AGE-RELATED CHANGES IN AUDITORY PERCEPTION

EDITED BY: Leah Fostick and Bruce A. Schneider
PUBLISHED IN: Frontiers in Psychology and Frontiers in Neuroscience

frontiers Research Topics

Frontiers eBook Copyright Statement

ISSN 1664-8714
ISBN 978-2-88976-994-0
DOI 10.3389/978-2-88976-994-0

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: frontiersin.org/about/contact

AGE-RELATED CHANGES IN AUDITORY PERCEPTION

Topic Editors:
Leah Fostick, Ariel University, Israel
Bruce A. Schneider, University of Toronto Mississauga, Canada

Citation: Fostick, L., Schneider, B. A., eds. (2022). Age-related Changes in Auditory Perception. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88976-994-0

Table of Contents

frontiers | Frontiers in Psychology

TYPE Editorial
PUBLISHED 08 August 2022
DOI 10.3389/fpsyg.2022.986586

OPEN ACCESS

EDITED AND REVIEWED BY
Rufin VanRullen,
Centre National de la Recherche Scientifique (CNRS), France

*CORRESPONDENCE
Leah Fostick
leah.fostick@ariel.ac.il

SPECIALTY SECTION
This article was submitted to Perception Science, a section of the journal Frontiers in Psychology

RECEIVED 05 July 2022
ACCEPTED 14 July 2022
PUBLISHED 08 August 2022

CITATION
Fostick L and Schneider BA (2022) Editorial: Age-related changes in auditory perception. *Front. Psychol.* 13:986586. doi: 10.3389/fpsyg.2022.986586

Editorial: Age-related changes in auditory perception

Leah Fostick[1]* and Bruce A. Schneider[2]

[1]Department of Communication Disorders, Ariel University, Ariel, Israel, [2]Department of Psychology, University of Toronto, Toronto, ON, Canada

KEYWORDS

auditory perception, aging, bottom-up, top-down, speech perception

Editorial on the Research Topic
Age-related changes in auditory perception

Auditory sensory systems have evolved to enable us to detect, locate, identify, and comprehend the various sound sources in our environment, in the form of auditory streams. In turn, the information available in these auditory streams has to be integrated with our world knowledge to determine how we should respond to our immediate environment, and to select or focus our attention on the auditory streams that are important to our short- or long-term survival. In other words, navigating in the real world requires both bottom-up processing of information to arrive at an accurate representation of our environment, and top-down control over the upward flow of information to focus attention on aspects of the environment that are critical to our survival. Hence, age-related changes in sensory and perceptual processes are quite likely to affect cognitive processing (e.g., speech comprehension), not only because such changes may degrade the perceptual representations of stimuli, but also because the degradation of the sensory information is likely to increase the demand on resources that are also required for efficient cognitive processing of the incoming information.

This present Research Topic of papers addresses a number of critical issues with regard to how aging might affect the pattern of interactions between bottom-up and top-down processes that are involved in the extraction of information about the world we live in. As a starting point, it should be noted that there is a huge body of evidence that the bottom-up processing of auditory information deteriorates with age. Pure-tone thresholds, temporal discriminability, the sensitivity of individuals to interaural cues, etc., decline with age. As a result, older individuals necessarily have to rely more than do younger individuals on higher-order, cognitive processes to extract information from the auditory signal when that signal is embedded in soundscapes consisting of a multitude of other auditory sources (for a review, see Schneider et al., 2010). Hence, we would expect that older adults who score high with respect to those cognitive abilities believed to be associated with the processing of speech would have better basic auditory capabilities that those older adults scoring lower with respect to those cognitive abilities. In this Research Topic of papers, Humes et al. found significant correlations between visually-assessed working memory and 7 of 8 tests of Basic Auditory Capabilities (TBAC) in a sample of 115 older adults, indicating that higher-order, age-related changes in cognitive level abilities can affect lower level processing of the acoustic scene. In other words, older adults with good working memories are better at extracting the bottom-up information in the auditory signal that allows them to detect, locate, and process attended sound sources.

The facilitative effect on cognitive-level processes, such as working memory on speech understanding, raises the question as to whether there are any age-related differences in the contribution of the cognitive abilities to speech understanding. Tamati et al. (this Research Topic) looked for age-related differences in the relative contributions of bottom-up and top-down processes in the perception of heard speech distorted by noise vocoding. They presented younger and older listeners with sentences preceded by a visual lexical prime that was ortho-graphically identical to that of the vocoded sentences or with a prime consisting of nonsense words. In addition, the lexical frequency and neighborhood density of the target words in the vocoded sentences were manipulated. Interestingly, although there were significant effects of both bottom-up (noise-vocoding) and top-down (priming, lexical frequency, neighborhood density) in the expected direction, no age-related differences were found, suggesting that the contribution of working memory to speech perception in this situation was equivalent in younger and older listeners.

The complexity of the interaction between top-down and bottom-up processes is illustrated in a third paper in this Research Topic. Weissgerber et al. found that measures on a German language age-standardized test of cognitive ability (Dem Tect) were correlated with speech perception in noise, a correlation that disappeared when corrected for high-frequency hearing loss. They note that a high-frequency hearing loss could have produced lower scores on the Dem Tect test which included acoustically presented test items that could have been misheard due to a high-frequency hearing loss. This result, along with a study by Fullgrabe (2020) that found that simulated hearing loss reduced scores on cognitive tests involving acoustically-presented material in young listeners, indicate that assessment of the higher-order cognitive processes thought to be important in speech perception tasks, can be affected by age-related changes in basic auditory processes. In addition, Zaltz and Kishon-Rabin (this Research Topic) found that even when a cognitive test does not involve aurally presented material, there are complex interactions between tests of basic auditory abilities and cognitive capacities. Specifically, these investigators found that the ability of older adults to take advantage of differences in fundamental frequency and formant structure to discriminate among different voices was related to both hearing sensitivity and a measure of cognitive ability based on a visual test (Trail Making Test).

Shvartzman et al. (this Research Topic), in addition to confirming the importance of the interaction of bottom-up and top-down processes in speech perception, also suggest that individual differences in the ability to rapidly reorganize perceptual processes to respond appropriately to a consistent set of auditory features of a sound source (perceptual learning), are correlated with performance in some types of difficult listening situations such as the processing of rapid speech. This suggests that individuals who are capable of rapid perceptual learning, are better able to adjust to the idiosyncrasies of a person's speech, thereby permitting them to adjust their speech-processing mechanisms to be better able to function in a complex auditory scene, especially one where they are exposed for the first time to a new speaker.

The preceding studies demonstrate that there are complex interactions between basic sensory processes and a number of cognitive processes involved in processing speech. This raises the question of how we might go about dissecting the nature of these processes. One way is to focus on the ability of an individual to make use of the information provided by a basic auditory process in performing a higher-order task where one might expect the performance of the higher-order task to be critically dependent on the information provided by a specific lower-order auditory ability. The study by Szelag et al. is a nice example of how limitations on a lower-level, bottom-up process, can affect the ways in which a higher-order, cognitive task is conducted. The lower-level task in this study was a temporal order judgment. After categorizing individuals into either low or high performers on this temporal order task, they then assessed how they performed a higher-order, but still temporally-based task, that might be expected to be affected by the individual's degree of lower-order temporal discriminability. In the higher-order task, the listener was presented with a series of clicks where the inter-click interval was fixed. The listener was instructed to mentally create a beat structure for this sequence by mentally accentuation some of the beats. These investigators found that the strategies used by participants to create a beat structure depended on their lower-order temporal order judgments, indicating that the strategies that listeners used in performing higher-order auditory tasks are conditional upon the lower-order processing capacities that might be useful in performing the task.

Of course, any task that places demands on a cognitive ability, such as working memory, has implications for situations where a person, in addition to attempting to understand speech in a noisy situation, has to simultaneously perform a different task that also draws on this cognitive level ability. The Nitsan et al. eye-tracking study (this Research Topic) found that the way in which older adults processed words presented in noise while performing a secondary task (digit recall) depended on their working memory capacity. The working memory load on the secondary task could be either low (1 digit presented) or high (4 digits presented). On trials in which the target word was correctly identified (by means of eye-tracking), the working-memory load on the secondary task affected the proportion of times the low-working-memory capacity individuals responded correctly. Their performance in the secondary digit recall task dropped significantly from when the working-memory load on the secondary task was low (a single digit) to when it was high (4 digits). No such decline was observed in the high-working-memory individuals. The pattern of eye movements during this task indicated that there were differences between high- and low-working-memory individuals in the ways in which top-down resources were allocated to this task, and that individuals with low working-memory capacity, when faced with a high-working-memory demand on the primary task might be unable to muster sufficient resources to perform well on the secondary task.

Clearly, demonstrating that there is a complex pattern of interaction between top-down and bottom-up processes involved in using auditory information to help us understand and navigate in the real world, doesn't specify precisely how, when, and where in sensory and perceptual processing such interactions take place. What is needed is the ability to examine this process through a moving temporal window to help us understand when and where such interactions occur. A number of studies have begun to use eye-tracking techniques to provide us with a temporal breakdown of this process. For example, Failes and Sommers (this Research Topic) used eye-tracking to identify age-related differences in the degree to which younger and older adults differed with respect to how preceding sentential context affected the correct identification of the final word in a spoken sentence. In some sentences, the preceding sentential context supported the sentence final word, in other sentences the preceding sentential context suggested a phonological competitor, while in a third type of sentence, the context prior to the sentence final word could not be readily used to predict the sentence-final word. Four images of objects were shown on a screen prior to the presentation of the test sentence with one of the objects corresponding to the sentence final word, another to a phonological competitor, along with two objects used as foils. By comparing the time-course of eye movements among these objects, they were able to identify intriguing differences between younger and older subjects with respect to how and when context influenced an individual's pattern of fixations during the presentation of sentences, supporting the notion that the manner in which top-down knowledge affects speech perception, can differ with age.

We also need to consider the contribution of non-auditory cues that contribute to speech understanding in everyday environments: namely the importance of visual cues to speech in difficult listening situations. Gordon-Salant et al. (this Research Topic) presented younger and older listeners (with and without hearing losses) with a visual image of the speaker they were attending to and varied the asynchrony of the visual and auditory components of speech. These investigators found that older adults (both with and without hearing loss) had higher thresholds for detecting an asynchrony between visible and audible speech. However, in all three groups, speech perception scores were equally affected by the degree of asynchrony, indicating that the contribution of visible speech to speech perception was the same in younger and older adults (once the signal-to-noise ratio was adjusted to produce equivalent speech recognition scores in all three groups) but that older adults had a higher threshold for detecting an asynchrony between the two. Here, as well as in other studies in this Research Topic, the effectiveness of top-down processes does not appear to be significantly affected by age when speech understanding is the primary task, suggesting that higher-order mechanisms remain effective in aging listeners. However, this does not necessarily mean that the manner in which this knowledge is used in aid of speech perception is the same in younger and older listeners (see Failes and Sommers above).

