Holyoak and R. G. Morrison (New York, NY: Oxford University Press), 67–89.

Moser, J. S., Hartwig, R., Moran, T. P., Jendrusina, A. A., and Kross, E. (2014). Neural markers of positive appraisal and their associations with trait reappraisal and worry. *J. Abnorm. Psychol.* 123, 91–105. doi: 10.1037/a0035817

Musil, R. (1930-43). *Der Mann Ohne Eigenschaften*. Vienna, AU: Rowohlt Verlag.

Nelson, D. L., Kitto, K., Galea, D., McEvoy, C. L., and Bruza, P. D. (2013). How activation, entanglement, and searching a semantic network contribute to event memory. *Mem. Cognit.* 41, 797–819. doi: 10.3758/s13421-013-0312-y

Nelson, E. A., Deacon, B. J., Lickel, J. J., and Sy, J. T. (2010a). Targeting the probability versus cost of feared outcomes in public speaking anxiety. *Behav. Res. Ther.* 48, 282–289. doi: 10.1016/j.brat.2009.11.007

Nelson, E. A., Lickel, J. J., Sy, J. T., Dixon, L. J., and Deacon, B. J. (2010b). Probability and cost biases in social phobia: nature, specificity, and relationship to treatment outcome. *J. Cogn. Psychol.* 24, 213–228. doi: 10.1891/0889-8391.24.3.213

Newman, M. G., and Llera, S. J. (2011). A novel theory of experiential avoidance in generalized anxiety disorder: a review and synthesis of research supporting a contrast avoidance model of worry. *Clin. Psychol. Rev.* 31, 371–382. doi: 10.1016/j.cpr.2011.01.008

Nicolis, G., and Prigogine, I. (1989). *Exploring Complexity*. New York, NY: W. H. Freeman and Company.

Noë, A. (2004). *Action in Perception*. Cambridge, MA: MIT Press.

Oaksford, M., and Chater, N. (2007). *Bayesian Rationality: The Probabilistic Approach to Human Reasoning*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780198524496.001.0001

Ohlsson, S. (2011). *Deep Learning: How the Mind Overrides Experience*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511780295

Olthuis, J. V., Stewart, S. H., Watt, M. C., Sabourin, B. C., and Keogh, E. (2012). Anxiety sensitivity and negative interpretation biases: their shared and unique associations with anxiety symptoms. *J. Psychopathol. Behav*. 34, 332–342. doi: 10.1007/s10862-012-9286-5

Patterson, R., and Barbey, A. K. (2013). "A multiple systems approach to causal reasoning," in *The Neural Basis of Human Belief Systems*, eds F. Krueger and J. Grafman (New York, NY: Psychology Press), 43–71.

Paulus, M. P., and Stein, M. B. (2010). Interoception in anxiety and depression. *Brain Struct. Funct.* 214, 451–463. doi: 10.1007/s00429-010-0258-9

Pavlov, I. (1927/2003). *Conditioned Reflexes*. Mineola, NY: Dover Publications.

Pecher, D. (2013). "The perceptual representation of mental categories," in *The Oxford Handbook of Cognitive Psychology*, ed D. Reisberg (Oxford: Oxford University Press), 358–373.

Perring, C. (2010). Mental illness. *Stanford Encyclopedia Philos*. Available online at: http://plato.stanford.edu/entries/mental-illness/

Persons, J. B., Roberts, N. A., Zalecki, C. A., and Brechwald, W. A. G. (2006). Naturalistic outcome of case formulation-driven cognitive-behavior therapy for anxious depressed outpatients. *Behav. Res. Ther.* 44, 1041–1051. doi: 10.1016/j.brat.2005.08.005

Pessoa, L. (2008). On the relationship between emotion and cognition. *Nat. Rev. Neurosci.* 9, 148–158. doi: 10.1038/nrn2317

Pessoa, L. (2014). "The impact of emotion on cognition," in *The Oxford Handbook of Cognitive Neuroscience*, vol. 2, eds K. N. Ochsner and S. M. Kosslyn (Oxford: Oxford University Press), 79–93.

Piantadosi, S. T., Tily, H., and Gibson, E. (2012). The communicative function of ambiguity in language. *Cognition* 122, 280–291. doi: 10.1016/j.cognition.2011.10.004

Poldrack, R. A., Kittur, A., Kalar, D., Miller, E., Seppa, C., Gil, Y., et al. (2011). The cognitive atlas: toward a knowledge foundation for cognitive neuroscience. *Front. Neuroinform*. 5:17. doi: 10.3389/fninf.2011.00017

Princeton University. (2010). *About WordNet.* Available online at: http://wordnet.princeton.edu/wordnet/

Quine, W. V. O. (1953/1980). "Two dogmas of empiricism," in *From a Logical Point of View, 2nd Revised Edn.* (New York, NY: Harper and Row), 20–46.

Quine, W. V. O. (1964). *Word and Object*. Cambridge, MA: MIT University Press.

Quine, W. V. O., and Ullian, J.S. (1978). *The Web of Belief, 2nd Edn*. Columbus, OH: McGraw-Hill.

Rabinovich, M. I., Afraimovich, V. S., Bick, C., and Varona, P. (2012a). Information flow dynamics in the brain. *Phys. Life Rev.* 9, 51–73. doi: 10.1016/j.plrev.2011.11.002

Rabinovich, M. I., Afraimovich, V. S., Bick, C., and Varona, P. (2012b). Instability, semantic dynamics and modeling brain data. *Phys. Life. Rev.* 9, 80–83. doi: 10.1016/j.plrev.2012.01.003

Rabinovich, M. I., Afraimovich, V. S., and Varona, P. (2010b). Heteroclinic binding. *Dynam. Syst.* 25, 433–442. doi: 10.1080/14689367.2010.515396

Rabinovich, M. I., Muezzinoglu, M. K., Strigo, I., and Bystritsky, A. (2010a). Dynamical principles of emotion-cognition interaction: mathematical images of mental disorders. *PLoS ONE* 5:e12547. doi: 10.1371/journal.pone.0012547

Rabinovich, M. I., Varona, P., Tristan, I., and Afraimovich, V. S. (2014). Chunking dynamics: heteroclinics in mind. *Front. Comput. Neurosci.* 8:1–22. doi: 10.3389/fncom.2014.00022

Reardon, S. (2014). Brain-mapping projects to join forces. *Nature* doi: 10.1038/nature.2014.14871. (in press).

Reisberg, D. (2014). "Mental images," in *The Oxford Handbook of Cognitive Psychology*, ed D. Reisberg (Oxford: Oxford University Press), 374–390. doi: 10.1093/oxfordhb/9780195376746.001.0001

Rescorla, R. A., and Wagner, A. R. (1972). "A theory of pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement," in *Classical Conditioning II: Current Research and Theory*, eds A. H. Black and W. F. Prokasy (New York, NY: Appleton-Century-Crofts), 64–99.

Resick, P. A., Nishith, P., Weaver, T. L., Astin, M. C., and Feuer, C. A. (2002). A comparison of cognitive-processing therapy with prolonged exposure and a waiting condition for the treatment of chronic posttraumatic stress disorder in female rape victims. *J. Consult. Clin. Psychol.* 70, 867–879. doi: 10.1037/0022-006X.70.4.867

Rhodes, J., and Gipps, R. G. T. (2008). Delusions, certainty, and the background. *Philos. Psychiatry Psychol.* 15, 295–310. doi: 10.1353/ppp.0.0202

Rips, L. J., Smith, E. E., and Medin, D. L. (2012). "Concepts and categories: memory, meaning, and metaphysics," in *The Oxford Handbook of Thinking and Reasoning*, eds K. J. Holyoak and R. G. Morrison (New York, NY: Oxford University Press), 177–209.

Roemer, L., Orsillo, S. M., and Barlow, D. H. (2002). "Generalized anxiety disorder," in *Anxiety and Its Disorders, 2nd Edn.*, ed D. H. Barlow (New York, NY: Guilford Press), 477–515.

Russell, B. (1921/2005). *The Analysis of Mind*. Mineola, NY: Dover Publications, Inc.

Ruths, J., and Ruths, D. (2014). Control profiles of complex networks. *Science* 343, 1373–1376. doi: 10.1126/science.1242063

Ryle, G. (1949/2009). *The Concept of Mind*. New York, NY: Routledge.

Samanez-Larkin, G. R., Hollon, N. G., Carstensen, L. L., and Knutson, B. (2008). Individual differences in insular sensitivity during loss anticipation predict avoidance learning. *Psychol. Sci.* 19, 320–323. doi: 10.1111/j.1467-9280.2008.02087.x

Samuels, S. (2012). "Massive modularity," in *The Oxford Handbook of Philosophy of Cognitive Science*, eds E. Margolis, R. Samuels, and S. P. Stich (New York, NY: Oxford University Press), 60–91. doi: 10.1093/oxfordhb/9780195309799.003.0004

Saxe, R., and Young, L. L. (2014). "Theory of mind: how brains think about thoughts," in *The Oxford Handbook of Cognitive Neuroscience*, Vol. 2, eds K. N. Ochsner and S. M. Kosslyn (Oxford: Oxford University Press), 204–213.

Scholz, B. C. (2011). Philosophy of linguistics. *Stanford Encyclopedia Philos.* Available online at: http://plato.stanford.edu/entries/linguistics/

Schraw, G., and Dennison, R. S. (1994). Assessing metacognitive awareness. *Contemp. Educ. Psychol.* 19, 460–475. doi: 10.1006/ceps.1994.1033

Schroeter, L. (2012). Two-dimensional semantics. *Stanford Encyclopedia Philos.* Available online at: http://plato.stanford.edu/archives/win2012/entries/two-dimensional-semantics/

Schultz, L. T., and Heimberg, R. G. (2008). Attentional focus in social anxiety disorder: potential for interactive processes. *Clin. Psychol. Rev*. 28, 1206–1221. doi: 10.1016/j.cpr.2008.04.003

Schwitzgebel, E. (2010). Belief. *Stanford Encyclopedia Philos.* Available online at: http://plato.stanford.edu/entries/belief/

Searle, J. R. (1983). *Intentionality*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139173452

Searle, J. R. (1992). *The Rediscovery of the Mind*. Cambridge, MA: MIT Press.

Searle, J. R. (1995). *The Construction of Social Reality*. New York, NY: Free Press.

Searle, J. R. (2007). "What is language: some preliminary remarks," in *Explorations in Pragmatics–Linguistic, Cognitive and Intercultural Aspects,* eds I. Kecskes, and L. R. Horn (Berlin, DE: Walter de Gruyter GmbH and Co.), 7–38. Reprinted in (2007). *John Searle's Philosophy of Language: Force, Meaning and Mind,* ed S. L. Tsohatzidis (Cambridge: Cambridge University Press), 15–46. Also reprinted in (2009). *Ethics Politics*, 9, 173–202.

Searle, J. R. (2010). *Making the Social World*. Oxford: Oxford University Press.

Segerberg, K., Meyer, J.-J., and Kracht, M. (2009). The logic of action. *Stanford Encyclopedia Philos.* Available online at: http://plato.stanford.edu/entries/logic-action/

Sengupta, S. (2013). Facebook unveils a new search tool. *N.Y. Times*. Available online at: http://bits.blogs.nytimes.com/2013/01/15/facebook-unveils-a-new-search-tool/

Shafer, G., and Tversky, A. (1985). Languages and designs for probability judgment. *Cogn. Sci*. 9, 309–339. doi: 10.1207/s15516709cog0903_2

Shanahan, M. (2009). The frame problem. *Stanford Encyclopedia Philos*. Available online at: http://plato.stanford.edu/entries/frame-problem/

Shea, N. (2013). "Neural mechanisms of decision-making and the personal level," in *The Oxford Handbook of Philosophy and Psychiatry*, eds K. W. M. Fulford, M. Davies, R. G. T. Gipps, G. Graham, J. Z. Sadler, G. Stanghellini, and T. Thornton (Oxford: Oxford University Press), 1063–1082. doi: 10.1093/oxfordhb/xya9780199579563.013.0062

Shermer, M. (2012). *The Believing Brain*. New York, NY: Henry Holt and Co.

Siddique, H. (2014). *World's First Computational Psychiatry Centre Opens in London. The Guardian.* Available online at: http://www.theguardian.com/science/2014/apr/02/worlds-first-computational-psychiatry-centre-london

Simmons, A., Strigo, I., Matthews, S. C., Paulus, M. P., and Stein, M. B. (2006). Anticipation of aversive visual stimuli is associated with increased insula activation in anxiety-prone subjects. *Biol. Psychiatry* 60, 402–409. doi: 10.1016/j.biopsych.2006.04.038

Skinner, B. F. (1947). Superstition in the pigeon. *J. Exp. Psychol.* 38, 168–172. doi: 10.1037/h0055873

Skinner, B. F. (1957/1991). *Verbal Behavior*. Acton, MA: Copley Publishing Group.

Smits, J. A., Berry, A. C., Tart, C. D., and Powers, M. B. (2008). The efficacy of cognitive-behavioral interventions for reducing anxiety sensitivity: a meta-analytic review. *Behav. Res. Ther*. 46, 1047–1054. doi: 10.1016/j.brat.2008.06.010

SNePS. Research Group. (2013). *SNePS*. Available online at: http://www.cse.buffalo.edu/sneps/

Snow, C. P. (1959/2012). *The Two Cultures*. Cambridge: Cambridge University Press.

Solé, R., and Seoane, L. (2014). Ambiguity in language networks. *Phys. Soc*. Available online at: http://arxiv.org/abs/1402.4802

Sonnemann, U., Camerer, C. F., Fox, C. R., and Langer, T. (2013). How psychological framing affects economic market prices in the lab and field. *Proc. Natl. Acad. Sci. U.S.A.* 110, 11779–11784. doi: 10.1073/pnas.1206326110

Spiegler, M. D., and Guevremont, D. C. (2009). *Contemporary Behavior Therapy*. Independence, MO: Wadsworth Publishing.

Spohn, W. (2009). "A survey of ranking theory," in *Degrees of Belief*, eds F. Huber and C. Schmidt-Petri (New York, NY: Springer Science+Business Media), 185–228. doi: 10.1007/978-1-4020-9198-8_8

Sporns, O., Tononi, G., and Kötter, R. (2005). The human connectome: a structural description of the human brain. *PLoS Comput. Biol.* 1:e42. doi: 10.1371/journal.pcbi.0010042

Squire, L. R., and Kandel, E. R. (1998). *Memory: From Mind to Molecules*. New York, NY: Scientific American Library.

Stone, M. (2014). "Semantics and computational semantics," in *Cambridge Handbook of Semantics (forthcoming)*, eds P. Dekker and M. Aloni (Cambridge: Cambridge University Press). Available online at: http://www.cs.rutgers.edu/~mdstone/pubs/compsem13.pdf

Studdert-Kennedy, M. (2005). "How did language go discrete?" in *Language Origins–Perspectives on Evolution*, ed M. Tallerman (Oxford: Oxford University Press), 48–67.

Stumpf, M. P., Thorne, T., de Silva, E., Stewart, R., An, H. J., Lappe, M., et al. (2008). Estimating the size of the human interactome. *Proc. Natl. Acad. Sci. U.S.A.* 105, 6959–6964. doi: 10.1073/pnas.0708078105

Swinburne, R. (2011). "Evidence," in *Evidentialism and Its Discontents*, ed T. Dougherty (Oxford: Oxford University Press), 195–206. doi: 10.1093/acprof:oso/9780199563500.003.0013

Swoyer, C. (2003). The linguistic relativity hypothesis. *Stanford Encyclopedia Philos.* Available online at: http://plato.stanford.edu/entries/relativism/supplement2.html

Taylor, C. T., and Alden, L. E. (2010). Safety behaviors and judgmental biases in social anxiety disorder. *Behav. Res. Ther.* 48, 226–237. doi: 10.1016/j.brat.2009.11.005

Thelen, E., and Smith, L. B. (2000). *A Dynamic Systems Approach to the Deelopment of Cognition and Action*. Cambridge, MA: MIT Press.

Trouche, S., Sasaki, J. M., Tu, T., and Reijmers, L. G. (2013). Fear extinction causes target-specific remodeling of perisomatic inhibitory synapses. *Neuron* 80, 1054–1065. doi: 10.1016/j.neuron.2013.07.047

Tryon, W. W., and McKay, D. (2009). Memory modification as an outcome variable in anxiety disorder treatment. *J. Anx. Disord.* 23, 546–556. doi: 10.1016/j.janxdis.2008.11.003

Tversky, A., and Kahneman, D. (1983). Extensional versus intuitive reasoning: the conjunction fallacy in probability judgment. *Psychol. Rev.* 90, 293–315. doi: 10.1037/0033-295X.90.4.293

University of Colorado Boulder. (1998). *What is LSA?* Available online at: http://lsa.colorado.edu

Vance, A. (2010). In pursuit of a mind map, slice by slice. *N.Y. Times*. Available online at: http://www.nytimes.com/2010/12/28/science/28brain.html?pagewanted=all

van Gulick, R. (2004). Consciousness. *Stanford Encyclopedia Philos*. Available online at: http://plato.stanford.edu/entries/consciousness/

Wallsten, T. S., Budescu, D. V., Erev, I., and Diederich, A. (1997). Evaluating and combining subjective probability estimates. *J. Behav. Decis. Making* 10, 243–268. doi: 10.1002/(SICI)1099-0771(199709)10:3%3C243::AID-BDM268%3E3.0.CO;2-M

Warman, D. M., Lysaker, P. H., Martin, J. M., Davis, L., and Haudenschield, S. L. (2007). Jumping to conclusions and the continuum of delusional beliefs. *Behav. Res. Ther.* 45, 1255–1269. doi: 10.1016/j.brat.2006.09.002

Weisberg, J. (2011). "Varieties of bayesianism," in *Handbook of the History of Logic*–Vol. 10–*Inductive Logic*, eds D. M. Gabbay, S. Hartmann, and J. Woods (Amsterdam: Elsevier), 477–551.

Wells, A. (2006). "Anxiety disorders, metacognition, and change," in *Roadblocks in Cognitive-Behavioral Therapy: Transforming Challenges into Opportunities for Change*, ed R. L. Leahy (New York, NY: Guilford Press), 69–90.

Wilson-Mendenhall, C. D., Simmons, W. K., Martin, A., and Barsalou, L. W. (2013). Contextual processing of abstract concepts reveals neural representations of nonlinguistic semantic content. *J. Cogn. Neurosci.* 25, 920–935. doi: 10.1162/jocn_a_00361

Woody, S. R., and Nosen, E. (2009). "Psychological models of phobic disorders and panic," in *Oxford Handbook of Anxiety and Related Disorders*, eds M. M. Anthony, and M. B. Stein (Oxford: Oxford University Press), 209–224.

Wortham, J. (2010). *Apple Buys A Start-Up For Its Voice Technology*. Available online at: http://www.nytimes.com/2010/04/29/technology/29apple.html

Yee, E., Chrysikou, E. G., and Thompson-Schill, S. L. (2014). "Semantic memory," in *The Oxford Handbook of Cognitive Neuroscience*, Vol. 1, eds K. N. Ochsner, and S. M. Kosslyn (New York, NY: Oxford University Press), 353–374.

Zalta, E. N. (2012). Gottlob Frege. *Stanford Enyclopedia Philos*. Available online at: http://plato.stanford.edu/entries/frege/

Zanov, M. V., and Davison, G. C. (2010). A conceptual and empirical review of 25 years of cognitive assessment using the articulated thoughts in simulated situations (ATSS) think-aloud paradigm. *Cognitive Ther. Res.* 34, 282–291. doi: 10.1007/s10608-009-9271-9

Zhao, J., Crupi, V., Tentori, K., Fitelson, B., and Osherson, D. (2012). Updating: learning versus supposing. *Cognition* 124, 373–378. doi: 10.1016/j.cognition.2012.05.001

Zhao, J., and Osherson, D. (2010). Updating beliefs in light of uncertain evidence: descriptive assessment of Jeffrey's rule. *Think. Reasoning.* 16, 288–307. doi: 10.1080/13546783.2010.521695

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.

Received: 30 January 2014; paper pending published: 21 February 2014; accepted: 24 April 2014; published online: 15 May 2014.

Citation: Kronemyer D and Bystritsky A (2014) A non-linear dynamical approach to belief revision in cognitive behavioral therapy. Front. Comput. Neurosci. ***8****:55. doi: 10.3389/fncom.2014.00055*

This article was submitted to the journal Frontiers in Computational Neuroscience.

frontiers in Computational Neuroscience

ORIGINAL RESEARCH
published: 25 March 2015
doi: 10.3389/fncom.2015.00037

Characterizing psychological dimensions in non-pathological subjects through autonomic nervous system dynamics

Mimma Nardelli[1], Gaetano Valenza[1], Ioana A. Cristea[2,3], Claudio Gentili[2], Carmen Cotet[3], Daniel David[3], Antonio Lanata[1] and Enzo P. Scilingo[1]*

[1] Department of Information Engineering & Research Centre E. Piaggio, Faculty of Engineering, University of Pisa, Pisa, Italy, [2] Section of Psychology, Department of Surgical, Medical, Molecular, and Critical Area Pathology, University of Pisa, Pisa, Italy, [3] Department of Clinical Psychology and Pychotherapy, Babes-Bolyai University, Cluj-Napoca, Romania

The objective assessment of psychological traits of healthy subjects and psychiatric patients has been growing interest in clinical and bioengineering research fields during the last decade. Several experimental evidences strongly suggest that a link between Autonomic Nervous System (ANS) dynamics and specific dimensions such as anxiety, social phobia, stress, and emotional regulation might exist. Nevertheless, an extensive investigation on a wide range of psycho-cognitive scales and ANS non-invasive markers gathered from standard and non-linear analysis still needs to be addressed. In this study, we analyzed the discerning and correlation capabilities of a comprehensive set of ANS features and psycho-cognitive scales in 29 non-pathological subjects monitored during resting conditions. In particular, the state of the art of standard and non-linear analysis was performed on Heart Rate Variability, InterBreath Interval series, and InterBeat Respiration series, which were considered as monovariate and multivariate measurements. Experimental results show that each ANS feature is linked to specific psychological traits. Moreover, non-linear analysis outperforms the psychological assessment with respect to standard analysis. Considering that the current clinical practice relies only on subjective scores from interviews and questionnaires, this study provides objective tools for the assessment of psychological dimensions.

Keywords: psychological scales, Heart Rate Variability, InterBreath Intervals series, nonlinear analysis, multiscale entropy, multivariate multiscale entropy

OPEN ACCESS

Edited by:
Tobias Alecio Mattei, Brain & Spine Center - InvisionHealth - Kenmore Mercy Hospital, USA

Reviewed by:
Jianbo Gao, Wright State University, USA
Sara Bottiroli, IRCCS Mondino, Italy

***Correspondence:**
Gaetano Valenza, Department of Information Engineering, University of Pisa, via Caruso 16, 56126 Pisa, Italy
g.valenza@ieee.org

Received: *11 November 2014*
Paper pending published: *22 December 2014*
Accepted: *06 March 2015*
Published: *25 March 2015*

Citation:
Nardelli M, Valenza G, Cristea IA, Gentili C, Cotet C, David D, Lanata A and Scilingo EP (2015) Characterizing psychological dimensions in non-pathological subjects through autonomic nervous system dynamics. Front. Comput. Neurosci. 9:37. doi: 10.3389/fncom.2015.00037

1. Introduction

Psychological assessment refers to the practice of standardized evaluation of performance or impairment in different domains of thinking, learning and behavior. Accordingly, such an assessment can be used to characterize and quantify different behaviors in healthy subjects or to reveal the presence of behavioral disorders such as anxiety and social phobia. Depending on the factors under observation, psychological assessment can be achieved via different routes: behavioral tasks, questionnaires, or interviews. The evaluation is done by a professional (i.e., certified psychologist) in order to obtain a standardized and quantifiable information of the subject under study

(Cohen et al., 1992). These approaches are useful in performing an individual assessment for which the performance of one person can be interpreted through pre-existing norms, as well as in group assessment which allows for different comparisons (within a single group or between groups) (Kenny et al., 2008). It is worthwhile noting that self-report questionnaires and interviews currently represent the standard clinical practice in diagnosing psychiatric disorders (Cohen et al., 1992; Valenza et al., 2013a, 2014c).

Nevertheless, several issues in these kinds of approaches still need to be addressed. First, the scores are obtained with subjective procedures which might be biased by possible social desirability thoughts of the subject and possible recent emotional events. Moreover, professionals need to choose the appropriate test for each psychological dimension and subject, and verify that it has good psychometric properties in order to adhere to the evidence-based paradigm (i.e., reliability and validity) (Groth-Marnat, 2003; Hunsley and Mash, 2010). To overcome these problems, several efforts have been made in psycho-physiological and bioengineering research fields to objectify the psychological assessment. In particular, physiological correlates of the central and autonomic nervous systems (CNS and ANS, respectively) have been extensively studied and taken into account (Taillard et al., 1990, 1993; Carney et al., 1995; Glassman, 1998; Stampfer, 1998; Iverson et al., 2002, 2005; Watkins et al., 2002; Calvo and D'Mello, 2010; Lin et al., 2010; Petrantonakis and Hadjileontiadis, 2011; Valenza et al., 2012a,b, 2013a,b, 2014c).

To give some significant examples, physiological correlates of mood disorders such as bipolar disorders have been found on sleep (Stampfer, 1998; Iverson et al., 2002, 2005), hormonal system (Carney et al., 1995; Glassman, 1998; Watkins et al., 2002), and ANS dynamics through heartbeat and respiratory dynamics (Taillard et al., 1990, 1993; Valenza et al., 2013a, 2014c). Moreover, as the psychological dimensions can be related to variations of emotional states, several computational methods for automatic emotion recognition have been developed using electroencephalogram (EEG) and ANS signal analysis (Taillard et al., 1990, 1993; Calvo and D'Mello, 2010; Lin et al., 2010; Petrantonakis and Hadjileontiadis, 2011; Valenza et al., 2012a,b, 2013a,b, 2014c).

Here we focus on the link between ANS dynamics and psychological dimensions. This choice is justified by the fact that ANS dynamics cannot be straightforwardly changed by the subject intention and is under direct control of CNS pathways such as the prefrontal cortex, amygdala, and brainstem (Ruiz-Padial et al., 2011). Of note, dysfunctions on these CNS recruitment circuits lead to pathological effects (Heller et al., 2009) such as anhedonia, i.e., the loss of pleasure or interest in previously rewarding stimuli, which is a core feature of major depression and other serious mood disorders. Moreover, ANS monitoring is widely available, cost-effective, and can be easily performed through wearable systems such as sensorized t-shirts (Valenza et al., 2008, 2014c) or gloves (Lanatà et al., 2012), and its dynamics is thought to be less sensitive to artifact events than in the EEG case.

ANS dynamics has been demonstrated to provide effective markers of typical psychological processes. As a matter of fact, previous studies (Freeman and Nixon, 1985; Yeragani et al., 1999; Virtanen et al., 2003; Cohen and Benjamin, 2006; Shinba et al., 2008; Licht et al., 2009; Thayer et al., 2010, 2012) suggest that patients with anxiety are at increased risk for heart disease (e.g., the association between phobic anxiety or panic disorder and somatic morbidity as coronary heart disease, coronary spasm and ventricular arrhythmia). ANS markers of anxiety and panic disorders can be found through the analysis of the Heart Rate Variability (HRV), revealing an increased heart rate and decreased power in low-frequency (LF) and high frequency (HF) bands. A decreased HF spectral power of HRV was also found in patients affected by generalized anxiety disorder (Thayer et al., 1996), whereas a decreased heart rate was also found in autism spectrum disorders (Jansen et al., 2006) in response to stress. This change could be related to abnormal high basal (nor)epinephrine levels. On the contrary, increased mean heart rate associated to a reduced variability has been observed in depressed patients (Carney et al., 2005). Moreover, it has been shown how subjects reporting excessive and persistent fear of social situations are characterized by atypical ANS dynamics which is evident in variables as HRV mean, respiration rate, tidal volume, and blood pressure (Grossman et al., 2001). ANS markers gathered from non-linear analysis were related to phycological dimensions as anxiety (Cohen and Benjamin, 2006) and panic disorder through symbolic analysis (Yeragani et al., 2000). Despite the elevated number of previous studies, none of these researches have reached an acceptable level of accuracy to effectively, reliably, and objectively characterize the psychological dimensions of healthy subjects and psychiatric patients, and to forecast a clinical course. A possible reason can be related to the limited amount of ANS features and specific psychological traits that were taken into account.

Therefore, here we present a detailed study on psychological assessments through an extensive analysis of the ANS dynamics. Psychological dimensions were quantified by means of the 6 psycho-cognitive scales (see details on the series definition, estimation, and parameter extraction in Section 2.3).

In order to perform a comprehensive study, the ANS non-linear dynamics has to be taken into account. Although the detailed physiology behind such complex dynamics has not been completely clarified, it is worthwhile noting that ANS non-linear dynamics plays a crucial role in most of the underlying biological processes, as they have been proven to be of prognostic value in aging and diseases, showing robust and effective discerning and characterizing properties (Poon and Merrill, 1997; Glass, 2001; Goldberger et al., 2002; Stiedl and Meyer, 2003; Tulppo et al., 2005; Atyabi et al., 2006; Glass, 2009; Wu et al., 2009; Citi et al., 2012; Valenza et al., 2014a). Indeed, physiological systems are intrinsically non-linear systems characterized by multi-feedback interactions associated to long-range correlations (Marmarelis, 2004), likely due to the enormous amount of structural units inside them and to the various non-linear neural interactions and integrations occurring at the neuron and receptor levels. The study of the complexity of physiological signals, in particular, has led to important results in recent decades in understanding the mechanisms underlying mental illness (Yang and Tsai, 2012). Several measures of complexity have also been proposed and applied to the study of mental illness based on various biomedical signals, from EEG (Hu et al., 2006; Takahashi et al.,

2010; Gao et al., 2011), to MEG (Fernandez et al., 2010), through HRV (Mujica-Parodi et al., 2005; Hu et al., 2009, 2010; Gao et al., 2013; Valenza et al., 2014b). Accordingly, in this study we investigate the role of ANS non-linear dynamics in performing the psychological assessment, with respect to the standard analysis, i.e., analysis in the time and frequency domain.

2. Materials and Methods

2.1. Subjects Recruitment, Experimental Protocol, and Acquisition Set-up

A group of 29 non-pathological subjects (5 males), i.e., not suffering from both cardiovascular and evident mental pathologies, was recruited to participate in the experiment. Subjects were students recruited from the Babes-Bolyai University, via an online screening questionnaire assessing their intention to take part in the study. Participation was voluntary and each subjects signed a written informed consent after the study procedure had been explained. No compensation for participation was offered. Subjects underwent a medical screening interview to assess the presence of any medical condition or medication that might have interfered with their cardiovascular data. Their age ranged from 21 to 35 and were naive to the purpose of the experiment. The group was as heterogeneous as possible in order to have a wide range of psycho-cognitive-behavioral dimensions. The experimental protocol was structured in the following two phases: (1) submission of self report psycho-behavioral tests; (2) recording of the physiological signs. More in detail, all participants were screened by 6 self-report questionnaires (see details below), which were comprised of a total of 25 sub-scales. Then, physiological signals such as ElectroCardioGram (ECG), Respiration (RSP) were simultaneously acquired during resting state condition for 25 min through the BIOPAC MP150 device. The sampling rate was 1000 Hz for all signals. We used the ECG100C Electrocardiogram Amplifier from BIOPAC inc., connected with pregelled Ag/AgCl electrodes placed following Einthoven triangle configuration. The dedicated module of BIOPAC MP150 used to record the respiration activity is RSP100C Respiration Amplifier with the TSD201 sensor, which is a piezo-resistive sensor with the output resistance within the range 5–125 KOhm and bandwidth of 0.05–10 Hz. This piezoresistive sensor changed its electrical resistance if stretched or shortened, and it was sensitive to the thoracic circumference variations occurring during respiration.

The ECG signal was used to extract the HRV series, which refer to the variation of the time intervals between consecutive heartbeats identified with R-waves (RR intervals). Two different time series were extracted from the respiration activity: Inter-Breath Interval time series (IBI) and InterBeat Respiration (IBR). The IBI series was obtained detecting the local maxima of each respiratory act, whereas IBR consists of the amplitude of the respiration activity signal when sampled at the R-peak times.

2.2. Scales for the Assessment of Psychological Dimensions

In this work, we used a total of 6 self-report questionnaires in which, for most of them, different sub-scales are considered. The total number of sub-scales used in this experiment was 25. A Cronbach's α measure (Bland and Altman, 1997) is assigned to each scale and represents the consistency of the test. Such an α index depends on the number and average inter-correlation among the test questions. Details on each scale and related sub-scales are as follows:

- Positive and Negative Affect Schedule (PANAS, Watson et al., 1988). The PANAS contains 2 sub-scales—positive affect (PA) and negative affect (NA)—of 10 items describing emotions each. The scale has good reliability (Cronbach's $\alpha = 0.88$ for the PA and 0.87 for the NA sub-scale, respectively) and good construct validity. Cronbach's α for this sample was 0.87 for PA and 0.90 for NA, supporting good internal consistency.
- Liebowitz Social Anxiety Scale (LSAS, Liebowitz, 1987). The LSAS is a self-assessment social phobia questionnaire containing 24 items describing actions done in social situations, grouped at first in 2 sub-scales (social interaction and performance). Subjects rate these situations in terms of fear/anxiety and avoidance, allowing for a total of 4 separate sub-scales. The scale presents a very good internal consistency (Cronbach's $\alpha = 0.96$) as well as good convergent and divergent validity (Heimberg et al., 1999). Cronbach's α for this sample was 0.92, showing a very good internal consistency.
- Difficulties in Emotion Regulation (DERS, Gratz and Roemer, 2004). The DERS is a 36-item self-report scale measuring emotion dysregulation. The scale offers an overall score as well as scores for each of the 6 sub-scales related to DERS (Non-acceptance of Emotional Responses, Difficulties Engaging in Goal-Directed Behavior, Impulse Control Difficulties, Lack of Emotional Awareness, Limited Access to Emotion Regulation Strategies, and Lack of Emotional Clarity). Internal consistency for this scale is excellent (Cronbach's $\alpha = 0.93$) and construct and predictive validity are considered adequate.
- Interpersonal Reactivity Index (IRI, Davis, 1980). The IRI is a 28-item questionnaire measuring empathy. The scale provides scores for 4 sub-scales (Fantasy, Perspective-taking, Empathic Concern, and Personal Distress), as well as a general score of empathy. Internal consistency of the four sub-scales is acceptable (ranging from $\alpha = 0.70$–0.78).
- Behavioral Inhibition/Behavioral activation Scales (BIS/BAS, Carver and White, 1994). The BIS/BAS scale is composed of 20 items comprised in 4 sub-scales (Inhibition, Reward Responsiveness, Drive, and Fun Seeking), measuring behavioral inhibition and activation sensitivity. The scale has been adapted on the Romanian population, showing good construct validity and acceptable internal consistency (ranging from $\alpha = 0.62$–0.81) (Sava and Sperneac, 2006).
- Zuckerman Kuhlman Personality Questionnaire (ZKPQ, Zuckerman et al., 1993). The ZKPQ represents a five-factor (Impulsive Sensation Seeking, Neuroticism-Anxiety, Aggresion-Hostility, Sociability, and Activity) personality inventory containing 99 true-false items, therefore we used 5 sub-scales. The Romanian adaptation of this scale presents adequate internal consistency (α ranging from 0.69 to 0.88) and good convergent validity (Sârbescu and Neguţ, 2012).

2.3. Methodology of Signal Processing

In this section, the methodology of signal processing applied to the Heart Rate Variability (HRV), InterBreath Interval (IBI), and InterBeat Respiration (IBR) series is reported in detail. HRV refers to the variability of the series comprised of the distances between two consecutive R-waves detected from the Electrocardiogram, i.e., the R-R intervals. IBI is the series comprised of the distances between two consecutive local maxima of the respiration activity (the two maxima within two respiratory acts), whereas IBR series is the respiratory activity sampled at times corresponding to the R-peaks. Standard and non-linear monovariate and multivariate measure are extracted from each series in order to investigate a wide set of parameters characterizing the ANS linear and non-linear dynamics acting of the cardio-respiratory control.

2.3.1. Standard Measures

Standard analysis was performed on HRV series in order to extract parameters defined in the time and frequency domain (Camm et al., 1996; Acharya et al., 2006; Valenza et al., 2012b). Time domain features include statistical parameters and morphological indexes. More specifically, concerning the time domain analysis, in addition to the first (meanRR) and second order moment (SDNN) of the RR intervals, so-called normal-to-normal (NN) intervals, the square root of the mean of the sum of the squares of differences between subsequent NN intervals ($RMSSD = \sqrt{\frac{1}{N-1}\sum_{j=1}^{N-1}(RR_{j+1} - RR_j^2)}$ and the number of successive differences of intervals which differ by more than 50 ms, expressed as a percentage of the total number of heartbeats analyzed ($pNN50 = \frac{NN50}{N-1}100\%$) were calculated. Moreover, the triangular index (TINN) was estimated as the base of a triangle which better approximated the NN interval distribution (the minimum square difference is used to find such a triangle).

Concerning the frequency domain analysis, several features were calculated from the Power Spectral Density (PSD) analysis. In this work, PSD was estimated by using the Welch's periodogram, which uses the FFT (Fast Fourier Transform) algorithm. Window's width and overlap were chosen as a best compromise between the frequency resolution and variance of the estimated spectrum. Given the PSD, three spectral bands are defined as follows: VLF (very low frequency) with spectral components below 0.04 Hz; LF (low frequency), ranging between 0.04 and 0.15 Hz; HF (high frequency), comprising frequencies between 0.15 and 0.4 Hz. For each of the three frequency bands, the frequency having maximum magnitude (VLF peak, LF peak, and HF peak), the power expressed as percentage of the total power (VLF power %, LF power %, and HF power %), and the power normalized to the sum of the LF and HF power (LF power nu and HF power nu) were also evaluated. Moreover, the LF/HF power ratio was calculated.