Spoken language, in addition to conveying semantic information from the talker to the listener also contains emotional information as well. There are two sources of emotional information in speech: the semantic content of the speech that conveys information about an emotional state (e.g., I am really sad about this vs. I am really mad about this); and/or its prosody (the emotional tone of the speech conveyed by suprasegmental information derived from the tone of the speech such as the stress pattern, rhythm, and pitch). In this Research Topic of papers, Dor et al. looked for age-related changes in the ability of listeners to identify the emotional content of the speech when it was being masked by speech-spectrum noise. These investigators found, that although older adults needed a higher signal-to-noise ratio than younger adults, for both age groups, the emotion conveyed by prosody required a smaller signal-to-noise ratio for detection than emotion conveyed by semantic content, and that the percentage of correct detection of emotion increases in the same way as a function of the signal-to-noise ratio for both age groups. This suggests that the cognitive mechanisms responsible for the detection of emotion do not change with age. However, when the semantic emotional content differs from the prosodic emotional content, and the listener was asked to base their judgment on only one of the two conveyers of emotion, there was some evidence that older adults' judgements were more affected by the not-to-be-attended channel than younger adults, indicating potential age-related capacity limitations on top-down resources.

In considering how age affects auditory processing, it is also reasonable to investigate how technological changes have affected our soundscapes (e.g., the increasing importance of broadcast media, the use of sound amplification, surround sound, and immersive environments), and whether older adults are as well-equipped as younger adults to navigate these environments. Russell, in this Research Topic, discusses our limited understanding of how chronological age affects the perception of space, when that perception is based on acoustic cues. This limitation extends to how spatial perception is altered by technological changes in our everyday soundscapes. One of these changes involves the broadcast of the audio and visual components of a scene to remote receivers with the result that delays are occasionally introduced between the audio and visual portions of the broadcasts (for the effects of such delays, see Gordon-Salant et al., this Research Topic). Another instance involves the often-ubiquitous use of surround sound, which removes some of the acoustic cues to the spatial location of a sound source and changes its timbre due to comb filtering. Hence, modern soundscapes can consist of a mixture of well-localized auditorily compact sources, as well as those that have a much more diffuse timbre and are less precisely localized. Avivi-Reich et al. (this Research Topic) studied the ability of young adults (both native and non-native speakers of English) and older native speakers of English to identify auditory targets in a background of competing sound sources. Both targets and competing masking sources could be either compact or diffuse. They found that aside from the usual signal-to-noise differences between younger and older native listeners, the effects of a difference in timbre between masker and target were the same for these two groups. However, the younger non-native listeners differed from the former two groups in that the young non-native group tended to perceive all four combinations of target and masker timbre equivalently. This suggests that the listener's knowledge of the language affected how well they could make use of timbre differences due to the use of surround sound, whereas their age did not affect their ability to respond to timbre differences.

A common theme that appears to be emerging from the studies in this Research Topic is that, provided that the draw on top-resources is not too extensive (as could happen in dual-task situations), older adults without cognitive impairment are as capable as younger adults in using top-down knowledge and top-down processing abilities to parse the auditory scene, and to extract targeted information from that scene. However, in everyday situations, older adults most likely have to draw more on top-down resources in order to maintain an acceptable level of speech understanding than do younger adults, even when the signal-to-noise ratios are adjusted to equate performance between these two groups. This suggests that listening in everyday situations is more effortful for older than for younger adults, which could result in greater fatigue and a withdrawal from social interactions. Therefore, there exists a need in the clinic to assess the degree of effort involved in listening in noisy settings. In laboratory settings, dual-task paradigms are typically used to assess listening effort (see the Nitzen et al. study, this Research Topic, for an example of the use of a dual-task situation). Neeman et al. (this Research Topic) have developed and tested a relatively simple test of listening effort using equipment that would be found in audiological settings. They gave this test to a sample of both young and middle-aged adults. In this task, they adjusted the signal-to-noise ratio to obtain individual speech reception thresholds of 80% correct while measuring the cost of the dual-task on the secondary task. They found listening-effort effects in both age groups. However, the cost of the secondary task was greater in the middle-aged listeners than in the younger listeners, indicating that listening effort increases with age, and that this increased effort can be assessed in audiological settings, and allow the audiologist to address the patients' concerns by discussing with them the cost of listening in noisy environments and ways in which such costs can be reduced (e.g., use of assistive devices such as directional microphones, etc.).

Finally, there is the question of how effective certain interventions are in improving the quality of the lives of older persons with hearing impairment, such as cochlear implant users. In this issue, Brumer et al. evaluated the health-related quality of life of individuals with cochlear implants. These investigators found that bimodal and bilateral cochlear implant users who were better able to function in noisy environments experienced a higher degree of life-satisfaction, as measured by the Glasgow Benefit Inventory. More studies are needed on how these and other types of interventions can improve the quality of life of individuals experiencing communication difficulties.

This Research Topic of papers clearly indicates the complexities involved in utilizing acoustic cues to extract information that is not only important to our degree of life-satisfaction but even to our survival. Understanding speech, for example, requires the integration of information registered on the cochlear with information coming in from other senses (primarily vision) and from our stored world knowledge. The studies in this Research Topic contribute to our understanding of this extremely complex process. They also illustrate that there is: (1) much more to learn with respect to how speech understanding is accomplished in the noisy environments typical of everyday life, and (2) how we can utilize the information coming from such studies to improve auditory environments, and to help those with diminished auditory and/or cognitive abilities to function in difficult listening situations.

Author contributions

LF and BS contributed to the conception of the Research Topic. Both authors contributed to the manuscript and approved the submitted version.

Conflict of interest

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.

Publisher's note

References

Fullgrabe, C. (2020). On the possible overestimation of cognitive decline: the impact of age-related hearing loss on cognitive-test performance. *Front. Neurosci.* 14, 454. doi: 10.3380/fnins.2020.00454

Schneider, B. A., Pichora-Fuller, M. K., and Daneman, M. (2010). "Effects of senescent changes in audition and cognition on spoken language comprehension," in *Springer Handbook of Auditory Research :The Aging Auditory System*, eds S. Gordon-Salant, R. D. Frisina, A. N. Popper, R. R. Fay (New York, Springer), 167–210.

frontiers in Psychology

ORIGINAL RESEARCH
published: 12 January 2022
doi: 10.3389/fpsyg.2021.804891

Differences Between Young and Older Adults in Working Memory and Performance on the Test of Basic Auditory Capabilities†

Larry E. Humes*, Gary R. Kidd and Jennifer J. Lentz

Department of Speech, Language, and Hearing Sciences, Indiana University, Bloomington, IN, United States

Edited by:
Leah Fostick,
Ariel University, Israel

Reviewed by:
Yael Zaltz,
Tel Aviv University, Israel
Gregg H. Recanzone,
University of California, Davis,
United States

****Correspondence:***
Larry E. Humes
humes@indiana.edu

† This article is dedicated to the memory of Charles S. Watson, the developer of the TBAC and a long-time colleague, mentor, and friend

Specialty section:
This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Psychology

Received: *29 October 2021*
Accepted: *13 December 2021*
Published: *12 January 2022*

Citation:
Humes LE, Kidd GR and Lentz JJ (2022) Differences Between Young and Older Adults in Working Memory and Performance on the Test of Basic Auditory Capabilities. Front. Psychol. 12:804891. doi: 10.3389/fpsyg.2021.804891

The Test of Basic Auditory Capabilities (TBAC) is a battery of auditory-discrimination tasks and speech-identification tasks that has been normed on several hundred young normal-hearing adults. Previous research with the TBAC suggested that cognitive function may impact the performance of older adults. Here, we examined differences in performance on several TBAC tasks between a group of 34 young adults with a mean age of 22.5 years (SD = 3.1 years) and a group of 115 older adults with a mean age of 69.2 years (SD = 6.2 years) recruited from the local community. Performance of the young adults was consistent with prior norms for this age group. Not surprisingly, the two groups differed significantly in hearing loss and working memory with the older adults having more hearing loss and poorer working memory than the young adults. The two age groups also differed significantly in performance on six of the nine measures extracted from the TBAC (eight test scores and one average test score) with the older adults consistently performing worse than the young adults. However, when these age-group comparisons were repeated with working memory and hearing loss as covariates, the groups differed in performance on only one of the nine auditory measures from the TBAC. For eight of the nine TBAC measures, working memory was a significant covariate and hearing loss never emerged as a significant factor. Thus, the age-group deficits observed initially on the TBAC most often appeared to be mediated by age-related differences in working memory rather than deficits in auditory processing. The results of these analyses of age-group differences were supported further by linear-regression analyses with each of the 9 TBAC scores serving as the dependent measure and age, hearing loss, and working memory as the predictors. Regression analyses were conducted for the full set of 149 adults and for just the 115 older adults. Working memory again emerged as the predominant factor impacting TBAC performance. It is concluded that working memory should be considered when comparing the performance of young and older adults on auditory tasks, including the TBAC.

Keywords: aging, auditory perception, cognition, hearing loss, auditory discrimination and identification

INTRODUCTION

The World Health Organization (2021) estimated that there are 162 million older adults worldwide with disabling age-related hearing loss. World Health Organization (2021) estimates the prevalence of such audiometrically defined disabling hearing loss to be 25% for those over 60, increasing from 15.4% globally among people aged in their 60s to 58.2% globally for those over 90 years old. Audiometric hearing loss for pure tones, however, captures just one aspect of auditory function in adults that can lead to limitations in activity and restrictions on participation in society, according to the widely applied World Health Organization (2001) model of healthy function. Other measures of auditory function beyond the audiogram may have implications for healthy living as well.

Humes et al. (2012), based on a review of 165 articles published in the peer-reviewed literature between 1988 and 2012, found evidence for declines in various measures of auditory abilities with advancing age. The bulk of the research over the review period in Humes et al. (2012) was on auditory temporal processing. Importantly, Humes et al. (2012) noted that it was difficult to ascertain whether the observed declines in auditory abilities with age reflected deficits in higher-level auditory processing or were driven by concomitant declines in hearing threshold, cognitive function, or both. A recent review by Gallun and Best (2020) provides support for the existence of age-related declines in auditory processing but also notes concerns about possible peripheral and cognitive confounds.