2.3.2. Non-Linear Analysis

From the HRV, IBI, and IBR series, several non-linear measures were calculated. Such indices refer to the estimation and characterization of the phase space (or state space) of the physiological system generating the series. The phase space estimation involved the Takens method (Takens, 1981; Casdagli et al., 1991) and three parameters: m, the embedding dimension, which is a positive integer, τ, the time delay, and r, which is a positive real number and represents the margin of tolerance of the trajectories within the space. Takens theory allows for the reconstruction of the dynamic systems of different nature from time series through the method of "delayed outputs." Starting from a time series

$$X = [u(T), u(2T), ...u(NT)]$$

the attractors of the discrete dynamical system are rebuilt in a m-dimensional space, operating a delay τ on the signal. This allows achieving $N - (m - 1)$ signals of length m starting from only one:

$$\begin{cases} X_1 = [u(T), u(2T), ...u(mT)] \\ X_2 = [u(2T), u(2T + 2\tau), ...u(2T + (m-1)\tau)] \\ ... \\ X_{N-(m-1)} = [u(N - (m-1))T), ...u(N - (m-1))T \\ \qquad\qquad +(m-1)\tau)] \end{cases}$$

The various vectors X_j are the "delayed coordinates" and the derived m-dimensional space is called "reconstructed space." From the state space theory, several ANS non-linear parameters can be derived using the following analyses:

- Poincaré Plot
- Recurrence Plot
- Correlation dimension, Approximate, and Sample Entropy
- Detrended Fluctuation Analysis
- Multiscale Entropy and Multivariate Multiscale Entropy Analysis

2.3.2.1. Poincaré Plot

This technique quantifies the fluctuations of the dynamics of the time series through a map of each point RR(n) of the RR series vs. the previous one. The quantitative analysis from the graph can be made by calculating the standard deviation of the points by the straight line $RR_{j+1} = RR_j$. The first standard deviation, SD1, is related to the points that are perpendicular to the line-of-identity and describes the short-term variability, whereas the second, SD2, describes the long-term variability.

2.3.2.2. Recurrence Plot

RP is a graphical method to investigate and quantify the time series complexity. The estimation starts from vectors

$$u_j = (RR_j, RR_{j+\tau}, ..., RR_{j+(m-1)\tau})$$
$$j = 1, 2, ..., N - (m-1)\tau.$$

RP is a symmetrical square matrix of zeros and ones, whose dimensions are $N - (m-1)\tau$, and each element is given by

$$RP(j,k) = \begin{cases} 1 & \text{if } d(u_j - u_k \leq r) \\ 0 & \text{otherwise} \end{cases}$$

where d is the Euclidean distance.

Several features can be extracted from the RP by means of the Recurrence Quantification Analysis (RQA). In particular, in this study the following RQA indices were taken into account: longest diagonal line (RP Lmax) and average diagonal line length (RP Lmean), divergence (RP DIV), the percentage of recurrence points which form diagonal lines recurrence rate, determinism (RP DET), trend (RP REC) and entropy (RP ShanEn) (Zbilut et al., 1990; Marwan et al., 2002, 2007).

2.3.2.3. Correlation dimension, Approximate, and Sample Entropy Measures

Starting from the vectors $X_1, X_2, ..., X_{N-m+1}$ in $\mathbb{R}^m$, the distance between two vectors X_i and X_j, according to the definition of Takens applied to high dimensional deterministic systems is given by Takens (1981) and Schouten et al. (1994):

$$d[X_i, X_j] = max_{k=1,2,...,m}|u(i+k-1) - u(j+k-1)| \quad (1)$$

For each i, with $1 \leq i \leq N-m+1$, we measured a parameter $C_i^m(r)$:

$$C_i^m(r) = \frac{\text{Number of j such that}(d[X_i, X_j] \leq r)}{N-m+1} \quad (2)$$

and we defined

$$C^m(r) = \frac{\sum_{i=1}^{N-m+1} \log C_i^m(r)}{N-m+1} \quad (3)$$

The correlation dimension (CD) is given by Theiler (1987)

$$CD = lim_{r\to 0} lim_{N\to \inf} \frac{logC^m(r)}{logr}$$

The calculation of ApEn used in this study refers to the expression (Pincus, 1991; Fusheng et al., 2001):

$$\text{ApEn(m,r,N)} = [C^m(r) - C^{m+1}(r)] \quad (4)$$

SampEn is a remake of *ApEn* and measures the number of pairs of vectors of length m considered "neighbors," i.e., whose distance is less than r, even if the dimension of pattern increases from m to $m+1$. Unlike *ApEn*(m, r, N), *SampEn* does not include the distance of vectors with themselves, i.e., self-matches, as suggested in the later work of Grassberger and co-workers (Grassberger and Procaccia, 1983; Grassberger, 1988) and it has the advantage of being less dependent on time series length, showing relative consistency over a broader range of possible r-, m-, and N-values. By renaming $C^m(r)$ parameters without self-matches with the notation $U^m(r)$, *SampEn* is calculated by the following expression (Richman and Moorman, 2000):

$$\text{SampEn(m,r,N)} = -\ln \frac{U^{m+1}}{U^m} \quad (5)$$

2.3.2.4. Detrended Fluctuation Analysis

The detrended fluctuation analysis features (DFA1 and DFA2) (Peng et al., 1995; Penzel et al., 2003) were evaluated to study short- and long-term autocorrelation of the HRV series. The algorithm foresaw the estimation of the series

$$y(k) = \sum_{j=1}^{k} (RR_j - \overline{RR})$$

$k = 1, ..., N$. This series was divided into segments of equal length n and for each segment the linear approximation (least square fit, y_n) was computed. Then root-mean-square fluctuation was calculated

$$F(n) = \sqrt{\frac{1}{N}\sum_{k=1}^{N}(y(k) - y_n(k))^2}$$

Making a double log graph between $log(F(n))$ and different values of n, the slope of the regression line is the α scaling exponent. DFA1 and DFA2 features represent this slope between the ranges $4 \leq n \leq 16$ and $16 \leq n \leq 64$.

2.3.2.5. Multiscale Entropy and Multivariate Multiscale Entropy Analysis

Multiscale Entropy Analysis (MSE) is a powerful methodology based on the SampEn estimation. MSE was applied in several fields such as study of human gait dynamics (Costa et al., 2003), enhancement of postural complexity (Costa et al., 2007), and synthetic RR time series (Costa et al., 2002). MSE can be an effective non-linear method to collect information about physiological systems whose dynamics is associated to multiple different scales. This method is based on the application of sample entropy method to course-grained time series constructed from a one-dimensional discrete time series by averaging the data points within non-overlapping windows of increasing length, σ. Given a time series $\{x_1, ..., x_i, ..., x_N\}$ and a scale factor σ, each element of a course-grained series $\left\{y^{(\sigma)}\right\}$ is calculated using the equation

$$y_j^{(\sigma)} = \frac{1}{\sigma} \sum_{i=(j-1)\sigma+1}^{j\sigma} x_i,\ 1 \leq j \leq N/\sigma \quad (6)$$

The length of each coarse-grained time series is equal to the length of the original time series divided by σ. The second step consists in the computation of *SampEn* (Richman and Moorman, 2000; Lake et al., 2002) algorithm on these series. Previous studies in which MSE algorithm was applied to physiological data use the standard value $m = 2$ for the pattern dimension (Costa et al., 2003; Leistedt et al., 2011). In this work the choice of the right r was performed by a method already used in the liter SampEn values were calculated for scale factors σ which were in a range from 1 to 20 and the same process was carried out on HRV, IBI, and IBR series. The complexity index (CI) was measured as the area under the curve of MSE graph and it can be calculated for short time scales, from 1 to 8 (short CI), and for higher time scales, up to 20 (long CI) (Leistedt et al., 2011).

Besides MSE analysis, we performed the Multivariate Multiscale Entropy (MMSE) (Ahmed and Mandic, 2011, 2012) analysis. This algorithm allows performing MSE analysis using multivariate time series. In this work, MMSE was used to quantify the complexity of the series derived from the electrocardiogram and breath. In particular, MMSE results were obtained on the bivariate series HRV-IBI, and HRV-IBR through the estimation of the CI indices (as described above on MSE). Before the MMSE calculation, the involved time series are scaled in the range between 0 and 1 to prevent that the different amplitudes may influence the complexity complexity (Ahmed and Mandic, 2011).

2.3.2.6. Symbolic Analysis

Symbolic analysis (Yeragani et al., 2000; Porta et al., 2001; Baumert et al., 2002; Guzzetti et al., 2005; Tobaldini et al., 2009; Caminal et al., 2010) is another powerful non-linear method which was applied on HRV data series. For each HRV series gathered from each subject, 6 levels were constructed evenly dividing the amplitude range of the samples, and a symbol (from 0 to 5) was assigned to each data sample according to the level of belonging. Then, a window of three consecutive points moves along the HRV series, and three possible configurations are identified when running all the signal: the three points belong to the same level, i.e., no variation (0V), two consecutive points belong to the same level and one to another, i.e., one variation (1V), and the remaining cases, i.e., two variations (2V). The number of patterns falling into each group (0V, 1V, 2V) and the percentage of the total (0V%, 1V%, 2V%) were calculated and used as features. Previous studies support the hypothesis that an increase of 0V patterns is related to an activation of the sympathetic activity, an increase of 2V patterns is related to an increase of the parasympathetic activity, and increases of 1V patterns is associated to a simultaneous increase of both parasympathetic and sympathetic activities.

3. Experimental Results

Experimental results are expressed in terms of statistical and correlation analysis. In the literature it can be found the threshold score of each questionnaire above which the behavior of the subject results to show altered psycho-cognitive-behavioral traits. Among all the sub-scales we only considered those where the subjects spread out over a wide range of scores in order to identify two groups, one below and the other above the threshold. For each scale we identified two groups of subjects separated by the median. In order to have two groups numerically equivalent, we selected and investigated only these scales where the median was congruent with the threshold reported in the literature. In addition, for each of the 16 scales we verified that maximum and minimum scores of each group were in the tails of the population distribution reported by the literature. In other words, for each psychological subscale, the median value of the subjects score is calculated to identify two groups: one comprised of the subjects having scores below the median, and one comprised of the subjects having scores above the median. Only 16 out of 25 sub-scales divided the subjects in two groups numerically comparable, therefore we performed the statistical analysis on the scores obtained in these 16 sub-scales. The reference values from the literature about these sub-scales are evaluated on the control groups used in several previous works. For example we considered a sample of 103 subjects (age $= 27.00 \pm 8.80$) for IRI Empathic Concern and IRI Personal Distress sub-scales, referring to a study which explored the relationship among psychological mindedness and several aspects of awareness which comprended this indices of empathy (Beitel et al., 2005) and a sample of 582 subjects for IRI fantasy sub-scale taking this data from a guide study on the empathy scales (Davis, 1980). For the two PANAS sub-scales, a group of 537 volunteers aged 18–91 was in a work that tried to evaluate the reliability and validity of the PANAS (Crawford and Henry, 2004), and 53 participants (age = 34.32 ± 10.50) were asked to answer to the LSAS questionnaires to demonstrate that this method may be employed in the assessment of social anxiety disorder (Fresco et al., 2001; Rytwinski et al., 2009). As a reference for the values of BIS and BAS sub-scales we chose a previous study where the answers of of 2725 individuals aged 18–79 were observed to validate the application of this scale to measure the behavioral inhibition and activation and its correlation with depression and anxiety (Jorm et al., 1998). The threshold value of the answers of a group of 639 participants in a study of the shortened form of the questionnaire, was taken in account for ZKPQ Impulsive Sensation Seeking and Activity subscales (age $= 22.31 \pm 5.08$) (Aluja et al., 2003). At last, as regards DERS subscales, a study on 260 subjects in order to explore the factor structure and psychometric properties of DERS measures (age $= 23.10 \pm 5.67$) was used as reference for DERS Awareness (Gratz and Roemer, 2004) and a reference sample of 42 individuals (age $= 24.24 \pm 4.38$) was considered for the other DERS sub-scales, extracted from a research which compared the values of the this psychological tests on depressed patients and healthy subjects (Ehring et al., 2008).

In the statistical analysis, for each psychological sub-scale and for each ANS feature, we applied the Mann-Whitney test in order to evaluate whether the two groups were statistically different. Moreover, the non-parametric Spearman correlation coefficient was calculated between each psychological sub-scale and ANS feature.

3.1. Statistical Analysis

As mentioned above, for each ANS feature, Mann-Whitney non-parametric U-tests were used to test the null hypothesis of having no statistical difference between two groups. The use of such a non-parametric test is justified by having non-gaussian distribution of the samples ($p < 0.05$ of the null hypothesis of having gaussian samples of the Kolmogorov-Smirnov test).

Concerning features from HRV standard analysis, 8 sub-scales (LSAS Anxiety of Performance , DERS Non-Acceptance, DERS Awareness, IRI fantasy, IRI Empathic Concern, ZKPQ Activity, ZKPQ Impulsive Seeking Sensation, BAS) showed significant discerning capability mostly through frequency domain parameters (see details in **Table 1**). Concerning ANS features coming from non-linear analysis, 9 sub-scales (PANAS Positive Affect, DERS non-Acceptance, DERS Impulse, DERS Awareness, DERS Strategies, IRI Empathic Concern, BIS, BAS, ZKPQ Activity) showed

TABLE 1 | Statistical results related to standard HRV features (*U*-test).

Scales	Sub-scales	Statistical results	
		Features	***p*-value**
LSAS	LSAS Anx P	↓ VLF peak	<0.05
		↓ HF peak	<0.03
DERS	DERS Non-Accept	↓ LF peak	<0.03
		↑ VLF power	<0.05
	DERS Awareness	↓ TINN	<0.05
		↑ LF power nu	<0.05
		↓ HF power	<0.03
		↓ HF power %	<0.01
		↓ HF power nu	<0.05
		↑ LF/HF	<0.05
IRI	IRI fantasy	↓ VLF power	<0.05
		↑ HF power %	<0.05
	IRI EC	↑ RMSSD	<0.01
		↑ Pnn50	<0.01
		↓ LF power nu	<0.03
		↑ HF power	<0.01
		↑ HF power %	<0.03
		↑ HF power nu	<0.03
		↓ LF/HF	<0.03
BIS/BAS	BAS	↓ LF power %	<0.01
ZKPQ	ZKPQ Impuls.S.S.	↓ LF power %	<0.03
	ZKPQ Activity	↓ LF power nu	<0.03
		↑ HF power nu	<0.03
		↓ LF/HF	<0.03

VLF, Very Low Frequency; LF, Low Frequency; HF, High Frequency; nu, normalized units; TINN, width of triangular approximation to NN-interval frequency distribution; RMSSD, square root of mean squared forward differences of successive NN intervals; Pnn50, proportion of successive NN interval differences>50 ms ↑ indicates that an increase of the test score is associated to an increase of the feature value. ↓ indicates that an increase of the test score is associated to a decrease of the feature value.

TABLE 2 | Statistical results related to non-linear features (*U*-test).

Scales	Sub-scales	Statistical results	
		Features	***p*-value**
PANAS	PANAS PA	↓ MSE IBI (long CI)	<0.05
DERS	DERS non-Accept	↓ MSE IBR (short CI)	<0.05
		↓ MSE IBR (longCI)	<0.05
		↓ MMSE HRV-IBI (shortCI)	<0.01
		↓ MMSE HRV-IBI (longCI)	<0.01
	DERS Impulse	↓ 2V	<0.05
	DERS Awareness	↓ MMSE HRV-IBR (longCI)	<0.03
		↓ MMSE HRV-IBR (shortCI)	<0.05
	DERS Strategies	↓ 2V%	<0.05
IRI	IRI EC	↑ CD	<0.01
		↓ SD1	<0.01
		↓ DFA1	<0.05
		↓ MMSE HRV-IBR (shortCI)	<0.03
		↓ MMSE HRV-IBR (longCI)	<0.03
		↑ 1V%	<0.05
BIS/BAS	BIS	↑ CD	<0.03
	BAS	↓ MMSE HRV-IBR (shortCI)	<0.03
		↓ 0V	<0.03
		↓ 0V%	<0.03
ZKPQ	ZKPQ_Activity	↑ ApEn	<0.01
		↑ SampEn	<0.01
		↑ RP Lmax	<0.05

MSE HRV, Multiscale Entropy on HRV series; MSE IBR, Multiscale Entropy on IBR series; MSE IBI, Multiscale Entropy on IBI series; MMSE HRV-IBR, Multivariate Multiscale Entropy on bivariate HRV and IBR series; MMSE HRV-IBI, Multivariate Multiscale Entropy on bivariate HRV and IBI series; CI, Complexity Index. ApEn, Approximate Entropy, SampEn; Sample Entropy, 0V, number of patterns with none variation in the amplitude; 0V%, 1V%, 2V%, percentage of the total patterns with zero, one or two variations in the amplitude; SD1, Standard Deviation of PoincarÃl' Plot related to the points that are perpendicular to the line-of-identity; DFA1, Detrended Fluctuation Analysis (first slope); RP Lmax, Recurrence Plot (longest diagonal line); CD, Correlation Dimension ↑ indicates that an increase of the test score is associated to an increase of the feature value. ↓ indicates that an increase of the test score is associated to a decrease of the feature value.

significant differences considering monovariate and multivariate measures (see details in **Table 2**). An exemplary plot showing the discerning capability of MMSE analysis on DERS Non-Accept sub-scale is shown in **Figure 1**.

To summarize the results, all the extracted features were able to discern the two groups in 12 out of 16 sub-scales. More specifically, standard HRV analysis provided exclusive information, i.e., not overlapped with that coming from the non-linear analysis, on the psychological assessment in only 2 sub-scales, whereas features from ANS non-linear dynamics exclusively discriminated the two groups in 4 sub-scales (see details in **Figure 2**).

3.2. Correlation Analysis

The Spearman correlation coefficient was used to show the relationship between the values of each features through all the subjects and the relative score for each sub-scale. Accordingly, the coefficient ρ and $p - value$ expressing the probability that no correlation between the two variables exist, were assigned for each sub-scale and each feature. Results are shown in **Tables 3**, **4**.

We found that ANS features related to the linear HRV dynamics are significantly correlated with 5 sub-scales, reaching absolute values of ρ up to 0.52 (BAS and ZKPQ Impulsive Sensation Seeking). Moreover, 10 sub-scales are significantly correlated with markers of ANS non-linear dynamics, reaching absolute values of ρ up to 0.55 (DERS Non-Acceptance).

Although the correlation coefficient is not very high, it is, however, a very interesting result to be further validated and confirmed.

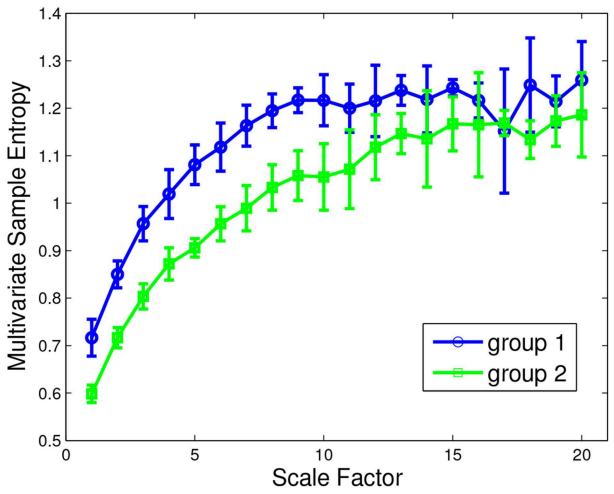


FIGURE 1 | Exemplary plot of Multivariate Multiscale Entropy analysis applied to HRV-IBI series in discerning the two groups (under the median-lower scores: group 1; over the median-higher scores: group 2) according to scores gathered from the DERS Non-Accept sub-scale.

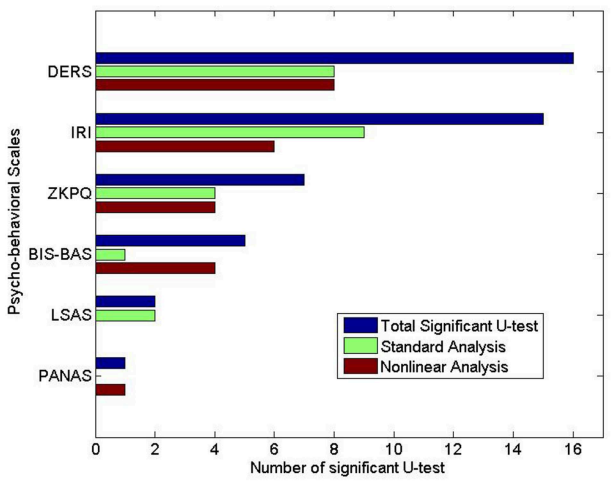


FIGURE 2 | Number of features with significant *p*-values ($p < 0.05$) given by the Mann-Whitney tests. For each phycological scale, the number of significant parameter of standard analysis, non-linear analysis, and the total are shown.

The number of features with significant p-values ($p < 0.05$) given by such a correlation coefficient is shown in **Figure 3** for each phycological dimension.

4. Discussion and Conclusion

In conclusion, we found several ANS biomarkers of psychological dimensions in non-pathological subjects. Such biomarkers are derived from the standard and complexity analysis of ANS measures such as HRV, IBI, and IBR series. We found that dimensions related to difficulties in emotion regulation (DERS), interpersonal reactivity (IRI), behavioral activation or inhibition (BIS/BAS), sensation-seeking and activity (ZKPQ), and anxiety performance (LSAS) are always associated to changes in the HRV dynamics, quantified using time and frequency domain indices (see **Table 1**). As all the scale define different psychological dimensions, it is very difficulty to give a common interpretation of features through them. The LF/HF ratio decrease, associated to increased questionnaires scores, characterizes the ZKPQ activity and IRI empathic concern, whereas an opposite trend is found for the awareness of difficulties in emotion regulation (DERS). HRV time domain indices such as TINN, Pnn50, and RMSSD are effective only to characterize the empathic concern and emotion regulation. These results, gathered from statistical analyses of standard HRV parameters, are further confirmed by the correlation analyses whose details are shown in **Table 3**.

TABLE 3 | Spearman correlation test results related to standard HRV features.

Sub-scales	Features	Spearman test results	
		rho	*p*-value
DERS Awareness	LF power nu	0.41	<0.03
	HF power	−0.41	<0.03
	HF power %	−0.49	<0.01
	HF power nu	−0.43	<0.03
	LF/HF	0.42	<0.03
IRI EC	Pnn50	0.43	<0.03
BAS	LF power %	−0.52	<0.01
ZKPQ Impuls.S.S.	LF power	−0.39	<0.05
	LF power %	0.52	<0.01
ZKPQ Activity	HF peak	0.41	<0.03
	LF power nu	−0.48	<0.01
	HF power %	0.44	<0.03
	HF power nu	0.48	<0.01
	LF/HF	−0.48	<0.01

VLF, Very Low Frequency; LF, Low Frequency; HF, High Frequency, nu, normalized units; TINN, width of triangular approximation to NN-interval frequency distribution; RMSSD, square root of mean squared forward differences of successive NN intervals; Pnn50, proportion of successive NN interval differences > 50 ms.

It is worthwhile noting that the HF power decreases with the DERS score. According to the literature (Porges, 1991, 1992), vagal tone is associated to the ability of emotional self-regulation and high flexibility and adaptability to environmental changes. According to our results, when an emotion dysregulation occurs, the sympathetic activity increases.

Other evidences supporting our results can be found in the current literature (Freeman and Nixon, 1985; Yeragani et al., 1999; Virtanen et al., 2003; Cohen and Benjamin, 2006; Shinba et al., 2008; Licht et al., 2009; Thayer et al., 2010, 2012) which suggest that patients with anxiety disorders revealed a decreased power in the HRV-LF bands.

TABLE 4 | Spearman correlation test results related to non-linear HRV, IBI, IBR features.

Sub-scales	Features	Spearman test results	
		rho	*p*-value
LSAS Anx P	0V %	0.40	<0.05
DERS Non-Accept	MSE IBI (short CI)	−0.45	<0.05
	MSE IBI (long CI)	−0.55	<0.01
DERS Goals	0V	−0.37	<0.05
	0V%	−0.42	<0.03
	2V	−0.39	<0.05
	2V%	−0.37	<0.05
DERS Impulse	MSE IBR (long CI)	−0.43	<0.05
	0V	0.37	<0.05
	0V%	0.39	<0.05
DERS Awareness	DFA1	0.37	<0.05
	MSE IBR (long CI)	0.50	<0.03
DERS Strategies	0V	−0.44	<0.03
	0V%	−0.47	<0.03
	2V%	−0.41	<0.03
IRI EC	CD	0.46	<0.03
	MMSE HRV-IBR (long CI)	−0.47	<0.03
BIS	CD	0.39	<0.05
ZKPQ Impuls.S.S.	MSE HRV (short CI)	−0.44	<0.03
ZKPQ Activity	ApEn	0.48	<0.01
	SampEn	0.52	<0.01
	DFA1	−0.47	<0.01
	RP Lmax,RP DET,RP REC	−0.49	<0.01
	1V%	0.45	<0.03

MSE HRV, Multiscale Entropy on HRV series; MSE IBR, Multiscale Entropy on IBR series; MSE IBI, Multiscale Entropy on IBI series; MMSE HRV-IBR, Multivariate Multiscale Entropy on bivariate HRV and IBR series; MMSE HRV-IBI, Multivariate Multiscale Entropy on bivariate HRV and IBI series; CI, Complexity Index. ApEn, Approximate Entropy; SampEn, Sample Entropy; 0V, number of patterns with none variation in the amplitude; 0V%, 1V%, 2V%, percentage of the total patterns with zero; one or two variations in the amplitude; SD1, Standard Deviation of PoincarÃl' Plot related to the points that are perpendicular to the line-of-identity; DFA1, Detrended Fluctuation Analysis (first slope); RP Lmax, Recurrence Plot (longest diagonal line); RP DET, Recurrence Plot (determinism); RP REC, Recurrence Plot (trend); CD, Correlation Dimension.

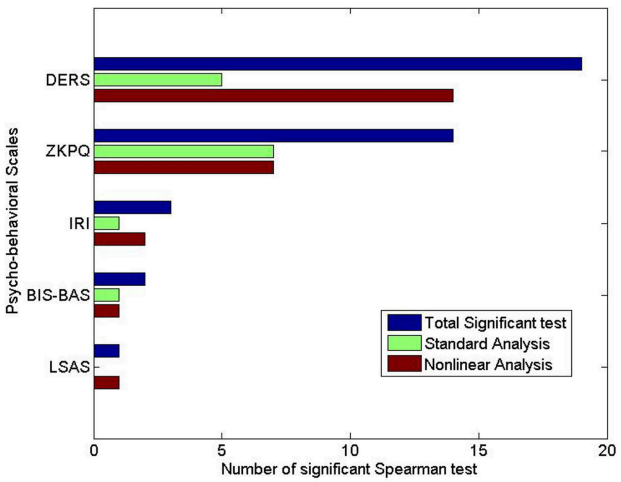


FIGURE 3 | Number of features with significant *p*-values ($p < 0.05$) given by the Spearman non-parametric correlation coefficient. For each phycological scale, the number of significant parameter of standard analysis, non-linear analysis, and the total are shown.

Concerning the ANS non-linear dynamics, several biomarkers of psychological dimensions were found in complexity measures such as sample entropy, monovariate and multivatiate multiscale entropy, short- an long-term correlations, correlation dimension, recurrence and symbolic analysis in characterizing dimensions as positive and negative affect (PANAS), social phobia (Liebowitz Social Anxiety Scale, LSAS), difficulties in emotion regulation (DERS), Interpersonal reactivity (IRI), behavioral inhibition or activation (BIS/BAS), and sensation-seeking and activity (ZKPQ). Our results on non-linear ANS markers for psychological dimensions confirm the previous findings (Yeragani et al., 2000; Cohen and Benjamin, 2006) and provide a wider portrait of the complexity modulation associated with behavioral characters.

Figures 2, **3** report the number of statistically significant features given by Mann-Whitney and Spearman non-parametric correlation, respectively. It is worthy to note that the non-linear features are overall more than those extracted from standard analysis, confirming that complexity dynamics measures play a relevant role in assessing the psycho-physiological dimensions.

Finally, some prudential considerations should be made. The patterns of physiological signals are acquired in rest conditions right after performing the test and the assumption behind the experiment is that the psychological assessment acted as an affective elicitation. Results have to be considered as preliminary to future experiments where subjects experience an actual affective dimension while they are monitored. Nevertheless, it is worthwhile pointing out that complexity measures can be considered promising markers to assess the psychological traits. Is important to underline that such interest is not diminished by the difficulty in giving a physiological meaning to complexity measurements. In this sense and more in generally, we underscored how our data suggested the possibility of an ANV fingerprinting of psychological dimensions. Therefore, beyond their precise physiological meaning, our results have interesting consequences for the psychometric and clinical fields. Our approach may be promising in describing the psychological dimensions as a combination of different features, providing a full classification of psychological characteristics through a baseline ECG acquisition. However, more studies with a much higher number of subjects are needed to test the reliability and the feasibility of these potential clinical

implications. Furthermore, to test if our methodology could also be extended to the extremes of the psychological dimensions, these studies should also include pathological samples (e.g., diagnosed subjects). Should that prove to be the case, this approach might hold promise as a tool for providing an external validation to psychological diagnosis.

Acknowledgments

The research leading to these results has received partial funding from the European Union Seventh Framework Programme FP7/2007Ű2013 under grant agreement n 601165 of the project "WEARHAP."

References

Acharya, U. R., Joseph, K. P., Kannathal, N., Lim, C. M., and Suri, J. S. (2006). Heart rate variability: a review. *Med. Biol. Eng. Comput.* 44, 1031–1051. doi: 10.1007/s11517-006-0119-0

Ahmed, M. U., and Mandic, D. P. (2011). Multivariate multiscale entropy: a tool for complexity analysis of multichannel data. *Phys. Rev. E Stat. Nonlin. Soft. Matter Phys.* 84:061918. doi: 10.1103/PhysRevE.84.061918

Ahmed, M. U., and Mandic, D. P. (2012). Multivariate multiscale entropy analysis. *IEEE Signal Process. Lett.* 19, 91–94. doi: 10.1109/LSP.2011.2180713

Aluja, A., Garcia, O., and Garcia, L. F. (2003). Psychometric properties of the zuckerman–kuhlman personality questionnaire (zkpq-iii-r): a study of a shortened form. *Pers. Individ. Dif.* 34, 1083–1097. doi: 10.1016/S0191-8869(02)00097-1

Atyabi, F., Livari, M., Kaviani, K., and Tabar, M. (2006). Two statistical methods for resolving healthy individuals and those with congestive heart failure based on extended self-similarity and a recursive method. *J. Biol. Phys.* 32, 489–495. doi: 10.1007/s10867-006-9031-y

Baumert, M., Walther, T., Hopfe, J., Stepan, H., Faber, R., and Voss, A. (2002). Joint symbolic dynamic analysis of beat-to-beat interactions of heart rate and systolic blood pressure in normal pregnancy. *Med. Biol. Eng. Comput.* 40, 241–245. doi: 10.1007/BF02348131

Beitel, M., Ferrer, E., and Cecero, J. J. (2005). Psychological mindedness and awareness of self and others. *J. Clin. Psychol.* 61, 739–750. doi: 10.1002/jclp.20095

Bland, J. M., and Altman, D. G. (1997). Statistics notes: Cronbach's alpha. *BMJ* 314:572. doi: 10.1136/bmj.314.7080.572

Calvo, R. A., and D'Mello, S. (2010). Affect detection: an interdisciplinary review of models, methods, and their applications. *IEEE Trans. Affect. Comput.* 1, 18–37. doi: 10.1109/T-AFFC.2010.1

Caminal, P., Giraldo, B., Vallverdú, M., Benito, S., Schroeder, R., and Voss, A. (2010). Symbolic dynamic analysis of relations between cardiac and breathing cycles in patients on weaning trials. *Ann. Biomed. Eng.* 38, 2542–2552. doi: 10.1007/s10439-010-0027-1

Camm, A., Malik, M., Bigger, J., Breithardt, G., Cerutti, S., Cohen, R., et al. (1996). Heart rate variability: standards of measurement, physiological interpretation, and clinical use. *Circulation* 93, 1043–1065. doi: 10.1161/01.CIR.93.5.1043

Carney, R., Freedland, K., Rich, M., and Jaffe, A. (1995). Depression as a risk factor for cardiac events in established coronary heart disease: a review of possible mechanisms. *Ann. Behav. Med.* 17, 142–149. doi: 10.1007/BF02895063

Carney, R. M., Freedland, K. E., and Veith, R. C. (2005). Depression, the autonomic nervous system, and coronary heart disease. *Psychosom. Med.* 67(Suppl. 1), S29–S33. doi: 10.1097/01.psy.0000162254.61556.d5

Carver, C. S., and White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: the bis/bas scales. *J. Pers. Soc. Psychol.* 67:319. doi: 10.1037/0022-3514.67.2.319

Casdagli, M., Eubank, S., Farmer, J. D., and Gibson, J. (1991). State space reconstruction in the presence of noise. *Phys. D* 51, 52–98. doi: 10.1016/0167-2789(91)90222-U

Citi, L., Valenza, G., and Barbieri, R. (2012). "Instantaneous estimation of high-order nonlinear heartbeat dynamics by lyapunov exponents," in *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE* (San Diego, CA: IEEE), 13–16.

Cohen, H., and Benjamin, J. (2006). Power spectrum analysis and cardiovascular morbidity in anxiety disorders. *Auton. Neurosci.* 128, 1–8. doi: 10.1016/j.autneu.2005.06.007

Cohen, R., Swerdlik, M., and Smith, D. (1992). *Psychological Testing and Assessment: An Introduction to Tests and Measurement.* Houston, TX: Mayfield Publishing Co.

Costa, M., Goldberger, A., and Peng, C. (2002). Multiscale entropy analysis of complex physiologic time series. *Phys. Rev. Lett.* 89:68102. doi: 10.1103/PhysRevLett.89.068102

Costa, M., Peng, C.-K., Goldberger, A.L., and Hausdorff, J. M. (2003). Multiscale entropy analysis of human gait dynamics. *Phys. A* 330, 53–60. doi: 10.1016/j.physa.2003.08.022

Costa, M., Priplata, A., Lipsitz, L., Wu, Z., Huang, N., Goldberger, A., et al. (2007). Noise and poise: enhancement of postural complexity in the elderly with a stochastic-resonance–based therapy. *Europhys. Lett.* 77:68008. doi: 10.1209/0295-5075/77/68008

Crawford, J. R., and Henry, J. D. (2004). The positive and negative affect schedule (panas): construct validity, measurement properties and normative data in a large non-clinical sample. *Br. J. Clin. Psychol.* 43, 245–265. doi: 10.1348/0144665031752934

Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. *JSAS Catal. Sel. Doc. Psychol.* 10:85.

Ehring, T., Fischer, S., Schnülle, J., Bösterling, A., and Tuschen-Caffier, B. (2008). Characteristics of emotion regulation in recovered depressed versus never depressed individuals. *Pers. Individ. Dif.* 44, 1574–1584. doi: 10.1016/j.paid.2008.01.013

Fernandez, A., Hornero, R., Gomez, C., Turrero, A., Gil-Gregorio, P., Matías-Santos, J., et al. (2010). Complexity analysis of spontaneous brain activity in alzheimer disease and mild cognitive impairment: an meg study. *Alzheimer Dis. Assoc. Disord.* 24, 182–189. doi: 10.1097/WAD.0b013e3181c727f7

Freeman, L. J., and Nixon, P. (1985). Are coronary artery spasm and progressive damage to the heart associated with the hyperventilation syndrome? *Br. Med. J. (Clin. Res. Ed.)* 291:851. doi: 10.1136/bmj.291.6499.851

Fresco, D. M., Coles, M. E., Heimberg, R. G., Liebowitz, M. R., Hami, S., and Stein, M. B. (2001). The liebowitz social anxiety scale: a comparison of the psychometric properties of self-report and clinician-administered formats. *Psychol. Med.* 31, 1025–1035. doi: 10.1017/S0033291701004056

Fusheng, Y., Bo, H., and Qingyu, T. (2001). "Approximate entropy and its application to biosignal analysis," in *Nonlinear Biomedical Signal Processing: Dynamic Analysis and Modeling*, Vol. 2, ed M. Akay (Hoboken, NJ: JohnWiley & Sons Inc), 72–91.

Gao, J., Gurbaxani, B. M., Hu, J., Heilman, K. J., Emanuele, II. V. A., Lewis, G. F., et al. (2013). Multiscale analysis of heart rate variability in non-stationary environments. *Front. Physiol.* 4:119. doi: 10.3389/fphys.2013.00119

Gao, J., Hu, J., and Tung, W.-w. (2011). Complexity measures of brain wave dynamics. *Cogn. Neurodynam.* 5, 171–182. doi: 10.1007/s11571-011-9151-3

Glass, L. (2001). Synchronization and rhythmic processes in physiology. *Nature* 410, 277–284. doi: 10.1038/35065745

Glass, L. (2009). Introduction to controversial topics in nonlinear science: is the normal heart rate chaotic? *Chaos* 19, 028501. doi: 10.1063/1.3156832

Glassman, A. (1998). Depression, cardiac death, and the central nervous system. *Neuropsychobiology* 37, 80–83. doi: 10.1159/000026482

Goldberger, A., Peng, C., and Lipsitz, L. (2002). What is physiologic complexity and how does it change with aging and disease? *Neurobiol. Aging* 23, 23–26. doi: 10.1016/S0197-4580(01)00266-4

Grassberger, P. (1988). Finite sample corrections to entropy and dimension estimates. *Phys. Lett. A* 128, 369–373. doi: 10.1016/0375-9601(88)90193-4

Grassberger, P., and Procaccia, I. (1983). Estimation of the kolmogorov entropy from a chaotic signal. *Phys. Rev. A* 28, 2591–2593. doi: 10.1103/PhysRevA.28.2591

Gratz, K. L., and Roemer, L. (2004). Multidimensional assessment of emotion regulation and dysregulation: development, factor structure, and initial validation of the difficulties in emotion regulation scale. *J. Psychopathol. Behav. Assess.* 26, 41–54. doi: 10.1023/B:JOBA.0000007455.08539.94

Grossman, P., Wilhelm, F. H., Kawachi, I., and Sparrow, D. (2001). Gender differences in psychophysiological responses to speech stress among older social phobics congruence and incongruence between self-evaluative and cardiovascular reactions. *Psychosom. Med.* 63, 765–777. doi: 10.1097/00006842-200109000-00010

Groth-Marnat, G. (2003). *Handbook of Psychological Assessment*. Hoboken; New York: John Wiley & Sons Inc.

Guzzetti, S., Borroni, E., Garbelli, P. E., Ceriani, E., Della Bella, P., Montano, N., et al. (2005). Symbolic dynamics of heart rate variability a probe to investigate cardiac autonomic modulation. *Circulation* 112, 465–470. doi: 10.1161/CIRCULATIONAHA.104.518449

Heimberg, R. G., Horner, K., Juster, H., Safren, S., Brown, E., Schneier, F., et al. (1999). Psychometric properties of the liebowitz social anxiety scale. *Psychol. Med.* 29, 199–212. doi: 10.1017/S0033291798007879

Heller, A. S., Johnstone, T., Shackman, A. J., Light, S. N., Peterson, M. J., Kolden, G. G., et al. (2009). Reduced capacity to sustain positive emotion in major depression reflects diminished maintenance of fronto-striatal brain activation. *Proc. Natl. Acad. Sci. U.S.A.* 106, 22445–22450. doi: 10.1073/pnas.0910651106

Hu, J., Gao, J., and Príncipe, J. C. (2006). Analysis of biomedical signals by the lempel-ziv complexity: the effect of finite data size. *IEEE Trans. Biomed. Eng.* 53, 2606–2609. doi: 10.1109/TBME.2006.883825

Hu, J., Gao, J., and Tung, W.-W. (2009). Characterizing heart rate variability by scale-dependent lyapunov exponent. *Chaos* 19, 028506. doi: 10.1063/1.3152007

Hu, J., Gao, J., Tung, W.-W., and Cao, Y. (2010). Multiscale analysis of heart rate variability: a comparison of different complexity measures. *Ann. Biomed. Eng.* 38, 854–864. doi: 10.1007/s10439-009-9863-2

Hunsley, J., and Mash, E. J. (2010). "The role of assessment in evidence-based practice," in *Handbook of Assessment and Treatment Planning for Psychological Disorders, 2nd Edn.*, eds M. M. Antony and D. H. Barlow (New York, NY: Guilford Press), 3–22.