The Test of Basic Auditory Capabilities (TBAC) was developed by Watson et al. (1982a,b; see Watson, 1987) as an easy-to-administer battery of auditory processing that tapped several auditory abilities. The original TBAC included three single-tone discrimination tests, three tests of temporal pattern discrimination, and two tests using syllables, one assessing temporal-order discrimination and the other syllable identification in noise. Several subsequent studies have employed versions of the TBAC with large numbers of young normal-hearing (YNH) adults (Watson and Miller, 1993; Surprenant and Watson, 2001; Kidd et al., 2007). The TBAC has also been found to be reliable in YNH listeners (Kidd et al., 2007) and in older adults with hearing impairment (OHI) of varying degrees (Christopherson and Humes, 1992).

The TBAC has been used to compare the auditory-processing performance of YNH and OHI listeners in some prior studies as well. Humes and Christopherson (1991), for example, compared the performance of 23 older adults, 65–86 years of age, to that of YNH adults listening either in quiet (N = 10; 19–36 years) or in a background of noise designed to simulate the average hearing loss of the OHI listeners (N = 12; 20–31 years). Significant deficits were observed in the performance of the OHI group compared to both YNH groups on 4 of the 8 TBAC tests: frequency discrimination and three measures of temporal processing (an embedded test-tone duration-discrimination task and two temporal-order discrimination tasks: one using pure tones and the other using syllables). In addition, although hearing loss was the primary factor affecting speech-identification, some TBAC measures (notably, frequency discrimination) accounted for small but significant improvements in predictions of speech-identification performance.

Humes and Christopherson (1991) did not obtain measures of cognitive function in their study and subsequent work by Watson and Miller (1993) showed a link between cognitive function and TBAC performance in a large group of YNH listeners. As noted in Humes (1996), this led to replication of the Humes and Christopherson (1991) study, but this time using YNH and OHI groups matched for hearing loss and cognitive function. When doing so, no differences in TBAC performance were observed between young and older adults. This suggested that the prior "age group" difference may have been driven by concomitant age-group differences in hearing loss, cognitive function, or both.

In another study with OHI listeners, Humes et al. (1994) examined individual differences in TBAC performance among a group of 50 older adults ranging in age from 63 to 83 years and having varying degrees of hearing loss. In addition to the TBAC, cognitive function was assessed with the Wechsler Adult Intelligence Scale-Revised (WAIS-R; Wechsler, 1981) and the Wechsler Memory Scale-Revised (Wechsler, 1983). The primary focus of the investigation by Humes et al. (1994) was on the association of both the TBAC and the cognitive measures with unaided speech-recognition performance. Although not the focus, moderate correlations were evident between performance on the TBAC and performance on the cognitive measures.

More recently, Humes et al. (2013b) measured auditory performance in a group of 98 older adults making use of 27 different stimulus conditions to measure seven main auditory psychophysical abilities. In addition, the TBAC was employed, but the results were only presented for the mean performance on 6 of the 8 TBAC tests; the six discrimination tasks making use of tonal stimuli. The two TBAC tests making use of syllables were omitted because the focus of the study by Humes et al. (2013b) was on the identification of factors underlying individual differences in aided speech perception and the authors felt it was inappropriate to use speech-based tests to predict performance on other speech-based tests. Given the large number of other auditory measures in Humes et al. (2013b), only the mean TBAC performance for the six tonal tests, referred to as TBAC6, was considered. This 6-test mean TBAC score was found to be reliable in a group of 31 older adults with a test-retest correlation of r = 0.76 but mean retest scores were slightly (79.1 vs 75.4% correct) and significantly ($p < 0.001$) higher than the test scores. When comparing the mean performance of the 98 older adults to that of a normative group of 27 YNH listeners, the older group had significantly ($p < 0.01$) lower TBAC6 scores (82.9 vs. 76.1% correct). Because performance on the TBAC was not well represented in the principal-components solution for the large set of auditory psychophysical measures in that study, it was dropped by Humes et al. (2013b) from subsequent regression analyses.

In the present study, we looked more carefully at the performance of older adults on the TBAC using the data collected originally by Humes et al. (2013b). Rather than only averaging across the six TBAC tests making use of tonal stimuli, performance on each of eight TBAC tests was examined separately, as in earlier studies with the TBAC. Because measures of hearing loss and cognitive function were also available from most of the study participants, we also examined the relative contributions of these factors to TBAC performance. Given the prior observations of differences in performance between YNH and OHI listeners on many of the TBAC tests, we addressed whether such age-group differences remained after statistically controlling for differences in hearing loss and cognitive function between the two age groups.

MATERIALS AND METHODS

Participants

There were 115 adults in the OHI group and 34 adults in the YNH group for these analyses. Both groups are larger than those in Humes et al. (2013b), as only those with complete data across the full set of psychophysical, cognitive, and speech-recognition measures were included in those prior analyses. Here, participants only needed to complete the tests used to determine study eligibility, including an audiogram, three brief cognitive tests and the TBAC. Because all these measures were obtained in the early part of the lengthy data-collection process, data were available for these measures from larger samples of older and younger adults than in Humes et al. (2013b).

The group of older adults included 56 women (48.7%) and 59 men (51.3%), with a mean age of 69.2 years (SD = 6.2 years). The group of young adults included 26 women (76.5%) and eight men (23.5%), with a mean age of 22.5 years (SD = 3.1 years). None of the participants were current hearing-aid users and 90% of the older adults had never worn hearing aids. All subjects had no evidence of middle-ear pathology (air-bone gaps < 10 dB and normal tympanograms bilaterally), no signs of dementia (Mini Mental Status Exam, MMSE, > 25; Folstein et al., 1975), and had English as his or her native language. Older subjects were recruited primarily via newspaper ads in the local paper and younger subjects by flyers and university online postings.

The study protocol was approved by the Indiana University-Bloomington Institutional Review Board prior to data collection. All subjects signed informed consent forms for the study and the use of their de-identified data for research purposes. All subjects were paid for their participation.

For the older adults, the primary audiometric inclusion criterion was bilaterally symmetrical hearing with the threshold at 4,000 Hz ≤ 60 dB HL (ANSI, 2004) in at least one ear. This maximum hearing loss at 4,000 Hz was established to ensure that the spectrally shaped speech stimuli used in other portions of the study would be fully audible through 4,000 Hz. For the young adults, hearing thresholds were ≤25 dB HL from 250 through 8,000 Hz in both ears. The means and standard deviations for the air-conduction hearing thresholds of each ear are shown for each group in **Figure 1**. When controlling for the effects of hearing loss on TBAC performance in the analyses, the average threshold for 500, 1,000, 2,000, and 4,000 Hz, PTA4, will be used. As noted below, the TBAC is presented diotically. As a result, the better-ear PTA4 was used in the analyses below when controlling for hearing loss. For the older adults, the mean better-ear PTA4 = 25.3 dB HL (SD = 11.1 dB HL), and, for the young adults, the mean better-ear PTA4 = 6.5 dB HL (SD = 3.4 dB HL) which was a significant difference [t (147) = 9.7, $p < 0.001$] with a very large effect size (Cohen's d = 1.9; Cohen, 1988). Although this difference is significant, 59% of the older adults had normal hearing, as defined by better-ear PTA4 ≤ 25 dB HL, and 42% when defined as better-ear PTA4 ≤ 20 dB HL.

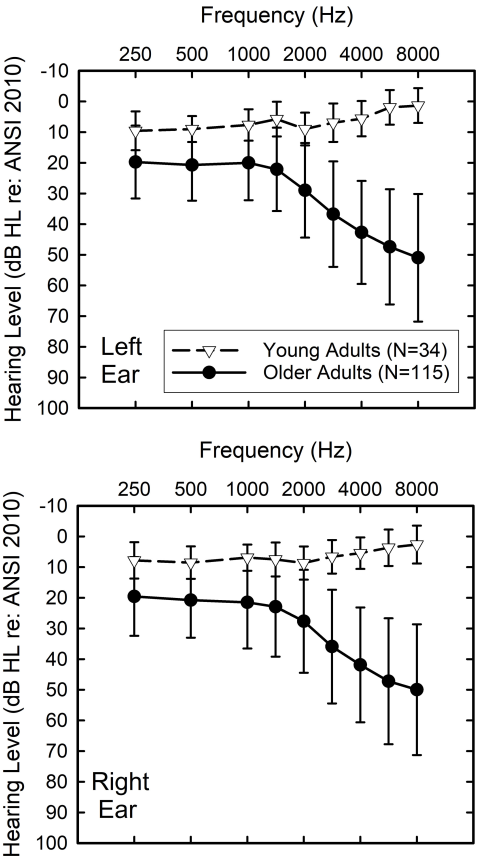


FIGURE 1 | Means and standard deviations for the air-conduction pure-tone thresholds from the audiogram for young (triangles) and older (circles) adults and the left (top) and right (bottom) ear.

Equipment and Materials

All testing was conducted with participants seated in a sound-attenuating room. All TBAC stimuli were played through a 16-bit high-quality sound card (Digital Audio Labs Card Deluxe) with a sampling rate of 44,100 Hz. The output was fed into Etymotic Research ER-3A insert earphones. TBAC stimuli were presented diotically at a level of 85 dB SPL as measured in a 2-cm^3 coupler. The relatively high presentation level was used to further minimize the impact of elevated hearing thresholds on TBAC performance for the older adults.

The Test of Basic Auditory Capabilities

The version of the TBAC used here was the TBAC-4, obtained from Communication Disorders Technology (CDT), Inc. The test battery, equivalent to that used in the earlier work with OHI listeners, includes six tests of auditory discrimination using tones, and two tests using speech sounds. The eight tests are briefly described below. For additional details see Kidd et al. (2007) and the TBAC information available on the CDT web site.[1]

Trials in each of the tests, except for the last test (syllable identification), were structured in a modified two-alternative forced-choice (2AFC) format in which a standard stimulus was followed by two test stimuli, one of which was different from the standard. The listeners used a computer keyboard to indicate which test stimulus was different from the standard. Trials were arranged in groups of six, and the level of difficulty was systematically increased from trial to trial, within each group, in logarithmic steps. For seven of the eight TBAC tests, eight levels of difficulty were tested over 72 trials, presenting the six easiest levels in the first 36 trials, followed by an increase in difficulty of two log steps for trials 37–72. For the penultimate test, the temporal-order task using syllables as stimuli, only five levels of difficulty were included with a total of 48 trials.

The TBAC administered here was comprised of eight tests. Each test is described briefly here.

> *Single-tone frequency discrimination (dF)* for which the standard was a 1,000-Hz 250-ms tone and frequency increments were used.
>
> *Single-tone intensity discrimination (dI)* for which the standard was a 1,000-Hz 250-ms tone and intensity increments were used.
>
> *Single-tone duration discrimination (dT)* for which the standard was a 1,000-Hz 100-ms tone and duration increments were used.
>
> *Pulse-train discrimination (dPT;* rhythm) with the standard consisting of six 20-ms pulses (1,000-Hz tone) arranged in three pairs, with a 40-ms pause within a pair and a 120-ms pause between pairs. The "different" sequence included an increase in the duration within a pair with a corresponding decrease in the duration between pairs, altering the rhythm of the sequence while keeping the total duration constant.
>
> *Embedded tone detection (dETT)* with the standard consisting of a sequence of eight tones of differing frequency with a temporal gap (ranging from 10 to 200 ms) in the middle of the sequence. The "different" sequence had a tone (also ranging from 10 to 200 ms in duration) filling the temporal gap in the middle position. A different sequence of frequencies (ranging from 300 to 3,000 Hz) was presented on each trial. The duration of the middle gap or tone was varied to manipulate task difficulty.
>
> *Temporal-order discrimination for tones (dTOpt)* for which the standard was a four-tone pattern consisting of two equal-duration tones (550 and 710 Hz) preceded and followed by a 100-ms 625-Hz tone. The middle tones were presented in reverse order in the "different" interval. The duration of the tones varied from 20 to 200 ms in equal-log steps. Shrivastav et al. (2008) found that the resulting variations in both rate of presentation and tone duration impact the performance of OHI listeners on this task.
>
> *Temporal-order discrimination for syllables (dTOsyl)* is similar to the preceding test, but with consonant-vowel (CV) syllables comprising the sequence instead of tones. The task is to discriminate /fa/-/ta/-/ka/-/pa/ from /fa/-ka/-/ta/-/pa/. The duration of the syllables was varied (by reducing the vowel duration) from 250 to 75 ms in five steps.
>
> *Syllable identification (SylID)* was a test of the recognition of nonsense CVC syllables in broadband noise. A 3AFC paradigm was used, with foils created by altering the vowel or one of the consonants. Five speech-to-noise ratios (SNRs) were used with decreasing SNRs within each set of five trials. A set of 100 stimuli was presented twice in separate blocks, with a different random order for each block.

Working Memory Tests

Three tests from a Matlab-based working memory test battery developed by Lewandowsky et al. (2010) were administered. For all tests, there were no time constraints on the recall task at the end of each trial and no feedback was provided. Each test took approximately 10 mins to complete. All testing took place with the participant comfortably seated in front of a computer monitor and keyboard inside a sound-attenuating booth. Procedural modifications to accommodate the older participants were implemented by Humes et al. (2013b) and are noted again here.

Memory Updating

At the start of each trial, subjects were presented with a sequence of three to five digits. Each digit was surrounded by a square to mark its position on the screen. After all digits were presented, the squares remained on the screen and a different sequence of arithmetic operations (addition or subtraction, with numbers ranging from +7 to -7) appeared in each of the squares, one at a time. The subject's task was to remember the digits that appeared in each square and then perform the sequence of arithmetic operations presented in each of the squares. The subject was asked to indicate (using the keyboard) the final resulting value in each square after a sequence of two to six sequential arithmetic operations. Consider the following example for a set size of 3. Three digits, 2 4 1, appear on the screen, one in each square. The digits are then replaced by +1 -2 +5 and these mathematical operations are applied to the digits retained in memory such that the new 3-digit sequence in memory is 3 2 6. Next, another set of three mathematical operations appear in the three squares on the screen: +2 +3-1. The new sequence in memory is now 5 5 5. For two sequential operations, the task ends, and the subject would enter 5 5 5 as the response. Otherwise, this process continues for up to a total of six sequential operations before the total from memory is requested as the response. The test consisted of

---

[1] http://comdistec.com/new/TBAC.html

15 trials with a randomly generated sequence of set size (3–5 co-occurring series of operations) and number of operations (2–6) on each trial.

Because this test was challenging for older adults, some adjustments were made to the procedures to ensure that the task was well understood, and to make it a bit less challenging. The number of practice trials was increased from two (the default) to four, and the time between items (to be added or subtracted) was increased from 250 to 500 ms. The first two practice trials used a 3-s inter-item time to allow the experimenter to explain the required operations during the trial. Also, the default instructions were supplemented with a verbal explanation of the task that included a subject-paced simulated trial using cue cards to present the stimuli.

Sentence Span

The "easy" version of the sentence-span task was used for this study. In this task, subjects were presented with an alternating sequence of simple sentences (3–6 words in length) and single letters on the computer screen. Subjects judged whether the sentence was true or false on each presentation, with 4 s allowed for responding. The letters required no response. After from four to eight sentence/letter presentations, subjects were asked to recall the letters in the order they were presented. The test consisted of 15 trials (after three practice trials) with three instances of each number of sentence/letter presentations.

Spatial Short-Term Memory

This test assessed a subject's ability to recall the location of dots (filled circles) in a 10 × 10 grid. On each trial, an empty grid was presented and then a sequence of dots appeared in the grid. Each dot remained on the screen for approximately 1 s before it was removed, and the next dot appeared. From two to six dots were presented on each trial. After all the dots had been presented (and removed), the subject was asked to indicate the relative position of the dots by touching (or pointing and clicking with a computer mouse) the cells within the grid. This test consisted of 30 trials (6 at each set size).

RESULTS

Reliability

Of the 115 older adults who completed the TBAC, 29 (25%) repeated the TBAC after completion of all other measures in the larger psychophysical study (Humes et al., 2013b) to provide an assessment of TBAC reliability. As noted, the test-retest data were only reported for the 6-test average of the TBAC, TBAC6, in Humes et al. (2013b). **Table 1** summarizes the results from the test-retest analyses for all 8 TBAC tests and the TBAC6 average. Performance on only two TBAC measures, the dF test and the TBAC6 average score, showed significant changes from test to retest with both showing a 4–5% point improvement on retest. Six of the nine test-retest correlations in **Table 1** are significant ($p < 0.05$, adjusted for multiple comparisons), SylID being the only test with poor test-retest correlation ($r = -0.03$). Of the remaining eight test-retest correlations in **Table 1**, all are moderate in strength and six of the eight are significant. Not surprisingly, the strongest test-retest correlation was observed for the score based on the most trials, the TBAC6 average score.

We also explored whether the reliability would be further enhanced by averaging all seven of the auditory discrimination measures, but the test-retest correlation for this 7-test average score decreased slightly to $r = 0.72$ ($p < 0.001$) compared to the 6-test average ($r = 0.76$, $p < 0.001$). Finally, we generated a 4-test average for the four pure-tone discrimination tasks which also had the four highest test-retest correlation in **Table 1**: dF, dI, dT, and dTOpt. The test-retest correlation for this TBAC4 average score was $r = 0.71$ ($p < 0.001$). In summary, individual test scores from the TBAC show moderate reliability among older adults and the reliability is enhanced when various average scores are used, with the TBAC6 average proving to be the most reliable, although the differences in r values among the various TBAC averages are not significant ($p > 0.1$).

The reliability of the working-memory tests had been established in older adults previously by Humes et al. (2013b). For the three working-memory tests, Humes et al. (2013b) reported

TABLE 1 | Test and retest means (M) and standard deviations (SD) for 29 of the 115 older adults.

TBAC Test	Test M%	Test SD%	Retest M%	Retest SD%	p_t*	r	p_r*
dF	**76.3**	**9.6**	**81.5**	**7.2**	**0.002**	**0.58**	**<0.001**
dI	84.0	10.4	88.0	8.6	0.026	**0.55**	**0.002**
dT	74.0	10.2	78.2	11.6	0.006	**0.68**	**<0.001**
dPT	80.7	12.6	85.6	8.4	0.012	0.45	0.015
dETT	71.4	10.8	74.6	7.8	0.063	**0.52**	**0.004**
dTOpt	67.3	9.7	66.6	9.5	0.503	**0.55**	**0.002**
dTOsyl	55.1	10.1	56.3	11.5	0.553	0.45	0.015
SylID	53.4	10.5	56.8	6.7	0.164	–0.03	0.860
TBAC6	**75.6**	**7.5**	**79.1**	**5.6**	**<0.001**	**0.76**	**<0.001**

Test-retest correlations (r), and their significance (p_r), are also shown. Significance of differences in means between test and retest and of the correlations is also shown (p_t). Entries in bold font indicate either significant differences between means (p_t) or correlations (p_r).

**p values adjusted for multiple comparisons with criterion $p < 0.05/9$ or $p < 0.0055$.*

that there were no significant changes in mean performance from test to retest and the test-retest correlations were $r = 0.69$, 0.83, and 0.83 for the spatial STM, sentence span, and memory updating tests, respectively.

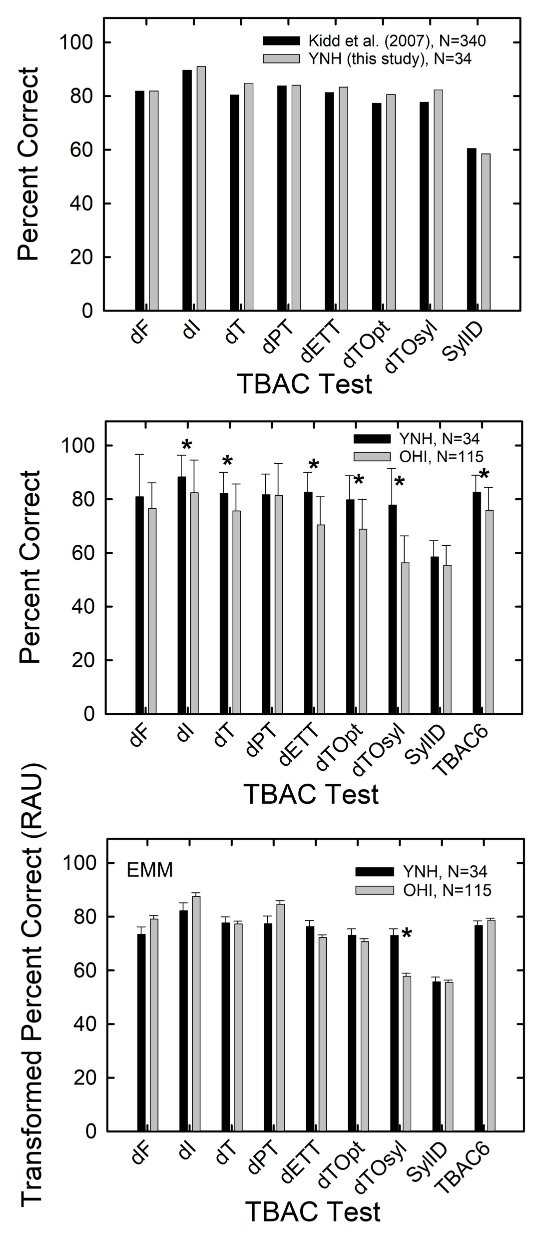


FIGURE 2 | The top panel compares the mean percent-correct performance for young normal-hearing (YNH) adults in this study ($N = 34$; gray bars) to the corresponding mean normative values from Kidd et al. (2007; $N = 340$; black bars). The middle panel shows the means and standard deviations for the percent-correct scores on the TBAC for the 34 YNH (black bars) and 115 older hearing-impaired (OHI; gray bars) adults in this study. The bottom panel provides the estimated marginal means (EMM, controlling for PTA4 and working memory) and standard errors for the YNH (black bars) and OHI (gray bars) groups in this study. The asterisks in the lower two panels mark significant effects (adjusted $p < 0.0055$) of subject group.