Iverson, G., Gaetz, M., Rzempoluck, E., McLean, P., Linden, W., and Remick, R. (2005). A new potential marker for abnormal cardiac physiology in depression. *J. Behav. Med.* 28, 507–511. doi: 10.1007/s10439-009-9863-2

Iverson, G., Stampfer, H., and Gaetz, M. (2002). Reliability of circadian heart pattern analysis in psychiatry. *Psychiatr. Q.* 73, 195–203. doi: 10.1023/A:1016036704524

Jansen, L. M., Gispen-de Wied, C. C., Wiegant, V. M., Westenberg, H. G., Lahuis, B. E., and van Engeland, H. (2006). Autonomic and neuroendocrine responses to a psychosocial stressor in adults with autistic spectrum disorder. *J. Autism Dev. Disord.* 36, 891–899. doi: 10.1007/s10803-006-0124-z

Jorm, A. F., Christensen, H., Henderson, A. S., Jacomb, P. A., Korten, A. E., and Rodgers, B. (1998). Using the bis/bas scales to measure behavioural inhibition and behavioural activation: factor structure, validity and norms in a large community sample. *Pers. Individ. Dif.* 26, 49–58. doi: 10.1016/S0191-8869(98)00143-3

Kenny, M. C., Alvarez, K., Donohue, B. C., and Winick, C. B. (2008). "Overview of behavioral assessment with adults," in *Handbook of Psychological Assessment, Case Conceptualization, and Treatment, Adults*, Vol. 1, eds M. Hersen and J. Rosqvist (Hoboken, NJ: John Wiley & Sons Inc).

Lake, D. E., Richman, J. S., Griffin, M. P., and Moorman, J. R. (2002). Sample entropy analysis of neonatal heart rate variability. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* 283, R789–R797. doi: 10.1152/ajpregu.00069.2002

Lanatà, A., Valenza, G., and Scilingo, E. P. (2012). A novel eda glove based on textile-integrated electrodes for affective computing. *Med. Biol. Eng. Comput.* 50, 1163–1172. doi: 10.1007/s11517-012-0921-9

Leistedt, S. J., Linkowski, P., Lanquart, J., Mietus, J., Davis, R. B., Goldberger, A. L., et al. (2011). Decreased neuroautonomic complexity in men during an acute major depressive episode: analysis of heart rate dynamics. *Transl. Psychiatry* 1, e27. doi: 10.1038/tp.2011.23

Licht, C. M., de Geus, E. J., van Dyck, R., and Penninx, B. W. (2009). Association between anxiety disorders and heart rate variability in the netherlands study of depression and anxiety (nesda). *Psychosom. Med.* 71, 508–518. doi: 10.1097/PSY.0b013e3181a292a6

Liebowitz, M. R. (1987). Social phobia. *Mod. Probl. Pharmacopsychiatry* 22, 141–173.

Lin, Y., Wang, C., Jung, T., Wu, T., Jeng, S., Duann, J., et al. (2010). EEG-based emotion recognition in music listening. *IEEE Trans. Biomed. Eng.* 57, 1798–1806. doi: 10.1109/TBME.2010.2048568

Marmarelis, V. Z. (2004). *Nonlinear Dynamic Modeling of Physiological Systems*. New York, NY: Wiley-Interscience.

Marwan, N., Carmen Romano, M., Thiel, M., and Kurths, J. (2007). Recurrence plots for the analysis of complex systems. *Phys. Rep.* 438, 237–329. doi: 10.1016/j.physrep.2006.11.001

Marwan, N., Wessel, N., Meyerfeldt, U., Schirdewan, A., and Kurths, J. (2002). Recurrence-plot-based measures of complexity and their application to heart-rate-variability data. *Phys. Rev. E* 66:026702. doi: 10.1103/PhysRevE.66.026702

Mujica-Parodi, L., Yeragani, V., and Malaspina, D. (2005). Nonlinear complexity and spectral analyses of heart rate variability in medicated and unmedicated patients with schizophrenia1. *Neuropsychobiology* 51, 10–15. doi: 10.1159/000082850

Peng, C., Havlin, S., Stanley, H., and Goldberger, A. (1995). Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series. *Chaos* 5:82. doi: 10.1063/1.166141

Penzel, T., Kantelhardt, J. W., Grote, L., Peter, J.-H., and Bunde, A. (2003). Comparison of detrended fluctuation analysis and spectral analysis for heart rate variability in sleep and sleep apnea. *IEEE Trans. Biomed. Eng.* 50, 1143–1151. doi: 10.1109/TBME.2003.817636

Petrantonakis, P. C., and Hadjileontiadis, L. J. (2011). A novel emotion elicitation index using frontal brain asymmetry for enhanced eeg-based emotion recognition. *IEEE Trans. Inf. Technol. Biomed.* 15, 737–746. doi: 10.1109/TITB.2011.2157933

Pincus, S. (1991). Approximate entropy as a measure of system complexity. *Proc. Natl. Acad. Sci. U.S.A.* 88:2297. doi: 10.1073/pnas.88.6.2297

Poon, C., and Merrill, C. (1997). Decrease of cardiac chaos in congestive heart failure. *Nature* 389, 492–495. doi: 10.1038/39043

Porges, S. W. (1991). *Vagal Tone: An Autonomic Mediator of Affect*. Cambridge, UK: Cambridge University Press.

Porges, S. W. (1992). "Autonomic regulation and attention," in *Attention and Information Processing in Infants and Adults: Perspectives from Human and Animal Research*, eds B. A. Campbell, H. Hayne, and R. Richardson (Hillsdale, NJ: Erlbaum), 201–223.

Porta, A., Guzzetti, S., Montano, N., Furlan, R., Pagani, M., Malliani, A., et al. (2001). Entropy, entropy rate, and pattern classification as tools to typify complexity in short heart period variability series. *IEEE Trans. Biomed. Eng.* 48, 1282–1291. doi: 10.1109/10.959324

Richman, J., and Moorman, J. (2000). Physiological time-series analysis using approximate entropy and sample entropy. *Am. J. Physiol. Heart Circ. Physiol.* 278, H2039–H2049.

Ruiz-Padial, E., Vila, J., and Thayer, J. F. (2011). The effect of conscious and non-conscious presentation of biologically relevant emotion pictures on emotion modulated startle and phasic heart rate. *Int. J. Psychophysiol* 79, 341–346. doi: 10.1016/j.ijpsycho.2010.12.001

Rytwinski, N. K., Fresco, D. M., Heimberg, R. G., Coles, M. E., Liebowitz, M. R., Cissell, S., et al. (2009). Screening for social anxiety disorder with the self-report version of the liebowitz social anxiety scale. *Depress. Anxiety* 26, 34–38. doi: 10.1002/da.20503

Sârbescu, P., and Neguţ, A. (2012). Psychometric properties of the romanian version of the Zuckerman-Kuhlman personality questionnaire. *Eur. J. Psychol. Assess.* 29, 241–252. doi: 10.1027/1015-5759/a000152

Sava, F. A., and Sperneac, A.-M. (2006). Sensitivity to reward and sensitivity to punishment rating scales: a validation study on the romanian population. *Pers. Individ. Dif.* 41, 1445–1456. doi: 10.1016/j.paid.2006.04.024

Schouten, J. C., Takens, F., and van den Bleek, C. M. (1994). Estimation of the dimension of a noisy attractor. *Phys. Rev. E* 50:1851. doi: 10.1103/PhysRevE.50.1851

Shinba, T., Kariya, N., Matsui, Y., Ozawa, N., Matsuda, Y., and Yamamoto, K.-I. (2008). Decrease in heart rate variability response to task is related to anxiety and depressiveness in normal subjects. *Psychiatry Clin. Neurosci.* 62, 603–609. doi: 10.1111/j.1440-1819.2008.01855.x

Stampfer, H. (1998). The relationship between psychiatric illness and the circadian pattern of heart rate. *Aust. Psychiatry* 32, 187–198. doi: 10.3109/00048679809062728

Stiedl, O., and Meyer, M. (2003). Fractal dynamics in circadian cardiac time series of corticotropin-releasing factor receptor subtype-2 deficient mice. *J. Math. Biol* 47, 169–197. doi: 10.1007/s00285-003-0197-7

Taillard, J., Lemoine, P., Boule, P., Drogue, M., and Mouret, J. (1993). Sleep and heart rate circadian rhythm in depression: the necessity to separate. *Chronobiol. Int.* 10, 63–72. doi: 10.3109/07420529309064483

Taillard, J., Sanchez, P., Lemoine, P., and Mouret, J. (1990). Heart rate circadian rhythm as a biological marker of desynchronization in major depression: a methodological and preliminary report. *Chronobiol. Int.* 7, 305–316. doi: 10.3109/07420529009064636

Takahashi, T., Cho, R. Y., Mizuno, T., Kikuchi, M., Murata, T., Takahashi, K., et al. (2010). Antipsychotics reverse abnormal eeg complexity in drug-naive schizophrenia: a multiscale entropy analysis. *Neuroimage* 51, 173–182. doi: 10.1016/j.neuroimage.2010.02.009

Takens, F. (1981). "Detecting strange attractors in turbulence," in *Dynamical Systems and Turbulence, Warwick 1980*, eds D. A. Rand and L.-S. Young (Warwick: Springer), 366–381.

Thayer, J. F., Åhs, F., Fredrikson, M., Sollers, J. J. III, and Wager, T. D. (2012). A meta-analysis of heart rate variability and neuroimaging studies: implications for heart rate variability as a marker of stress and health. *Neurosci. Biobehav. Rev.* 36, 747–756. doi: 10.1016/j.neubiorev.2011.11.009

Thayer, J. F., Friedman, B. H., and Borkovec, T. D. (1996). Autonomic characteristics of generalized anxiety disorder and worry. *Biol. Psychiatry* 39, 255–266. doi: 10.1016/0006-3223(95)00136-0

Thayer, J. F., Yamamoto, S. S., and Brosschot, J. F. (2010). The relationship of autonomic imbalance, heart rate variability and cardiovascular disease risk factors. *Int. J. Cardiol.* 141, 122–131. doi: 10.1016/j.ijcard.2009.09.543

Theiler, J. (1987). Efficient algorithm for estimating the correlation dimension from a set of discrete points. *Phys. Rev. A* 36:4456. doi: 10.1103/PhysRevA.36.4456

Tobaldini, E., Porta, A., Wei, S.-G., Zhang, Z.-H., Francis, J., Casali, K. R., et al. (2009). Symbolic analysis detects alterations of cardiac autonomic modulation in congestive heart failure rats. *Auton. Neurosci.* 150, 21–26. doi: 10.1016/j.autneu.2009.03.009

Tulppo, M., Kiviniemi, A., Hautala, A., Kallio, M., Seppanen, T., Makikallio, T., et al. (2005). Physiological background of the loss of fractal heart rate dynamics. *Circulation* 112, 314. doi: 10.1161/CIRCULATIONAHA.104.523712

Valenza, G., Allegrini, P., Lanatà, A., and Scilingo, E. P. (2012a). Dominant lyapunov exponent and approximate entropy in heart rate variability during emotional visual elicitation. *Front. Neuroeng.* 5:3. doi: 10.3389/fneng.2012.00003

Valenza, G., Citi, L., Gentili, C., Lanatá, A., Scilingo, E., and Barbieri, R. (2014a). Point-process nonlinear autonomic assessment of depressive states in bipolar patients. *Methods Inf. Med.* 53, 296–302. doi: 10.3414/ME13-02-0036

Valenza, G., Gentili, C., Lanatà, A., and Scilingo, E. P. (2013a). Mood recognition in bipolar patients through the psyche platform: preliminary evaluations and perspectives. *Artif. Intell. Med.* 57, 49–58. doi: 10.1016/j.artmed.2012.12.001

Valenza, G., Lanatà, A., Ferro, M., and Scilingo, E. P. (2008). "Real-time discrimination of multiple cardiac arrhythmias for wearable systems based on neural networks," in *Computers in Cardiology, 2008* (Bologna: IEEE), 1053–1056.

Valenza, G., Lanata, A., and Scilingo, E. P. (2012b). The role of nonlinear dynamics in affective valence and arousal recognition. *IEEE Trans. Affect. Comput.* 3, 237–249. doi: 10.1109/T-AFFC.2011.30

Valenza, G., Lanatá, A., and Scilingo, E. P. (2013b). Improving emotion recognition systems by embedding cardiorespiratory coupling. *Physiol. Meas.* 34:449. doi: 10.1088/0967-3334/34/4/449

Valenza, G., Nardelli, M., Bertschy, G., Lanata, A., and Scilingo, E. (2014b). Mood states modulate complexity in heartbeat dynamics: a multiscale entropy analysis. *Europhys. Lett.* 107:18003. doi: 10.1209/0295-5075/107/18003

Valenza, G., Nardelli, M., Lanata, A., Gentili, C., Bertschy, G., Paradiso, R., et al. (2014c). Wearable monitoring for mood recognition in bipolar disorder based on history-dependent long-term heart rate variability analysis. *IEEE J. Biomed. Health Inform.* 18, 1625–1635. doi: 10.1109/JBHI.2013.2290382

Virtanen, R., Jula, A., Salminen, J. K., Voipio-Pulkki, L.-M., Helenius, H., Kuusela, T., et al. (2003). Anxiety and hostility are associated with reduced baroreflex sensitivity and increased beat-to-beat blood pressure variability. *Psychosom. Med.* 65, 751–756. doi: 10.1097/01.PSY.0000088760.65046.CF

Watkins, L., Blumenthal, J., and Carney, R. (2002). Association of anxiety with reduced baroreflex cardiac control in patients after acute myocardial infarction. *Am. Heart J.* 143, 460–466. doi: 10.1067/mhj.2002.120404

Watson, D., Clark, L. A., and Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the panas scales. *J. Pers. Soc. Psychol.* 54:1063. doi: 10.1037/0022-3514.54.6.1063

Wu, G., Arzeno, N., Shen, L., Tang, D., Zheng, D., Zhao, N., et al. (2009). Chaotic signatures of heart rate variability and its power spectrum in health, aging and heart failure. *PLoS ONE* 4:e4323. doi: 10.1371/journal.pone.0004323

Yang, A. C., and Tsai, S.-J. (2012). Is mental illness complex? from behavior to brain. *Prog. Neuropsychopharmacol. Biol. Psychiatry* 45, 253–257. doi: 10.1016/j.pnpbp.2012.09.015

Yeragani, V. K., Jampala, V., Sobelewski, E., Kay, J., and Igel, G. (1999). Effects of paroxetine on heart period variability in patients with panic disorder: a study of holter ecg records. *Neuropsychobiology* 40, 124–128. doi: 10.1159/000026608

Yeragani, V. K., Nadella, R., Hinze, B., Yeragani, S., and Jampala, V. (2000). Nonlinear measures of heart period variability: decreased measures of symbolic dynamics in patients with panic disorder. *Depress. Anxiety* 12, 67–77. doi: 10.1002/1520-6394(2000)12:2<67::AID-DA2>3.0.CO;2-C

Zbilut, J. P., Koebbe, M., Loeb, H., and Mayer-Kress, G. (1990). "Use of recurrence plots in the analysis of heart beat intervals," in *Proceedings of Computers in Cardiology 1990* (Chicago, IL: IEEE), 263–266.

Zuckerman, M., Kuhlman, D. M., Joireman, J., Teta, P., and Kraft, M. (1993). A comparison of three structural models for personality: the big three, the big five, and the alternative five. *J. Pers. Soc. Psychol.* 65:757. doi: 10.1037/0022-3514.65.4.757

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.

frontiers in
COMPUTATIONAL NEUROSCIENCE

GENERAL COMMENTARY
published: 12 August 2013
doi: 10.3389/fncom.2013.00106

What is the mathematical description of the treated mood pattern in bipolar disorder?

Fatemeh Hadaeghi*, Mohammad R. Hashemi Golpayegani and Shahriar Gharibzadeh

Biomedical Engineering Faculty, Amirkabir University of Technology, Tehran, Iran
**Correspondence: f_hadaeghi@aut.ac.ir*

Edited by:
Tobias A. Mattei, Ohio State University, USA

A commentary on

Mathematical models of bipolar disorder
by Daugherty, D., Roque-Urrea, T., Urrea-Roque, J., Troyer, J., Wirkus, S., and Porter, M. A. (2009). Commun. Nonlinear Sci. Numer. Simulat. 14, 2897–2908.

In their innovative article, Daugherty et al. (2009) have modeled the mood swings of a patient with bipolar disorder as a Liénard oscillator with autonomous forcing. They proposed that emotional state of untreated and treated bipolar type-II patient could be mathematically represented by the Equation (1), in which $x(t)$, represents emotional state in time t. In this equation, by adjusting the parameter ρ, both treated and untreated person could be modeled.

$$\ddot{x} - 0.38\dot{x} + 180x = \rho\dot{x}^3 + \mu\dot{x}^5 - \nu\dot{x}^{11} \quad (1)$$

The phase space of Equation (1) which is shown in **Figure 1B**, includes an unstable limit cycle encircled by a large stable limit cycle. The authors have supposed that after treatment, the smaller stable limit cycle with sufficiently small amplitude would correspond to the ultimate emotional pattern to be achieved.

Nevertheless, we believe with basis of previous studies (Gottschalk et al., 1995; Huber et al., 1999) that both in normal persons and treated patients, mood variations and emotional states do not exhibit such a periodic pattern (After 300 months in **Figure 1A**) and could be better described by a low amplitude chaotic time series. Some of our evidences for this supposition are: (1) the spatial complexity of brain components. In the brain, there are a large number of interacting neurons connected by synapses and interacting networks connected functionally or structurally. As already demonstrated in studies in complex systems, the existence of multiple and interdependent connections acting in complex positive and negative feedback loops is very likely to lead to apparently random and unpredictable states (Korn and Faure, 2003). This unpredictability is a fundamental feature of chaotic patterns. (2) The temporal complexity of brain behavior. Besides the complex structural pattern in the brain, recordings from nerve cells as well as electroencephalograms have showed the chaotic temporal function of the brain in its interaction with the environment (Korn and Faure, 2003; Rabinovich et al., 2012).

In the case of mood as a state of the mind, therefore, it can be expected that mood variation in normal individuals would be more complex rather than being ordered. In addition, the environment is in constant modification and therefore, expecting that it would generate standard and fixed emotional states or moods in such a periodic manner seems to be quite unrealistic. Indeed, in the case of bipolar disorder, it has already been demonstrated that we are dealing with an intermittent behavior (Gottschalk et al., 1995) which can be simplified to a stable periodic pattern, in contrast with the highly chaotic patterns in normal individuals. Therefore, we believe that in treated patients, it would not be adequate to reach a state with periodic oscillation with low amplitude. In fact, in abnormal states, as changes in the complexity of brain dynamics occur, therapeutic strategies would attempt to compensate these changes (Bahrami et al., 2005; Mendez et al., 2012).

Based on the above-mentioned view, we propose to modify the aforementioned model by inserting a time dependent term which reflects the momentary interactions of brain with time varying environment as well as interpersonal relationship. The proposed equation for untreated person could be considered as follows in which $\rho = -0.03302$, $\mu = 0.078$, $\nu = 0.00093$, and $\eta = 0.1$.

$$\ddot{x} - 0.038\dot{x} + 0.180x = \rho\dot{x}^3 + \mu\dot{x}^5 - \nu\dot{x}^{11} - \eta x^3 \quad (2)$$

The effect of treatments could be inserted through a sinusoidal function which results to Equation (3).

$$\ddot{x} - 0.038\dot{x} + 0.180x = \rho\dot{x}^3 + \mu\dot{x}^5 - \nu\dot{x}^{11} - \eta x^3 + q\cos(\omega t) \quad (3)$$

Changing the parameters of this equation, especially, ω, q, and, η, would yield diverse patterns such as periodic, quasi-periodic, chaotic, and intermittent behaviors. Considering $\eta = 1$, $\omega = 2$, and $q = 1.2$ the Equation (3) has a chaotic solution. In order to provide a deeper insight in to such dynamics, we represent this time series and the chaotic attractor in phase plane in **Figures 1C,D**. In such example, we present a mathematical representation of an untreated 20-year-old patient Equation (2) as well as the effects of treatment, which is represented by Equation (3). In phase space portrait (**Figure 1D**), a small amplitude stable chaotic attractor which is encircled by the large unstable periodic orbit (not shown in the figure) represents the desired attractor of emotional state for treated person.

It is obvious that our modified model can represent both rhythmic pattern of mood variation in patients and the complex pattern of mood states in treated subjects. Additionally, our equation seems to be more consistent with observed evidences from empirical studies because its adjustable parameters could reflect the effect of therapeutic strategies (Huber et al., 1999); however, theoretically, the

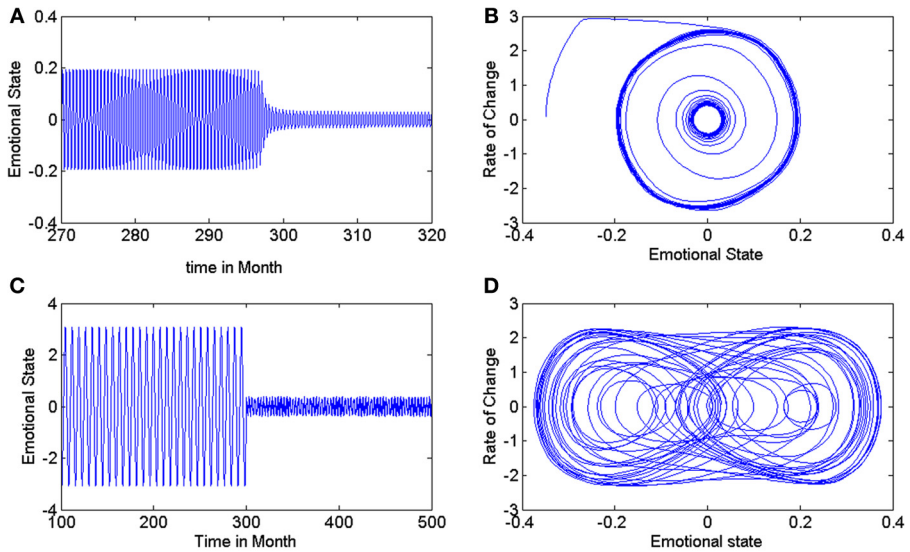


FIGURE 1 | (A,B) Time series of mood and phase space of treated patient in model of Equation (1). It has been supposed that smaller stable limit cycle with small amplitude is the desired emotional pattern of the patient after treatment (Daugherty et al., 2009). **(C)** Time series of mood pattern in modified model, before and after treatment. **(D)** Bounded chaotic attractor as a representation of relative variations in emotional state and the rate of its changes in a treated patient using modified model.

occurrence of a tangent bifurcation in the equation by change in one of the parameters would be required in order to transit from a periodic pattern to a chaotic behavior. The exact meaning of such event in clinical terms still remains to be elucidated in future studies.

Finally, it is important to emphasize that, ultimately, the validity of all these theoretical models and predictions will rely on empirical studies employing qualitative analysis of self-rated mood records (life charts) based on psychological tests or using complexity measures extracted from functional test time series such EEG, fMRI, or PET scan.

REFERENCES

Bahrami, B., Seyedsadjadi, R., Babadi, B., and Noroozian, M. (2005). Brain complexity increases in mania.*Neuroreport* 16, 187–191. doi: 10.1097/00001756-200502080-00025

Daugherty, D., Roque-Urrea, T., Urrea-Roque, J., Troyer, J., Wirkus, S., and Porter, M. A. (2009). Mathematical models of bipolar disorder. *Commun. Nonlin. Sci. Numer. Simulat.* 14, 2897–2908. doi: 10.1016/j.cnsns.2008.10.027

Gottschalk, A., Bauer, M. S., and Whybrow, P. C. (1995). Evidence of chaotic mood variation in bipolar disorder. *Arch. Gen. Psychiatry* 52, 947–959. doi: 10.1001/archpsyc.1995.03950230061009

Huber, M. T., Braun, H. A., and Krieg, J. C. (1999). Consequences of deterministic and random dynamics for the course of affective disorders. *Biol. Psychiatry* 46, 256–262. doi: 10.1016/S0006-3223(98)00311-4

Korn, H., and Faure, P. (2003). Is there chaos in the brain. II. Experimental evidence and related models. *C. R. Biol.* 326, 1210–1213.

Mendez, M. A., Zuluaga, P., Hornero, R., Gomez, C., Escudero, J., Rodriquez-Palancas, A., et al. (2012). Complexity analysis of spontaneous brain activity: effects of depression and antidepressant treatment. *J. Psychopharmacol.* 26, 636–643. doi: 10.1177/0269881111408966

Rabinovich, M. I.., Afraimovich, V. S., Bick, C., and Varona, P. (2012). Information flow dynamics in the brain. *Phys. Life Rev*. 123C, 76–84.

Received: 18 July 2013; accepted: 19 July 2013; published online: 12 August 2013.
Citation: Hadaeghi F, Hashemi Golpayegani MR and Gharibzadeh S (2013) What is the mathematical description of the treated mood pattern in bipolar disorder? Front. Comput. Neurosci. 7:106. doi: 10.3389/fncom.2013.00106

frontiers in
COMPUTATIONAL NEUROSCIENCE

OPINION ARTICLE
published: 23 August 2013
doi: 10.3389/fncom.2013.00116

Does "crisis-induced intermittency" explain bipolar disorder dynamics?

Fatemeh Hadaeghi*, Mohammad R. Hashemi Golpayegani and Keivan Moradi

Biomedical Engineering Faculty, Amirkabir University of Technology, Tehran, Iran
**Correspondence: f_hadaeghi@aut.ac.ir*

Edited by:
Tobias A. Mattei, Ohio State University, USA

Keywords: bipolar disorder, chaos theory, crisis, intermittency, mathematical modeling, multisatability, strange attractor

The brain presents a large number of spatially connected and interacting neurons and synapses that form many positive and negative feedback circuits. These complex networks in interaction with the environment have been experimentally demonstrated to produce temporally chaotic behavior which may be detected in recordings from individual nerve cells or neural ensembles (Korn and Faure, 2003). According to such paradigm, the brain could be considered as a complex system with chaos as its predominant dynamics. As a result, concepts of complex system and chaos theory could be applied to the studies of normal and abnormal brain functions.

One of the fundamental features of some complex systems is "multistability," which can be understood as the coexistence of several interacting attractors (Chian et al., 2006). These interactions results in various complex behaviors in the long term dynamics of the system. Previous studies in several research areas, including neuroscience, have already reported the existence of multistability in natural systems (Chian et al., 2006; Goldbeter, 2011; Rabinovich et al., 2012).

From the perspective of chaos theory, irregular alternation between episodes of various forms of chaotic or periodic behaviors is known as "intermittency" (Tanaka et al., 2005; Chian et al., 2006). In a "global bifurcation," an "attractor-merging crisis" could yield to intermittent behavior. This crisis occurs through the collision of two or more attractors with the boundaries of the basin of the attraction of other attractors (Tanaka et al., 2005; Chian et al., 2006). In this case, by crossing the boundary, the trajectory of the system would be attracted by the other attractor. Such trajectory would then, remain there until another crossing which may lead to a returning to the first attractor. Chaotic intermittency has been reported in circuit oscillators, economic variables, non-periodic associative dynamics in chaotic neural networks as well as in psychiatric disorders like obsessive–compulsive disorder (Tanaka et al., 2005; Chian et al., 2006; Rabinovich and Varona, 2011). However, we believe that such concept also could be applied to mood variation pattern in bipolar disorder.

According to physiological studies, neuroplastic variations may be the underlying mechanism which explain the misregulation of the main circuits involved in the emotional processing (Kandel et al., 2000; Berns and Nemeroff, 2003). This emotional dysregulation is somatically represented as irregular mood swings. Therefore, we believe that the clinical course of bipolar disorder, which is characterized by repeated erratic cycles of mania, depression and episodes of randomly appeared chaotic transitional states (Gottschalk et al., 1995; Berns and Nemeroff, 2003; Rabinovich et al., 2012), may also be understood based on the concept of chaotic intermittency. Manic, depressive and transitional states could be considered as stable or unstable attractors of a dynamical system through which the mood trajectory moves. Therefore, such accidental and abrupt changes of the mood state in bipolar disorder can result from the collision of the initial mood trajectory with the boundary of the basin of the attraction of the another mood attractors. According to chaos theory, this intermittent behavioral pattern could be considered as "crisis-induced intermittency." Following such viewpoint, in healthy subjects, there would be only one "strange attractor" related to the mood states. Time series of such strange attractor represents both positive and negative emotions, unpredictably and in response to internal (for example thought, attention and memory) or external (environment) stimulus. In a bipolar person, however, initial emotional trigger of disease results in a type of "exterior crisis" in the system, in which the destruction of strange attractor is accompanied with formation of two abnormal attractors (mania and depression) and chaotic transients between them.

In order to model such scenario, models of chaotic systems which demonstrate various kind of crisis by changing their parameters (such as "forced Duffing" oscillator and "Ikeda" iterated map), could be utilized to characterize the basic

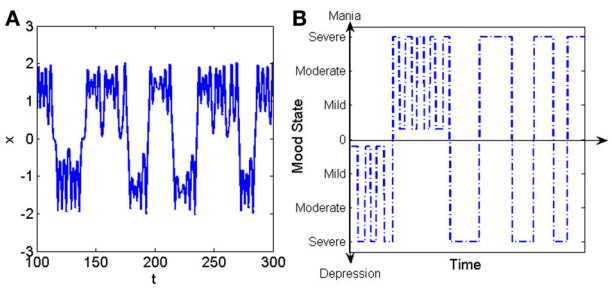


FIGURE 1 | (A) Example of crisis induced intermittency in the forced Duffing oscillator. **(B)** Example of temporal pattern of mood variation in a patient with bipolar disorder (Tretter et al., 2011).

features of human emotional states, when they are presenting multistable and intermittent behaviors, as in the case of bipolar disorder. In order to provide a deeper insight in to such dynamics, we represent the time series of forced Duffing oscillator in its crisis-induced intermittent mode in **Figure 1A** and an example of temporal pattern of self-rated mood records (life charts) in a person with bipolar disorder in **Figure 1B**. The proposed theoretical model would be useful in order to predict the evolution of such emotional states in bipolar disorder and to investigate the effects of psychopharmacological therapies. The experimental data for such investigations would most likely come from psychological tests, life chart recordings, or functional studies, such as EEG, fMRI, or PET-scan.

REFERENCES

Berns, G. S., and Nemeroff, C. B. (2003). The neurobiology of bipolar disorder. *Am. J. Med. Genet. C Semin. Med. Genet.* 123C, 76–84.

Chian, A. C.-L., Rempel, E. L., and Rogers, C. (2006). Complex economic dynamics: chaotic saddle, crisis and intermittency. *Chaos Soliton. Fract.* 29, 1194–1218.

Goldbeter, A. (2011). A model for the dynamics of bipolar disorders. *Prog. Biophys. Mol. Biol.* 105, 119–127.

Gottschalk, A., Bauer, M. S., and Whybrow, P. C. (1995). Evidence of chaotic mood variation in bipolar disorder. *Arch. Gen. Psychiatry* 52, 947–959.

Kandel, E. R., Schwartz, J. H., and Jessell, T. M. (2000). *Principles of Neural Science, 4th Edn*. New York, NY: McGraw-Hill.

Korn, H., and Faure, P. (2003). Is there chaos in the brain? II. Experimental evidence and related models. *C. R. Biol.* 326, 1210–1213.

Rabinovich, M. I., Afraimovich, V. S., Bick, C., and Varona, P. (2012). Information flow dynamics in the brain. *Phys. Life Rev.* 123C, 76–84.

Rabinovich, M. I., and Varona, P. (2011). Robust transient dynamics and brain functions. *Front. Comput. Neurosci.* 5:24. doi: 10.3389/fncom.2011.00024

Tanaka, G., Sanjuan, M. A., and Aihara, K. (2005). Crisis-induced intermittency in two coupled chaotic maps: towards understanding chaotic itinerancy. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 71, 016219.

Tretter, F., Gebicke-Haerter, P. J., an der Heiden, U., Rujescu, D., Mewes, H. W., and Turck, C. W. (2011). Affective disorders as complex dynamic diseases – a perspective from systems biology. *Pharmacopsychiatry* 44(Suppl. 1), S2–S8.

Received: 21 July 2013; accepted: 29 July 2013; published online: 23 August 2013.

Citation: Hadaeghi F, Hashemi Golpayegani MR and Moradi K (2013) Does "crisis-induced intermittency" explain bipolar disorder dynamics? Front. Comput. Neurosci. 7:116. doi: 10.3389/fncom.2013.00116

This article was submitted to the journal Frontiers in Computational Neuroscience.

frontiers in
COMPUTATIONAL NEUROSCIENCE

OPINION ARTICLE
published: 30 August 2013
doi: 10.3389/fncom.2013.00121

Is there any geometrical information in the nervous system?

Sajad Jafari *, Seyed M. R. Hashemi Golpayegani and Shahriar Gharibzadeh

Biomedical Engineering Faculty, Amirkabir University of Technology, Tehran, Iran
**Correspondence: sajadjafari@aut.ac.ir*

Edited by:
Tobias A. Mattei, Ohio State University, USA

Keywords: trains of impulses, chaotic systems, sensitivity to initial conditions, geometry, phase space

There has been an increasing interest in analyzing neurophysiology from complex and chaotic systems viewpoint in recent years. For example, although the famous Hodgkin and Huxley model (Hodgkin and Huxley, 1952) has been the basis of almost all of the proposed models for neural firing, the Rose-Hindmarsh model (Hindmarsh and Rose, 1984) is known to be a more refined model because as it has the ability of showing different firing patterns, especially chaotic bursts of action potential, which causes a proper matching between this model behavior and many real experimental data.

It is believed that information is transferred in the brain by trains of impulses, or action potentials, often organized in sequences of bursts; therefore, it is useful to determine the temporal patterns of such trains (Korn and Faure, 2003). Since chaotic systems are sensitive to initial conditions (Hilborn, 2000), lots of signals with minimum similarity in time domain could have a same source; such behavior might be better understood by analyzing those signals in the phase space and from geometrical viewpoint (Jafari et al., 2013d), as although chaotic signals have pseudorandom behavior in time, they are ordered in phase space (i.e., if one plots the signals as a trajectory in a coordinate of system variables, he will encounter an ordered and specific topology which is called strange attractor) (Hilborn, 2000).

In fact in many applications of chaotic signals and systems, using temporal properties without being careful about this sensitivity to initial conditions, could lead to important misinterpretations (Jafari et al., 2012, 2013a,c,d). Hence, it seems that more than temporal patterns, it is of paramount importance to investigate topological patterns in such impulse trains. In order to accomplish such tasks several we have recently proposed some interesting tools for geometrical analysis (Jafari et al., in press; Shekofteh et al., in press).

In order to show the benefit of using geometry and topology in the phase space (state space), a simple example is provided in the sequence. Consider the famous Logistic map which is a very simple and well investigated chaotic map:

$$x_{k+1} = Ax_k\,(1 - x_k) \tag{1}$$

Suppose that we have two different maps with different values of parameter A:

$$x_{k+1} = 3.8x_k\,(1 - x_k) \tag{2}$$
$$x_{k+1} = 3.9x_k\,(1 - x_k)$$

If we obtain one time series from each of them, as can be seen in **Figure 1A**, they are both random-like and recognizing the difference between them seems difficult in the time domain. However, they have two ordered and easily distinguishable patterns in the state space (**Figure 1B**).

Since looking at neurophysiology from dynamical and geometrical points of view has already been successfully investigated

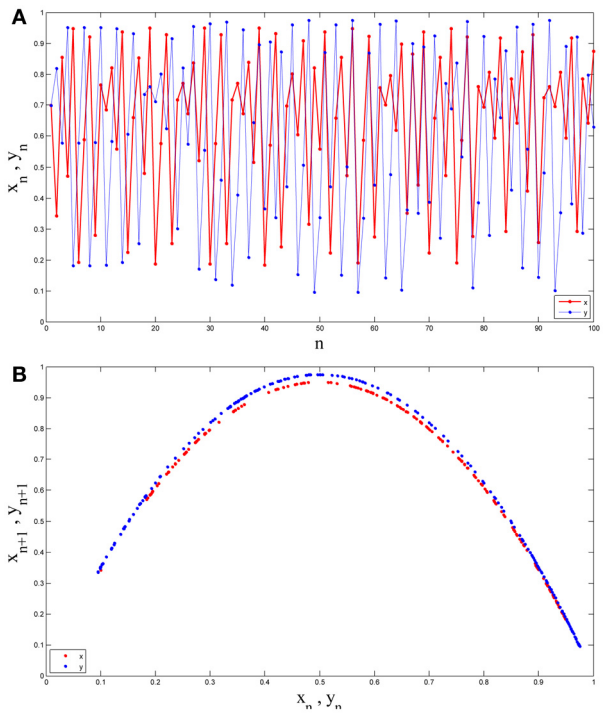


FIGURE 1 | (A) Two time series obtained from two different Logistic maps. **(B)** Those two time series embedded in the state space. As can be seen while recognizing the difference between them is not such easy in the time domain (both are random-like), they have two ordered and easily distinguishable pattern in the state space.

in some previous works (Sauer, 1994; Christini and Collins, 1995; Gottschalk et al., 1995; Milton and Black, 1995; Sarbadhikari and Chakrabarty, 2001; Korn and Faure, 2003; Hadaeghi et al., 2013; Jafari et al., 2013a), we believe that future investigations, especially using real clinical data, will be able evaluate our hypothesis and prove the benefit of such geometrical analysis of non-linear data. Ultimately, a better understanding of neuronal information transportation from the nonlinear dynamics standpoint is expected to provide a better understanding of the basic pathophysiology of neurological disorders, possibly fostering new future therapeutic approaches.