Age-Group Differences in Working Memory

As noted, the YNH and OHI groups not only differed significantly in age but also in average hearing loss (better-ear PTA4, **Figure 1**). Age-group differences were also expected for the measures of working memory (e.g., Salthouse, 2010). This was confirmed here for the data from the 34 YNH and 115 OHI participants. The means (and standard deviations) for the percent-correct scores from the young adults were 74.7 (11.0), 72.2 (11.7), and 84.8 (5.0) for memory updating, sentence span, and spatial STM tasks, respectively. For the older adults, the means (and standard deviations) were 49.8 (20.5), 52.6 (16.5), 72.7 (6.8). Independent-sample *t*-tests were significant for all three working-memory tests [all $t\ (147) > 6.4$, $p < 0.001$]. This was also true for a single principal-component score (accounting for 73.6% of the variance) derived from principal-component analysis of the three working-memory scores. The means (and standard deviations) for the working-memory principal component (PCwm) were 1.13 (0.45) and -0.36 (0.87) for the young and older adults, respectively. The independent-samples *t*-test resulted in $t\ (147) = 9.3$, $p < 0.001$.

Age-Group Differences on the Test of Basic Auditory Capabilities

Figure 2 shows the TBAC scores for the YNH and OHI groups compared in various ways. In the top panel, the mean scores from the 34 YNH adults in these analyses (gray bars) are compared to the largest set of normative data obtained by Kidd et al. (2007) from 340 YNH adults (black bars). No statistical analyses were performed on the data in the top panel. Rather, the similarity of the means for both groups of YNH listeners is just offered as evidence that the performance of our group of 34 YNH adults on the TBAC appears to be representative or typical for YNH adults generally for this battery of tests.

The middle panel of **Figure 2** shows the means and standard deviations for the YNH and OHI groups in these analyses. Data are shown for each of the eight TBAC tests as well as the mean percent-correct score for the six tonal auditory-discrimination tasks (TBAC6). The TBAC percent-correct scores were transformed to rationalized arcsine units (RAU; Studebaker, 1985) to stabilize the error variance prior to analysis of variance (ANOVA) for the group effects. With a Bonferroni-adjusted *p* value of 0.0055 (0.05/9), 6 of the 9 group differences were significant with the young adults outperforming the older adults in each case [all $F\ (1{,}147) > 7.96$, $p < 0.005$]. For the six significant group effects, all eta-squared effect sizes were >0.05 indicating that all effect sizes were at least medium effects (Cohen, 1988). The three non-significant group differences were for the dF, dPT, and SylID TBAC tests.

As noted, the YNH and OHI groups not only differed significantly in age but also in average hearing loss (better-ear PTA4, **Figure 1**) and working memory. When analysis of covariance (ANCOVA) was performed on each of the nine TBAC measures, with covariates of better-ear PTA4 and overall working-memory performance (PCwm), significant group differences ($p < 0.05$ with Bonferroni adjustment to

$p < 0.0055$) were observed for only one of the nine TBAC measures. This is shown in the bottom panel of **Figure 2** which depicts the estimated marginal means (EMMs) for the rau-transformed TBAC scores after adjustment for the PTA4 and PCwm covariates. The lone significant difference between the YNH and OHI groups that remained after controlling for PTA4 and working memory was the temporal-order task with syllables [F (1,145) = 26.6, $p < 0.001$; eta squared = 0.15, large effect size). These analyses also found that the better-ear PTA4 covariate never had a significant effect on TBAC scores [all F (1,145) < 5.8, $p > 0.02$]. In contrast, the PCwm covariate was found to be significant in 8 of the 9 ANCOVAs [all F (1,145) > 12.2, $p < 0.001$] with medium to large effect sizes based on eta squared. The only TBAC measure that did not show a significant effect of working memory on performance was the syllable identification task [SylID; F (1,145) = 5.8, $p > 0.01$].

Individual Differences in the Test of Basic Auditory Capabilities Scores

To further evaluate the effects of age, hearing loss, and working memory on TBAC performance, linear-regression analyses were also performed with each TBAC measure serving as the dependent variable. **Table 2** summarizes the results for each of the nine regression analyses for the entire sample of 149 adults. The F values in the third column show that significant regression solutions emerged for all nine TBAC measures with the variance explained in each case shown by the r^2 values in the preceding column. The rows in bold font in **Table 2** mark those predictors found to be significant ($p < 0.05$, unadjusted). The final three columns in **Table 2** provide the zero-order, partial, and part correlations for each predictor in each regression analysis. The partial correlation examines the association between an independent variable and a dependent variable after controlling for the influence of other variables on both the independent and dependent variable. The part or semi-partial correlation examines the association between the independent and dependent variable after controlling for the effects of the other variables on just the independent variable. Review of those last two columns of correlations reveals a very clear pattern. For 8 of the 9 TBAC measures, working-memory performance was either the only (seven times) or predominant (one time) significant predictor. The only exception was the temporal-order task for syllables, dTOsyl, for which working-memory was again significant but age was the predominant predictor. Recall that it was only this task that showed a significant difference between age groups in the previously presented ANCOVA. Thus, the linear-regression analyses of the individual data, summarized in **Table 2**, support the group analyses summarized previously in the bottom panel of **Figure 2**.

A second set of linear-regression analyses was completed for all 9 TBAC measures as the dependent variable; this time for only the older adults. The age range, 60–88 years, was sufficient to expect some age-related changes in performance. The use of a narrower age range in such analyses can also provide stronger evidence of age-related effects on performance (e.g., Hofer and Sliwinski, 2001). **Table 3** summarizes the results from the second set of regression analyses for the 115 older adults. The F values reveal that significant regression solutions were observed for all but the syllable-identification task (sylID). For the other 8 TBAC measures in **Table 3**, the partial and part correlations in the far-right columns indicate that working memory was always the predominant predictor, and in 6 of the 8 cases was the sole significant factor. For the two TBAC tests for which a second significant predictor emerged (dF, dETT), in both cases, the other predictor was the better-ear PTA4. In summary, among the older adults, ranging in age from 60 to 88 years, working memory was the sole or primary predictor of performance on the TBAC.

DISCUSSION

As was demonstrated in the top panel of **Figure 2**, the 34 YNH listeners in this study performed as expected, based on the normative data for the TBAC from Kidd et al. (2007). In addition, as had been found in Christopherson and Humes (1992) and Kidd et al. (2007), the TBAC scores were fairly reliable in older adults, although the reliability was enhanced considerably by averaging the scores for the 6 tonal auditory-discrimination tasks (TBAC6), as had been done by Humes et al. (2013b).

Differences in performance on the TBAC between the YNH and OHI groups, reported in the middle panel of **Figure 2**, were consistent with age-group differences reported previously by Humes and Christopherson (1991). Older adults performed significantly worse than young adults on several TBAC tests. Subsequent ANCOVA analyses with hearing loss (better-ear PTA4) and cognition (working memory, as indexed by PCwm) as covariates (**Figure 2**, bottom), however, suggested that the difference in TBAC scores between age groups was primarily due to group differences in working memory, rather than some unspecified age-related factor. Age-group effects disappeared when the covariates were used as statistical controls in ANCOVAs with working-memory performance being the lone significant factor in 7 of the 9 analyses, and one of two significant factors in one of the remaining two analyses. That is, working memory was a significant covariate for 8 of the 9 TBAC measures with age group being significant for only the temporal-order task using syllables. This is in line with the analyses for the TBAC described in Humes (1996) in which performance on the TBAC for two groups of adults differing in age but matched for hearing loss and cognition did not differ significantly.

The regression analyses summarized for all participants (**Table 2**) and for only the older adults (**Table 3**) provided further support for the predominant importance of working memory to TBAC performance across the adult lifespan. Significant regression solutions emerged for most TBAC measures in both sets of linear-regression analyses and the partial and part correlations supported the predominance of working memory in those analyses.

These findings should not be misunderstood as indicating older adults are expected to perform equivalently to young adults on the TBAC. Older adults performed worse than young adults on many of the TBAC tests (**Figure 2**, middle). Rather,

TABLE 2 | Results of the linear-regression analyses for each the test of basic auditory capabilities (TBAC) score (in RAU) for 149 young and older adults.

TBAC Test	r^2	*F* (3,145)	Ind Var	Std Beta	*t*	*p*	*r*	Partial *r*	Part *r*
dF	0.162	9.32*	**PC WM**	**0.442**	**4.347**	**<0.001**	**0.383**	**0.340**	**0.331**
			zAge	0.192	1.550	0.123	−0.194	0.128	0.118
			zPTA4	−0.133	−1.218	0.225	−0.224	−0.101	−0.093
dI	0.218	13.43*	**PC WM**	**0.455**	**4.635**	**<0.001**	**0.455**	**0.359**	**0.340**
			zAge	0.117	0.971	0.333	−0.290	0.080	0.071
			zPTA4	−0.148	−1.405	0.162	−0.300	−0.116	−0.103
dT	0.235	14.87*	**PC WM**	**0.511**	**5.262**	**<0.001**	**0.483**	**0.400**	**0.382**
			zAge	0.010	0.084	0.933	−0.299	0.007	0.006
			zPTA4	0.040	0.389	0.698	−0.216	0.032	0.028
dPT	0.087	4.60**	**PC WM**	**0.391**	**3.687**	**<0.001**	**0.227**	**0.293**	**0.293**
			zAge	**0.261**	**2.011**	**0.046**	**−0.009**	**0.165**	**0.160**
			zPTA4	−0.015	−0.132	0.895	−0.031	−0.011	−0.010
dETT	0.355	26.62*	**PC WM**	**0.366**	**4.105**	**<0.001**	**0.552**	**0.323**	**0.274**
			zAge	−0.155	−1.426	0.156	−0.512	−0.118	−0.095
			zPTA4	−0.161	−1.682	0.095	−0.460	−0.138	−0.112
dTOpt	0.312	21.90*	**PC WM**	**0.425**	**4.611**	**<0.001**	**0.542**	**0.358**	**0.318**
			zAge	−0.122	−1.088	0.278	−0.453	−0.090	−0.075
			zPTA4	−0.070	−0.708	0.480	−0.376	−0.059	−0.049
dTOsyl	0.488	46.03*	**PC WM**	**0.241**	**3.035**	**0.003**	**0.581**	**0.244**	**0.180**
			zAge	**−0.519**	**−5.347**	**<0.001**	**−0.675**	**−0.406**	**−0.318**
			zPTA4	0.005	0.061	0.951	−0.489	0.005	0.004
sylID	0.070	3.61***	**PC WM**	**0.219**	**2.042**	**0.043**	**0.259**	**0.167**	**0.164**
			zAge	−0.029	−0.219	0.827	−0.202	−0.018	−0.018
			zPTA4	−0.041	−0.357	0.722	−0.174	−0.030	−0.029
TBAC6	0.345	25.51*	**PC WM**	**0.575**	**6.398**	**<0.001**	**0.582**	**0.469**	**0.430**
			zAge	0.076	0.696	0.488	−0.384	0.058	0.047
			zPTA4	−0.112	−1.170	0.244	−0.355	−0.097	−0.079

*Bold font highlights those independent variables having significant ($p < 0.05$) standardized Beta coefficients in significant regression solution. Asterisks mark significant F values for the regression solution: *$p < 0.001$; **$p < 0.01$; ***$p < 0.05$.*

this finding helps identify the factors underlying that observed age-group difference. It is the age-group difference in cognitive function, specifically working memory as measured here, that appears to underlie the poorer performance of older adults relative to young adults on the TBAC.

Although the focus here is on the TBAC, the link between auditory performance and working memory in older adults is not unique to the TBAC. Recently, Lentz, Humes and Kidd (in press), demonstrated similar links in this same study sample for over 20 psychoacoustic measurements spanning a much wider range of tasks than the TBAC. In those analyses, as was observed here, age alone seldom emerged as a significant predictor of psychoacoustic performance with working memory being the predominant predictor of performance. Unlike here, however, hearing loss was found to be a significant predictor on several psychoacoustic tasks as well, especially those tasks making use of stimuli that extended further into the high-frequency region of hearing loss than most stimuli in the TBAC.

Another important finding is that hearing loss, PTA4, was not related to TBAC performance for any of the tests, except for minor contributions to two TBAC measures within the group of 115 older adults (**Table 3**). The relative unimportance of hearing loss to TBAC performance had been noted previously (Humes and Christopherson, 1991) as a potential advantage in using the TBAC to assess auditory function in older adults, many of whom have significant hearing loss in the higher frequencies. Clearly, based on the audiograms in **Figure 1**, many of the older adults in this study had measurable hearing loss, especially in the higher frequencies. For the six tonal auditory-discrimination tasks in the TBAC, the stimuli are all in the mid-frequencies. Except for the embedded test-tone task, which makes use of frequencies varying between 300 and 3,000 Hz, the other five tone-based discrimination tasks in the TBAC use stimuli that are generally confined to 500–1,500 Hz which corresponds to the region of best hearing in older adults. Interestingly, among the older adults, the embedded test-tone task was one of two TBAC measures for which the better-ear PTA4 was found to be a significant secondary predictor (**Table 3**). To further minimize potential confounds of stimulus audibility in this study, a relatively high presentation level of 85 dB SPL was used for the TBAC. The absence of a significant effect of PTA4 on most of the TBAC tests further documents its utility as a measure of auditory function in older adults, including those with typical age-related hearing loss.

How do the age-group differences in percent-correct TBAC scores, such as those in the middle panel of **Figure 2**, translate to acoustical differences between the standard and comparison stimuli used in the various TBAC tests? To evaluate this, the median percent-correct scores for the YNH and OHI listeners

TABLE 3 | Results of the linear-regression analyses for each TBAC score (in RAU) for 115 older adults only.

TBAC Test	r^2	*F* (3,111)	Ind Var	Std Beta	*t*	*p*	*r*	Partial *r*	Part *r*
dF	0.257	12.77*	**PC WM**	**0.459**	**5.226**	**<0.001**	**0.472**	**0.444**	**0.428**
			zAge	0.099	1.002	0.318	−0.173	0.095	0.082
			zPTA4	**−0.214**	**−2.253**	**0.026**	**−0.268**	**−0.209**	**−0.184**
dI	0.203	9.45*	**PC WM**	**0.354**	**3.897**	**<0.001**	**0.417**	**0.347**	**0.330**
			zAge	−0.119	−1.161	0.248	−0.291	−0.110	−0.098
			zPTA4	−0.091	−0.921	0.359	−0.232	−0.087	−0.078
dT	0.200	9.24*	**PC WM**	**0.459**	**5.048**	**<0.001**	**0.442**	**0.432**	**0.429**
			zAge	0.007	0.072	0.943	−0.122	0.007	0.006
			zPTA4	0.067	0.681	0.497	−0.033	0.064	0.058
dPT	0.097	3.97***	**PC WM**	**0.313**	**3.234**	**0.002**	**0.310**	**0.293**	**0.292**
			zAge	−0.012	−0.108	0.914	−0.108	−0.010	−0.010
			zPTA4	0.030	0.282	0.779	−0.047	0.027	0.025
dETT	0.184	8.36*	**PC WM**	**0.333**	**3.627**	**<0.001**	**0.379**	**0.325**	**0.311**
			zAge	0.003	0.027	0.979	−0.221	0.003	0.002
			zPTA4	**−0.207**	**−20.80**	**0.040**	**−0.282**	**−0.194**	**−0.178**
dTOpt	0.172	7.69*	**PC WM**	**0.360**	**3.887**	**<0.001**	**0.399**	**0.346**	**0.336**
			zAge	−0.063	−0.605	0.546	−0.229	−0.057	−0.052
			zPTA4	−0.074	−0.736	0.463	−0.188	−0.070	−0.064
dTOsyl	0.119	4.99**	**PC WM**	**0.222**	**2.324**	**0.022**	**0.290**	**0.215**	**0.207**
			zAge	−0.127	−1.178	0.241	−0.258	−0.111	−0.105
			zPTA4	−0.101	−0.971	0.334	−0.216	−0.092	−0.087
sylID	0.048	1.85	PC WM	0.219	2.042	0.043	0.259	0.167	0.164
			zAge	−0.029	−0.219	0.827	−0.202	−0.018	−0.018
			zPTA4	−0.041	−0.357	0.722	−0.174	−0.030	−0.029
TBAC6	0.282	14.52*	**PC WM**	**0.487**	**5.651**	**<0.001**	**0.519**	**0.473**	**0.455**
			zAge	−0.027	−0.276	0.783	−0.251	−0.026	−0.022
			zPTA4	−0.099	−1.056	0.293	−0.223	−0.100	−0.085

*Bold font highlights those independent variables having significant ($p < 0.05$) standardized Beta coefficients in significant regression solution. Asterisks mark significant F values for the regression solution: *$p < 0.001$; **$p < 0.01$; ***$p < 0.05$.*

were generated. Next, the group ($N = 340$) psychometric functions from Kidd et al. (2007), relating the proportion correct to the physical stimulus parameter manipulated on each TBAC test, were used as transfer functions to convert each median percent-correct score to a physical stimulus difference. The normative group psychometric functions for each of the nine TBAC tests are shown in the upper two panels of **Figure 3**. The median proportion-correct scores for the YNH and OHI groups were then converted to stimulus values in Hz, ms, or dB, depending on the test. The transformed medians appear as the black and gray vertical bars in the lower panels of **Figure 3**. For all TBAC tasks, lower values represent better performance. Except for the pulse-train (rhythm) discrimination task (dPT), the OHI listeners clearly required a larger difference between the standard and comparison stimuli at the median threshold for that group. The superior performance of the YNH group is probably most apparent for the two temporal-order tasks, dTOpt and dTOsyl, in the lower left panel. Here, on average, the OHI listeners required the durations of stimuli comprising a stimulus sequence to be 2–3 times longer (and the resulting rate of presentation to be slower) than that of the YNH group to discriminate between the standard and comparison sequences.

Fogerty et al. (2010) and Humes et al. (2010), in analyses of data from subject samples not overlapping with the present samples, reported age-group differences for temporal-order identification of brief vowel sequences. The temporal-order task was a monaural closed-set sequence-identification task rather than a diotic temporal-order discrimination task as in the dTOsyl test of the TBAC. Fogerty et al. (2010) included data from 35 young and 151 older adults for both a two-item and four-item vowel sequence. As in the present study, stimulus manipulations were applied to minimize the impact of age-group differences in hearing thresholds on temporal-order identification performance. Older adults were found to have significantly poorer temporal-order thresholds for both the two-vowel and four-vowel sequences. For the two-item temporal-order task, the median threshold of the older adults was more than three times greater than that of the younger adults. For the four-item task, the temporal-order threshold was 1.7 times longer than that for the young adults. Thus, the magnitude of this age-group difference in the thresholds for the temporal-order identification of syllable sequences in Fogerty et al. (2010) is comparable to the magnitudes of the differences shown for dTOsyl in **Figure 3**. Also consistent with the present findings for the TBAC in **Table 3**, Fogerty et al. (2010) found cognitive

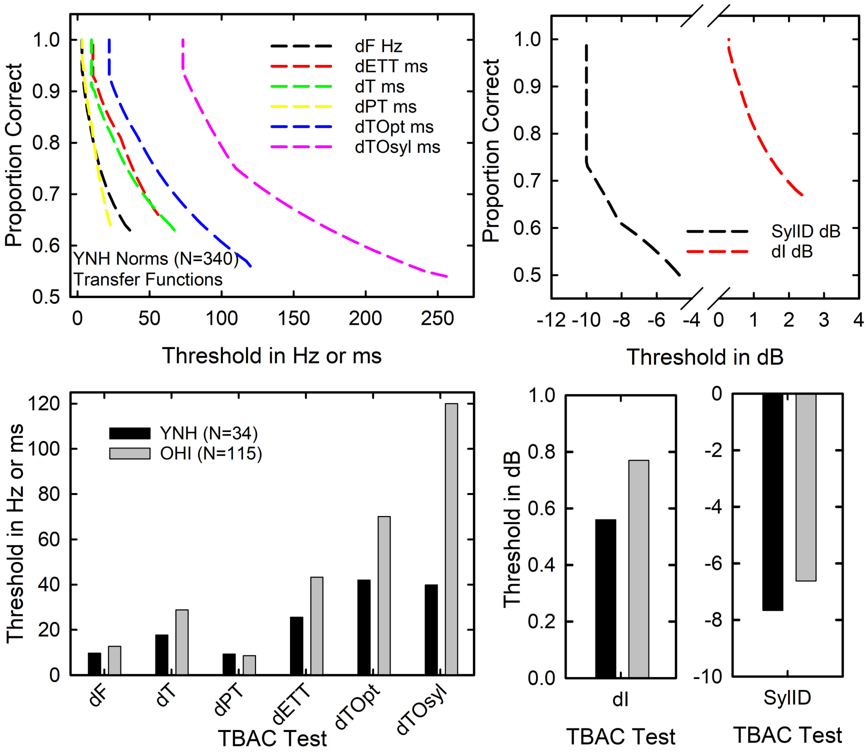


FIGURE 3 | The top two panels show the transfer functions from the group data (*N* = 340) of Kidd et al. (2007) relating the proportion correct on a given TBAC test to the underlying stimulus dimension manipulated on that test. The lower two panels show the physical stimulus change needed at the median percent-correct performance on each TBAC test for the YNH (black bars) and OHI (gray bars) adults.

function to be the only significant predictor associated with individual differences in temporal-order identification within the group of older adults. Age-group differences in temporal-order identification performance were not examined with covariates by Fogerty et al. (2010) to determine the extent to which age-group differences in cognitive function may have mediated the observed age-group differences in temporal-order thresholds.