REFERENCES

Christini, D. J., and Collins, J. J. (1995). Controlling nonchaotic neuronal noise using chaos control techniques. *Phys. Rev. Lett.* 75, 2782–2785. doi: 10.1103/PhysRevLett.75.2782

Gottschalk, A., Bauer, M. S., and Whybrow, P. C. (1995). Evidence of chaotic mood variation in bipolar disorder. *Arch. Gen. Psychiatry* 52, 947–959. doi: 10.1001/archpsyc.1995.03950230061009

Hadaeghi, F., Hashemi Golpayegani, M., and Moradi, K. (2013). Does "Crisis-Induced Intermittency" explain bipolar disorder dynamics. *Front. Comput. Neurosci.* 7:116. doi: 10.3389/fncom.2013.00116

Hilborn, R. C. (2000). *Chaos and Nonlinear Dynamics, 2nd Edn*. New York, NY: Oxford University Press Inc. doi: 10.1093/acprof:oso/9780198507239.001.0001

Hindmarsh, J. L., and Rose, R. M. (1984). A model of neuronal bursting using three coupled first order differential equations. *Proc. R. Soc. Lond. B Biol. Sci.* 221, 87–102. doi: 10.1098/rspb.1984.0024

Hodgkin, A. L., and Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol.* 117, 500–544.

Jafari, S., Baghdadi, G., Hashemi Golpayegani, S. M. R., Towhidkhah, F., and Gharibzadeh, S. (2013a). Is attention deficit hyperactivity disorder a kind of intermittent chaos. *J. Neuropsychiatry Clin. Neurosci.* 25:E02. doi: 10.1176/appi.neuropsych.12040079

Jafari, S., Hashemi Golpayegani, S. M. R., and Daliri, A. (2013b). Comment on 'Parameters identification of chaotic systems by quantum-behaved particle swarm optimization' [Int. J. Comput. Math. 86 (12) (2009), pp. 2225–2235]. *Int. J. Comput. Math.* 90, 903–905. doi: 10.1080/00207160.2012.743651

Jafari, S., Hashemi Golpayegani, S. M. R., and Darabad, M. R. (2013c). Comment on "Parameter identification and synchronization of fractional-order chaotic systems" (Commun Nonlinear Sci Numer Simulat 2012;17:305–16). *Commun. Nonlinear Sci. Numer. Simulat.* 18, 811–814. doi: 10.1016/j.cnsns.2012.07.020

Jafari, S., Hashemi Golpayegani, S. M. R., Jafari, A. H., and Gharibzadeh, S. (2013d). A novel viewpoint on the parameter estimation in a chaotic neuron model. *J. Neuropsychiatry Clin. Neurosci.* 25:E19. doi: 10.1176/appi.neuropsych.12010012

Jafari, S., Hashemi Golpayegani, S. M. R., Sprott, J. C., Jafari, A. H., and Abdolmohammadi, H. R. (in press). A new cost function for parameter estimation of chaotic maps. *Int. J. Bifurcation Chaos*.

Jafari, S., Hashemi Golpayegani, S. M. R., Jafari, A. H., and Gharibzadeh, S. (2012). Some remarks on chaotic systems. *Int. J. Gen. Syst.* 41, 329–330. doi: 10.1080/03081079.2012.664855

Korn, H., and Faure, P. (2003). Is there chaos in the brain. II. Experimental evidence and related models. *C. R. Biol.* 326, 787–840. doi: 10.1016/j.crvi.2003.09.011

Milton, J., and Black, D. (1995). Dynamic diseases in neurology and psychiatry. *Chaos* 5, 8–13. doi: 10.1063/1.166103

Sarbadhikari, S. N., and Chakrabarty, K. (2001). Chaos in the brain: a short review alluding to epilepsy, depression, exercise and lateralization. *Med. Eng. Phys.* 23, 445–455. doi: 10.1016/S1350-4533(01)00075-3

Sauer, T. (1994). Reconstruction of dynamical systems from interspike intervals. *Phys. Rev. Lett.* 72, 3811–3814. doi: 10.1103/PhysRevLett.72.3811

Shekofteh, Y., Jafari, S., Sprott, J. C., Hashemi Golpayegani, S. M. R., and Almasganj, F. (in press). A gaussian mixture model based cost function for parameter estimation of chaotic biological systems. *Commun. Nonlinear Sci. Numer. Simulat*.

Received: 03 August 2013; accepted: 15 August 2013; published online: 30 August 2013.

Citation: Jafari S, Hashemi Golpayegani SMR and Gharibzadeh S (2013) Is there any geometrical information in the nervous system? Front. Comput. Neurosci. 7:121. doi: 10.3389/fncom.2013.00121

This article was submitted to the journal Frontiers in Computational Neuroscience.

frontiers in
COMPUTATIONAL NEUROSCIENCE

OPINION ARTICLE
published: 01 October 2013
doi: 10.3389/fncom.2013.00130

Can cellular automata be a representative model for visual perception dynamics?

Maryam Beigzadeh, Seyyed Mohammad R. Hashemi Golpayegani and Shahriar Gharibzadeh*

Department of Biomedical Engineering, Amirkabir University of Technology, Tehran, Iran
**Correspondence: gharibzadeh@aut.ac.ir*

Edited by:
Tobias A. Mattei, Ohio State University, USA

Keywords: cellular automata, visual perception modeling, complex systems, chaos and nonlinear dynamics, EEG

"Cellular automata (CA) are mathematical models for systems in which many simple components act together to produce complicated patterns of behavior" (Packard and Wolfram, 1985). Applying the CA theoretical framework in the field of neuroscience has shown successful results in the interpretation of some cognitive aspects (Adams et al., 1992; Pashaie and Farhat, 2009; Kozma and Puljic, 2013; Lopez-Ruiz and Fournier-Prunaret, 2013; Mattei, 2013). In this short analysis, we suggest that CA can be a very reasonable tool to model both dynamical and structural aspects of visual perception. As Wolfram declared in his book, A New Kind of Science, visual perception is a kind of modeling and reducing the input visual sensory data into a more summary but still informative representation in the brain (Wolfram, 2002).

Studying the visual system can be very useful, because as already previously demonstrated, "The visual system has the most complex neural circuitry of all the sensory systems (Kandel et al., 2000)" and at least 20% of human cerebral cortex is related to the visual part (Olshausen, 2002). Additionally, trying to understand visual perception may lead us to a better understanding of how other cognitive processes in the brains work.

It has already been demonstrated that brain dynamics (which are reflected in EEG, MEG, and ECoG signals) are inherently chaotic (Freeman, 1991). As we perceive different sensory information (i.e., images, sounds, odors, etc.) and recognize different patterns, these dynamical processes tend to turn into a more regular pattern. This stage has been referred by other researchers as: "the transients between gas-like randomness and liquid-like order (Kozma et al., 2012)." According to such paradigm, each stimulus would tend to lead the system to its own "liquid-like attractor" which is different from the other one. So, after the sensorial stimuli, the brain dynamics would exhibit a temporary switching between these different states.

But what would be the advantage of using such CA model? There are millions of neurons in the visual system that are highly interactive, each one demonstrating its own complex behavior. Their combined and integrated functions lead to the overall process of perception. The CA framework provides a model, in which a collection of many interactive agents (cells) relate to each other according to specific "interaction rules" in space and time. The number of agents, their dynamical properties, and their interactions with each other, determine which kind of behavior (chaotic, periodic, etc.) the CA will adopt.

Compared to other alternative multi-agent modeling tools (such as artificial neural networks), in CA the researcher is able to determine the local behavior of individuals as well as their interaction rules and connectivity patterns, both locally and globally in space. In CA model it is also possible to analyze the behavior of the system from both the micro to macro levels. But how could the analysis of the space properties of CA make visual perception modeling more realistic? It has already been demonstrated that, in the visual system (at least in the primary processing areas such as V1) there are specialized cells which, because of their own specific structure and function, become more sensitive to specific properties of the perceived visual scene (such as image edges, textures, orientation, spatial frequencies) that are inherently space related features.

In such sense, CA would fit as a very appropriate model, as it exhibits close theoretical similarities with other methods which use graph theory and small world networks analysis (Sporns, 2006; Stam and Reijneveld, 2007). Additionally, it has already been suggested that probabilistic CA can be successfully employed to model the olfactory perception (Kozma et al., 2012). Nevertheless, using CA for modeling visual perception from the dynamical and structural standpoints has not yet been reported before, although CA has already been used for modeling simpler visual-related tasks, such as retina function, or as a computational tool for implementing image processing tasks in computer vision applications (like edge detection, texture detection, noise reduction, etc.) (Wolfram, 2002; Dhillon, 2012). In this short commentary we defend that CA can be used as a holistic model for the integration of local visual aspects in a broader multimodal integration of the global aspects of visual perception.

One possible strategy in order to implement such paradigm would be to use specific objective measures (such as the number of active neurons, or the mean activation value of a specific network), and afterwards attempt to match the behavior of such time series (by comparing its phase space and strange attractors) with real EEG recordings related to specific visual tasks (such as the classic "face/non face discrimination").

In summary, the dynamic behavior of CA has been shown to be a power tool for modeling several types of neuronal activity and we believe that it can be successfully used to study global features of visual perception. In fact, future studies on this area may be able to demonstrate how perceptual deficits commonly observed in clinical practice (such as face recognition deficits in autistic patients) may be represented by a change in the basic parameters of CA models of visual representation.

REFERENCES

Adams, F. R., Nguyen, H. T., Raghavan, R., and Slawny, J. (1992). A parallel network for visual cognition. *IEEE Trans. Neural Netw*. 3, 906–922. doi: 10.1109/72.165593

Dhillon, P. K. (2012). A novel framework to image edge detection using cellular automata. *Int. J. Comput. Appl*. 1, 1–5.

Freeman, W. J. (1991). The physiology of perception. *Sci. Am*. 264, 78–85. doi: 10.1038/scientificamerican0291-78

Kandel, E. R., Schwartz, J. H., and Jessell, T. M. (2000). *Principles of Neural Science, 4th Edn.* New York, NY: McGraw-Hill Companies.

Kozma, R., and Puljic, M. (2013). Hierarchical random cellular neural networks for system-level brain-like signal processing. *Neural Netw.* 45, 101–110. doi: 10.1016/j.neunet.2013.02.010

Kozma, R., Puljic, M., and Freeman, W. J. (2012). "Thermodynamic model of criticality in the cortex based on EEG/ECOG data," in *Criticality in Neural Systems*, ed P. Dietmar (Berkeley, CA: John Wiley and Sons, Inc.).

Lopez-Ruiz, R., and Fournier-Prunaret, D. (2013). "The bistable brain: a neuronal model with symbiotic interactions," in *Symbiosis, Evolution, Biology and Ecological Effects*, eds A. F. Camiso and C. C. Pedroso (New York, NY: NOVA Science Publications).

Mattei, T. A. (2013). The fuzzy logic of degenerative disc disease: from a lorenz attractor to a percolation threshold model. *World Neurosurg.* 80, 8–12. doi: 10.1016/j.wneu.2013.05.007

Olshausen, B. A. (2002). *Sensory Processes, Psychology 129* [Online]. // Visual Cortex. - berkley, 2002. - 2013. Available online at: http://redwood.berkeley.edu/bruno/psc129/lecture-notes/visual-cortex.html lecture notes: http://redwood.berkeley.edu/bruno/psc129.

Packard, N. H., and Wolfram, S. (1985). Two-dimensional cellular automata. *J. Stat. Phys.* 38, 901–946. doi: 10.1007/BF01010423

Pashaie, R., and Farhat, N. H. (2009). Self-organization in a parametrically coupled logistic map network: a model for information processing in the visual cortex. *IEEE Trans. Neural Netw*. 20, 597–608. doi: 10.1109/TNN.2008.2010703

Sporns, O. (2006). Small-world connectivity, motif composition, and complexity of fractal neuronal connections. *Biosystems* 85, 55–64. doi: 10.1016/j.biosystems.2006.02.008

Stam, C. J., and Reijneveld, J. C. (2007). Graph theoretical analysis of complex networks in the brain. *Nonlinear Biomed. Phys.* 1, 1–19. doi: 10.1186/1753-4631-1-3

Wolfram, S. (2002). *A New Kind of Science*. Champaign, IL: Wolfram Media Inc.

Received: 01 September 2013; accepted: 09 September 2013; published online: 01 October 2013.

Citation: Beigzadeh M, Hashemi Golpayegani SMR and Gharibzadeh S (2013) Can cellular automata be a representative model for visual perception dynamics? Front. Comput. Neurosci. 7:130. doi: 10.3389/fncom.2013.00130

This article was submitted to the journal Frontiers in Computational Neuroscience.

frontiers in
COMPUTATIONAL NEUROSCIENCE

OPINION ARTICLE
published: 06 February 2014
doi: 10.3389/fncom.2014.00011

Bifurcation analysis of "synchronization fluctuation": a diagnostic measure of brain epileptic states

Fatemeh Bakouie[1,2]*, Keivan Moradi[1], Shahriar Gharibzadeh[1] and Farzad Towhidkhah[2]

[1] *Neural and Cognitive Sciences Lab, Biomedical Engineering Faculty, Amirkabir University of Technology, Tehran, Iran*
[2] *Cybernetics and Modeling of Biological Systems Laboratory, Biomedical Engineering Faculty, Amirkabir University of Technology, Tehran, Iran*
**Correspondence: fbakouie@aut.ac.ir*

Edited and reviewed by:
Tobias A. Mattei, Ohio State University, USA

Keywords: complex network, synchronization fluctuation, dynamical system, bifurcation, control parameter

The brain is a complex network with functional elements spatially distributed in different regions. One suggested mechanism for communication among these distributed elements is synchronization (Singer, 1993).

Two oscillating neural groups are called to be "synchronized" if over the time, their phase difference does not remarkably increase. In a real system composed of some oscillators, synchronization level is a computable parameter. According to this paradigm, depending on the functional state of the brain, the level of synchronization among brain regions may vary over time. This variation is called "synchronization fluctuation" (SF). Regarding brain's higher functions such as consciousness and memory, for instance, SF patterns are important features of normal brain states (Schnitzler and Gross, 2005; Watrous et al., 2013).

In some pathological brain states such as epilepsy, however, hypersynchronization is a major problem (Lehnertz et al., 2009). In such situations, synchronization occurs without fluctuations. Therefore, in epilepsy, SF may lose its dynamicity, producing a narrow-dynamics signal. The question which arises is: "how is it possible to manage diseases related to the poor dynamics of SF in the brain?"

Dynamical systems approach may be able to provide some answers to this question: Based on dynamical systems theory, even slight modification of a parameter (so-called "control parameter") is able to lead to a significant qualitative change in the system's behavior. This change is called a "bifurcation" (Guckenheimer, 2007). Dynamical approach has already been successfully used to the study of the functional status of epileptic states. For example, Babloyantz and Destexhe reported the nonlinearity of absence (Babloyantz and Destexhe, 1986). Moreover, Stam claims that epilepsy is the most important application of nonlinear EEG study (Stam, 2005). In another research Perez Velazquez et al. suggested that the interictal ictal transition may be the result of bifurcation due to alteration in control parameters like the balance between excitation an inhibition in the underlying neuronal networks (Perez Velazquez et al., 2003).

We hypothesize that SF may be a representative parameter of brain dynamics, which have identifiable bifurcations according to specific brain states. According to such approach, SF dynamics is supposed to change from a rich state to a narrower one, when brain changes from normal conscious to abnormal unconscious epileptic conditions.

Biologically, different mechanisms have already been suggested as the underlying basis of brain synchronization. For instance, it has been shown that gap junctions, coupling of neurons via long-term synaptic plasticity, interneurons, and rhythm generators of the brain such as the medial septum-diagonal band of Broca (MSDBB) may play a role in the synchronization between two neurons or more neuronal networks (Buzsáki, 2002). Such biological mechanisms that control synchronization can be considered as control parameters of SF in brain dynamics. For example, among these parameters, variations may exist in the number and permeability of gap junctions, the synaptic strength between two neurons, the distribution, frequency and strength of the GABA inhibition by interneurons, and the distribution, frequency and strength of excitation and inhibition of the cholinergic and GABAergic neurons of the MSDBB. Moreover, Margineanu and Klitgaard have already demonstrated that levetiracetam (LEV) antagonizes neuronal (hyper) synchronization, in the CA3 area of rat brain slices which is prone to epilepsy (Georg Margineanu and Klitgaard, 2000). In another research, Clemens showed that Valproate decreases EEG synchronization in idiopathic generalized epilepsy (Clemens, 2008).

Concerning connectivity among brain regions, Kay et al. explained that in treatment-responsive epileptic patients, compared to healthy controls, default Mode Network (DMN) connectivity does not reduce significantly; however, in treatment-resistant epileptic patients, there exists connectivity reduction compared to control group (Kay et al., 2013). In another study, the researchers showed DMN alterations in mesial temporal lobe epilepsy. Furthermore, Liao et al. have showed that in mesial temporal lobe epilepsy (mTLE) patients with hippocampal sclerosis (HS), there are reductions in functional and structural connectivity between hippocampal structures and their adjacent regions (Liao et al., 2011). Compared to the controls, it was shown that there is significant reduction in functional and structural connectivity between the posterior cingulate cortex (PCC)/precuneus (PCUN) and bilateral mesial temporal lobes (mTLs). Resting functional magnetic resonance imaging studies showed that in drug-resistant temporal lobe epilepsy, functional connectivity between the hippocampus, anterior

temporal, precentral cortices and the default mode and sensorimotor networks reduces Based on their findings it would be claimed that the reduction in functional connectivity within the DMN in mTLE may be the result of the connection density reduction, leading to degeneration of structural connectivity (Voets et al., 2012). These finding showed that in epilepsy, connectivity reduction occurred, while pharmacological treatment tend to drive this change in connectivity back to normal state. The mechanism of such therapeutic action, however, is still relatively unknown (Jin and Zhong, 2011).

In the future, it would be interesting to analyze the efficacy of therapeutic strategies addressing diseases caused by SF dynamicity changes (such as antiepileptic drugs) according to their capacity to carefully tune the control parameters of SF in order to set the brain back to its normal states. As an evidence, Krystal et al. hypothesized that Lyapunov exponent ($\lambda 1$) may decrease during the electroconvulsive therapy (ECT) seizures (Krystal et al., 1996). It seems that despite they did not assess synchronization directly, decreased $\lambda 1$ corresponds to decreased EEG complexity. In another experimental treatment strategy for epilepsy, researchers have implemented an "automated, just-in-time stimulation seizure control method" in epileptic rats. Interestingly, the successful control of seizures with such therapy highly correlated with desynchronization of brain dynamics (Good et al., 2009).

Such experimental researches support the idea that, by tuning control parameters of SF, it may be possible to drive pathological brain states into normal ones. Therefore, we suggest that SF may be an important measure that represents the brain dynamics and that SF dynamics may be a potential subject of future experimental studies aiming to uncover the underlying mechanisms of pathological brain states.

REFERENCES

Babloyantz, A., and Destexhe, A. (1986). Low-dimensional chaos in an instance of epilepsy. *Proc. Natl. Acad. Sci. U.S.A.* 83, 3513–3517. doi: 10.1073/pnas.83.10.3513

Buzsáki, G. (2002). Theta oscillations in the hippocampus. *Neuron* 33, 325–340. doi: 10.1016/S0896-6273(02)00586-X

Clemens, B. (2008). Valproate decreases EEG synchronization in a use-dependent manner in idiopathic generalized epilepsy. *Seizure* 17, 224–233. doi: 10.1016/j.seizure.2007.07.005

Georg Margineanu, D., and Klitgaard, H. (2000). Inhibition of neuronal hypersynchrony *in vitro* differentiates levetiracetam from classical antiepileptic drugs. *Pharmacol. Res.* 42, 281–285. doi: 10.1006/phrs.2000.0689

Good, L. B., Sabesan, S., Marsh, S. T., Tsakalis, K., Treiman, D., and Iasemidis, L. (2009). Control of synchronization of brain dynamics leads to control of epileptic seizures in rodents. *Int. J. Neural Syst.* 19, 173–196. doi: 10.1142/S0129065709001951

Guckenheimer, J. (2007). Bifurcation. *Scholarpedia* 2, 1517. doi: 10.4249/scholarpedia.1517

Jin, M.-M., and Zhong, C. (2011). Role of gap junctions in epilepsy. *Neurosci. Bull.* 27, 389–406. doi: 10.1007/s12264-011-1944-1

Kay, B. P., Difrancesco, M. W., Privitera, M. D., Gotman, J., Holland, S. K., and Szaflarski, J. P. (2013). Reduced default mode network connectivity in treatment–resistant idiopathic generalized epilepsy. *Epilepsia* 54, 461–470. doi: 10.1111/epi.12057

Krystal, A. D., Greenside, H. S., Weiner, R. D., and Gassert, D. (1996). A comparison of EEG signal dynamics in waking, after anesthesia induction and during electroconvulsive therapy seizures. *Electroencephalogr. Clin. Neurophysiol.* 99, 129–140. doi: 10.1016/0013-4694(96)95090-7

Lehnertz, K., Bialonski, S., Horstmann, M.-T., Krug, D., Rothkegel, A., Staniek, M., and Wagner, T. (2009). Synchronization phenomena in human epileptic brain networks. *J. Neurosci. Methods* 183, 42–48. doi: 10.1016/j.jneumeth.2009.05.015

Liao, W., Zhang, Z., Pan, Z., Mantini, D., Ding, J., Duan, X., et al. (2011). Default mode network abnormalities in mesial temporal lobe epilepsy: a study combining fMRI and DTI. *Hum. Brain Mapp.* 32, 883–895. doi: 10.1002/hbm.21076

Perez Velazquez, J. L., Cortez, M. A., Snead Iii, O. C., and Wennberg, R. (2003). Dynamical regimes underlying epileptiform events: role of instabilities and bifurcations in brain activity. *Physica D* 186, 205–220. doi: 10.1016/j.physd.2003.07.002

Schnitzler, A., and Gross, J. (2005). Normal and pathological oscillatory communication in the brain. *Nat. Rev. Neurosci.* 6, 285–296. doi: 10.1038/nrn1650

Singer, W. (1993). Synchronization of cortical activity and its putative role in information processing and learning. *Annu. Rev. Physiol.* 55, 349–374. doi: 10.1146/annurev.ph.55.030193.002025

Stam, C. J. (2005). Nonlinear dynamical analysis of EEG and MEG: review of an emerging field. *Clin. Neurophysiol.* 116, 2266–2301. doi: 10.1016/j.clinph.2005.06.011

Voets, N. L., Beckmann, C. F., Cole, D. M., Hong, S., Bernasconi, A., and Bernasconi, N. (2012). Structural substrates for resting network disruption in temporal lobe epilepsy. *Brain* 135, 2350–2357. doi: 10.1093/brain/aws137

Watrous, A. J., Tandon, N., Conner, C. R., Pieters, T., and Ekstrom, A. D. (2013). Frequency-specific network connectivity increases underlie accurate spatiotemporal memory retrieval. *Nat. Neurosci.* 16, 349–356. doi: 10.1038/nn.3315

Received: 23 November 2013; accepted: 21 January 2014; published online: 06 February 2014.

Citation: Bakouie F, Moradi K, Gharibzadeh S and Towhidkhah F (2014) Bifurcation analysis of "synchronization fluctuation": a diagnostic measure of brain epileptic states. Front. Comput. Neurosci. ***8****:11. doi: 10.3389/fncom.2014.00011*

This article was submitted to the journal Frontiers in Computational Neuroscience.

frontiers in
COMPUTATIONAL NEUROSCIENCE

GENERAL COMMENTARY
published: 20 February 2014
doi: 10.3389/fncom.2014.00015

A more realistic quantum mechanical model of conscious perception during binocular rivalry

***Mohammad Reza Paraan*[1], *Fatemeh Bakouie*[2]* *and Shahriar Gharibzadeh*[2]**

[1] *Energy Engineering and Physics Department, Amirkabir University of Technology, Tehran, Iran*
[2] *Neural and Cognitive Sciences Lab, Biomedical Engineering Department, Amirkabir University of Technology, Tehran, Iran*
**Correspondence: fbakouie@aut.ac.ir*

Edited by:
Tobias Alecio Mattei, Ohio State University, USA

Keywords: consciousness, quantum state, mixed state, probability distribution, dominance duration

A commentary on

Quantum formalism to describe binocular rivalry
by Manousakis, E. (2009). Biosystems 98, 57–66. doi: 10.1016/j.biosystems.2009.05.012

Since the first systematic description of binocular rivalry by Wheatstone, this fascinating phenomenon has provided several new insights into the mechanisms of visual awareness (Leopold and Logothetis, 1999). Binocular rivalry (BR) is the subjective experience of randomly alternating perceptions pertaining to the two eyes when they are presented with conflicting stimuli. Because of its nature, BR enables consciousness researchers to separately investigate the mechanisms of perception and conscious experience (Gazzaniga et al., 2009). Among various descriptions of this phenomenon, quantum mechanical descriptions stand out as the most radical.

In a recent innovative work by Manousakis, the formalism of quantum mechanics is utilized to describe the conscious experience during BR. Although the author has successfully derived the observed probability distribution of dominance durations (PDDD), his approach undermines some essential features of conscious perception during BR. Generally, two kinds of perception dominate during BR: (1) full dominance of one eye's stimulus, (2) composite or mixed dominance of the two monocular stimuli (Yang et al., 1992). Our argument revolves around the latter kind of perception which is also referred to as transition phase or transition state.

Classically, simplifications imposed experimental conditions in which only full dominance was perceived by subjects and mixed state's (MS) duration was minimized. However, many experiments reveal the diversity in rivalry's temporal dynamics and specifically the important role of MS (Hollins, 1980; Blake et al., 1992; Bossink et al., 1993; Wilson et al., 2001). Regarding the neural correlates of MS, it has been shown that the frontoparietal areas of brain trigger rivalry transitions (Lumer et al., 1998; Knapen et al., 2011). It must be emphasized that various studies on the neural concomitants of BR suggest that no single neural site or neural mechanism is at work during BR, rather multiple stages and brain areas are involved (Blake and Logothetis, 2002).

Many attempts have been made to model the dynamical behavior of BR, most of which try to reproduce the temporal dynamics of BR by reconstructing specific neural mechanisms (Kalarickal and Marshall, 2000; Laing and Chow, 2002; Stollenwerk and Bode, 2003; Freeman, 2005). A major number of these models ignore MS in order to avoid crippling complications, yet Brascamp and colleagues show that none of the previous models is capable of reproducing the full range of observed dynamics which include MS (Brascamp et al., 2006b) and hence try to develop a new model (Brascamp et al., 2006a; Noest and van Ee, 2006). Another group of models of which Manousakis' model is an example capture certain aspects of rivalry's dynamics without resorting to the underlying neural circuits (Mamassian and Goutcher, 2005). However, in order to obtain the PDDD, Manousakis employs some temporal parameters characterizing neuronal firings. This is an interesting achievement because it ties the dynamics of conscious perception to specific firing patterns.

Like the classical models, Manousakis' model only treats the two dominance states which are represented by two quantum states, while MS is ignored. The author compares his theoretical PDDD with the observed PDDD of classical experiments (Levelt, 1968; Lehky, 1995) which did not record the mixed states' duration separately. We believe that the quantum states are only symbols which are manipulated according to the quantum formalism, and bear no resemblance to the perception they represent. Therefore, in Manousakis' approach, only the number of states and their associated probabilities determine the favored PDDD. Therefore, unlike classical models, the scope of the quantum mechanical model can be readily extended by introducing a third quantum state which represents MS. In order to test the new model, its PDDD should be calculated and compared against that of experimental data which are separate recordings of dominance durations of the three states. It must be emphasized that the probability distribution is not a complete description of the dynamics of BR, and it is necessary to extract other relative quantities from the model in future works.

It is worthwhile discussing another work by Conte and colleagues who showed that mental states follow quantum mechanics during the conscious bi-stable perception of ambiguous figures (Conte et al., 2009). Their model shares a lot of features with that of Manousakis, with the exception that they take into account the periods when their subjects report indeterminate perception. Indeterminate perception resembles MS in that they are

both mental states and are mediated by specific neural correlates. But Conte et al. represent indeterminacy state by the wave-function of the two-state system rather than an additional third quantum state. Technically, a wave-function is a superposition of all the real possible states of a quantum system. We believe that this is an inappropriate take on the problem which leads to inconsistencies within the model. The developers of these two quantum mechanical models believe that the actualization of each quantum state is equal to the activation of neural correlates of consciousness (NCC) of the corresponding perception; a state is actualized when a quantum system is measured (observed) and subsequently its wave-function "collapses" to that constituent state. Therefore, we believe that wave-function is not a legitimate representation, because it does not describe a real state of a system and is doomed to collapse, and on the other hand, specific NCC of MS or that of indeterminate perception demands a distinct associated quantum state.

Manousakis' neglect of MS might be justified by the presumption that this state only functions as a bridge between the two dominance states. That is, MS does not compete with the other two and is not involved in rivalry. It is noteworthy that the term "transition" has led to a misunderstanding, namely that the MS occurs only when the perception is being switched from one eye to another. But as is often the case with BR experiments, subjects report the same perception as the one that was dominant before MS. Hence, there is no particular regular periodic alternation between dominance and suppression (Mueller and Blake, 1989; Brascamp et al., 2006b). We believe these indicate that MS is not a mere bridge connecting the two dominant states, but a state which dominates consciousness randomly and therefore, enters statistical calculations of quantum mechanics.

REFERENCES

Blake, R., and Logothetis, N. K. (2002). Visual competition. *Nat. Rev. Neurosci.* 3, 13–21. doi: 10.1038/nrn701

Blake, R., O'Shea, R. P., and Mueller, T. (1992). Spatial zones of binocular rivalry in central and peripheral vision. *Vis. Neurosci.* 8, 469–478. doi: 10.1017/S0952523800004971

Bossink, C., Stalmeier, P., and De Weert, C. M. (1993). A test of Levelt's second proposition for binocular rivalry. *Vision Res.* 33, 1413–1419. doi: 10.1016/0042-6989(93)90047-Z

Brascamp, J. W., Noest, A. J., van Ee, R., and Van Den Berg, A. V. (2006a). Transition phases show the importance of noise in binocular rivalry. *J. Vis.* 6, 845–845. doi: 10.1167/6.6.845

Brascamp, J. W., van Ee, R., Noest, A. J., Jacobs, R. H., and Van Den Berg, A. V. (2006b). The time course of binocular rivalry reveals a fundamental role of noise. *J. Vis.* 6, 1244–1256. doi: 10.1167/6.11.8

Conte, E., Khrennikov, A. Y., Todarello, O., Federici, A., Mendolicchio, L., and Zbilut, J. P. (2009). Mental states follow quantum mechanics during perception and cognition of ambiguous figures. *Open Syst. Inf. Dyn.* 16, 85–100. doi: 10.1142/S1230161209000074

Freeman, A. W. (2005). Multistage model for binocular rivalry. *J. Neurophysiol.* 94, 4412–4420. doi: 10.1152/jn.00557.2005

Gazzaniga, M. S., Bizzi, E., Caramazza, A., Chalupa, L. M., Grafton, S. T., Heatherton, T. F., et al. (2009). *The Cognitive Neurosciences*. Cambridge, MA: The MIT Press.

Hollins, M. (1980). The effect of contrast on the completeness of binocular rivalry suppression. *Percept. Psychophys.* 27, 550–556. doi: 10.3758/BF03198684

Kalarickal, G. J., and Marshall, J. A. (2000). Neural model of temporal and stochastic properties of binocular rivalry. *Neurocomputing* 32, 843–853. doi: 10.1016/S0925-2312(00)00252-6

Knapen, T., Brascamp, J., Pearson, J., van Ee, R., and Blake, R. (2011). The role of frontal and parietal brain areas in bistable perception. *J. Neurosci.* 31, 10293–10301. doi: 10.1523/JNEUROSCI.1727-11.2011

Laing, C. R., and Chow, C. C. (2002). A spiking neuron model for binocular rivalry. *J. Comput. Neurosci.* 12, 39–53. doi: 10.1023/A:1014942129705

Lehky, S. R. (1995). Binocular rivalry is not chaotic. *Proc. Biol. Sci.* 259, 71–76. doi: 10.1098/rspb.1995.0011

Leopold, D. A., and Logothetis, N. K. (1999). Multistable phenomena: changing views in perception. *Trends Cogn. Sci.* 3, 254–264. doi: 10.1016/S1364-6613(99)01332-7

Levelt, W. J. (1968). *On Binocular Rivalry*. The Hague: Mouton.

Lumer, E. D., Friston, K. J., and Rees, G. (1998). Neural correlates of perceptual rivalry in the human brain. *Science* 280, 1930–1934. doi: 10.1126/science.280.5371.1930

Mamassian, P., and Goutcher, R. (2005). Temporal dynamics in bistable perception. *J. Vis.* 5, 361–375. doi: 10.1167/5.4.7

Mueller, T., and Blake, R. (1989). A fresh look at the temporal dynamics of binocular rivalry. *Biol. Cybern.* 61, 223–232. doi: 10.1007/BF00198769

Noest, A., and van Ee, R. (2006). Statistical-mechanics modeling of rivalrous dominance and transition durations. *Perception* 35, 53. doi: 10.1068/v060618

Stollenwerk, L., and Bode, M. (2003). Lateral neural model of binocular rivalry. *Neural Comput.* 15, 2863–2882. doi: 10.1162/089976603322518777

Wilson, H. R., Blake, R., and Lee, S.-H. (2001). Dynamics of travelling waves in visual perception. *Nature* 412, 907–910. doi: 10.1038/35091066

Yang, Y., Rose, D., and Blake, R. (1992). On the variety of percepts associated with dichoptic viewing of dissimilar monocular stimuli. *Perception* 21, 47–62. doi: 10.1068/p210047

Received: 12 January 2014; accepted: 02 February 2014; published online: 20 February 2014.

Citation: Paraan MR, Bakouie F and Gharibzadeh S (2014) A more realistic quantum mechanical model of conscious perception during binocular rivalry. Front. Comput. Neurosci. ***8****:15. doi: 10.3389/fncom.2014.00015*

This article was submitted to the journal Frontiers in Computational Neuroscience.

frontiers in
COMPUTATIONAL NEUROSCIENCE

OPINION ARTICLE
published: 11 March 2014
doi: 10.3389/fncom.2014.00028

A hypothesis on the role of perturbation size on the human sensorimotor adaptation

Fatemeh Yavari[1], Farzad Towhidkhah[1]* and Mohammad Darainy[2]

[1] *Biomedical Engineering Department, Amirkabir University of Technology, Tehran, Iran*
[2] *Department of Psychology, McGill University, Montreal, QC, Canada*
**Correspondence: towhidkhah@aut.ac.it*

Edited and reviewed by:
Tobias Alecio Mattei, Ohio State University, USA

Keywords: adaptation, perturbation amplitude, error size, sensory recalibration, internal model

INTRODUCTION

Some evidence suggests that depending on the size of error produced by a perturbation, distinct learning mechanisms and neural structures are employed in the brain (Kluzik et al., 2008; Criscimagna-Hemminger et al., 2010; Gibo et al., 2013). Here, based on some existing evidence, we propose a hypothesis about the potential adaptation mechanisms which may be employed in the brain based on the perturbation magnitude. In the following sections, we first briefly explain the proposed hypothesis. Then a short description about the resolution of hand proprioceptive sensory is presented. In this hypothesis, the size of error is assessed relative to the resolution of proprioceptive sensory. Next, the empirical evidence supporting the proposed hypothesis are shortly described.

THE HYPOTHESIS

Our hypothesis schematically represented in **Figure 1** is as follows:

1- For small perturbation amplitude compared to proprioceptive sensory resolution, the produced movement error (Err. in **Figure 1**) will be small as well. Small error does not often result in subject's awareness (Cressman and Henriques, 2009; Criscimagna-Hemminger et al., 2010). In this condition, the brain may consider the perturbation resulting from an internal source and compensate it with recalibration of proprioceptive sensory. This may be expressed by shifting the input-output relationship of proprioceptive sensory module (i.e., Proprioceptive block in **Figure 1**). The input-output relationship of this module has been modeled with a quantization (staircase) function to represent the limited resolution.

2- For large perturbation amplitude, the produced movement error will be large as well, which typically make subject aware of the perturbation (Malfait and Ostry, 2004). In this case the assumption is that the perturbation may be caused by an external source and the brain may need to form/update internal forward and/or inverse models of the new dynamics to reduce movement errors.

RESOLUTION OF PROPRIOCEPTIVE SENSORY

It is possible to infer about the resolution of proprioceptive sensory based on some of previous studies. Diedrichsen et al. (2010) moved the subject's hand passively using a robotic arm along a trajectory deviated 8° to the left or right of the subjects' body midline. In the absence of visual feedback, subjects were not able to guess the direction of this deviation. In another study (Farrer et al., 2003), the experimenter moved subject's hand by pulling a rod connected to a joystick. Subjects had no direct view of their hand; instead a virtual hand image provided the visual feedback for them. The visual feedback was deviated either to the right or left relative to the actual hand movement by a certain angular value (0, 5, 10, 15, 20, 30, 40, or 50°) in each trial. At the end of each movement, subjects had to indicate if their movement and the visual feedback were at the same place. They were not able to detect the deviation when it was less than 5° (Figure 2. in Farrer et al., 2003). Also, Darainy et al. (2013) observed that during passive hand movements perceptual boundary was at the left of the midline. Based on the observations in the above mentioned studies and some others (Cressman and Henriques, 2009; Fuentes et al., 2011), it can be suggested that resolution of proprioceptive sensory is about 5° (in the midline direction). On the other hand, there are some evidence supporting this notion that proprioceptive sensory is more precise in front-back direction than left-right (van Beers et al., 2002; Wilson et al., 2010). Therefore it seems plausible to infer that maximum resolution of proprioceptive sensory is in the midline direction.

EVIDENCE SUPPORTING THE PROPOSED HYPOTHESIS

Some of the observations which can be explained based on this hypothesis are given in the following:

- Based on the proposed hypothesis, adaptation to an abrupt perturbation, which produces large errors, results in formation of an IM in the brain, while adaptation to a gradual perturbation is probably not dependent on IMs. Cerebellum is one of the main candidate brain regions to contain IMs, specifically internal forward models (see Yavari et al., 2013 for a review). It has already been demonstrated that individuals with cerebellar damage have difficulties in adapting to an abrupt force field during hand reaching movements (Smith and Shadmehr, 2005); however when that perturbation was imposed gradually they are usually able to adapt their movements (Criscimagna-Hemminger et al., 2010; Izawa et al., 2012). These observations

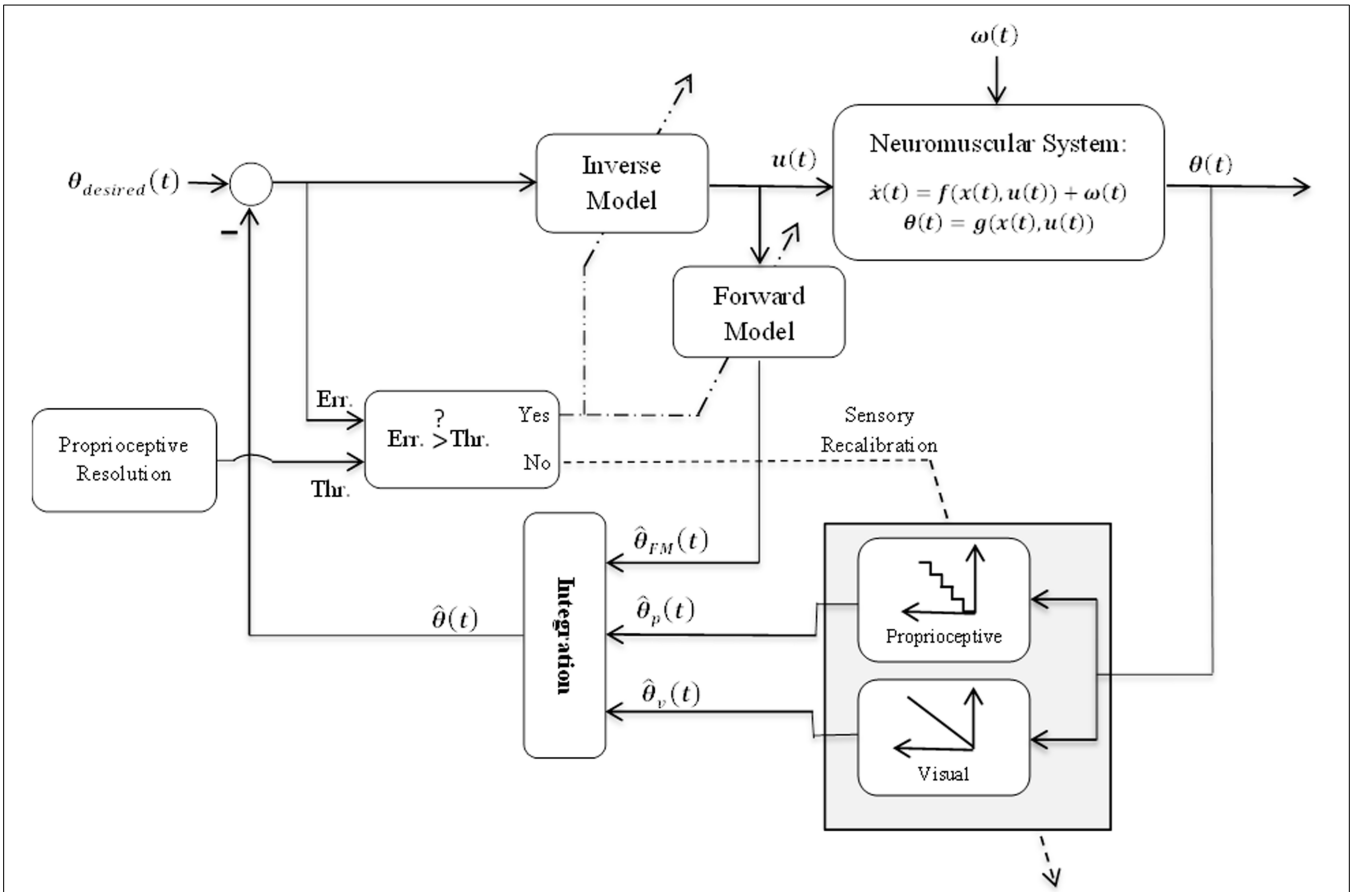


FIGURE 1 | Schematic representation of the proposed hypothesis. The general structure of this model has been borrowed from other studies e.g., (Shadmehr and Krakauer, 2008). $\theta(t)$, $\hat{\theta}_{FM}(t)$, $\hat{\theta}_p(t)$, and $\hat{\theta}_v(t)$ are respectively system output and its estimations by forward model, proprioceptive sensory, and visual sensory. $\hat{\theta}(t)$ is final estimation of system output obtained from integration. Dashed and dot-dashed lines show sensory recalibration and internal models' (IMs') adaptation, respectively.

confirm dependency of adaptation in presence of large, but not small errors on cerebellum and are in line with the proposed hypothesis.

- It has been observed that sudden and gradual introduction of perturbations—which result in large and small errors, respectively—produce different generalization patterns. Motor memories produced by abrupt perturbations are in an extrinsic coordinate system and generalize to the untrained arm (Criscimagna-Hemminger et al., 2003; Malfait and Ostry, 2004), whereas gradual presentation of perturbations cause adaptation in intrinsic arm coordinates that does not transfer to the other arm (Malfait and Ostry, 2004; Wilson et al., 2010). Also it has been observed that gradual perturbations lead to more robust generalization when using the trained arm in a different context, while this generalization is smaller in response to a sudden perturbation (Kluzik et al., 2008). These observations can be explained based on the proposed hypothesis as follows: the brain forms an IM of the perturbation in response to large errors (in an extrinsic coordinate system). The created model would be applicable in performing movements with another hand in the presence of the same perturbation. On the other hand, gradual presentation of the perturbation results in sensory recalibration which is specific to the trained arm (intrinsic arm coordinates). This explains the generalization pattern produced by small errors.
- Subjects showed almost the same size of aftereffect when adapted to gradual and abrupt perturbations; however washout rate was significantly higher in the abrupt group (Kluzik et al., 2008). On the other hand, functional imaging and computational studies support the existence of multiple IMs in the brain which are activated based on the context (Haruno et al., 2001; Imamizu et al., 2003, 2004). Having this point in mind, the mentioned observation may be explained as follows: adaptation to an abrupt perturbation results in formation of an IM in the brain. Eliminating the perturbation causes aftereffects which will not last for long because the brain rapidly switches back to the suitable IM for the condition with no perturbation. This may not be the case for small errors.
- Sensory recalibration due to adaptation to small errors has been observed in some previous studies (Cressman and Henriques, 2009).

SUMMARY

We presented a hypothesis about the possible adaptation mechanisms employed in the brain based on error size. The proposed hypothesis can help to provide a better understanding of motor adaptation mechanism in brain. Further validation of the hypothesis requires more investigations and experiments. For example, adaptation in response to a gradual perturbation can be compared in deafferented subjects, cerebellar patients, and healthy individuals. This comparison may be performed regarding generalization patterns to untrained hand or to other contexts with the same hand, adaptation rate, wash-out rate, etc. It has been shown that deafferented individuals were able to adapt their reaches to altered visual feedback of the hand (Ingram et al., 2000; Bernier et al., 2006; Miall and Cole, 2007). Adaptation in these subjects may show different features compared to healthy ones.

REFERENCES

Bernier, P.-M., Chua, R., Bard, C., and Franks, I. M. (2006). Updating of an internal model without proprioception: a deafferentation study. *Neuroreport* 17, 1421–1425. doi: 10.1097/01.wnr.0000233096.13032.34

Cressman, E. K., and Henriques, D. Y. (2009). Sensory recalibration of hand position following visuomotor adaptation. *J. Neurophysiol.* 102, 3505–3518. doi: 10.1152/jn.00514.2009

Criscimagna-Hemminger, S. E., Bastian, A. J., and Shadmehr, R. (2010). Size of error affects cerebellar contributions to motor learning. *J. Neurophysiol.* 103, 2275–2284. doi: 10.1152/jn.00822.2009

Criscimagna-Hemminger, S. E., Donchin, O., Gazzaniga, M. S., and Shadmehr, R. (2003). Learned dynamics of reaching movements generalize from dominant to nondominant arm. *J. Neurophysiol.* 89, 168–176. doi: 10.1152/jn.00622.2002

Darainy, M., Vahdat, S., and Ostry, D. J. (2013). Perceptual learning in sensorimotor adaptation. *J. Neurophysiol.* 110, 2152–2162. doi: 10.1152/jn.00439.2013

Diedrichsen, J., White, O., Newman, D., and Lally, N. (2010). Use-dependent and error-based learning of motor behaviors. *J. Neurosci.* 30, 5159–5166. doi: 10.1523/JNEUROSCI.5406-09.2010

Farrer, C., Franck, N., Paillard, J., and Jeannerod, M. (2003). The role of proprioception in action recognition. *Conscious. Cogn.* 12, 609–619. doi: 10.1016/S1053-8100(03)00047-3

Fuentes, C. T., Mostofsky, S. H., and Bastian, A. J. (2011). No proprioceptive deficits in autism despite movement-related sensory and execution impairments. *J. Autism Dev. Disord.* 41, 1352–1361. doi: 10.1007/s10803-010-1161-1

Gibo, T. L., Criscimagna-Hemminger, S. E., Okamura, A. M., and Bastian, A. J. (2013). Cerebellar motor learning: are environment dynamics more important than error size? *J. Neurophysiol.* 110, 322–333. doi: 10.1152/jn.00745.2012

Haruno, M., Wolpert, D. M., and Kawato, M. (2001). Mosaic model for sensorimotor learning and control. *Neural Comput.* 13, 2201–2220. doi: 10.1162/089976601750541778

Imamizu, H., Kuroda, T., Miyauchi, S., Yoshioka, T., and Kawato, M. (2003). Modular organization of internal models of tools in the human cerebellum. *Proc. Natl. Acad. Sci. U.S.A.* 100, 5461–5466. doi: 10.1073/pnas.0835746100

Imamizu, H., Kuroda, T., Yoshioka, T., and Kawato, M. (2004). Functional magnetic resonance imaging examination of two modular architectures for switching multiple internal models. *J. Neurosci.* 24, 1173–1181. doi: 10.1523/JNEUROSCI.4011-03.2004

Ingram, H. A., van Donkelaar, P., Cole, J., Vercher, J. L., Gauthier, G. M., and Miall, R. C. (2000). The role of proprioception and attention in a visuomotor adaptation task. *Exp. Brain Res.* 132, 114–126. doi: 10.1007/s002219900322

Izawa, J., Criscimagna-Hemminger, S. E., and Shadmehr, R. (2012). Cerebellar contributions to reach adaptation and learning sensory consequences of action. *J. Neurosci.* 32, 4230–4239. doi: 10.1523/JNEUROSCI.6353-11.2012

Kluzik, J., Diedrichsen, J., Shadmehr, R., and Bastian, A. J. (2008). Reach adaptation: what determines whether we learn an internal model of the tool or adapt the model of our arm? *J Neurophysiol* 100, 1455–1464. doi: 10.1152/jn.90334.2008

Malfait, N., and Ostry, D. J. (2004). Is interlimb transfer of force-field adaptation a cognitive response to the sudden introduction of load? *J. Neurosci.* 24, 8084–8089. doi: 10.1523/JNEUROSCI.1742-04.2004

Miall, R. C., and Cole, J. (2007). Evidence for stronger visuo-motor than visuo-proprioceptive conflict during mirror drawing performed by a deafferented subject and control subjects. *Exp. Brain Res.* 176, 432–439. doi: 10.1007/s00221-006-0626-0

Shadmehr, R., and Krakauer, J. W. (2008). A computational neuroanatomy for motor control. *Exp. Brain Res.* 185, 359–381. doi: 10.1007/s00221-008-1280-5

Smith, M. A., and Shadmehr, R. (2005). Intact ability to learn internal models of arm dynamics in Huntington's disease but not cerebellar degeneration. *J. Neurophysiol.* 93, 2809–2821. doi: 10.1152/jn.00943.2004

van Beers, R. J., Wolpert, D. M., and Haggard, P. (2002). When feeling is more important than seeing in sensorimotor adaptation. *Curr. Biol.* 12, 834–837. doi: 10.1016/S0960-9822(02)00836-9

Wilson, E. T., Wong, J., and Gribble, P. L. (2010). Mapping proprioception across a 2D horizontal workspace. *PLoS ONE* 5:e11851. doi: 10.1371/journal.pone.0011851

Yavari, F., Towhidkhah, F., and Ahmadi-Pajouh, M. A. (2013). Are fast/slow process in motor adaptation and forward/inverse internal model two sides of the same coin? *Med. Hypotheses* 81, 592–600. doi: 10.1016/j.mehy.2013.07.009

Received: 22 January 2014; accepted: 22 February 2014; published online: 11 March 2014.

Citation: Yavari F, Towhidkhah F and Darainy M (2014) A hypothesis on the role of perturbation size on the human sensorimotor adaptation. Front. Comput. Neurosci. ***8****:28. doi: 10.3389/fncom.2014.00028*

This article was submitted to the journal Frontiers in Computational Neuroscience.

frontiers in
COMPUTATIONAL NEUROSCIENCE

OPINION ARTICLE
published: 09 April 2014
doi: 10.3389/fncom.2014.00040

Artificial neural networks: powerful tools for modeling chaotic behavior in the nervous system

Malihe Molaie[1], Razieh Falahian[1], Shahriar Gharibzadeh[1], Sajad Jafari[1]* and Julien C. Sprott[2]

[1] *Department of Bioelectric, Biomedical Engineering Faculty, Amirkabir University of Technology, Tehran, Iran*
[2] *Department of Physics, University of Wisconsin, Madison, Wisconsin, WI, USA*
**Correspondence: sajadjafari@aut.ac.ir*

Edited and reviewed by:
Tobias Alecio Mattei, Ohio State University, USA

Keywords: artificial neural networks, biological systems, electroretinogram, chaos, bifurcation diagram

Modeling real-world systems plays a pivotal role in their analysis and contributes to a better understanding of their behavior and performance. Classification, optimization, control, and pattern recognition problems rely heavily on modeling techniques. Such models can be categorized into three classes: white-box, black-box, and gray-box (Nelles, 2001). White-box models are fully derived from first principles, i.e., physical, chemical, biological, economical, etc. laws. All equations and parameters are determined from theory. Black-box models are based solely on experimental data, and their structure and parameters are determined by experimental modeling. Building black-box models requires little or no prior knowledge of the system. Gray-box models represent a compromise or combination of white-box and black-box models (Nelles, 2001).

In the modeling of highly nonlinear and complex phenomena, we may not have a good understanding of the processes, and thus black-box models may be our best (or even our only) choice. Artificial neural networks (ANNs) are one of the most powerful and popular tools for black-box modeling and are designed and inspired by real biological neural networks.

There has been an increasing interest in analyzing neurophysiology from a nonlinear and chaotic systems viewpoint in recent years (Christini and Collins, 1995; Sarbadhikari and Chakrabarty, 2001; Korn and Faure, 2003; Hadaeghi et al., 2013; Jafari et al., 2013; Mattei, 2013). For example, although the famous Hodgkin and Huxley model (Hodgkin and Huxley, 1952) has been the basis of almost all of the proposed models for neural firing, the Rose-Hindmarsh model (Hindmarsh and Rose, 1984) is known to be a more refined model since it can show different firing patterns, especially chaotic bursts of action potential, which enable a proper matching between this model behavior and experimental data. Another example of the observation of chaotic behavior in the nervous system is the period-doubling route to chaos in flicker vision (Crevier and Meister, 1998), which is the focus of this letter.

Stimulation with periodic flashes of light is useful for distinguishing some disorders of the human visual system (Crevier and Meister, 1998). It has been shown by Crevier and Meister (1998) that during electroretinogram (ERG) recordings of the visual system, period-doubling can occur. It is well-known that period-doubling occurs in nonlinear dynamical systems, and it is often associated with the onset of chaos. In one study (Crevier and Meister, 1998) the retina of a salamander was stimulated with a periodic square-wave flashes, and the ERG was recorded. The flash frequency was changed between zero and 30 Hz, while the contrast was constant. In another record, the contrast was changed while the frequency was fixed at 16 Hz. All the ERG signals were filtered at 1–1000 Hz. Using a common approach to obtain a discrete time series from a continuous recorded signal, successive local maxima of the signal were extracted as a time series (**Figure 1A**). As shown in **Figures 1B,C**, both the parameters (flash frequency and contrast) have a great effect on the recorded ERG signals and cause bifurcations resulting in a period-doubling route to chaos.

However, it is difficult to understand the exact relations between the parameters and their effects. In other words, it is not easy to build a white-box model that can regenerate the signals and diagrams accurately. That may be because of the highly complex and nonlinear dynamics involved. We have used the ability of an ANN in learning highly nonlinear dynamics as a black-box model of this system. We used a four hidden layer feed-forward neural network with (7/4/8/5) neurons in the layers (**Figure 1D**) and nonlinear transfer functions hyperbolic tangent function that help the network learn the complex relationships between input and output. The activation function of the last layer of the network is linear transfer function. We used two parameters (contrast and frequency) and three time delays (x_{n-1}, x_{n-2}, and x_{n-3}) as the inputs of the ANN to fit each data point of the time series (x_n) as the output of the network.

As shown in **Figures 1E,F**, this model can generate bifurcation diagrams similar to those obtained from real data. As the result, we believe that ANNs are powerful tools for modeling highly nonlinear behavior in the nervous system. We plan to construct ANN models in future work including extension to more cases and details, extension of the ideas in Hadaeghi et al. (2013) to patients with bipolar disorder, and extension of the ideas in Jafari et al. (2013) to patients with attention deficit hyperactivity disorder (ADHD).

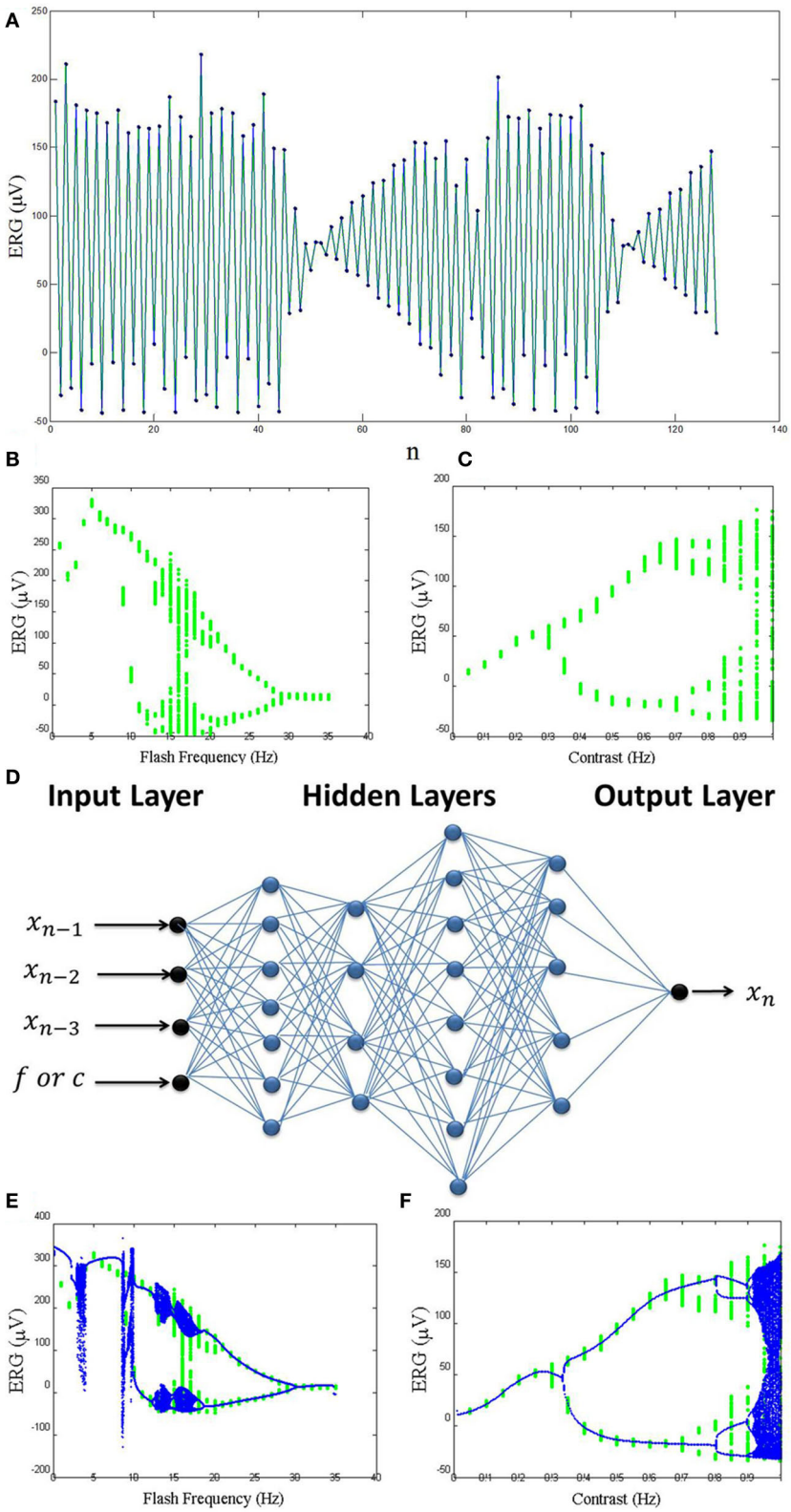


FIGURE 1 | (A) One example of the local maxima of the ERG signals. **(B)** Real bifurcation diagram resulted from varying flash frequency. **(C)** Real bifurcation diagram resulted from varying contrast. **(D)** The structure of the ANNs that were used. **(E)** Artificial bifurcation diagram resulted from varying the flash frequency input in the ANN. **(F)** Artificial bifurcation diagram resulted from varying the contrast input in the ANN.

ACKNOWLEDGMENTS

The authors would like to thank Professor Markus Meister for allowing us to use his data.

REFERENCES

Christini, D. J., and Collins, J. J. (1995). Controlling nonchaotic neuronal noise using chaos control techniques. *Phys. Rev. Lett.* 75, 2782–2785. doi: 10.1103/PhysRevLett.75.2782

Crevier, D. W., and Meister, M. (1998). Synchronous period-doubling in flicker vision of salamander and man. *J. Neurophysiol.* 79, 1869–1878.

Hadaeghi, F., Hashemi Golpayegani, M., and Moradi, K. (2013). Does "Crisis-Induced Intermittency" explain bipolar disorder dynamics? *Front. Comput. Neurosci.* 7:116. doi: 10.3389/fncom.2013.00116

Hindmarsh, J. L., and Rose, R. M. (1984). A model of neuronal bursting using three coupled first order differential equations. *Proc. R. Soc. Lond. B Biol. Sci.* 221, 87–102. doi: 10.1098/rspb.1984.0024

Hodgkin, A. L., and Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol.* 117, 500–544.

Jafari, S., Baghdadi, G., Hashemi Golpayegani, S. M. R., Towhidkhah, F., and Gharibzadeh, S. (2013). Is attention deficit hyperactivity disorder a kind of intermittent chaos? *J. Neuropsychiatry Clin. Neurosci.* 25, E02. doi: 10.1176/appi.neuropsych.12040079

Korn, H., and Faure, P. (2003). Is there chaos in the brain? II. Experimental evidence and related models. *C.R. Biol.* 326, 787–840. doi: 10.1016/j.crvi.2003.09.011

Mattei, T. A. (2013). Nonlinear (chaotic) dynamics and fractal analysis: new applications to the study of the microvascularity of gliomas. *World Neurosurg.* 79, 4–7 doi: 10.1016/j.wneu.2012.11.047

Nelles, O. (2001). *Nonlinear System Identification: From Classical Approaches to Neural Networks and Fuzzy Models*. Berlin, Heidelberg: Springer-Verlag. doi: 10.1007/978-3-662-04323-3_1

Sarbadhikari, S. N., and Chakrabarty, K. (2001). Chaos in the brain: a short review alluding to epilepsy, depression, exercise and lateralization. *Med. Eng. Phys.* 23, 445–455. doi: 10.1016/S1350-4533(01)00075-3

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.

Received: 13 March 2014; accepted: 21 March 2014; published online: 09 April 2014.

Citation: Molaie M, Falahian R, Gharibzadeh S, Jafari S and Sprott JC (2014) Artificial neural networks: powerful tools for modeling chaotic behavior in the nervous system. Front. Comput. Neurosci. ***8****:40. doi: 10.3389/fncom.2014.00040*

This article was submitted to the journal Frontiers in Computational Neuroscience.

frontiers in
COMPUTATIONAL NEUROSCIENCE

OPINION ARTICLE
published: 21 July 2014
doi: 10.3389/fncom.2014.00073

Synchrony analysis: application in early diagnosis, staging and prognosis of multiple sclerosis

Zahra Ghanbari and Shahriar Gharibzadeh*

Department of Biomedical Engineering, Amirkabir University of Technology, Tehran, Iran
**Correspondence: gharibzadeh@aut.ac.ir*

Edited by:
Tobias Alecio Mattei, Ohio State University, USA

Reviewed by:
Al-Rahim Abbasali Tailor, The Ohio State University Wexner Medical Center, USA

Keywords: multiple sclerosis (MS), synchrony, early diagnosis, staging, prognosis, prediction

Multiple Sclerosis (MS) is an autoimmune disease caused by degeneration of the myelin sheath of large diameter fibers in the central nervous system. This will cause deficits in the conducting properties of nerves and also affect electrical signaling. As a result, in MS patients, nerve conduction will be slower than normal (Kandel et al., 2000).

Neural synchrony has been of great interest in neuroscience recently. In signal processing, synchrony refers to quantifying similarity, coherence or correlation among signals and could be measured using a variety of methods (Dauwels et al., 2010). Neural synchrony represents how synchronous the neurons are firing (Vialatte et al., 2008). It is proven that synchrony is an important feature of brain signals. Many neurological diseases are accompanied by abnormalities in neural synchrony (Dauwels et al., 2008).

For example, loss of synchrony among brain signals has been observed in disorders such as Parkinson's and Alzheimer's disease (AD) and was used for the purpose of diagnosis. On the other hand, increasing synchrony has been reported in disorders such as epileptic seizures (Vialatte et al., 2009).

Since perturbation in electrical signaling and slowing of nerve conduction are common among MS and the aforementioned diseases, it brings up the idea of using synchrony for MS as well. In addition, previous works on MS have reported loss of connectivity and synchronous function among different parts of patients' brains. It should be mentioned that most of the previous works were concentrated on the cognitive impairments caused by the disease, and they applied their methods on MEG signals (Arrondo et al., 2009; Hardmeier et al., 2012).

The other point which should be noted is that although MRI and ERP are both common tools in MS diagnosis and follow up, definite diagnosis cannot be made based on these criteria individually. In addition, MRI needs to be repeated (Greenberg et al., 2009; Longo et al., 2012) and it is not affordable and available in many situations. So, we should try to find a reliable solution.

According to the aforementioned points, we believe that recording electrical brain signals (particularly EEG and ERP) and calculating local and global synchrony among their channels may provide us with an individual tool for diagnosing MS. Actually, the idea we put forward is using calculated synchrony indices for the purpose of detection, classification and prediction on electrical brain signals. Of course, the previous results which investigated connectivity and synchronous function of brain parts support our idea (Arrondo et al., 2009; Hardmeier et al., 2012).

The proposed idea may also help us to detect MS in early stages. Additionally, we believe as impairments will increase by progression of the disease, synchrony measures may have significant differences in different stages of the disease. So, they could be useful for staging of the disease as well.

We also propose measuring synchrony among brain signals in the onset periods. It seems that there should be a correlation between the changes in synchrony measures and disease prognosis. In better words, based on the calculated synchrony indices, we can predict the trend of the disease. This would provide us with a clearer perspective of the possible efficiency of different management modalities (including medical and surgical). Additionally, based on the potential level of neural dyssynchrony the proposed idea can be useful in order to assess the efficiency of the selected treatments for both the patient and the physician. Surely experimental evaluations are needed to validate our hypothesis.

REFERENCES

Arrondo, G., Alegre, M., Sepulcre, J., Iriarte, J., Artieda, J., and Villoslada, P. (2009). Abnormalities in brain synchronization are correlated with cognitive impairment in multiple sclerosis. *Mult. Scler.* 15, 509–516. doi: 10.1177/1352458508101321

Dauwels, J., Vialatte, F., and Cichocki, A. (2008). "Quantifying statistical synchrony: algorithms and applications to brain data analysis and early prediction of Alzheimer's disease," in *Proceedings of the 3rd INFORMS Workshop on Data Mining and Health Informatics (DM-HI 2008)*, eds J. Li, D. Aleman, and R. Sikora.

Dauwels, J., Vialatte, F., Musha, T., and Cichocki, A. (2010). A comparative study of synchrony measures for the early diagnosis of Alzheimer's disease based on EEG. *Neuroimage* 49, 668–693. doi: 10.1016/j.neuroimage.2009.06.056

Greenberg, D. A., Aminoff, M., and Simon, M. R. (2009). *Clinical Neurology*. New York, NY: McGraw Hill.

Hardmeier, H., Schoonheim, M., Geurts, J., Hillebrand, A., Polman, C., Barkhof, F., et al. (2012). Cognitive dysfunction in early multiple sclerosis: altered centrality derived from resting-state functional connectivity using magneto-encephalography. *PLoS ONE* 7:e42087. doi: 10.1371/journal.pone.0042087

Kandel, E. R., Schwartz, J. H., and Jessell, T. M. (2000). *Principles of Neural Science, 4th Edn.* New York, NY: McGraw-Hill Companies.

Longo, D. L., Fauci, A., Kasper, D., Hauser, S., and Jameson, J. (2012). *Harrison's Principles of Internal Medicine*. New York, NY: McGraw Hill.

Vialatte, F., Dauwels, J., Rutkowski, T., and Cichocki, A. (2008). "Measuring neural synchrony by message passing," in *Advances in Neural Information Processing Systems* (Washington, DC), 361–368.

Vialatte, F., Sole-Casals, J., Dauwels, J., Maurice, M., and Cichocki, A. (2009). Bump time-frequency toolbox: a toolbox for time-frequency oscillatory bursts extraction in electrophysiological signals. *BMC Neurosci.* 10:46. doi: 10.1186/1471-2202-10-46

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.

Received: 26 April 2014; accepted: 27 June 2014; published online: 21 July 2014.
Citation: Ghanbari Z and Gharibzadeh S (2014) Synchrony analysis: application in early diagnosis, staging and prognosis of multiple sclerosis. Front. Comput. Neurosci. ***8****:73. doi: 10.3389/fncom.2014.00073*

This article was submitted to the journal Frontiers in Computational Neuroscience.

frontiers in
COMPUTATIONAL NEUROSCIENCE

OPINION ARTICLE
published: 21 July 2014
doi: 10.3389/fncom.2014.00077

The hypothetical cost-conflict monitor: is it a possible trigger for conflict-driven control mechanisms in the human brain?

Sareh Zendehrouh *, Shahriar Gharibzadeh and Farzad Towhidkhah

Biomedical Engineering Faculty, Amirkabir University of Technology, Tehran, Iran
**Correspondence: sareh.zendehrouh@gmail.com*

Edited by:
Tobias Alecio Mattei, Ohio State University, USA

Reviewed by:
Tobias Alecio Mattei, Ohio State University, USA
Carlos Rodrigo Goulart, Ohio State University, USA

Keywords: performance monitoring, cognitive control, conflict-driven control, monitor-controller networks, feedback-related negativity

Flexible goal-directed behavior requires a performance monitoring system to monitor behavioral consequences in order to detect the need for further adjustments and control. When a failure in performance is detected by the monitoring system, some signals are transmitted to the brain structures responsible for control implementation. Evidences suggest the anterior cingulate cortex (ACC) (Carter et al., 1998; Gehring and Knight, 2000; MacDonald et al., 2000; Ferdinand et al., 2012) and the lateral prefrontal cortex (lPFC) (MacDonald et al., 2000; Ridderinkhof et al., 2004a,b) as the neural correlates of performance monitoring and control implementation systems, respectively. The interaction of these two systems appears to modulate some components of event-related brain potentials (ERPs) linked with performance monitoring such as the error-related negativity (ERN), the N200, and the feedback-related negativity (FRN) (Gruendler et al., 2011). The ERN is an ERP component that begins close to the time of the erroneous response in speeded response time tasks and peaks about 100 ms later (Gehring et al., 1993). The N200 is another negative deflection in ERP that peaks between 200 and 400 ms after stimulus onset, prior to the response execution, on correct trials of cognitive control experiments (Olvet and Hajcak, 2008). The FRN as one of the most studied components is a negative-going deflection observed 230–330 ms following outcome presentation (Miltner et al., 1997) in gambling and trial-and-error learning tasks (Holroyd et al., 2006). Source localization studies show the neural source of the FRN to be located most probably in the ACC (Miltner et al., 1997; Gehring and Willoughby, 2002; Bellebaum and Daum, 2008; Hauser et al., 2014).

The central question in the interaction of performance monitoring and control systems is how the brain determines the need to recruit the intervention of control structures. The reinforcement learning (RL) account of performance monitoring and control is one of the influential theories to the field (Holroyd and Coles, 2002; Holroyd et al., 2005). The theory is based on the physiological evidences that reveal the similarity of the phasic activity of the mesencephalic dopamine system and reward prediction errors (RPEs) in temporal difference models of learning (Suri, 2002). The theory holds that the monitor is located in the basal ganglia, which produces RPE signals that indicate when events are better or worse than expected. These RPEs are used by the ACC to improve performance on the task at hand (Holroyd et al., 2005). According to the RL model, negative RPEs sent to the ACC generate the ERN and the FRN. Another prominent theory, the conflict-monitoring theory (CMT) proposes that the performance monitoring system monitors for the coactivation of mutually incompatible response tendencies or conflict during response selection. The CMT suggests that the ACC detects response-conflict signal and sends this information to the dorsolateral prefrontal cortex for further adjustment and control (Botvinick et al., 2001; Yeung et al., 2004). Based on this theory, the N2 and the ERN can be described using conflict signal. The CMT argues that the N2 and the ERN are electrophysiologically correlated with pre-response and post-response conflict signals, respectively. However, since no motor response exists after external feedback presentation, the CMT cannot account for the phenomena commencing after feedback onset, e.g., the FRN (Ullsperger et al., 2014). In our previous studies, we have explained the significance of integrating the computational models associated with the RL and the CMT (Zendehrouh et al., 2013, 2014). Since the unification of these two theories depends centrally on conflict signal definition, we propose a hypothetical cost-conflict monitor in the brain that extends the CMT theory to account for post feedback activities in feedback-based learning tasks. Based on this proposal, the FRN can be described using a cost-conflict signal.

The basis for our hypothetical cost-conflict monitor is that: (1) Theoretically, conflict can occur anywhere within the information processing system (Carter and van Veen, 2007). (2) Conflict-driven control is domain-specific suggested to be mediated by multiple, independent, and parallel-operating conflict monitor-controller loops in the brain (Egner, 2008). (3) The appraisal of costs and benefits associated with different candidate actions is a key aspect of decision-making.

The Delay-based and the effort-based costs (effort needed to perform an action in order to obtain a reward) are two types of costs that bias decision making (Floresco et al., 2008). In delay-based

tasks, as the time passes, the subjective value of a reward is discounted hyperbolically (Green and Myerson, 2004). Also, the aversiveness of a negative event decreases hyperbolically with time (Murphy et al., 2001). Evidences suggest that discounting can happen across many reward types, reward magnitudes, and several timescales even in the order of tens of milliseconds (Haith et al., 2012). In this paper, it is hypothesized that in feedback-based learning tasks, the participants are faced with delay-based evaluations. Therefore, in these tasks, the time interval between response selection and feedback presentation gives rise to a cost. This delay elevates the cost of the rewarded outcome and reduces the cost of the non-rewarded outcome associated with the selected action. In fact, the conflict can be produced by simultaneous activation of the expected costs of possible outcomes that are mutually exclusive. Therefore, when a cost-conflict is detected by the monitoring system, the regulatory mechanism implements the required control, e.g., by modifying the excitatory weights to the response units. The cost-conflict signal that may occur between expected costs can show the amount of subjective transient uncertainty about what will happen that increases with time (delay) until receiving the actual outcome. The cost-conflict signal can also be viewed in the context of the emerging field of neuroeconomics as an ambiguity signal that may be present during decision-making. Ambiguity is defined as a lack of confidence in probability assignment to the possible outcomes (Kishida et al., 2010). This is consistent with investigations suggesting the existence of an ambiguity-sensitive mechanism in the ventromedial prefrontal cortex (vmPFC) (Glimcher and Rustichini, 2004), and also with the role of this area in delay cost coding (Prévost et al., 2010; Rushworth et al., 2011; Dreher, 2013).

This proposal can be validated by performing simple gambling games or probabilistic reinforcement learning tasks with feedback-timing manipulations at the timescale of milliseconds while measuring the brain responses with functional magnetic resonance imaging (fMRI) and electroencephalography (EEG) to identify the contributions of the ACC and the vmPFC in those tasks. Especially, the behaviors of addicted and depressed individuals in these tasks that show anomalies in value based decision making (Sharp et al., 2012) can be beneficial.

Therefore, the cost-conflict monitor as an independent and parallel loop to the response-conflict monitor detects the conflict between the costs of likely outcomes of the selected action and uses this information to adjust the behavior for the future, thereby implements trial-by-trial adjustments. Surely, this proposal is speculative and further experimental studies and research is needed to evaluate its merit. However, the proposal can provide promising avenues toward the unification of computational models associated with the RL and the CMT.