More recently, Humes et al. (2013a), again making use of a study sample independent of that in the present study, found temporal-order processing in hearing, vision, and touch to be strongly associated with cognitive function in a cross-sectional study of 245 young, middle-age, and older adults. Recently, in longitudinal follow-up analyses of 98 of the original 195 middle-age and older adults included in the cross-sectional study of Humes et al. (2013a), independent from this study sample, auditory temporal-order identification for brief syllables emerged as the most significant auditory measure in regression analyses predicting cognitive function, both for brief clinical cognitive assessments (Humes, 2020) and for comprehensive cognitive assessments (Humes, 2021). Both monaural and dichotic temporal-order identification measures were found to decline longitudinally, dichotic longitudinal declines also having been observed by Babkoff and Fostick (2017). Humes (2020, 2021) found both temporal-order identification measures to be predictive of declines in cognitive function in older adults over a 9-year period. Temporal-order processing typically explained 10–20% of the variance in cognitive function among middle-age and older adults using either form of cognitive assessment.

These prior and current findings reinforce the need to evaluate cognition when comparing the auditory performance of young and older adults as their declines in auditory abilities may be driven by differences in cognitive function. The focus in this study was on working memory and three different visual working-memory tasks were completed by all subjects. Although the tasks are considered working-memory measures, they may be considered relatively complex working-memory tasks compared to simpler measures such as forward or backward digit span. As a result of the complexity, other aspects of higher-level processing may be tapped beyond working memory alone. For example, one task required completion of a sequence of arithmetic operations, another required the reading and evaluation of sentences between items in the recall set (letters), and the third required spatial processing. The principal-components analysis was designed to capture the common working-memory component shared by all three tasks, thus providing a measure that excludes the task-specific variance. However, the task-specific abilities are also of interest because they are cognitive abilities that may be related to performance on the TBAC tasks, independent of the contribution of working memory. The relative strength of the correlations between specific working-memory tasks and TBAC performance among the older adults may vary across tasks, revealing selective influences of math, linguistic, and spatial abilities on the TBAC tasks. To evaluate this possibility, for the 115 older adults, correlations between the z-transformed TBAC and the three z-transformed working-memory scores were calculated and compared to that between the TBAC score and the overall working-memory principal component (the latter was not z-transformed because it already has a mean of 0 and standard deviation of 1 as a principal component score). Among the three working-memory tasks, performance on the sentence-span task had slightly but consistently higher correlations than the other two working-memory tasks across all TBAC measures. Moreover, the correlation between TBAC performance and performance on the sentence-span task was slightly but consistently higher than that for the overall working-memory principal-component. For example, averaged across all 8 TBAC scores and the TBAC6 average measure, 9 correlations in total, the mean correlation with the z-transformed sentence-span score was $r = 0.42$ versus $r = 0.39$ for the working-memory principal component. The difference between correlations was largest for the two speech-based TBAC tests, dTOsyl and sylID, and for two tasks with longer, more complex, sound sequences, dPT and dETT. For these four TBAC tests, the correlations with sentence-span score were 0.06–0.11 higher than those with the working-memory principal component (PCwm). The higher correlations for TBAC scores with sentence-span scores versus the overall working-memory principal component are relatively slight improvements. Nonetheless, these differences suggest that a working-memory measure involving the linguistic processing of visual stimuli correlates a bit more strongly with performance on the TBAC than a spatially based, arithmetic-based, or overall (PCwm) measure of working memory.

The mechanisms that underlie the observed correlations between performance on the TBAC and measures of working memory may also be related to task-specific aspects of the TBAC measures. That is, the implementation of the standard-two-alternative stimulus presentation format of the TBAC shares some characteristics of many working memory tasks. For all seven auditory-discrimination tasks, the standard stimulus is always presented first and then two additional stimuli are presented sequentially after that standard with only one differing from the standard. The task is to select the stimulus that differed from the standard. To do so, one must hold the standard in memory while performing comparisons with two subsequent stimuli. Further, an individual trial has sound durations for the standard and comparison stimuli that vary from task to task, with the longest stimuli occurring in the dPT, dETT, dTOpt, and dTOsyl tasks. Thus, although working memory is involved in the specific psychophysical procedure used in all TBAC tests, working memory may be taxed to a greater extent for those tasks with longer standard and comparison stimuli.

On the other hand, the concomitant decline in auditory abilities and cognition among older adults may offer insights into the factors underlying the well-established cognitive declines as adults advance in age (Humes et al., 2013a; Humes, 2020, 2021). That is, there may be shared underlying mechanisms that negatively impact both sensory and cognitive processing with increasing age either concomitantly or sequentially (Humes and Young, 2016).

Finally, as was noted previously, explaining the underlying mechanisms responsible for age-group differences in auditory abilities does not mean that older adults have auditory processing typical of that found in young adults. Older adults have difficulty processing many aspects of auditory stimuli. For sound sequences, older adults have considerable difficulty with rapid sequences (**Figure 3**). Knowing that this may be driven by underlying deficits in cognitive function does not change the fact that older adults have more difficulty processing fast sound sequences, it just explains why that difficulty is observed. Of course, running speech is a rapid sequence of sounds and the observed age-related deficit in temporal-order processing may underlie some of the speech-recognition difficulties experienced by older adults. However, temporal-processing measures were not significant predictors of aided speech understanding in Humes et al. (2013b) and were largely independent of speech measures in a study of auditory abilities in young listeners, using an expanded version of the TBAC (Kidd et al., 2007). The auditory task that accounted for most of the variance in speech understanding in both of those studies was the recognition (in noise) of familiar non-speech sounds. It may be that the cognitive changes that are associated with reduced temporal-processing abilities also have a negative impact on the recognition of spectrally and temporally complex familiar sounds, beyond their influence on the temporal-order TBAC measures.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by IU Bloomington IRB. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

LH oversaw data analyses and was primarily responsible for the preparation of the manuscript. All authors shared responsibility equally for the design of the study, contributed to the writing of the manuscript, and approved the submitted version.

FUNDING

This work was supported, in part, by NIH grant R01 AG008293.

ACKNOWLEDGMENTS

The authors thank Dan Fogerty, Charles Brandt, Dana Kinney, Megan Chaney, Hannah Fehlberg, Ellie Barlow, Nick Humes, and Thomas Gennaro for their assistance with various aspects of data collection for this study.

REFERENCES

ANSI (2004). *S3.6-2004, Specification for Audiometers*. New York, NY: American National Standards Institute.

Babkoff, H., and Fostick, L. (2017). Age-related changes in auditory processing and speech perception: cross-sectional and longitudinal analyses. *Euro. J. Ageing* 14, 269–281. doi: 10.1007/s10433-017-0410-y

Christopherson, L. A., and Humes, L. E. (1992). Some psychometric properties of the Test of Basic Auditory Capabilities (TBAC). *J. Speech Hear. Res.* 35, 929–935. doi: 10.1044/jshr.3504.929

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. US: Lawrence Erlbaum Associates.

Fogerty, D., Humes, L. E., and Kewley-Port, D. (2010). Auditory temporalorder processing of vowel sequences by young and older adults. *J. Acoust. Soc. Am.* 127, 2509–2520.

Folstein, M. F., Folstein, S. E., and McHugh, P. R. (1975). Mini-Mental State: a practical method for grading the cognitive state of patients for the clinician. *J. Psychiatr. Res.* 12, 189–198.

Gallun, F. J., and Best, V. (2020). "Age-related changes in segregation of sound sources," in *Springer Handbook of Auditory Research, Aging and Hearing-Causes and Consequences*, ed. K. Helfer (Switzerland: Springer).

Hofer, S. M., and Sliwinski, M. J. (2001). Understanding Ageing. An evaluation of research designs for assessing the interdependence of ageing-related changes. *Gerontology* 47, 341–352. doi: 10.1159/000052825

Humes, L. E. (1996). Speech understanding in the elderly. *J. Am. Acad. Audiol.* 7, 161–167.

Humes, L. E. (2020). Associations between measures of auditory function and brief assessments of cognition. *Am. J. Audiol.* 29, 825–837. doi: 10.1044/2020_AJA-20-00077

Humes, L. E. (2021). Longitudinal changes in auditory and cognitive function in middle-aged and older adults. *J. Speech Lang. Hear. Res.* 64, 230–249. doi: 10.1044/2020_JSLHR-20-00274

Humes, L. E., and Christopherson, L. (1991). Speech-identification difficulties of the hearing-impaired elderly: the contributions of auditory-processing deficits. *J. Speech Hear. Res.* 34, 686–693. doi: 10.1044/jshr.3403.686

Humes, L. E., and Young, L. A. (2016). Sensory-cognitive interactions in older adults. *Ear Hear.* 37, 52S–61S. doi: 10.1097/AUD.0000000000000303

Humes, L. E., Kidd, G. R., and Lentz, J. J. (2013b). Auditory and cognitive factors underlying individual differences in aided speech-understanding among older adults. *Front. Syst. Neurosci.* 7:55.