REFERENCES

Bellebaum, C., and Daum, I. (2008). Learning-related changes in reward expectancy are reflected in the feedback-related negativity. *Eur. J. Neurosci.* 27, 1823–1835. doi: 10.1111/j.1460-9568.2008.06138.x

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., and Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychol. Rev.* 108, 624–652. doi: 10.1037/0033-295X.108.3.624

Carter, C. S., Braver, T. S., Barch, D. M., Botvinick, M. M., Noll, D., and Cohen, J. D. (1998). Anterior cingulate cortex, error detection, and the online monitoring of performance. *Science* 280, 747–749. doi: 10.1126/science.280.5364.747

Carter, C. S., and van Veen, V. (2007). Anterior cingulate cortex and conflict detection: an update of theory and data. *Cogn. Affect. Behav. Neurosci.* 7, 367–379. doi: 10.3758/CABN.7.4.367

Dreher, J.-C. (2013). Neural coding of computational factors affecting decision making. *Prog. Brain Res.* 202, 289–320. doi: 10.1016/B978-0-444-62604-2.00016-2

Egner, T. (2008). Multiple conflict-driven control mechanisms in the human brain. *Trends Cogn. Sci.* 12, 374–380. doi: 10.1016/j.tics.2008.07.001

Ferdinand, N. K., Mecklinger, A., Kray, J., and Gehring, W. J. (2012). The processing of unexpected positive response outcomes in the mediofrontal cortex. *J. Neurosci.* 32, 12087–12092. doi: 10.1523/JNEUROSCI.1410-12.2012

Floresco, S. B., St Onge, J. R., Ghods-Sharifi, S., and Winstanley, C. A. (2008). Cortico-limbic-striatal circuits subserving different forms of cost-benefit decision making. *Cogn. Affect. Behav. Neurosci.* 8, 375–389. doi: 10.3758/CABN.8.4.375

Gehring, W. J., Goss, B., Coles, M. G. H., Meyer, D. E., and Donchin, E. (1993). A neural system for error detection and compensation. *Psychol. Sci.* 4, 385–390. doi: 10.1111/j.1467-9280.1993.tb00586.x

Gehring, W. J., and Knight, R. T. (2000). Prefrontal-cingulate interactions in action monitoring. *Nat. Neurosci.* 3, 516–520. doi: 10.1038/74899

Gehring, W. J., and Willoughby, A. R. (2002). The medial frontal cortex and the rapid processing of monetary gains and losses. *Science* 295, 2279–2282. doi: 10.1126/science.1066893

Glimcher, P. W., and Rustichini, A. (2004). Neuroeconomics: the consilience of brain and decision. *Science* 306, 447–452. doi: 10.1126/science.1102566

Green, L., and Myerson, J. (2004). A discounting framework for choice with delayed and probabilistic rewards. *Psychol. Bull.* 130, 769–792. doi: 10.1037/0033-2909.130.5.769

Gruendler, T. O. J., Ullsperger, M., and Huster, R. J. (2011). Event-related potential correlates of performance-monitoring in a lateralized time-estimation task. *PloS ONE* 6:e25591. doi: 10.1371/journal.pone.0025591

Haith, A. M., Reppert, T. R., and Shadmehr, R. (2012). Evidence for hyperbolic temporal discounting of reward in control of movements. *J. Neurosci.* 32, 11727–11736. doi: 10.1523/JNEUROSCI.0424-12.2012

Hauser, T. U., Iannaccone, R., Stämpfli, P., Drechsler, R., Brandeis, D., Walitza, S., et al. (2014). The feedback-related negativity (FRN) revisited: new insights into the localization, meaning and network organization. *Neuroimage* 84, 159–168. doi: 10.1016/j.neuroimage.2013.08.028

Holroyd, C. B., and Coles, M. G. H. (2002). The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. *Psychol. Rev.* 109, 679–709. doi: 10.1037/0033-295X.109.4.679

Holroyd, C. B., Hajcak, G., and Larsen, J. T. (2006). The good, the bad and the neutral: electrophysiological responses to feedback stimuli. *Brain Res.* 1105, 93–101. doi: 10.1016/j.brainres.2005.12.015

Holroyd, C. B., Yeung, N., Coles, M. G. H., and Cohen, J. D. (2005). A mechanism for error detection in speeded response time tasks. *J. Exp. Psychol. Gen.* 134, 163–191. doi: 10.1037/0096-3445.134.2.163

Kishida, K. T., King-Casas, B., and Montague, P. R. (2010). Neuroeconomic approaches to mental disorders. *Neuron* 67, 543–554. doi: 10.1016/j.neuron.2010.07.021

MacDonald, A. W. 3rd., Cohen, J. D., Stenger, V. A., and Carter, C. S. (2000). Dissociating the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control. *Science* 288, 1835–1838. doi: 10.1126/science.288.5472.1835

Miltner, W. H. R., Braun, C. H., and Coles, M. G. H. (1997). Event-related brain potentials following incorrect feedback in a time-estimation task: evidence for a generic neural system for error detection. *J. Cogn. Neurosci.* 9, 788–798. doi: 10.1162/jocn.1997.9.6.788

Murphy, J., Vuchinich, R., and Simpson, C. (2001). Delayed reward and cost discounting. *Psychol. Rec.* 51, 571–588.

Olvet, D. M., and Hajcak, G. (2008). The error-related negativity (ERN) and psychopathology: toward an endophenotype. *Clin. Psychol. Rev.* 28, 1343–1354. doi: 10.1016/j.cpr.2008.07.003

Prévost, C., Pessiglione, M., Météreau, E., Cléry-Melin, M.-L., and Dreher, J.-C. (2010). Separate valuation subsystems for delay and effort decision costs. *J. Neurosci.* 30, 14080–14090. doi: 10.1523/JNEUROSCI.2752-10.2010

Ridderinkhof, K. R., Ullsperger, M., Crone, E. A., and Nieuwenhuis, S. (2004a). The role of the medial frontal cortex in cognitive control. *Science* 306, 443–447. doi: 10.1126/science.1100301

Ridderinkhof, K. R., van den Wildenberg, W. P. M., Segalowitz, S. J., and Carter, C. S. (2004b). Neurocognitive mechanisms of cognitive control: the role of prefrontal cortex in action selection, response inhibition, performance monitoring, and reward-based learning. *Brain Cogn.* 56, 129–140. doi: 10.1016/j.bandc.2004.09.016

Rushworth, M. F. S., Noonan, M. P., Boorman, E. D., Walton, M. E., and Behrens, T. E. (2011). Frontal cortex and reward-guided learning and decision-making. *Neuron* 70, 1054–1069. doi: 10.1016/j.neuron.2011.05.014

Sharp, C., Monterosso, J., and Montague, P. R. (2012). Neuroeconomics: a bridge for translational research. *Biol. Psychiatry* 72, 87–92. doi: 10.1016/j.biopsych.2012.02.029

Suri, R. E. (2002). TD models of reward predictive responses in dopamine neurons. *Neural Netw.* 15, 523–533. doi: 10.1016/S0893-6080(02)00046-1

Ullsperger, M., Danielmeier, C., and Jocham, G. (2014). Neurophysiology of performance monitoring and adaptive behavior. *Physiol. Rev.* 94, 35–79. doi: 10.1152/physrev.00041.2012

Yeung, N., Cohen, J. D., and Botvinick, M. M. (2004). The neural basis of error detection: conflict monitoring and the error-related negativity. *Psychol. Rev.* 111, 931–959. doi: 10.1037/0033-295X.111.4.931

Zendehrouh, S., Gharibzadeh, S., and Towhidkhah, F. (2013). Modeling error detection in human brain: a preliminary unification of reinforcement learning and conflict monitoring theories. *Neurocomputing* 103, 1–13. doi: 10.1016/j.neucom.2012.04.026

Zendehrouh, S., Gharibzadeh, S., and Towhidkhah, F. (2014). Reinforcement-conflict based control: an integrative model of error detection in anterior cingulate cortex. *Neurocomputing* 123, 140–149. doi: 10.1016/j.neucom.2013.06.020

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.

Received: 17 May 2014; accepted: 30 June 2014; published online: 21 July 2014.

Citation: Zendehrouh S, Gharibzadeh S and Towhidkhah F (2014) The hypothetical cost-conflict monitor: is it a possible trigger for conflict-driven control mechanisms in the human brain? Front. Comput. Neurosci. ***8****:77. doi: 10.3389/fncom.2014.00077*

This article was submitted to the journal Frontiers in Computational Neuroscience.

frontiers in
COMPUTATIONAL NEUROSCIENCE

OPINION ARTICLE
published: 22 July 2014
doi: 10.3389/fncom.2014.00072

Modeling studies for designing transcranial direct current stimulation protocol in Alzheimer's disease

Shirin Mahdavi, Fatemeh Yavari, Shahriar Gharibzadeh* and Farzad Towhidkhah

Department of Biomedical Engineering, Amirkabir University of Technology, Tehran, Iran
**Correspondence: gharibzadeh@aut.ac.ir*

Edited by:
Tobias Alecio Mattei, Ohio State University, USA

Reviewed by:
Al-Rahim Abbasali Tailor, The Ohio State University Wexner Medical Center, USA

Keywords: brain stimulation, transcranial direct current stimulation (tDCS), computational modeling, finite element model, human head model, Alzheimer's disease

Transcranial direct current stimulation (tDCS) has been proposed as a technique for brain activity modulation. In this technique, a weak current (usually 1–2 mA) is delivered to scalp through two sponge electrodes. There are two types of tDCS stimulation: cathodal and anodal, which inhibit and facilitate neuronal activity, respectively (Hansen, 2012).

tDCS has been shown to be effective in Alzheimer's disease (AD). Several studies have revealed that tDCS application can improve memory performance in Alzheimer's patients (APs) (Ferrucci et al., 2008; Boggio et al., 2009, 2012). For example, results of a single session tDCS study (Ferrucci et al., 2008) revealed that anodal/cathodal tDCS significantly enhanced/worsened word recognition in AD patients. In another study, application of anodal stimulation over DLPFC of APs has led to recognition memory improvement in a visual memory task (Boggio et al., 2009). These effects seem to be persistent, as in a multi-session tDCS study (Boggio et al., 2012), improvement in patients' visual recognition lasted for 4 weeks.

Current pathway through brain plays a key role in the observed effects. Currently, modeling studies provide the only way for determining the pattern of current flow during tDCS. In recent years, finite element modeling has been suggested as a reliable and helpful tool in clinical therapeutic applications (Bikson et al., 2012).

A critical issue which is required to be considered in modeling studies is the inter-individual anatomical variations. A modeling study has shown the profound role of individual cortical morphology in determination of current flow distribution for healthy people (Datta et al., 2012). Also the impact of pathologic anatomy (skull defects and lesions) on modulation of current flow has been examined in some previous studies (Datta et al., 2010, 2011). Specifically, in AD loss of neuronal structures and synaptic damages result in cortex shrinkage and ventricular enlargement (Frisoni et al., 2010). This changes the volume of CSF- referred as "super highway" for current flow- and therefore can significantly alters current pathway in these patients' head compared to healthy subjects (Bikson et al., 2012). These studies suggest that it is not precise to determine the dosage of applied current only based on healthy human modeling or clinical trial outcomes.

We hypothesize that change in cortical thickness due to brain atrophy has significant effects on current flow pattern. These anatomical alterations may shift the stimulated areas and peak current density location in head. They may even alter the expected results from tDCS application.

We suggest that cortical thickness is required to be considered in modeling studies to obtain more precise pattern of current flow in head and the stimulated brain regions. Specifically, AD affects differently on each patient's brain structure. We suggest developing individualized models based on each patient's MRI data. These models can be used by clinicians to find the optimal electrode montage and current amplitude for each patient.

Using Individual-based models for designing clinical protocols could provide us with better interpretation of the results.

REFERENCES

Bikson, M., Rahman, A., Datta, A., Fregni, F., and Merabet, L. (2012). High-resolution modeling assisted design of customized and individualized transcranial direct current stimulation protocols. *Neuromodulation* 15, 306–315. doi: 10.1111/j.1525-1403.2012.00481.x

Boggio, P. S., Ferrucci, R., Mameli, F., Martins, D., Martins, O., Vergari, M., et al. (2012). Prolonged visual memory enhancement after direct current stimulation in Alzheimer's disease. *Brain Stimul.* 5, 223–230. doi: 10.1016/j.brs.2011.06.006

Boggio, P. S., Khoury, L. P., Martins, D. C., Martins, O. E., De Macedo, E. C., and Fregni, F. (2009). Temporal cortex direct current stimulation enhances performance on a visual recognition memory task in Alzheimer disease. *J. Neurol. Neurosurg. Psychiatry* 80, 444–447. doi: 10.1136/jnnp.2007.141853

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., Bikson, M., and Fregni, F. (2010). Transcranial direct current stimulation in patients with skull defects and skull plates: high-resolution computational FEM study of factors altering cortical current flow. *Neuroimage* 52, 1268–1278. doi: 10.1016/j.neuroimage.2010.04.252

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

Ferrucci, R., Mameli, F., Guidi, I., Mrakic-Sposta, S., Vergari, M., Marceglia, S., et al. (2008). Transcranial direct current stimulation improves recognition memory in Alzheimer disease. *Neurology* 71,

493–498. doi: 10.1212/01.wnl.0000317060.43722.a3

Frisoni, G. B., Fox, N. C., Jack, C. R. Jr., Scheltens, P., and Thompson, P. M. (2010). The clinical use of structural MRI in Alzheimer disease. *Nat. Rev. Neurol.* 6, 67–77. doi: 10.1038/nrneurol.2009.215

Hansen, N. (2012). Action mechanisms of transcranial direct current stimulation in Alzheimer's disease and memory loss. *Front. Psychiatry* 3:48. doi: 10.3389/fpsyt.2012.00048

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.

Received: 30 January 2014; accepted: 27 June 2014; published online: 22 July 2014.

Citation: Mahdavi S, Yavari F, Gharibzadeh S and Towhidkhah F (2014) Modeling studies for designing transcranial direct current stimulation protocol in Alzheimer's disease. Front. Comput. Neurosci. ***8****:72. doi: 10.3389/fncom.2014.00072*

This article was submitted to the journal Frontiers in Computational Neuroscience.

frontiers in
COMPUTATIONAL NEUROSCIENCE

OPINION ARTICLE
published: 05 January 2015
doi: 10.3389/fncom.2014.00170

Does our brain use the same policy for interacting with people and manipulating different objects?

Fatemeh Yavari*

Biomedical Engineering Department, Amirkabir University of Technology, Tehran, Iran
**Correspondence: f-yavari@aut.ac.ir*

Edited by:
Tobias Alecio Mattei, Ohio State University, USA

Reviewed by:
Da-Hui Wang, Beijing Normal University, China
Sen Song, Tsinghua University, USA

Keywords: internal forward model, internal inverse model, Modular organization, Schematic processing, Stereotypes

INTRODUCTION

Our first impression of other people is greatly affected by our previous experiences. Schematic processing, proposed in social psychology, explains our behavior in interacting with other people. It suggests existence of different schemas in our brain for different groups of people, e.g., extroverts, introverts, shy, women, men, etc. and also schemas related to special people like our parents, close friends, supervisor, and even ourselves. Each schema is recalled when we meet the corresponding person/personality (Atkinson, 1996).

On the other hand there is a relatively well accepted theory–model based theory- in motor control and learning studies (Daw and Dayan, 2014; Dayan and Berridge, 2014). It suggests existence of some internal models (forward and/or inverse) in the brain which help us for planning and execution of the actions.

Although these two viewpoints may seem very distinct, there are some interesting similarities between them, which are explained in the following section. I hypothesize that these correspondences may suggest that the brain employs same algorithms in dealing with both situations.

Understanding the brain function is a great challenge for many scientists. Further evaluation of the proposed hypothesis may be helpful to achieve better understanding of the brain function, as advances in each field may encourage new ideas in the other one.

In the following sections each of the two viewpoints and then their similarities are explained.

STEREOTYPES IN SOCIAL PSYCHOLOGY

Stereotype is defined as "a fixed, often simplistic generalization about a particular group or class of people (Cardwell, 2014)". Stereotypes, schemas, and schematic processing enable us to efficiently organize and process the huge volume of input information to our brain. Instead of processing every little detail about a new person, we can just recall the most similar schemas and generally categorize the person e.g., based on his most obvious physical features (Atkinson, 1996). Stereotypes enable us to respond rapidly in situations which we have had similar experience. Despite all the benefits, stereotypes may also result in prejudice. Since they bias our impressions, they can have very negative and even mortal (e.g., Amadou Diallo case) consequences (Atkinson, 1996).

INTERNAL MODELS IN MOTOR CONTROL AND LEARNING STUDIES

Internal models are defined as representations of external objects and/or our body organs in the brain (Kawato, 1999) (see Yavari et al., 2013 for a review). They are categorized into "forward" and "inverse" which mimic the "input-output" and "output-input" relationship of the related object/organ, respectively. Model-based theory suggests that motor learning/adaptation leads to formation/modification of internal models (Hunter et al., 2009). Kawato et al. have proposed co-existence of multiple pairs of internal forward-inverse models in the brain and therefore, a modular structure for motor control and learning (Wolpert and Kawato, 1998; Haruno et al., 2001, 2003; Doya et al., 2002; Imamizu et al., 2003; Wada et al., 2003). Based on this idea, which has been supported by different behavioral and imaging evidence (Wolpert and Kawato, 1998; Imamizu et al., 2003), there are an inverse (controller) and a forward (predictor) internal model within each pair. Contribution of each controller to the final motor command is determined based on accuracy of the linked forward model. This modular structure can explain our remarkable ability in motor learning, adaptation, and behavioral switching (Haruno et al., 2003).

SIMILARITIES OF THE TWO MENTIONED VIEWPOINTS

Some similarities between the two mentioned viewpoints are described here:

- Both processes are implicit and unconscious. Associations which are activated through stereotypes can be deeply learned and become automatic (as shown by priming-based experiments) (Rudman and Borgida, 1995; Atkinson, 1996; Bargh et al., 1996). Similarly, after enough practice, a motor skill (such as driving) can be performed unconsciously and without need to attention (Schmahmann, 1997).
- Based on primary effect, the initial information which we receive (e.g., hear) about a person significantly bias our impression of him/her. This effect has been explained using schematic processing as follows: we try to achieve a general impression about the person by searching for the most consistent schema or stereotype with the input information. This schema

determines our judgment about his/her personality.

There is a same process about internal models in motor control: when manipulating a new tool the most suitable FM/IM pair is activated based on context, e.g., by looking at the object's appearance, and the corresponding IM is used as controller. In the next trial, the pair which produced the least minimum prediction error will be activated and used (Wolpert et al., 2003).

- Stereotypes help us in inference, i.e., making judgment beyond the given information. For instance when we hear that someone is affectionate, we will probably consider him/her also a generous person (Atkinson, 1996). There is conceptually similar to generalization in motor learning which has been proposed to be resulted from internal models. An internal model formed by practicing a motor action under a special condition can partly be generalized to other circumstances. For example, practicing a movement with the right hand generates a learning which partly generalizes to the left hand (Sainburg and Wang, 2002; Wang and Sainburg, 2006; Balitsky Thompson and Henriques, 2010).
- One of the famous models in impression formation is the continuum model (Fiske et al., 1999) which describes the whole range of processes from stereotypes to individuation. Based on this model, automatic stereotypes are the first psychological process activated when we meet someone for the first time. We categorize this person unconsciously and automatically in terms of age, sex, and ethnicity. This is called initial categorization. If the person is important for us, we obtain more information about him (piecemeal integration) and finally judge him based on his individual characteristics (individuation). Proceeding from stereotypes toward individuation happens slowly (Atkinson, 1996).

Based on internal model theory learning a new motor skill goes through an almost similar process: When we try to manipulate a new object, in the early stage, CNS combines output signals from internal models of most similar (and familiar) objects. After some practice we learn to manipulate the new object skillfully and the reason is the special internal model which has been formed for it (Imamizu and Kawato, 2012). Depending on the complexity of the new motor task, its learning would need different time. It could take even years (e.g., for professional athletes).

As it can be seen in both situations, in a new condition reliance is more on previous experience, while gathering more information over time leads to formation of special new internal model/stereotype.

CONCLUSION

Human brain is probably the most fascinating creation in the world. Many scientists in different fields are trying to understand its function. Here I hypothesized that maybe our brain applies the same policy for some distinct applications, e.g., social interaction and manipulating different objects.

It worth mentioning that internal models have been proposed not only in motor control and learning, but also in some other fields such as control of mental activities (Ito, 2008), cognitive planning (Dayan and Yu, 2006), and decision making (Daw et al., 2011). These processes may even have more in common with stereotypes.

It would be interesting to also compare the corresponding neural substrates for stereotypes and internal models. Cell recording in some animal studies (Liu et al., 2003; Cerminara et al., 2009; Laurens et al., 2013) and also imaging studies (Imamizu et al., 2000, 2003; Blakemore et al., 2001; Kawato et al., 2003; Higuchi et al., 2007; Milner et al., 2007) suggest lateral and anterior cerebellum as the probable site of formation or storage of internal models. Some studies have suggested that motor cortex and other frontal motor areas have important roles in computation of internal models (Li et al., 2001; Shadmehr, 2004; Richardson et al., 2006; Shadmehr and Krakauer, 2008). Medial prefrontal cortex (mPFC) has been proposed as a candidate region in model-based evaluation (Hampton et al., 2006, 2008; Valentin et al., 2007; Daw et al., 2011). On the other hand, some neuroimaging studies have shown mPFC as a crucial region in social inferences, (Mitchell et al., 2005a,b, 2006), and judgments of warmth and competence (Harris and Fiske, 2006). Activity in middle mPFC is shown to be associated with thinking about either the self or a similar other (Ida Gobbini et al., 2004; Mitchell et al., 2006); while activity in dorsal mPFC is associated with thinking about a dissimilar other. Therefore, mPFC seems to be important for ingroup and outgroup perception (Amodio and Lieberman, 2009). Perceiving a person as a social being, which has been proposed to form the basis of prejudice (Qiu, 2006), has been suggested to be dependent on dorsal mPFC (Amodio and Lieberman, 2009). Therefore, PFC seems to be a crucial brain region for both internal models and stereotypes.

Further evaluation of the proposed hypothesis may be helpful to achieve better understanding of the brain function. For example as it was mentioned, stereotypes have significant effect on our social life and undeniable effect on impression formation. They sometimes have negative (even mortal) impact on our judgments, because they bias our impressions. The more we increase our knowledge about this concept, the more we can modify our thoughts in a good manner.

Discoveries in each field may lead to new findings in the other. For instance it has been shown that stereotypes may be activated through unconscious priming; e.g., in an experiment by Bargh et al. (1996) seeing images of young African American men triggered more aggressive behavior compared to images of young Caucasian men, even though the images were displayed for less than thirty thousandths seconds (subliminally) (Atkinson, 1996). This observation can be verified about motor actions as well. For example to investigate if seeing a special tool, such as a piano, can prime the piano playing skill. This can be both useful for better understanding the motor related mechanisms in the brain and also in practical applications such as preparing the athletes before their match to achieve better results.

The proposed hypothesis needs to be verified by some specially-designed experiments.

REFERENCES

Amodio, D. M., and Lieberman, M. D. (2009). "Pictures in our heads: Contributions of fMRI to the study of prejudice and stereotyping,"

in *Handbook of Prejudice, Stereotyping, and Discrimination* (New York, NY: Earlbaum), 347–366.

Atkinson, R. L. (1996). *Hilgard's Introduction to Psychology.* Philadelphia, PA: Harcourt Brace College Publishers.

Balitsky Thompson, A. K., and Henriques, D. Y. (2010). Visuomotor adaptation and intermanual transfer under different viewing conditions. *Exp. Brain Res.* 202, 543–552. doi: 10.1007/s00221-010-2155-0

Bargh, J. A., Chen, M., and Burrows, L. (1996). Automaticity of social behavior: direct effects of trait construct and stereotype activation on action. *J. Pers. Soc. Psychol.* 71:230. doi: 10.1037/0022-3514.71.2.230

Blakemore, S.-J., Frith, C. D., and Wolpert, D. M. (2001). The cerebellum is involved in predicting the sensory consequences of action. *Neuroreport* 12, 1879–1884. doi: 10.1097/00001756-200107030-00023

Cardwell, M. (2014). *Dictionary of Psychology.* Oxon: Routledge.

Cerminara, N. L., Apps, R., and Marple-Horvat, D. E. (2009). An internal model of a moving visual target in the lateral cerebellum. *J. Physiol.* 587, 429–442. doi: 10.1113/jphysiol.2008.163337

Daw, N. D., and Dayan, P. (2014). The algorithmic anatomy of model-based evaluation. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 369:20130478. doi: 10.1098/rstb.2013.0478

Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., and Dolan, R. J. (2011). Model-based influences on humans' choices and striatal prediction errors. *Neuron* 69, 1204–1215. doi: 10.1016/j.neuron.2011.02.027

Dayan, P., and Berridge, K. C. (2014). Model-based and model-free Pavlovian reward learning: revaluation, revision, and revelation. *Cogn. Affect. Behav. Neurosci.* 14, 473–492. doi: 10.3758/s13415-014-0277-8

Dayan, P., and Yu, A. J. (2006). Phasic norepinephrine: a neural interrupt signal for unexpected events. *Network* 17, 335–350. doi: 10.1080/09548980601004024

Doya, K., Samejima, K., Katagiri, K., and Kawato, M. (2002). Multiple model-based reinforcement learning. *Neural Comput.* 14, 1347–1369. doi: 10.1162/089976602753712972

Fiske, S. T., Lin, M., and Neuberg, S. (1999). "The continuum model," in *Dual-process Theories in Social Psychology*, eds S. Chaiken and Y. Trope (New York, NY: Guilford Press), xiii, 657, 231–254.

Hampton, A. N., Bossaerts, P., and O'Doherty, J. P. (2006). The role of the ventromedial prefrontal cortex in abstract state-based inference during decision making in humans. *J. Neurosci.* 26, 8360–8367. doi: 10.1523/JNEUROSCI.1010-06.2006

Hampton, A. N., Bossaerts, P., and O'Doherty, J. P. (2008). Neural correlates of mentalizing-related computations during strategic interactions in humans. *Proc. Natl. Acad. Sci. U.S.A.* 105, 6741–6746. doi: 10.1073/pnas.0711099105

Harris, L. T., and Fiske, S. T. (2006). Dehumanizing the lowest of the low: neuroimaging responses to extreme out-groups. *Psychol. Sci.* 17, 847–853. doi: 10.1111/j.1467-9280.2006.01793.x

Haruno, M., Wolpert, D. M., and Kawato, M. (2001). Mosaic model for sensorimotor learning and control. *Neural Comput.* 13, 2201–2220. doi: 10.1162/089976601750541778

Haruno, M., Wolpert, D. M., and Kawato, M. (2003). Hierarchical MOSAIC for movement generation. *Int. Cong. Ser.* 1250, 575–590. doi: 10.1016/S0531-5131(03)00190-0

Higuchi, S., Imamizu, H., and Kawato, M. (2007). Cerebellar activity evoked by common tool-use execution and imagery tasks: an fMRI study. *Cortex* 43, 350–358. doi: 10.1016/S0010-9452(08)70460-X

Hunter, T., Sacco, P., Nitsche, M. A., and Turner, D. L. (2009). Modulation of internal model formation during force field-induced motor learning by anodal transcranial direct current stimulation of primary motor cortex. *J. Physiol.* 587, 2949–2961. doi: 10.1113/jphysiol.2009.169284

Ida Gobbini, M., Leibenluft, E., Santiago, N., and Haxby, J. V. (2004). Social and emotional attachment in the neural representation of faces. *Neuroimage* 22, 1628–1635. doi: 10.1016/j.neuroimage.2004.03.049

Imamizu, H., and Kawato, M. (2012). Cerebellar internal models: implications for the dexterous use of tools. *Cerebellum* 11, 325–335. doi: 10.1007/s12311-010-0241-2

Imamizu, H., Kuroda, T., Miyauchi, S., Yoshioka, T., and Kawato, M. (2003). Modular organization of internal models of tools in the human cerebellum. *Proc. Natl. Acad. Sci. U.S.A.* 100, 5461–5466. doi: 10.1073/pnas.0835746100

Imamizu, H., Miyauchi, S., Tamada, T., Sasaki, Y., Takino, R., Pütz, B., et al. (2000). Human cerebellar activity reflecting an acquired internal model of a new tool. *Nature* 403, 192–195. doi: 10.1038/35003194

Ito, M. (2008). Control of mental activities by internal models in the cerebellum. *Nat. Rev. Neurosci.* 9, 304–313. doi: 10.1038/nrn2332

Kawato, M. (1999). Internal models for motor control and trajectory planning. *Curr. Opin. Neurobiol.* 9, 718–727. doi: 10.1016/S0959-4388(99)00028-8

Kawato, M., Kuroda, T., Imamizu, H., Nakano, E., Miyauchi, S., and Yoshioka, T. (2003). Internal forward models in the cerebellum: fMRI study on grip force and load force coupling. *Prog. Brain Res.* 142, 171–188. doi: 10.1016/S0079-6123(03)42013-X

Laurens, J., Meng, H., and Angelaki, D. E. (2013). Computation of linear acceleration through an internal model in the macaque cerebellum. *Nat. Neurosci.* 16, 1701–1708. doi: 10.1038/nn.3530

Li, C.-S. R., Padoa-Schioppa, C., and Bizzi, E. (2001). Neuronal correlates of motor performance and motor learning in the primary motor cortex of monkeys adapting to an external force field. *Neuron* 30, 593–607. doi: 10.1016/S0896-6273(01)00301-4

Liu, X., Robertson, E., and Miall, R. C. (2003). Neuronal activity related to the visual representation of arm movements in the lateral cerebellar cortex. *J. Neurophysiol.* 89, 1223–1237. doi: 10.1152/jn.00817.2002

Milner, T. E., Franklin, D. W., Imamizu, H., and Kawato, M. (2007). Central control of grasp: manipulation of objects with complex and simple dynamics. *Neuroimage* 36, 388–395. doi: 10.1016/j.neuroimage.2007.01.057

Mitchell, J. P., Banaji, M. R., and Macrae, C. N. (2005a). The link between social cognition and self-referential thought in the medial prefrontal cortex. *J. Cogn. Neurosci.* 17, 1306–1315. doi: 10.1162/0898929055002418

Mitchell, J. P., Macrae, C. N., and Banaji, M. R. (2006). Dissociable medial prefrontal contributions to judgments of similar and dissimilar others. *Neuron* 50, 655–663. doi: 10.1016/j.neuron.2006.03.040

Mitchell, J. P., Neil Macrae, C., and Banaji, M. R. (2005b). Forming impressions of people versus inanimate objects: social-cognitive processing in the medial prefrontal cortex. *Neuroimage* 26, 251–257. doi: 10.1016/j.neuroimage.2005.01.031

Qiu, J. (2006). Peering into the root of prejudice. *Nat. Rev. Neurosci.* 7, 508–509. doi: 10.1038/nrn1959

Richardson, A. G., Overduin, S. A., Valero-Cabré, A., Padoa-Schioppa, C., Pascual-Leone, A., Bizzi, E., et al. (2006). Disruption of primary motor cortex before learning impairs memory of movement dynamics. *J. Neurosci.* 26, 12466–12470. doi: 10.1523/JNEUROSCI.1139-06.2006

Rudman, L. A., and Borgida, E. (1995). The afterglow of construct accessibility: the behavioral consequences of priming men to view women as sexual objects. *J. Exp. Soc. Psychol.* 31, 493–517. doi: 10.1006/jesp.1995.1022

Sainburg, R. L., and Wang, J. (2002). Interlimb transfer of visuomotor rotations: independence of direction and final position information. *Exp. Brain Res.* 145, 437–447. doi: 10.1007/s00221-002-1140-7

Schmahmann, J. D. (1997). *The Cerebellum and Cognition: The Cerebellum and Cognition*. San Diego, CA: Academic Press.

Shadmehr, R. (2004). Generalization as a behavioral window to the neural mechanisms of learning internal models. *Hum. Mov. Sci.* 23, 543–568. doi: 10.1016/j.humov.2004.04.003

Shadmehr, R., and Krakauer, J. W. (2008). A computational neuroanatomy for motor control. *Exp. Brain Res.* 185, 359–381. doi: 10.1007/s00221-008-1280-5

Valentin, V. V., Dickinson, A., and O'Doherty, J. P. (2007). Determining the neural substrates of goal-directed learning in the human brain. *J. Neurosci.* 27, 4019–4026. doi: 10.1523/JNEUROSCI.0564-07.2007

Wada, Y., Kawabata, Y., Kotosaka, S., Yamamoto, K., Kitazawa, S., and Kawato, M. (2003). Acquisition and contextual switching of multiple internal models for different viscous force fields. *Neurosci. Res.* 46, 319–331. doi: 10.1016/S0168-0102(03)00094-4

Wang, J., and Sainburg, R. L. (2006). The symmetry of interlimb transfer depends on workspace locations. *Exp. Brain Res.* 170, 464–471. doi: 10.1007/s00221-005-0230-8

Wolpert, D. M., Doya, K., and Kawato, M. (2003). A unifying computational framework for motor control and social interaction. *Philos. Trans.*

R. Soc. Lond. B Biol. Sci. 358, 593–602. doi: 10.1098/rstb.2002.1238

Wolpert, D. M., and Kawato, M. (1998). Multiple paired forward and inverse models for motor control. *Neural Netw.* 11, 1317–1329. doi: 10.1016/S0893-6080(98)00066-5

Yavari, F., Towhidkhah, F., and Ahmadi-Pajouh, M. A. (2013). Are fast/slow process in motor adaptation and forward/inverse internal model two sides of the same coin? *Med. Hypotheses* 81, 592–600. doi: 10.1016/j.mehy.2013.07.009

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

Received: 18 September 2014; accepted: 09 December 2014; published online: 05 January 2015.

Citation: Yavari F (2015) Does our brain use the same policy for interacting with people and manipulating different objects? Front. Comput. Neurosci. ***8****:170. doi: 10.3389/fncom.2014.00170*

This article was submitted to the journal Frontiers in Computational Neuroscience.

frontiers in Computational Neuroscience

ORIGINAL RESEARCH
published: 24 April 2015
doi: 10.3389/fncom.2015.00048

Stochastic non-linear oscillator models of EEG: the Alzheimer's disease case

Parham Ghorbanian[1], *Subramanian Ramakrishnan*[2] *and Hashem Ashrafiuon*[1]*

[1] Department of Mechanical Engineering, Center for Nonlinear Dynamics and Control, Villanova University, Villanova, PA, USA, [2] Department of Mechanical and Industrial Engineering, University of Minnesota Duluth, Duluth, MN, USA

In this article, the Electroencephalography (EEG) signal of the human brain is modeled as the output of stochastic non-linear coupled oscillator networks. It is shown that EEG signals recorded under different brain states in healthy as well as Alzheimer's disease (AD) patients may be understood as distinct, statistically significant realizations of the model. EEG signals recorded during resting eyes-open (EO) and eyes-closed (EC) resting conditions in a pilot study with AD patients and age-matched healthy control subjects (CTL) are employed. An optimization scheme is then utilized to match the output of the stochastic Duffing—van der Pol double oscillator network with EEG signals recorded during each condition for AD and CTL subjects by selecting the model physical parameters and noise intensity. The selected signal characteristics are power spectral densities in major brain frequency bands Shannon and sample entropies. These measures allow matching of linear time varying frequency content as well as non-linear signal information content and complexity. The main finding of the work is that statistically significant unique models represent the EC and EO conditions for both CTL and AD subjects. However, it is also shown that the inclusion of sample entropy in the optimization process, to match the complexity of the EEG signal, enhances the stochastic non-linear oscillator model performance.

Keywords: EEG, Alzheimer's disease, stochastic differential equations, duffing—van der Pol, entropy

OPEN ACCESS

Edited by:
Tobias Alecio Mattei, Brain & Spine Center - InvisionHealth - Kenmore Mercy Hospital, USA

Reviewed by:
Robin A. A. Ince, University of Manchester, UK
Fan Liao, Washington University in St louis, USA

***Correspondence:**
Hashem Ashrafiuon, Department of Mechanical Engineering, Center for Nonlinear Dynamics and Control, Villanova University, 800 E. Lancaster Ave., Villanova, PA 19085, USA
hashem.ashrafiuon@villanova.edu

Received: *28 January 2015*
Accepted: *07 April 2015*
Published: *24 April 2015*

Citation:
Ghorbanian P, Ramakrishnan S and Ashrafiuon H (2015) Stochastic non-linear oscillator models of EEG: the Alzheimer's disease case. Front. Comput. Neurosci. 9:48. doi: 10.3389/fncom.2015.00048

1. Introduction

Quantitative analysis of human brain electroencephalography (EEG) recordings aimed at enhancing our understanding of brain injuries and disorders is currently an important research area. In addition to being useful in diagnosis, such analysis can provide insights into the underlying neurophysiology of the injury or disorder, thereby leading to better treatment and preventive strategies. Alzheimer's disease (AD) is the most common form of dementia and is the subject of intense research. While no known cure exists, certain medications have shown promise in delaying the symptoms (Dauwels et al., 2010) prompting researchers to seek early diagnosis and intervention strategies. In this context, analysis of the EEG is a potential non-invasive tool that may aid early diagnosis of AD. However, the use of EEG signal analysis in order to improve the diagnosis of AD is a complex problem where, despite significant advances, a number of fundamental questions remain open (Elgendi et al., 2011).

Considering now the characteristics of the EEG, since the non-stationary nature of the signal is generally well-recognized (see, for instance Akin, 2002), decomposition using a fast Fourier

transform (FFT) with sliding windows and the wavelet transforms have been the most popular techniques employed to capture the spectral properties of EEG (Darvishi and Al-Ani, 2007; Dauwels et al., 2010). However, linear transformation methods fail to address the non-linear characteristics of the EEG signal (Stam, 2005). Therefore, non-linear dynamic approaches have been attempted as well, mostly involving computationally complex time series analysis (Jeong, 2004). Several other aspects of non-linear modeling and analysis in this context have also been studied in the literature (see, for instance, Stam, 2005 for a review). These include frameworks based on a neural mass model (Valdes et al., 1999; Huang et al., 2011), coupled oscillators (Baier et al., 2005; Leistritz et al., 2007), continuum models (Kim et al., 2007), non-linear non-stationary models (Celka and Colditz, 2002; Rankine et al., 2007), random neural networks (Acedo and Morano, 2013), and chaotic phenomena and stability aspects (Rodrigues et al., 2007; Dafilis et al., 2009). Stochastic approaches based on Markov chain Monte Carlo methods (Hettiarachchi et al., 2012) and Markov process amplitude (Wang et al., 2011) that take into account the inherent randomness of the EEG signal have also been reported. In the same vein, limit cycle oscillators (Hernandez et al., 1996; Burke and Paor, 2004) as well as stochastic synchronization (Bressloff and Lai, 2011) and stochastic approximation (Fell et al., 2000; Sun et al., 2008) methods have been considered in EEG modeling. Notably, limit cycle behavior at each of the brain frequency bands appears to provide a more accurate representation of the EEG signal than one based on chaotic phenomena.

Some of the most important features in non-linear dynamic and stochastic approaches are signal information content and complexity as measured using various forms of information entropy. Measures such as Shannon entropy (Shannon, 1948) characterize the information content in a signal and higher entropy corresponds to increased randomness and chaotic behavior (Abasolo et al., 2006). Importantly, one observes that, with respect to the EEG signal, higher information content correlates with better brain function (Shin et al., 2006). Furthermore, it has been reported that variations in information entropic measures may be used to detect functional abnormalities in the brain caused by disorders or injuries (Slobounov et al., 2009). Hence, information content of the EEG signal, characterized by information-entropic measures, may be expected to be important in identifying distinct states of the brain. This is further reinforced by the recent results of McBride and colleagues on the role of information entropic and spectral analysis in the study of the early stages of Alzheimer's disease and mild Traumatic Brain Injury McBride et al. (2013a,b, 2014).

Entropy may also be utilized to measure signal complexity. For instance, embedding entropy provides information about how the EEG signal fluctuates in time by comparing the time series with a delayed version of itself (Abasolo et al., 2006). Moreover, the concept of approximate entropy was introduced as a measure of system complexity (Pincus, 1991) and has been applied to brain wave signals (Quiroga et al., 2001). However, the approximate entropy measure suffers from drawbacks such as bias and inconsistency (Xu et al., 2010). Hence, the notion of sample entropy was introduced (Richman and Moorman, 2000) as an improvement over approximate entropy.

In recent work, the authors proposed a phenomenological model of the EEG signal based on the dynamics of a stochastic, coupled, Duffing- van der Pol oscillator network (Ghorbanian et al., 2015). An optimization scheme was adopted to match model output with actual EEG data obtained from healthy subjects in the two distinct resting eyes-open (EO) and eyes-closed (EC) conditions and it was shown that the actual EEG signals in both cases were distinct realizations of the model with qualitatively different non-linear dynamic characteristics. Moreover, the model output and the actual EEG data were shown to be in good agreement in terms of both the power spectra (frequency content) and Shannon entropy (information content).

In the present effort, we improve the model introduced in Ghorbanian et al. (2015) by matching the sample entropy of the model output and EEG signal to capture its complexity. A global optimization routine is employed in order to match the output of with EEG recordings in terms of power spectrum, Shannon entropy, and sample entropy. The EEG signals were recorded under resting EC and EO conditions in an earlier pilot study of Alzheimer's disease (AD) patients vs. age-matched healthy control (CTL) subjects (Ghorbanian et al., 2013). The model parameters obtained for the oscillators representing EC and EO EEG signals for CTL and AD patients are compared in order to establish statistically significant, distinct models for AD and CTL subjects under each condition. In addition, we present new results from a phase space reconstruction analysis of the model output to match the actual EEG signal. The results indicate that the analytical model effectively captures the frequency spectrum and non-linear characteristics of the EEG signal in terms of complexity and information content. Furthermore, it is shown that the addition of sample entropy significantly enhances the model performance in terms of complexity and non-linear dynamic characteristics, as demonstrated by phase space reconstruction. The results suggest exciting new pathways to develop better tools for distinguishing pathological and normal brain states in AD and perhaps other neurological diseases and disorders.

The rest of the article is set as follows. Details of the EEG recordings, the analytical model, the optimization scheme and the phase space reconstruction technique are provided in Section 2. The results are presented in Section 3 and discussed in Section 4. The articles concludes with comments on further research in Section 5.

2. Materials and Methods

2.1. EEG Recording Blocks

Twenty six AD patients and healthy age-matched CTL subjects were selected for this study ("A Brain-Computer Interface for Diagnosing Brain Function," Aspire IRB, Human Subject Protocol Number PDMC-001, approved on October 7, 2010). Of the 26 subjects selected, one withdrew and one did not qualify as AD or CTL. Subjects were asked to relax and wear an EEG recording headset during alternating blocks of EC and EO followed by a variety of cognitive and auditory tasks and a final EC-EO resting period.

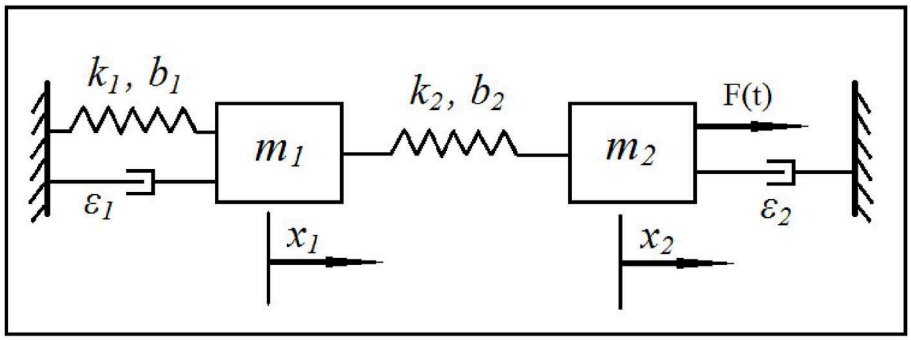


FIGURE 1 | Schematic of the stochastic coupled Duffing—van der Pol oscillators.

TABLE 1 | Optimal parameters of the Duffing—van der Pol oscillator for EC and EO of CTL subjects ($N = 40$) and the p-values from unpaired t-test, Wilcoxon rank sum test, and Bonferonni correction.

Parameter	Eyes-Closed (EC)	Eyes-Open (EO)	t-test	Wilcoxon	Bonferroni
k_1	7286.5 ± 192.4	2427.2 ± 448.91	1e-15	1e-8	1e-7
k_2	4523.5 ± 282.3	499.92 ± 84.04	1e-15	1e-8	1e-7
b_1	232.05 ± 18.3	95.61 ± 24.20	1e-15	1e-8	1e-7
b_2	10.78 ± 2.3	103.36 ± 9.22	1e-15	1e-8	1e-7
ϵ_1	33.60 ± 5.4	48.89 ± 9.49	1e-15	1e-8	1e-7
ϵ_2	0.97 ± 0.19	28.75 ± 1.74	1e-15	1e-8	1e-7
μ	2.34 ± 0.47	1.82 ± 0.78	0.01	0.06	0.06

The EEG signals were recorded through a single-dry electrode device at position Fp1 (based on a 10–20 electrode placement system) with a Bluetooth enabled telemetric headset. The headset's effective sample rate is 125 Hz. Frequencies below 1 Hz and above 60 Hz (near Nyquist frequency) were filtered out by the device hardware. On comparison of the EEG recordings by the device with those from other widely accepted devices, frequencies within 2–30 Hz were deemed to be very accurate.

The recording device eliminated frequently observed artifacts including line noise. Other artifacts were mainly due to eye- and muscle-movements, which are common at Fp1 location and can be clearly identified by their high amplitudes compared to true EEG signal recordings during resting states. These artifacts were removed using a simple artifact detection that eliminated any part of the signal greater than 4.5σ (standard deviation). The algorithm also reconstructed the nulled samples using FFT interpolation of the trailing and subsequent recorded data (Ghorbanian et al., 2013).

The EEG recordings in this study were obtained from subjects in an AD pilot study with 14 control (CTL) subjects and 10 Alzheimer's Disease (AD) patients presented in our earlier work (Ghorbanian et al., 2013). Recording blocks of 40-s duration (approximately 5000 sample signals) from resting eyes-closed (EC) and eyes-open (EO) conditions were selected. In all, 60 random blocks were selected from the pilot study: 40 blocks from control CTL subjects (20 EC and 20 EO) and 20 blocks from AD subjects (10 EC and 10 EO). Note that, the smaller number of AD patients along with smaller number of AD patient recording sessions that were were not dominated by artifacts resulted in the selection of smaller AD sample size.

2.2. EEG Features

The time-varying power spectrum in each of the major brain EEG frequency bands was calculated using short time fast Fourier transform (FFT) with sliding window, since a good model must produce signals that can match EEG's frequency content. Specifically, the power spectrum was computed in seven ranges: lower δ (1–2 Hz), upper δ (2–4 Hz), θ (4–8 Hz), α (8–13 Hz), lower β (13–20 Hz), upper β (20–30 Hz), and γ (30–60 Hz). However, lower δ and γ band powers, which happen to have little power, were ignored due to unreliability of the device in those frequency ranges.

Shannon entropy was measured based on a sliding temporal window technique. A temporal window was defined to slide along the signal time representation with a sliding step (interval or bin) to sample a part of the signal. A discrete entropy estimator was applied, in which 10 uniform intervals equally divided the range of the normalized observed signal. Then the probability that the sampled signal belongs to the interval is the ratio between the number of the samples found within each interval and the total number of samples of the signal. The Shannon entropy is then calculated based on these probabilities (Shin et al., 2006), separately for each 40-s EEG recording block (5000 samples).

Sample entropy (SE) is the negative natural logarithm of the conditional probability that two sequences of a time series, similar for m points, remain similar at the next point. For given N data points from a time series, $[x(1), x(2), \cdots, x(N)]$, we calculated SE of each 40-s EEG recording block (5000 samples) by the statistic (Abasolo et al., 2006):

$$\mathrm{SE}(m, r, N) = \left\{ -\ln \left[\frac{U^{m+1}(r)}{U^m(r)} \right] \right\}, \tag{1}$$

where m is the run length, r is the tolerance window size, and

$$U^m(r) = \frac{1}{(N-m)(N-m-1)} \sum_{i=1}^{N-m} U_i. \tag{2}$$

In the above equation, U_i indicates the number of k's ($1 \leq k \leq N - m$) such that the Euclidean distance between $X_m(i)$ and $X_m(k)$, $k \neq i$, is less than or equal r and $X_m(i) = [x(i), x(i+1), \cdots, x(i+m-1)]$.

Generally, large m or small r values result in number of matches being too small for confident estimation of the conditional probability and vice versa (Lake and Moorman, 2011). In this study, we used $m = 2$ and $r = 0.25\sigma$ based on the consistency of the results and recommended ranges in previous studies (Richman and Moorman, 2000; Xu et al., 2010).

2.3. Stochastic Coupled Non-linear Oscillators

We recall that the EEG has been modeled in the literature taking into account characteristics including non-linearity (both chaotic and non-chaotic), non-stationarity, and randomness of the signal (Fell et al., 2000; Rankine et al., 2007; Sun et al., 2008). The EEG has also been studied as the manifestation of underlying limit cycle oscillations at a given frequency and other such periodic solutions (Hernandez et al., 1996; Burke and Paor, 2004). While inspired by the above, the authors were fundamentally motivated to develop models that can better reproduce the significant linear and non-linear characteristics of actual EEG signals. Hence, we proposed a phenomenological model of the EEG based on a coupled system of Duffing—van der Pol oscillators subjected to white noise excitation (Ghorbanian et al., 2015). This particular oscillator was selected because the Duffing non-linearity allows a system with only two oscillators capture the major brain frequency spectra and van der Pol non-linearity provides self-excited limit cycle behavior which have been previously reported for each major brain frequency bands (Burke and Paor, 2004).

We consider a phenomenological model of the EEG based on a coupled system of Duffing—van der Pol oscillators subject to white noise excitation, as shown in **Figure 1**. The equations for the model may be written as:

$$\begin{cases} \ddot{x}_1 + (k_1 + k_2)x_1 - k_2 x_2 = -b_1 x_1^3 - b_2(x_1 - x_2)^3 \\ \qquad\qquad\qquad\qquad +\epsilon_1 \dot{x}_1(1 - x_1^2), \\ \ddot{x}_2 - k_2 x_1 + k_2 x_2 = b_2(x_1 - x_2)^3 \\ \qquad\qquad\qquad\qquad +\epsilon_2 \dot{x}_2(1 - x_2^2) + \mu\, dW, \end{cases} \tag{3}$$

where $x_i, \dot{x}_i, \ddot{x}_i, i = 1, 2$ are positions, velocities, and accelerations of the two oscillators, respectively. Parameters $k_i, b_i, \epsilon_i, i = 1, 2$ are, respectively, linear stiffness, cubic stiffness, and van der Pol damping coefficient of the two oscillators. Parameters b_is indicate the strength of the Duffing non-linearity resulting in multiple resonant frequencies while ϵ_is indicate the strength of van der Pol non-linearity and determine the extent of self-excitation and the shape of the resulting limit cycle. Parameter μ represents the intensity of white noise and dW is a Wiener process (Gardiner, 1985; Higham, 2001) representing the additive noise in the stochastic differential system. The input excitation to the system is provided through μdW. The output

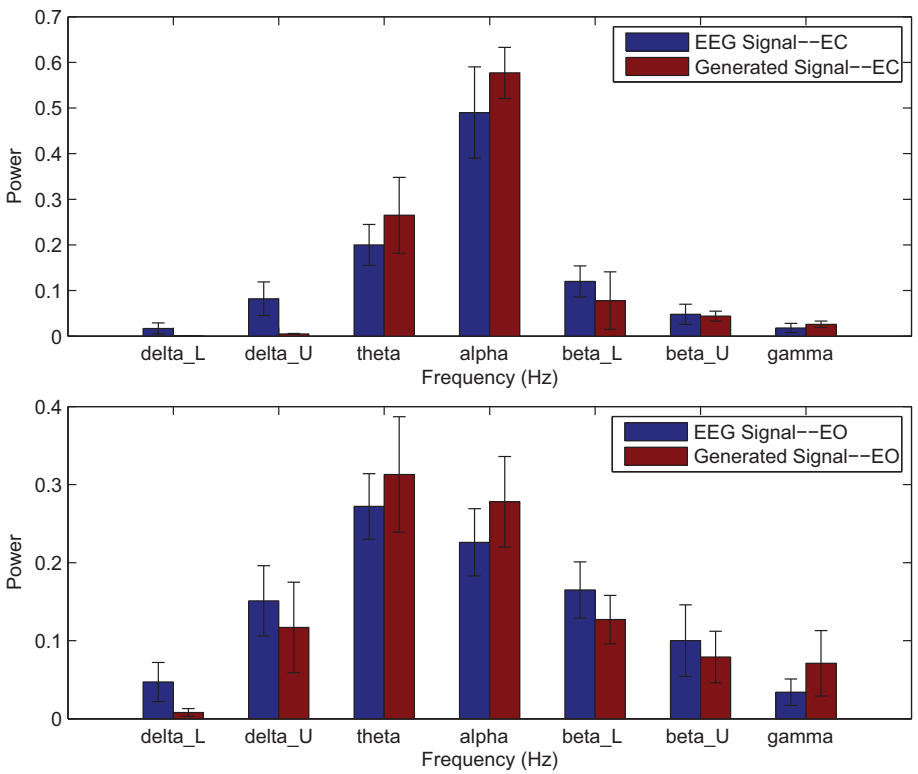


FIGURE 2 | Comparison of major brain frequency band mean powers of CTL EEG signals and optimal oscillator model output; EC (top), EO (bottom).

may be selected as any combination of the positions and velocities to mimic an EEG signal. Note that, the Euler-Maruyama method (Higham, 2001) was selected to integrate the stochastic differential equations in Equation (3) since standard numerical integration methods are not applicable.

2.4. Optimization Formulation

We have selected the velocity of the second oscillator as the system output approximating the EEG signal since it is directly impacted by the noise. A global optimization search method based on a multi-start algorithm (Ugray et al., 2007) was adopted to determine the oscillator model parameters that can produce the output matching various EEG signals. The optimization objective function was selected as the root mean squared of the errors in power spectrum of each selected brain frequency bands plus weighted values of the errors in absolute Shannon and sample entropies. Hence, the optimization goal is error minimization:

$$\min_{p} J = \sqrt{\sum_{j=1}^{m}(P_{Ej} - P_{Oj})^2 + w_1|S_E - S_O| + w_2|SP_E - SP_O|}, \quad (4)$$

where J is the objective function, $p = [k_1, k_2, b_1, b_2, \epsilon_1, \epsilon_2, \mu]$ the decision variables, P_{Ej} and P_{Oj} the powers in the major brain frequency bands for the normalized EEG signal and the model output, respectively, m is number of frequency bands ($m = 7$), S_E and S_O the Shannon entropies of the EEG signal and the model output, respectively, SP_E and SP_O the sample entropies of the EEG signal and the model output, respectively, w_1 and w_2 are weighting factor for absolute Shannon and sample entropies, respectively, and $|\ |$ represents absolute value. The weighting factors w_1 and w_2 are required to give equal importance to power spectrum and entropy characteristics of the signal. Note that the magnitude of the output signals are matched through normalization of both the model output and the EEG signal with respect to their standard deviations.

The objective function minimization is subject to equality constraints represented by the state (Equation 3) and inequality constraints represented by the decision variable lower and upper bounds:

$$0 < k_i \leq 1e4, \quad 0 < b_i \leq \tfrac{1}{2}k_i, \quad 0 < \epsilon_i \leq \tfrac{1}{3}k_i, \quad i = 1, 2, \quad 0 \leq \mu \leq 2. \quad (5)$$

The constraints for b_i's and ϵ_i's were imposed to avoid the chaotic regime (Li et al., 2006) and provide a periodic stochastic response.

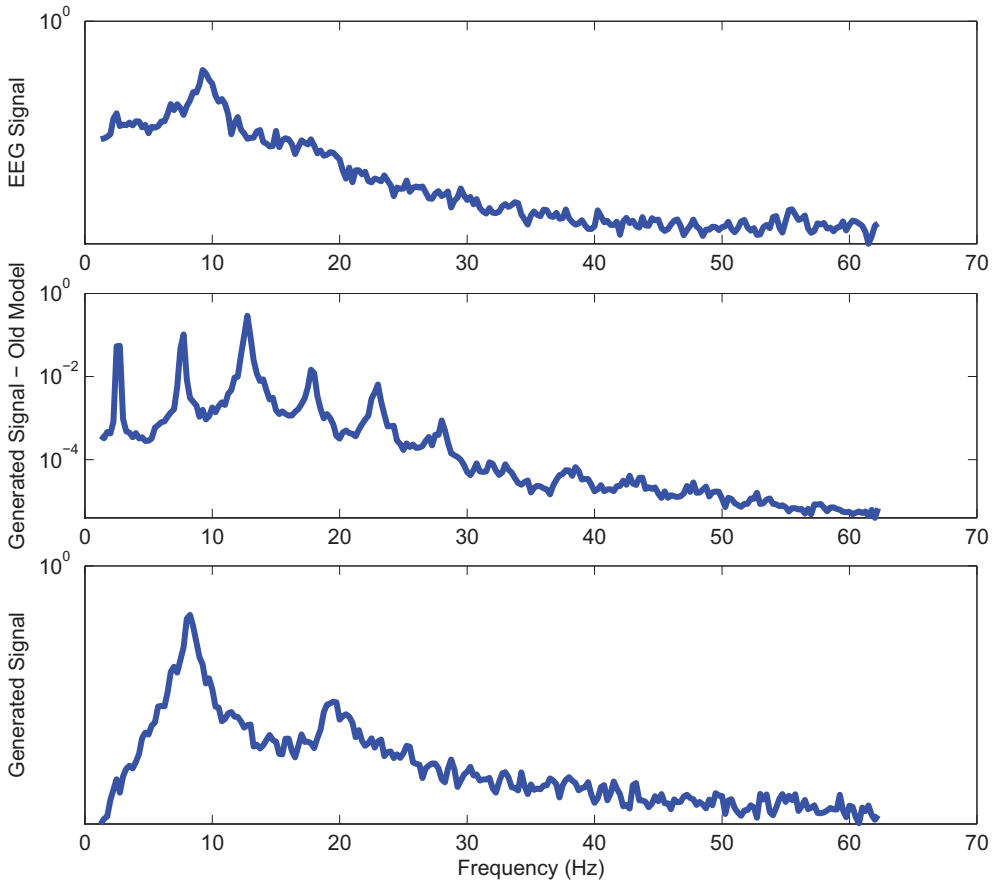


FIGURE 3 | Power spectrum of a sample CTL EC (top) EEG signal; (middle) output of stochastic oscillator model using Shannon entropy; (bottom) output of stochastic oscillator model using Shannon and sample entropies.

Noise intensity is also constrained to avoid a response dominated by random noise. The initial guesses for the global optimization search are randomly generated within the bounds defined in Equation (5).

The stochastic component was introduced as white noise, which was generated through a normally distributed random variable and applied to the model via Wiener process. A new random process was generated and applied to the model during integration of the equations, at each iteration of the optimization algorithm.

2.5. Statistical Analysis

A key objective of the phenomenological modeling in this work is the ability to establish a correspondence between variations in model parameters and the variations in the data obtained from different physiological conditions. Hence, the parametric unpaired t-test and non-parametric Wilcoxon rank sum statistical testing methods were employed to determine the relative significance of the model parameters. Furthermore, Bonferroni correction was applied due to multiple comparisons problem and adequacy of sample sizes for statistical tests were established using power analysis.

2.6. Phase Space Reconstruction

In addition to matching Shannon and sample entropies of the model output and EEG signal through the optimization process, it is of interest to investigate matching other features such as the phase plot which plays a significant role in non-linear time series analysis (Kantz and Schreiber, 2004). It is known that any dynamic system can be completely recovered in the phase space, which maybe reconstructed from the measured time domain response of the system (Nie et al., 2013). While phase space consists of velocity and position variables for a mechanical system, in the case where just the time representation of a signal is available, a phase space reconstruction technique based on the method of delays is used (Kantz and Schreiber, 2004).

The main idea is that one does not need the derivatives to form a coordinate system in which to capture the structure of phase space, but instead one could directly use the lagged variables:

$$x(n+T) = x(t_0 + (n+T)\Delta\tau_s), \tag{6}$$

where $x(n)$ is the nth sample of the time series, $\Delta\tau_s$ the time step, and T the delay integer to be determined. Then, a vector

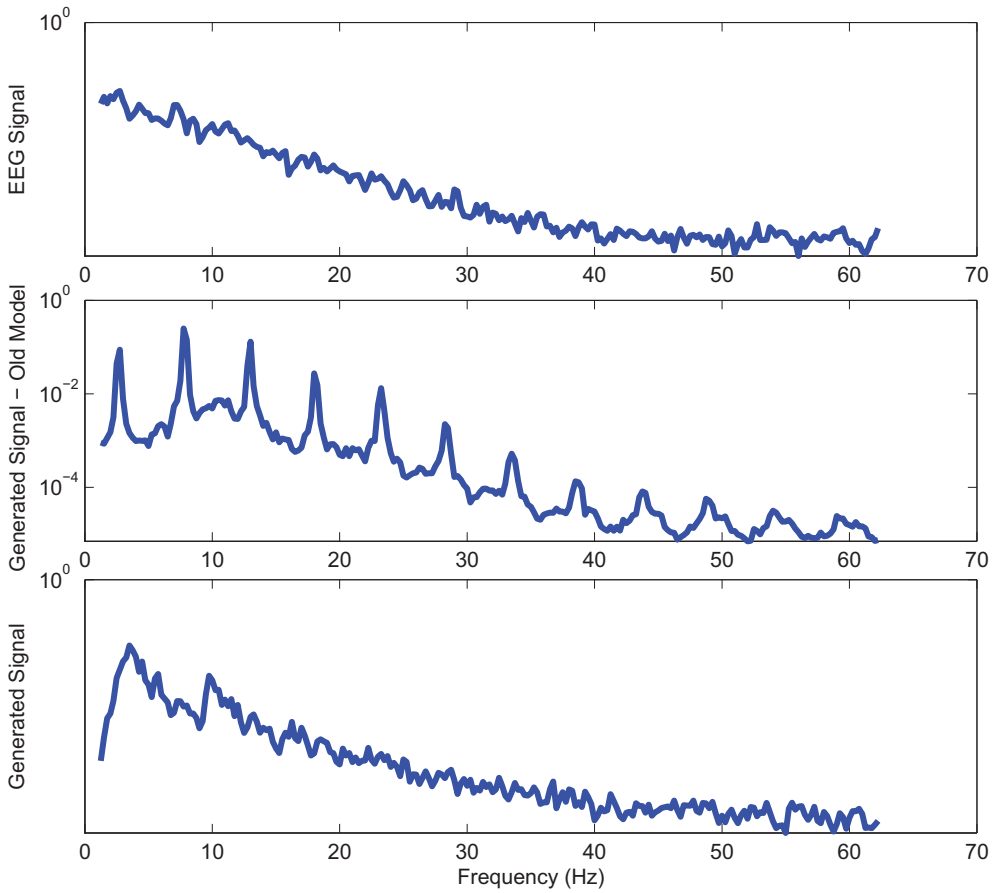


FIGURE 4 | Power spectrum of a sample CTL EO (top) EEG signal; (middle) output of stochastic oscillator model using Shannon entropy; (bottom) output of stochastic oscillator model using Shannon and sample entropies.

of (embedding) dimension d may be constructed using the time lags as:

$$[x(n), x(n+T), x(n+2T), \cdots, x(n+(d-1)T)]. \quad (7)$$

Time-delay embedding is probably one of the best systematic methods for converting scalar data to multidimensional phase space (Abarbanel et al., 1993; Burke and Paor, 2004; Nie et al., 2013). An appropriate and successful reconstruction depends on the choice of both time delay T and the embedding dimension d (Nie et al., 2013).

In this study, the appropriate value of the time lag was determined using the average mutual information method applied to each EEG recording block. The idea behind mutual information is to identify the amount of information that can be learned about a measurement at one time from a measurement taken at another time. Consider the time series nth sample $x(n)$ and its value after time delay T with the associated probability distributions of $P(x(n))$ and $P(x(n+T))$, respectively. The average information which can be obtained about $x(n+T)$ from $x(n)$ is given by the mutual information of the two measurements (Abarbanel et al., 1993; Mizrach, 1996):

$$I(x(n), x(n+T)) = \log_2 \left[\frac{P(x(n), x(n+T))}{P(x(n))P(x(n+T))} \right], \quad (8)$$

where $P(x(n), x(n+T))$ is the joint probability of the measurements $x(n)$ and $x(n+T)$ calculated using a binning-based method, in which 20 uniform intervals divided the range of the measurements equally. The average mutual information between

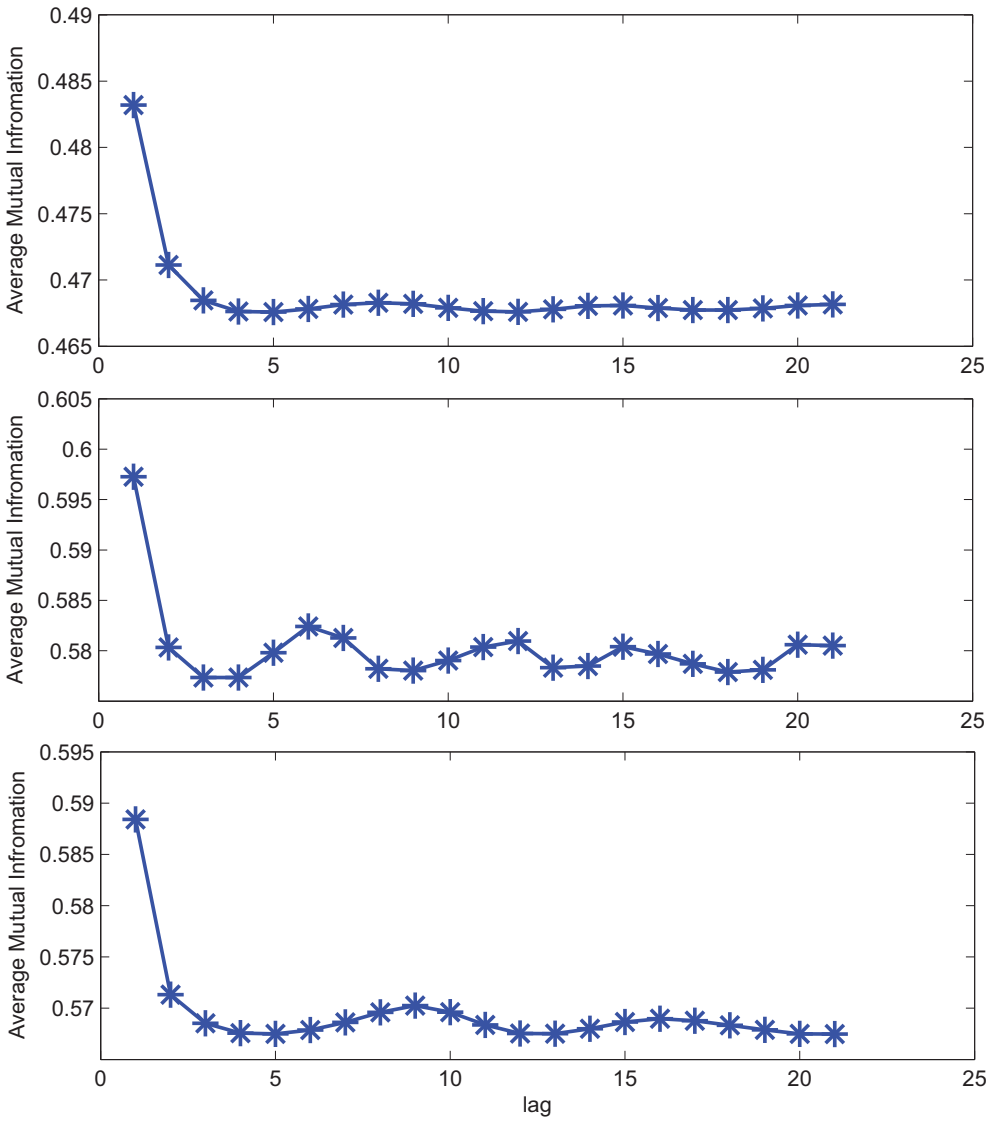


FIGURE 5 | Average mutual information for a sample CTL EC (top) EEG signal; (middle) output of stochastic oscillator model using Shannon entropy; (bottom) output of stochastic oscillator model using Shannon and sample entropies.

measurements of any value $x(n)$ and $x(n+T)$ is the average over all possible measurements of $I(x(n), x(n+T))$ (Abarbanel et al., 1993):

$$I(T) = \sum_{x(n),x(n+T)} P(x(n), x(n+T))I(x(n), x(n+T)). \quad (9)$$

If T is too small, the measurements $x(n)$ and $x(n+T)$ will have too much overlap. However, if T is too large, then $I(T)$ will approach zero and nothing relates $x(n)$ to $x(n+T)$. It is suggested that the proper T can be chosen as the first minimum of $I(T)$ which is not necessarily optimal but has been shown to work well (Abarbanel et al., 1993; Nie et al., 2013). If in a case, no minima exists for $I(T)$, the choice of $T = 1$ or 2 has been suggested (Abarbanel et al., 1993).

After specifying the correct time delay T, an appropriate embedding dimension, d, should also be found for the phase space reconstruction. If d is too small, the trajectories will not be unique. On the other hand, too large a d will result in additional computational cost by requiring extra dimensions (Nie et al., 2013).

3. Results

The optimization algorithm was separately applied to determine the model parameters (decision variables) for each of the 60 selected EEG signals using the weighting factors $w_1 = w_2 = 0.35$. These weighting factors give equal importance to the entropy measures and power spectrum. We then categorized the resulting 60 set of model parameters into four groups based recording

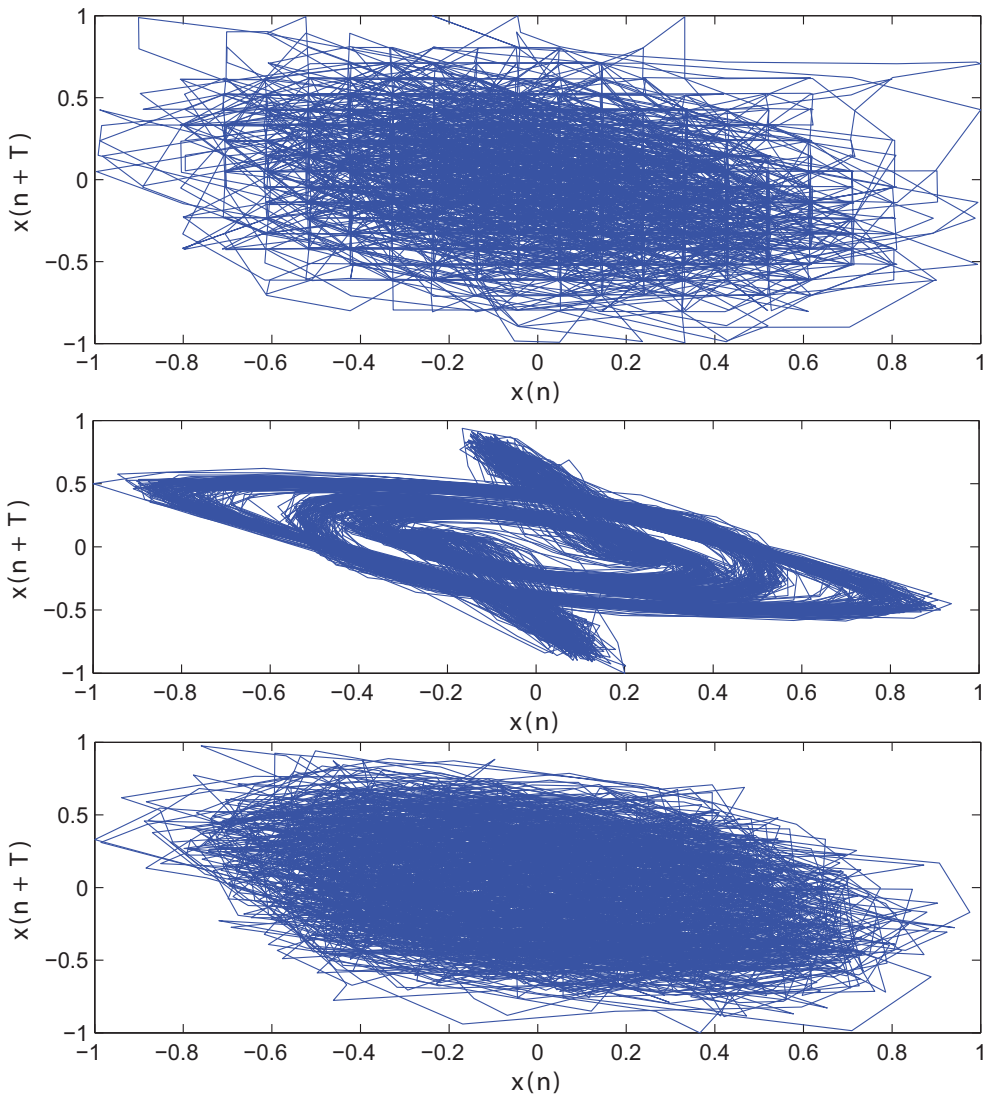


FIGURE 6 | Reconstructed phase plot of a sample CTL EC (top) EEG signal; (middle) output of stochastic oscillator model using Shannon entropy; (bottom) output of stochastic oscillator model using Shannon and sample entropies.

conditions and subject diagnosis: EC-CTL, EO-CTL, EC-AD, and EO-AD.

3.1. Healthy Eyes-Closed and Eyes-Open Results

Initially, we studied the models derived for the EC and EO EEG signals of CTL subjects for validation purposes. The means and standard deviations of the optimal values of the model parameters for EC-CTL and EO-CTL EEG signals are listed in **Table 1**. The p-values from the two statistical tests and non-parametric method after Bonferroni corrections indicate that the differences between of all parameters of the two models are strongly statistically significant with the exception of noise intensity. Note that, μ is also found statistically significant using t-test but is slightly off when the non-parametric method is used.

In order to ensure that adequate sample sizes are used, the minimum required difference between means of two groups of data for each parameter are computed. As expected due to very small p-values, the sample size for statistical testing is found to be sufficient with more than 99.9% power for all parameters except noise intensity μ, which was not found to be statistically significant using the non-parametric method.

Power spectrums of the optimal stochastic oscillator model output and EEG signals for the EC and EO cases of CTL subjects are presented in **Figure 2** where θ, α, and β band powers show excellent agreement. The comparison revealed that, as expected, the optimal model is closely following the α-band dominance in the EC cases. While, in the EO cases, the optimal model follows a more flat frequency distribution from upper δ to lower β frequency bands. Furthermore, Shannon and SE values of the EEG signals and the model outputs for the EC and EO cases show close agreement. Shannon entropy values were 1.80 ± 0.08 and 1.92 ± 0.08 for EC EEG and model output, respectively, and 1.71 ± 0.11 and 1.57 ± 0.15 for EO EEG and model output, respectively. While, SE values were 1.04 ± 0.20 and 1.17 ± 0.22 for EC EEG and model output, respectively, and 0.97 ± 0.20 and 1.20 ± 0.18 for EO EEG and model output, respectively. These results show a significant improvement over our previous model where only Shannon entropy was used (Ghorbanian et al., 2015). The improvement is clearly observed in the the power spectra of sample EC and EO EEG signals and their corresponding optimal model outputs, respectively shown in **Figures 3**, **4**. Both figures demonstrate more distributed spectra of the model outputs with similar noise complexities to the actual EEG signals when SE is added to the objective function; i.e., power spectra of the signals without matching of SE have very discrete peaks unlike the EEG.

The impact of SE to match signal complexity is further demonstrated through phase plot reconstruction of the time series. Average mutual information for a sample EC EEG signal and outputs of the optimal stochastic oscillator models are shown in **Figure 5** as a function of lag time. The first minimum occurs at $T = 5$ lag samples for both the EEG signal and the optimal model derived with both Shannon and sample entropies while $T = 3$ for the output of the model derived solely based on Shannon entropy. The resulting reconstructed phase plots of the EC EEG signal and the outputs of the two optimal models are presented in **Figure 6**. Clearly, the reconstructed phase plots of the EEG and the output of the model derived using both Shannon and sample entropies, display similar behavior. While the output of the model derived using only Shannon entropy is qualitatively different form the EEG signal in terms of complexity and noise. Indeed this result provides further affirmation that the stochastic Duffing—van der Pol model yields an output that matches the actual EEG data in terms of non-linear characteristics observed in the phase space.

TABLE 2 | Optimal parameters of the Duffing—van der Pol oscillator model for EC and EO of AD subjects ($N = 20$) and the p-values from unpaired t-test, Wilcoxon rank sum test, and Bonferonni correction.

Parameter	Eyes-Closed (EC)	Eyes-Open (EO)	t-test	Wilcoxon	Bonferroni
k_1	1742.1 ± 197.91	3139.9 ± 1040.9	0.0005	0.0025	0.009
k_2	1270.8 ± 277.13	650.32 ± 175.76	1e-5	0.0005	0.002
b_1	771.99 ± 126.81	101.1 ± 27.86	1e-12	0.0001	0.001
b_2	1.91 ± 0.22	81.3 ± 9.76	1e-15	0.0001	0.001
ϵ_1	63.7 ± 11.64	56.3 ± 5.75	0.0884	0.021	0.063
ϵ_2	20.7 ± 5.64	19.12 ± 2.87	0.4234	0.879	0.95
μ	1.78 ± 0.8	1.74 ± 0.67	0.905	0.879	0.95

3.2. Alzheimer's Disease vs. Control Results

Next, we studied the models derived for the EC and EO EEG signals of AD subjects. The mean and standard deviation of the optimal values of the model parameters for EC-AD and EO-AD EEG signals are listed in **Table 2** along with the p-values from the two statistical tests and the non-parametric test after Bonferroni corrections indicating that the differences between only the

TABLE 3 | The p-values from unpaired t-test, Wilcoxon rank sum test, and Bonferonni correction for comparison of model parameters between AD ($N = 20$) and CTL ($N = 40$) subjects.

Parameter	t-test (EC)	Wilcoxon (EC)	Bonf. (EC)	t-test (EO)	Wilcoxon (EO)	Bonf. (EO)
k_1	1e-30	1e-5	4e-5	0.013	0.027	0.08
k_2	1e-23	1e-5	3e-5	0.0034	0.0015	0.007
b_1	1e-17	1e-5	5e-5	0.58	0.027	0.08
b_2	1e-12	1e-5	6e-5	1e-6	1e-5	9e-5
ϵ_1	1e-10	6e-5	1e-5	0.031	0.0018	0.007
ϵ_2	1e-15	5e-5	7e-6	4e-12	1e-5	7e-5
μ	0.02	0.06	0.06	0.80	0.70	0.7

first four parameters of the two models are statistically significant. Next, we separately compared the model parameters of EC and EO EEG signals of CTL subjects with those AD patients.

Table 3 lists the p-values from the two statistical testing methods and the non-parametric method after Bonferroni corrections comparing CTL vs. AD subjects under separate EC and EO conditions. The results indicate that the difference between all model parameters of CTL and AD subjects under EC condition are strongly statistically significant except for noise intensity. Again, μ is also found statistically significant using t-test but is slightly off when non-parametric method is used. The difference between the model parameters of CTL and AD subjects under EO condition are not, however, as strong, though they are still mostly statistically significant. In the EO case, parameter μ is not statistically significant using either method and t-test does not find b_1 to be statistically significant either.

The power analysis results for 90%, 95%, 99%, and 99.9% for two statistical are listed in **Tables 4**, **5** for EC and EO cases, respectively. The actual difference between means are given within parentheses following each parameter. The results indicated that our sample size for statistical testing in EC case between AD and CTL subjects was sufficient for all parameters

TABLE 4 | Minimum required difference between model parameter mean values of EC AD vs. EC CTL for various desired powers of statistical tests.

Parameter	90%	95%	99%	99.9%
Δ_{k1} (5544.3)	281.39	313.53	371.72	439.46
Δ_{k2} (3252.7)	406.65	453.09	537.18	635.08
Δ_{b1} (539.94)	106.44	118.60	140.61	166.24
Δ_{b2} (8.87)	2.82	3.14	3.72	4.40
Δ_{e1} (30.09)	11.54	12.86	15.25	18.03
Δ_{e2} (19.79)	4.64	5.17	6.13	7.25
Δ_{μ} (0.56)	0.87	0.97	1.15	1.36

TABLE 5 | Minimum required difference between model parameter mean values of EO AD vs. EO CTL for various desired powers of statistical tests.

Parameter	90%	95%	99%	99.9%
Δ_{k1} (712.78)	1009.12	1124.36	1333.04	1575.97
Δ_{k2} (175.81)	175.81	195.89	232.25	274.57
Δ_{b1} (5.49)	36.86	41.07	48.69	57.56
Δ_{b2} (22.02)	13.61	15.17	17.98	21.26
Δ_{e1} (7.41)	12.27	13.68	16.22	19.17
Δ_{e2} (9.62)	3.14	3.50	4.15	4.91
Δ_{μ} (0.07)	1.08	1.21	1.43	1.70

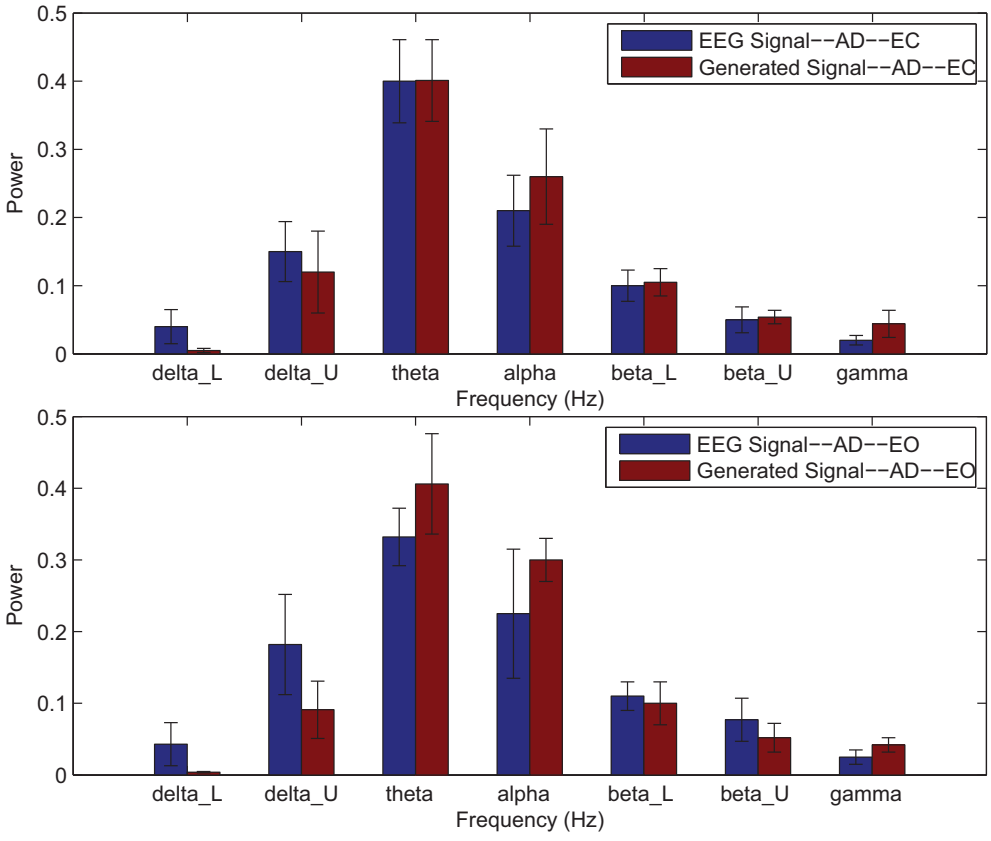


FIGURE 7 | Comparison of major brain frequency band mean powers of AD EEG signals and optimal oscillator model output; EC (top), EO (bottom).

except μ with more than 99.9% power. However, in the EO case, only sample size for parameters b_2 and ϵ_2 has more than 99.9% confidence and k_2 shows a 90% confidence. The sample size for the remaining parameters did not provide sufficient confidence.

Power spectrums of the optimal stochastic oscillator model output and EEG signals for the EC and EO cases of AD subjects are presented in **Figure 7** where again θ, α, and β band powers show excellent agreement. The comparison revealed that the optimal model was closely and correctly slightly θ-band dominated in the EC cases for AD subjects (Ghorbanian et al., 2013). While, in the EO cases, the optimal model followed the more flat frequency distribution. Again, it should be noted that the higher error rates are related to those frequency bands with lower powers. Furthermore, Shannon and SE values of the EEG signals and the model outputs for the EC and EO cases show close agreement. Shannon entropy values were 1.78 ± 0.04 and 1.70 ± 0.10 for EC EEG and model output, respectively, and 1.63 ± 0.32 and 1.62 ± 0.27 for EO EEG and model output, respectively. While, SE values were 1.06 ± 0.19 and 1.17 ± 0.21 for EC EEG and model output, respectively, and 1.02 ± 0.39 and 1.29 ± 0.24 for EO EEG and model output, respectively.

Power spectra of outputs of the optimal stochastic oscillator models and EEG signals for sample EC and EO cases of AD subjects are presented in **Figures 8**, **9**. Again, it is clear that the addition of SE to the objective function results in output signals with power spectra patterns which are much more similar to the EEG signal in terms of distribution and noise complexity. As expected, the power spectrum plots demonstrated that the EC EEG signals from AD subjects were slightly θ band dominated unlike α band dominance of EC EEG recordings from CTL subjects.

4. Discussion

Power spectra of the optimal stochastic oscillator model output and EEG signals show excellent agreement in the brain's major frequency bands. The comparison revealed that the optimal model is closely following the α-band dominance in EC recordings for the control subjects. Furthermore, the model for EC recordings of AD patients closely followed θ-band power dominance indicating the slowing of the EEG signal for these patients. In the EO cases, the optimal model, as expected, followed a more flat frequency distribution from upper δ to lower β frequency bands for both AD and CTL subjects. Further evidence

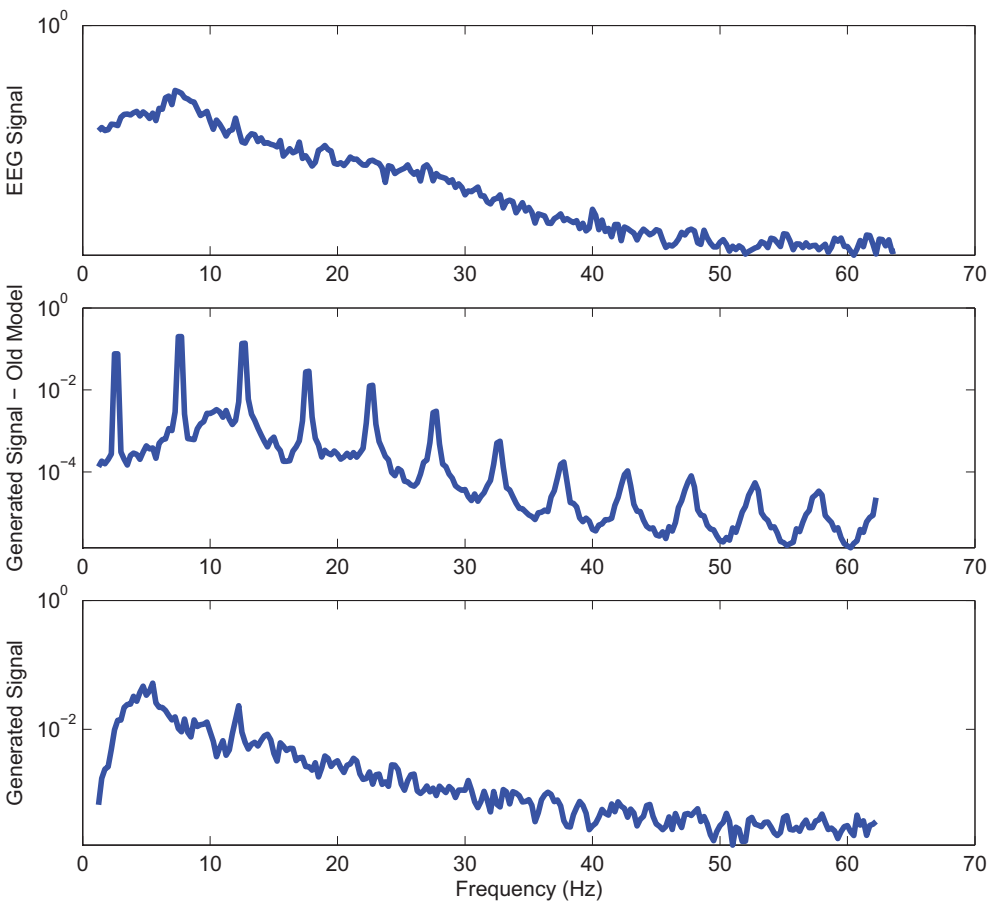


FIGURE 8 | Power spectrum of a sample AD EC (top) EEG signal; (middle) output of stochastic oscillator model using Shannon entropy; (bottom) output of stochastic oscillator model using Shannon and sample entropies.

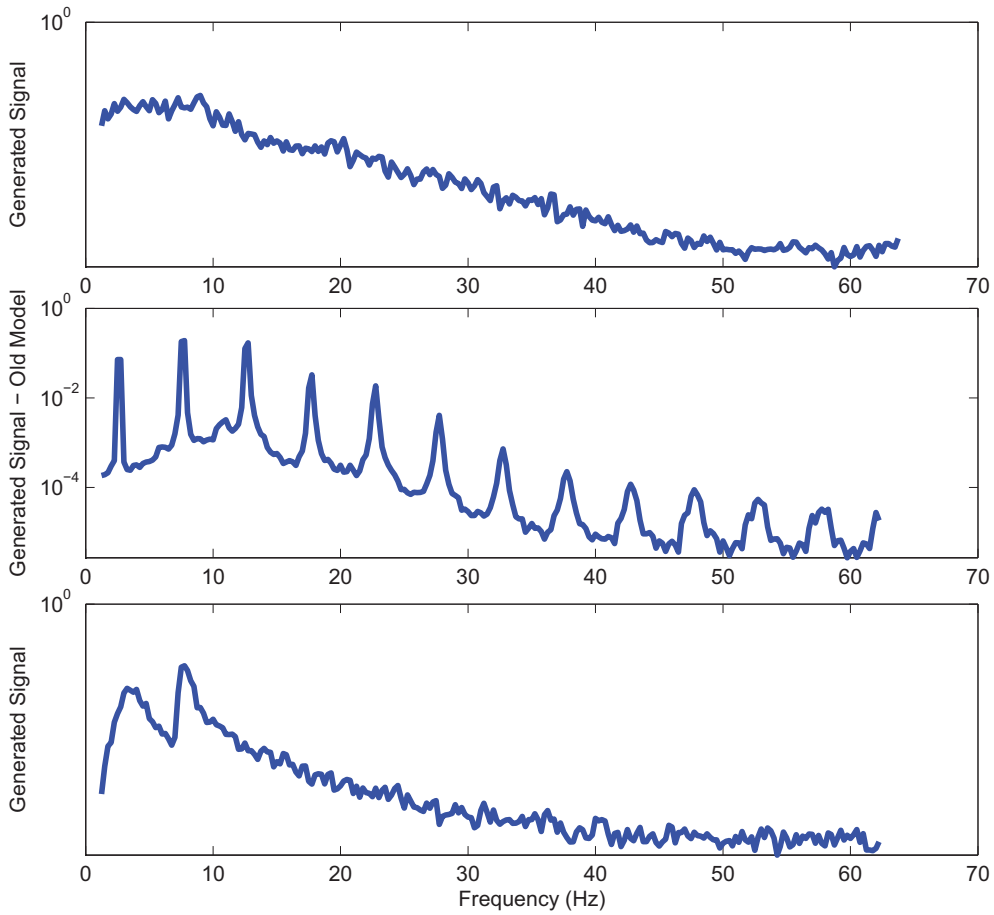


FIGURE 9 | Power spectrum of a sample AD EO (top) EEG signal; (middle) output of stochastic oscillator model using Shannon entropy; (bottom) output of stochastic oscillator model using Shannon and sample entropies.

of robustness of the the models derived in this study is that the models derived for healthy subject EC and EO EEG signals in our earlier study (Ghorbanian et al., 2015) fall within the same distributions obtained for the CTL subjects in the clinical study.

Moreover, Shannon and SE values of the EEG signals and the model outputs for the EC and EO cases show close agreement for both CTL and AD subjects. However, the difference between the entropy values of the CTL subjects and AD patients were not statistically significant for neither the EEG signal nor the model output. This aspect needs to be further studied since EEG signals from AD patients may be expected to have lower complexity and thus lower entropy values.

The contributions of the article are as follows. Firstly, the objective function of the optimization scheme that yields model parameters based on comparison with actual EEG data in our previous work was extended to include both Shannon and sample entropies, with the latter being a measure of signal complexity. The procedure yielded model outputs that were in agreement with the actual EEG signals with respect to the frequency content (power spectra), information content (Shannon entropy) and complexity (sample entropy). It was shown that the addition of SE significantly enhances the performance of the optimal model in terms of both power spectrum and non-linear characteristics displayed through phase space reconstruction. The results demonstrate the feasibility of stochastic non-linear oscillator models which can be further studied for greater insight into EEG signal dynamic characteristics.

Secondly, the model parameter differences for EC and EO EEG recordings were statistically significant leading to qualitatively and quantitatively distinct realizations of the underlying models for the cases considered. This is a key result of the work since it verifies that distinct models represent the EEG signals recorded under different brain states. Potentially, this could lead to unique models for different brain disorders and injuries.

Thirdly, the study provided unique models for EC and EO EEG recordings from AD patients. The results showed that almost all of the model parameters were statistically significant for the EC and EO cases when comparing the AD and CTL subjects. Moreover, the power spectrum plots showed a good match between the generated signal from the stochastic model and the actual EEG signal from AD patients. However, the results for the EC case of AD were more accurate and reasonable than the results

of EO cases mainly due to the ability of the optimization scheme to provide a better match in EC cases. The important conclusion here is that unique stochastic non-linear oscillator models can be developed to represent EEG signals from patients with a brain disorder.

Of particular interest is the potential connection between our model and the neural mass models studied in the literature. For instance, characterization of functional connectivity between remote cortical areas has been studied using neural mass models (David and Friston, 2003; David et al., 2004). These and other efforts (Sotero et al., 2007) represent intriguing attempts to capture actual neural dynamics using coupled oscillator models and suggest that, after all, models such as the one discussed in this article may be of broader scope than being purely phenomenological. Extrapolating further, it would then be of immense interest to understand the manifestation of phenomena such as synchronization (Mirollo and Strogatz, 1990) within the framework of our model and the implications for EEG characterization.

5. Conclusions

In this article, we presented results that further develop our recent work on modeling the EEG signal as the response of a stochastic, coupled Duffing—van der Pol system of two oscillators. The results presented verify that unique and statistically significant stochastic Duffing—van der Pol oscillator models represent EEG recorded from AD patients vs. health controls. Overall, the results presented in this article further affirm the efficacy of a stochastic Duffing—van der Pol oscillator network model in capturing the key characteristics of actual EEG data under different brain states as well as brain conditions in terms of healthy controls vs. patients with a brain disorder. The validation provided by the results certainly motivates further research toward improving the analytical model and testing it against larger data sets. Furthermore, the results suggest that the modeling approach could potentially help develop novel diagnostic and interventional tools for neurological diseases and disorders.

References

Abarbanel, H., Brown, R., Sidorowich, J., and Tsimring, L. S. (1993). Analysis of observed chaotic data. *Rev. Mod. Phys.* 65, 1331–1391.

Abasolo, D., Hornero, R., Espino, P., Alvarez, D., and Poza, J. (2006). Entropy analysis of the EEG background activity in Alzheimer's disease patients. *Physiol. Meas.* 27, 241–253. doi: 10.1088/0967-3334/27/3/003

Acedo, L., and Morano, J. (2013). Brain oscillations in a random neural network. *Math. Comput. Model.* 57, 1768–1772. doi: 10.1016/j.mcm.2011.11.028

Akin, M. (2002). Comparison of wavelet transform and FFT methods in the analysis of EEG signals. *J. Med. Syst.* 26, 241–247. doi: 10.1023/A:1015075101937

Baier, G., Hermann, T., and Muller, M. (2005). "Polyrhythmic organization of coupled nonlinear oscillators," in *International Conference on Information Visualization* (Washington, DC), 5–10.

Bressloff, P., and Lai, Y. (2011). Stochastic synchronization of neuronal populations with intrinsic and extrinsic noise. *J. Math. Neurosci.* 1, 1–28. doi: 10.1186/2190-8567-1-2

Burke, D., and Paor, A. D. (2004). A stochastic limit cycle oscillator model of the EEG. *Biol. Cyber.* 91, 221–230. doi: 10.1007/s00422-004-0509-z

Celka, P., and Colditz, P. (2002). Nonlinear nonstationary wiener model of infant EEG seizures. *IEEE Trans. Biomed. Eng.* 49, 556–564. doi: 10.1109/TBME.2002.1001970

Dafilis, M., Frascoli, F., Cadusch, P., and Liley, D. (2009). Chaos and generalised multistability in a mesoscopic model of the electroencephalogram. *Physica D* 238, 1056–1060. doi: 10.1016/j.physd.2009.03.003

Darvishi, S., and Al-Ani, A. (2007). "Brain-computer interface analysis using continuous wavelet transform and adaptive neuro-fuzzy classifier," in *International Conference of the IEEE Engineering in Medicine and Biology Society* (Lyon), 3220–3223.

Dauwels, J., Vialatte, F., and Cichocki, A. (2010). Diagnosis of Alzheimer's disease from EEG signals: where are we standing. *Curr. Alzheimer Res.* 7, 487–505. doi: 10.2174/156720510792231720

David, O., Cosmelli, D., and Friston, K. J. (2004). Evaluation of different measures of functional connectivity using a neural mass model. *Neuroimage* 21, 659–673. doi: 10.1016/j.neuroimage.2003.10.006

David, O., and Friston, K. J. (2003). A neural mass model for meg/eeg: coupling and neuronal dynamics. *Neuroimage* 20, 1743–1755. doi: 10.1016/j.neuroimage.2003.07.015

Elgendi, M., Vialatte, F., Cichocki, A., Latchoumane, C., Jeong, J., and Dauwels, J. (2011). "Optimization of EEG frequency bands for improved diagnosis of Alzheimer disease," in *International Conference of the IEEE Engineering in Medicine and Biology Society* (Boston: MA), 6087–6091.

Fell, J., Kaplan, A., Darkhovsky, B., and Roschke, J. (2000). EEG analysis with nonlinear deterministic and stochastic methods: a combined strategy. *Acta Neurobiol. Exp.* 60, 87–108.

Gardiner, C. (1985). *Handbook of stochastic methods for Physics, Chemistry and the Natural Sciences*. New York, NY: Springer.

Ghorbanian, P., Devilbiss, D., Verma, A., Bernstein, A., Hess, T., Simon, A., et al. (2013). Identification of resting and active state EEG features of Alzheimer's disease using discrete wavelet transform. *Ann. Biomed. Eng.* 41, 1243–1257. doi: 10.1007/s10439-013-0795-5

Ghorbanian, P., Ramakrishnan, S., Whitman, A., and Ashrafiuon, H. (2015). A phenomenological model of EEG based on the dynamics of a stochastic Duffing–van der Pol oscillator network. *Biomed. Signal Process. Control* 15, 1–10. doi: 10.1016/j.bspc.2014.08.013

Hernandez, J., Valdes, P., and Vila, P. (1996). EEG spike and wave modeled by a stochastic limit cycle. *Neuroreport* 7, 2246–2250.

Hettiarachchi, I., Mohamed, S., and Nahavandi, S. (2012). "A marginalised Markov chain Monte Carlo approach for model based analysis of EEG data," in *IEEE International Symposium on Biomedical Imaging* (Barcelona), 1539–1542.

Higham, D. J. (2001). An algorithmic introduction to numerical simulation of stochastic differential equations. *SIAM Rev.* 43, 525–546. doi: 10.1137/S0036144500378302

Huang, G., Zhang, D., Meng, J., and Zhu, X. (2011). Interactions between two neural populations: a mechanism of chaos and oscillation in neural mass model. *Neurocomputing* 74, 1026–1034. doi: 10.1016/j.neucom.2010.11.019

Jeong, J. (2004). EEG dynamics in patients with Alzheimer's disease. *Clin. Neurophysiol.* 15, 1490–1505. doi: 10.1016/j.clinph.2004.01.001

Kantz, H., and Schreiber, T. (2004). *Nonlinear Time Series Analysis*. New York, NY: Cambridge University Press.

Kim, J., Shin, H., and Robinson, P. (2007). Compact continuum brain model for human Electroencephalogram. *Complex Syst. II* 6802, T1–T8. doi: 10.1117/12.759005

Lake, D., and Moorman, J. (2011). Accurate estimation of entropy in very short physiological time series: the problem of atrial fibrillation detection in implanted ventricular devices. *Am. J. Physiol. Heart Circ. Physiol.* 300, H319–H325. doi: 10.1152/ajpheart.00561.2010

Leistritz, L., Putsche, P., Schwab, K., Hesse, W., Susse, T., Haueisen, J., et al. (2007). Coupled oscillators for modeling and analysis of EEG/MEG oscillation. *Biomed. Tech.* 52, 83–89. doi: 10.1515/BMT.2007.016

Li, Z., Xu, W., and Zhang, X. (2006). Analysis of chaotic behavior in the extended duffing–van der Pol system subject to additive non-symmetry biharmonical excitation. *Appl. Math. Comput.* 183, 858–871. doi: 10.1016/j.amc.2006.06.033

McBride, J., Zhao, X., Munro, N., Smith, C., Jicha, G., Hively, L., et al. (2014). Spectral and complexity analysis of scalp EEG characteristics for mild cognitive impairment and early Alzheimer's disease. *Comput. Methods Programs Biomed.* 114, 153–163. doi: 10.1016/j.cmpb.2014.01.019

McBride, J., Zhao, X., Munro, N., Smith, C., Jicha, G., and Jiang, Y. (2013a). Resting EEG discrimination of early stage Alzheimer's disease from normal aging using inter-channel coherence network graphs. *Ann. Biomed. Eng.* 41, 1233–1242. doi: 10.1007/s10439-013-0788-4

McBride, J., Zhao, X., Nichols, T., Vagnini, V., Munro, N., Berry, D., et al. (2013b). Scalp eeg-based discrimination of cognitive deficits after traumatic brain injury using event-related tsallis entropy analysis. *IEEE Trans. Biomed. Eng.* 60, 90–96. doi: 10.1109/TBME.2012.2223698

Mirollo, R. E., and Strogatz, S. H. (1990). Synchronization of pulse-coupled biological oscillators. *SIAM J. Appl. Math.* 50, 1645–1662.

Mizrach, B. (1996). Determining delay times for phase space reconstruction with application to the FF/DM exchange rate. *J. Econ. Behav. Organ.* 30, 369–381.

Nie, Z., Hao, H., and Ma, H. (2013). Structural damage detection based on the reconstructed phase space for reinforced concrete slab: experimental study. *J. Sound Vibrat.* 332, 1061–1078. doi: 10.1016/j.jsv.2012.08.024

Pincus, S. (1991). Approximate entropy as a measure of system complexity. *Proc. Natl. Acad. Sci. U.S.A.* 88, 2297–2301.

Quiroga, R., Rosso, O., Basar, E., and Schurmann, M. (2001). Wavelet entropy in event-related potentials: a new method shows ordering of EEG oscillations. *Biol. Cyber.* 84, 291–299. doi: 10.1007/s004220000212

Rankine, L., Stevenson, N., Mesbah, M., and Boashash, B. (2007). A nonstationary model of newborn EEG. *IEEE Trans. Biomed. Eng.* 54, 19–28. doi: 10.1109/TBME.2006.886667

Richman, J., and Moorman, J. (2000). Physiological time-series analysis using approximate entropy and sample entropy. *Am. J. Physiol. Heart Circ. Physiol.* 278, H2039–H2049.

Rodrigues, S., Goncalves, J., and Terry, J. (2007). Existence and stability of limit cycles in a macroscopic neuronal population model. *Physica D* 233, 39–65. doi: 10.1016/j.physd.2007.06.010

Shannon, C. (1948). A mathematical theory of communication. *Bell Syst. Tech. J.* 27, 623–656.

Shin, H., Tong, S., Yamashita, S., Jia, X., Geocadin, R., and Thakor, N. (2006). Quantitative EEG and effect of hypothermia on brain recovery after cardiac arrest. *IEEE Trans. Biomed. Eng.* 53, 1016–1023. doi: 10.1109/TBME.2006.873394

Slobounov, S., Cao, C., and Sebastianelli, W. (2009). Differential effect of first versus second concussive episodes on wavelet information quality of EEG. *Clin. Neurophysiol.* 120, 862–867. doi: 10.1016/j.clinph.2009.03.009

Sotero, R. C., Trugillo-Barreto, N. J., Itturia-Medina, Y., Carbonell, F., and Jimenez, J. C. (2007). Realistically coupled neural mass models can generate eeg rythms. *Neural Comput.* 19, 478–512. doi: 10.1162/neco.2007.19.2.478

Stam, C. J. (2005). Nonlinear dynamical analysis of EEG and MEG: review of an emerging field. *Clin. Neurophysiol.* 116, 2266–2301. doi: 10.1016/j.clinph.2005.06.011

Sun, S., Lan, M., and Lu, Y. (2008). "Adaptive EEG signal classification using stochastic approximation methods," in *International Conference Acoustics, Speech and Signal Process* (Las Vegas, NV), 413–416.

Ugray, Z., Lasdon, L., Plummer, J., Glover, F., Kelly, J., and Marti, R. (2007). Scatter search and local nlp solvers: a multistart framework for global optimization. *Inf. J. Comput.* 19, 328–340. doi: 10.1287/ijoc.1060.0175

Valdes, P., Jimenez, J., Riera, J., Biscay, R., and Ozaki, T. (1999). Nonlinear EEG analysis based on a neural mass model. *Biol. Cyber.* 81, 415–424.

Wang, J., Wang, B., Wang, X., and Nakamura, M. (2011). "MPA EEG model–based vigilance level estimation by artificial neural network," in *International Conference on Biomedical Engineering and Informatics* (Shanghai), 795–799.

Xu, G., Zhang, X., Yu, H., Ho, S., Yang, Q., Fu, W., et al. (2010). Complexity analysis of EEG under magnetic stimulation at acupoints. *IEEE Trans. Appl. Superconductivity* 20, 1029–1032. doi: 10.1109/TASC.2010.2040726

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.

frontiers in Computational Neuroscience

OPINION
published: 22 May 2015
doi: 10.3389/fncom.2015.00058

Multisensory integration using dynamical Bayesian networks

Taher Abbas Shangari[1], *Mohsen Falahi*[1], *Fatemeh Bakouie*[2]* and *Shahriar Gharibzadeh*[2]*

[1] *Amirkabir Robotic Center, Amirkabir University of Technology, Tehran, Iran,* [2] *Institute for Cognitive and Brain Sciences, Shahid Beheshti University, Tehran, Iran*

Keywords: multisensory integration, Dynamic Bayesian Networks, modeling, sensory processing disorder, Bayesian Models

Multisensory Integration (MSI) is the study of how information coming from different sensory modalities, such as vision, audition and etc. are being integrated by the nervous system (Stein et al., 2009) as a complex system. MSI is one of the most important aspects of neuroscience which has a great influence on our decision making system. It plays a key role in our understanding of surrounding environment which makes a coherent representation of the world for us (Lewkowicz and Ghazanfar, 2009). Since signals in our sensory systems are corrupted by variability or noise, the nervous system combines different kinds of sensory information like sound, touch etc. to achieve a meaningful and continuous stream of percepts (Kording and Wolpert, 2006; Lewkowicz and Ghazanfar, 2009). Recently, researchers have shown an increased interest in MSI modeling, to discover the causes of related disorders such as under-sensitivity or hyposensitivity (Knill and Pouget, 2004). Moreover individuals with Autism Spectrum Disorder (ASD) have an impaired ability to integrate multisensory information to make a unified percept (Stevenson et al., 2014).

Different researches have modeled MSI in a variety of ways. Computational methods, such as Kalman Filter (KF) and Bayesian Networks (BN) are used widely to model probabilistic functions of the nervous system including MSI (Van Der Kooij et al., 1999; Kording and Wolpert, 2004). In KF-based models there is a basic assumption on accuracy of the sensory input data. This assumption says that the error's Probability Density Function (PDF) of each sensor is Gaussian. According to KF, it is provable that data fusion of two different kinds of data for one variable measurement leads to more accurate results (Kalman, 1960). A serious weakness with this method, however, is its basic assumption. Assuming a Gaussian form of the PDF of the sensory systems' error is in contradiction with the brain's internal models and prior knowledge about human sensory system and environmental models which are not necessarily Gaussian-like. Additionally, as different formats are used by each sensory modality to encode the same properties of the environment or body, MSI cannot be as simple as an averaging between sensory inputs (Deneve and Pouget, 2004). Hence, it is clear that KF-based models are not valid for many MSI studies and therefore researchers tried to modify this method (Van der Zijpp and Hamerslag, 1994; Julier and Jeffrey, 2004).

Since BNs have not any assumption on accuracy of the input data, they have attracted much attention recently. A BN is a graphical model that represents probabilistic relationships among variables of interest. By using graphical models in conjunction with statistical techniques, several advantages for data analysis will be obtained: Firstly, because a BN represents conditional dependencies among all variables, it is able to handle situations where some data entries are missing. Secondly, the model can be used to learn causal relationships, so it can be used to understand a problem domain and to predict the consequences of intervention. Thirdly, because BNs have both causal and probabilistic semantics, they represent combining prior knowledge and data ideally (Heckerman, 1998; Wasserman, 2011).

OPEN ACCESS

Edited by:
Tobias Alecio Mattei, Brain & Spine Center - InvisionHealth - Kenmore Mercy Hospital, USA

Reviewed by:
Malte J. Rasch, Beijing Normal University, China

***Correspondence:**
Fatemeh Bakouie, f_bakouie@sbu.ac.ir; Shahriar Gharibzadeh, gharibzadeh@aut.ac.ir

Received: *21 February 2015*
Accepted: *29 April 2015*
Published: *22 May 2015*

Citation:
Abbas Shangari T, Falahi M, Bakouie F and Gharibzadeh S (2015) Multisensory integration using dynamical Bayesian networks. Front. Comput. Neurosci. 9:58. doi: 10.3389/fncom.2015.00058

Generally, there are three main inference tasks for BNs: inferring unobserved variables, parameter learning, and structure learning. They are used widely for modeling knowledge in computational biology, bioinformatics, etc. For example, a BN could represent the probabilistic relationships between diseases and symptoms. Given symptoms, the network can be used to compute the probabilities of the presence of various diseases.

As it mentioned before, the brain needs using different resources of information altogether to be able to make a sound decision about a situation. In such cases BNs can be used to model brain's function in many studies (Seilheimer et al., 2014). It is worth mentioning that in BNs, relationship between different nodes is not as simple as an averaging and we can model more complex probabilistic problems by using BNs (Bishop and Nasser, 2006).

However, it is obvious that the reliability of sensory modalities varies widely according to the context and in a BN the effect of one node on the other one can vary from one task or situation to another one. But it is clear that when we assume a node as a parent node for another one, this relation could not be changed and new experiences would not cause new links between separated nodes. The main weakness of the BN based models is the failure to address the way it uses to reconstruct the network, based on new observed experiences. Most studies in MSI modeling have only focused on one task in which the effective sensory resources are known before, therefore, the structure of the network is known too, and we only need to train the network. By contrast, when we want to model MSI, we should not restrain it only in some certain tasks but the model should instead be generalizable to other tasks. It means that the model should be more dynamic and task independent. In addition, it is clear that time has a great influence in our decision making and reasoning and unfortunately, BN fails to code the time directly (Mihajlovic and Petkovic, 2001).

We suggest that, MSI models will be more generalized if we use Dynamic Bayesian Networks (DBN) which describes a system that dynamically changes over time. In a BN that models the interactions between sensory modalities, the nodes are associated with activated sensory modalities and the edges represent the interactions among sensory modalities. Sensory modalities of a neural system including n sensory modalities are indexed in a set $I = \{i : i = 1, 2, \ldots n\}$. Consider activation of a sensory modality measured by fMRI time-series or EEG over the sensory modality. Let x_i be the activation measuring the response of sensory modality *i*.

BNs describe the PDF over the activation of sensory modalities, where the graphical structure provides an easy way to specify conditional interdependencies for a compact parameterization of the distribution. A BN defined by a structure S is a directed acyclic graph (DAG) and a joint distribution over the set of time-series $x = \{x_i : i \in I\}$. The set of activations of the parents of sensory modality i is denoted by a_i, and a DAG offers a simple and unique way to decompose the likelihood of activation in terms of conditional probabilities: where $\theta = \{\theta_i : i \in I\}$ represents the parameters of the conditional probabilities (Rajapakse and Zhou, 2007).

DBNs extend BNs to incorporate temporal characteristics of the time-series x. $x(t) = \{x_i(t) : i \in I\}$ represents the activations of n sensory modalities at time t, where the instances $t = 1, 2, \ldots T$ correspond to the times when sensory modality measures are taken and T denotes the total number of measures. In order to model the temporal dynamics of brain processes, we need to model a probability distribution over the set of random variables $\bigcup_{t=1}^{T} x(t)$ which is complex and practically hard.

To avoid an explosion of the model complexity, one can assume that the temporal changes of activations of brain regions are stationary and first-order Markovian. This assumption provides a tractable causal model that explicitly takes into account the temporal dependencies of brain processes. When facing more complex temporal processes and connectivity patterns, higher-order and non-stationary Markov models can be used to overcome the complexity.

The connectivity structure between two consecutive data sampling is represented by the transition network, which renders the joint distribution of all possible trajectories of temporal processes. The structure of the DBN is obtained by unrolling the transition network over consecutive scans for all $t = 1, 2, \ldots, T$ (Rajapakse and Zhou, 2007).

In an overview, we here suggest that DBN may be a more useful method to model MSI in comparison to prior methods because of three reasons. Firstly, as DBN changes dynamically, initial structure of the network does not lead to an unreliable result and we can use the network in various kinds of studies (because this method is task-independent). Secondly, in cases which we are not sure about the relation and interaction between different sensory modalities, DBN output can help us to achieve a more accurate understanding about MSI processes. Moreover, there exist cyclic functional networks in the brain, such as cortico-subcortical loops which BNs are not capable to model. Unlike BN, DBN has the capability of modeling recurrent networks while still satisfying the acyclic constraint of the transition network (Rajapakse and Zhou, 2007). This is an important advantage of modeling neural system with DBN as these key features of DBN help us to obtain a proper viewpoint about MSI in different tasks and it makes the study of related disorders easier and closer to reality.

References

Bishop, C. M., and Nasser, M. N. (2006). *Pattern Recognition and Machine Learning*, Vol. 1. New York, NY: Springer.

Deneve, S., and Pouget, A. (2004). Bayesian multisensory integration and cross-modal spatial links. *J. Physiol. Paris* 98, 249–258. doi: 10.1016/j.jphysparis.2004.03.011

Heckerman, D. (1998). *A Tutorial on Learning with Bayesian Networks*. Springer Netherlands.

Julier, S. J., and Jeffrey, K. U. (2004). Unscented filtering and nonlinear estimation. *Proc. IEEE* 92, 401–422. doi: 10.1109/JPROC.2003.823141

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *J. Basic Eng.* 82, 35–45.

Knill, D. C., and Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends Neurosci.* 27, 712–719. doi: 10.1016/j.tins.2004.10.007

Kording, K. P., and Wolpert, D. M. (2004). Bayesian integration in sensorimotor learning. *Nature* 427, 244–247. doi: 10.1038/nature02169

Kording, K. P., and Wolpert, D. M. (2006). Bayesian decision theory in sensorimotor control. *Trends Cogn. Sci.* 10, 319–326. doi: 10.1016/j.tics.2006.05.003

Lewkowicz, D. J., and Ghazanfar, A. A. (2009). The emergence of multisensory systems through perceptual narrowing. *Trends Cogn. Sci.* 13, 470–478. doi: 10.1016/j.tics.2009.08.004

Mihajlovic, V., and Petkovic, M. (2001). *Dynamic Bayesian Networks: A State of the Art*. Enschede: University of Twente, Centre for Telematics and Information Technology.

Rajapakse, J. C., and Zhou, J. (2007). Learning effective brain connectivity with dynamic Bayesian networks. *Neuroimage* 37, 749–760. doi: 10.1016/j.neuroimage.2007.06.003

Seilheimer, R. L., Rosenberg, A., and Angelaki, D. E. (2014). Models and processes of multisensory cue combination. *Curr. Opin. Neurobiol.* 25, 38–46. doi: 10.1016/j.conb.2013.11.008

Stein, B. E., Stanford, T. R., and Rowland, B. A. (2009). The neural basis of multisensory integration in the midbrain: its organization and maturation. *Hear Res.* 258, 4–15. doi: 10.1016/j.heares.2009.03.012

Stevenson, R. A., Siemann, J. K., Schneider, B. C., Eberly, H. E., Woynaroski, T. G., Camarata, S. M., et al. (2014). Multisensory temporal integration in autism spectrum disorders. *J. Neurosci.* 34, 691–697. doi: 10.1523/JNEUROSCI.3615-13.2014

Van Der Kooij, H., Jacobs, R., Koopman, B., and Grootenboer, H. (1999). A multisensory integration model of human stance control. *Biol. Cybern.* 80, 299–308.

Van der Zijpp, N. J., and Hamerslag, R. (1994). Improved Kalman filtering approach for estimating origin-destination matrices for freeway corridors. *Trans. Res. Record* 1443, 100–123.

Wasserman, L. (2011). *All of Statistics*. New York, NY: Springer Science & Business Media.

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.

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