Humes, L. E., Busey, T. A., Craig, J., and Kewley-Port, D. (2013a). Are age-related changes in cognitive function driven by age-related changes in sensory processing? *Atten. Percept. Psychophys.* 75, 508–524. doi: 10.3758/s13414-012-0406-9

Humes, L. E., Dubno, J. R., Gordon-Salant, S., Lister, J. J., Cacace, A. T., Cruickshanks, K. J., et al. (2012). Central presbycusis: a review and evaluation of the evidence. *J. Am. Acad. Audiol.* 23, 635–666. doi: 10.3766/jaaa.23.8.5

Humes, L. E., Kewley-Port, D., Fogerty, D., and Kinney, D. (2010). Measures of hearing threshold and temporal processing across the adult lifespan. *Hear. Res.* 264, 30–40. doi: 10.1016/j.heares.2009.09.010

Humes, L. E., Watson, B. U., Christensen, L. A., Cokely, C. A., Halling, D. A., and Lee, L. (1994). Factors associated with individual differences in clinical measures of speech recognition among the elderly. *J. Speech Hear. Res.* 37, 465–474. doi: 10.1044/jshr.3702.465

Kidd, G. R., Watson, C. S., and Gygi, B. (2007). Individual differences in auditory abilities. *J. Acoust. Soc. Am.* 122, 418–435. doi: 10.1121/1.2743154

Lewandowsky, S., Oberauer, K., Yang, L.-X., and Ecker, U. K. H. (2010). A working memory test battery for MATLAB. *Behav. Res. Methods* 42, 571–585. doi: 10.3758/BRM.42.2.571

Salthouse, T. A. (2010). *Major Issues in Cognitive Aging*. New York, NY: Oxford University Press.

Shrivastav, M. N., Humes, L. E., and Aylsworth, L. (2008). Temporal-order discrimination of tonal sequences by younger and older adults: the role of duration and rate. *J. Acoust. Soc. Am.* 124, 462–471. doi: 10.1121/1.2932089

Studebaker, G. A. (1985). A "rationalized" arcsine transform. *J. Speech Hear. Res.* 28, 455–462. doi: 10.1044/jshr.2803.455

Surprenant, A. M., and Watson, C. S. (2001). Individual differences in the processing of speech and nonspeech sounds by normal-hearing listeners. *J. Acoust. Soc. Am.* 110, 2085–2095. doi: 10.1121/1.1404973

Watson, B. U., and Miller, T. K. (1993). Auditory perception, phonological processing and reading ability/disability. *J. Speech Hear. Res.* 36, 850–863. doi: 10.1044/jshr.3604.850

Watson, C. S. (1987). "Uncertainty, informational masking, and the capacity of immediate auditory memory," in *Auditory Processing of Complex Sounds*, eds W. A. Yost and C. S. Watson (Hillsdale, NJ: Lawrence Erlbaum), 267–277.

Watson, C. S., Jensen, J. K., Foyle, D. C., Leek, M. R., and Goldgar, D. E. (1982a). Performance of 146 normal adult listeners on a battery of auditory discrimination tasks. *J. Acoust. Soc. Am.* 71:S73. doi: 10.1121/1.5134059

Watson, C. S., Johnson, D. M., Lehman, J. R., Kelly, W. J., and Jensen, J. K. (1982b). An auditory discrimination test battery. *J. Acoust. Soc. Am.* 71:S73.

Wechsler, D. (1981). *The Wechsler Adult Intelligence Scale-Revised*. New York: The Psychological Corporation.

Wechsler, D. (1983). *The Wechsler Memory Scale-Revised*. New York: The Psychological Corporation.

World Health Organization (2001). *International Classification of Functioning, Disability and Health*. Geneva: World Health Organization.

World Health Organization (2021). *World Report on Hearing*. Geneva: World Health Organization.

Conflict of Interest: 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 Neuroscience

ORIGINAL RESEARCH
published: 12 January 2022
doi: 10.3389/fnins.2021.812533

Sub- and Supra-Second Timing in Auditory Perception: Evidence for Cross-Domain Relationships

Elzbieta Szelag*, Magdalena Stanczyk and Aneta Szymaszek

Laboratory of Neuropsychology, Nencki Institute of Experimental Biology of the Polish Academy of Sciences, Warsaw, Poland

Previous studies indicate that there are at least two levels of temporal processing: the sub- and supra-second domains. The relationship between these domains remains unclear. The aim of this study was to test whether performance on the sub-second level is related to that on the supra-second one, or whether these two domains operate independently. Participants were 118 healthy adults (mean age = 23 years). The sub-second level was studied with a temporal-order judgment task and indexed by the Temporal Order Threshold (TOT), on which lower values corresponded to better performance. On the basis of TOT results, the initial sample was classified into two groups characterized by either higher temporal efficiency (HTE) or lower temporal efficiency (LTE). Next, the efficiency of performance on the supra-second level was studied in these two groups using the subjective accentuation task, in which participants listened to monotonous sequences of beats and were asked to mentally accentuate every n-th beat to create individual rhythmic patterns. The extent of temporal integration was assessed on the basis of the number of beats being united and better performance corresponded to longer units. The novel results are differences between groups in this temporal integration. The HTE group integrated beats in significantly longer units than did the LTE group. Moreover, for tasks with higher mental load, the HTE group relied more on a constant time strategy, whereas the LTE group relied more on mental counting, probably because of less efficient temporal integration. These findings provide insight into associations between sub- and supra-second levels of processing and point to a common time keeping system, which is active independently of temporal domain.

Keywords: auditory perception, temporal information processing, timing, sub-second timing, supra-second timing

OPEN ACCESS

Edited by:
Leah Fostick,
Ariel University, Israel

Reviewed by:
Marc Wittmann,
Institute for Frontier Areas of Psychology and Mental Health (IGPP), Germany
Claude Alain,
Rotman Research Institute (RRI), Canada

****Correspondence:***
Elzbieta Szelag
e.szelag@nencki.edu.pl

Specialty section:
This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Neuroscience

Received: *10 November 2021*
Accepted: *13 December 2021*
Published: *12 January 2022*

Citation:
Szelag E, Stanczyk M and Szymaszek A (2022) Sub- and Supra-Second Timing in Auditory Perception: Evidence for Cross-Domain Relationships. Front. Neurosci. 15:812533. doi: 10.3389/fnins.2021.812533

INTRODUCTION

Temporal Constraints of Cognitive Functions

Time may be considered as the frontier in cognitive sciences and a fundamental property of working human brains. Much evidence from both everyday observations and extensive research studies has consistently indicated that many cognitive functions—such as language, perception, short-term and working memory, attention, motor activity, decision making, executive functions, etc.,—are temporally segmented in specific time intervals and are rooted in a defined temporal template (e.g., Szelag et al., 2004a, 2008, 2010, 2011, 2014; Szymaszek et al., 2009, 2018; Ulbrich et al., 2009; Oron et al., 2015; Nowak et al., 2016; Buhusi et al., 2018; Szelag, 2018; Choinski et al., 2020; Jablonska et al., 2020). Temporal information processing (TIP) is omnipresent, for example in every verbal or perceptual act, in movement control, in learning, and in planning. Efficient intrinsic timing mechanisms are necessary to effectively execute these activities. Patterning in time, therefore, is considered as providing a structure for complex cognition. Because of such ubiquity of time in mind, the basic question arises: how is temporal information processed in our brains?

As the first step toward answering this question, we can refer to the taxonomy system of our time experiences. Based on the data from a variety of studies, we can postulate that temporal perception is not veridical and subjective time is not a linear function of clock time. Deviations from the timed template have been commonly observed in psychological studies. Moreover, a large body of experimental studies has consistently indicated that humans can process temporal information over several time scales, or operational processing windows, which may be categorized into major groups (Pöppel, 1994; Mauk and Buonomano, 2004).

According to the framework model of TIP proposed by Pöppel (1997, 2004, 2009); see also Wittmann, 2009, 2013; Montemayor and Wittmann, 2014), one can distinguish a few hierarchically ordered timescales controlling our mental operations. This model offers new ways to explain different temporal phenomena, such as perception of order (corresponding to temporal resolution), the feeling of "nowness," and so on. Of crucial importance in this taxonomy system are two distinct domains: one operating in a range of some tens of milliseconds and the other in a range of a few seconds. These two domains are involved in different mental processes and may be studied with different experimental paradigms. Hence, they seem to be controlled by different neuronal mechanisms.

It should be stressed that in this paper we do not refer to any "constant numerical values" obtained in particular experimental studies, but rather indicate two temporal domains or operational processing windows: one of sub-seconds, the other of supra-seconds with, of course, intra- and inter-individual variability.

Hierarchical Multi-Timescale Framework Model for Cognitive Functions

Sub-Second Processing and Temporal Resolution

The sub-second level is related to perception of succession and identification of the temporal order of incoming events. Mental operations at this processing level are rooted in temporal resolution—the ability to perceive the order of stimuli presented in rapid sequences, thus, the identification of their before–after relation. Such temporal resolution requires, first, the identification of particular incoming stimuli within a sequence and, then, the perception of their order. This enables efficient flow of cognitive processes and coordination of every perceptual or motor act (Buhusi and Meck, 2005; Fostick and Babkoff, 2013, 2017; Wittmann, 2013). The efficiency of a subject's performance on this processing level can be evaluated with temporal order judgment tasks and indexed by the individual temporal-order threshold value (TOT). This is defined as the shortest interval between two successive stimuli within a rapid sequence that is necessary for a listener to report their order with at least 75% correctness (Fink et al., 2005, 2006; Szymaszek et al., 2009, 2017; Szelag et al., 2011, 2018; Bao et al., 2013, 2014; Nowak et al., 2016; Jablonska et al., 2020).

Of course, shorter gaps (lower TOTs) correspond to higher temporal resolution and thus to better performance in the millisecond domain. Experimental studies, including previous studies conducted in our laboratory, have consistently indicated that the typical TOT in young healthy listeners is about 30–80 ms (Mills and Rollman, 1980; von Steinbüchel et al., 1999a,b; Szymaszek et al., 2009; Heinrich et al., 2014; Fostick and Babkoff, 2017). The data show also huge inter- and intra-individual variability in TOT. Some people are often not able to report temporal order correctly at shorter gaps, needing longer intervals between sequential stimuli. These people are characterized by lower temporal resolution. Interestingly, such lower resolution often corresponds with poorer cognitive functioning in comparison with those characterized by better resolution (Jablonska et al., 2020 in working memory; Nowak et al., 2016 in executive functions and Grondin et al., 2007; Oron et al., 2015; Szymaszek et al., 2017, 2018 in language capacity). Huge intra-individual differences in these cognitive functions have been frequently reported in previous studies and may be marked in every-day situations, as well as in various clinical samples (Szelag et al., 2004a, 2014; Teixeira et al., 2013).

In the present study, temporal resolution in temporal order judgment was used to assess the participants' efficiency in the sub-second range (see below).

Supra-Second Processing and Temporal Integration

It has been long known that the supra-second processing level plays an important role in cognitive functioning. The mental operations at this processing level are temporally segmented into intervals of a few seconds. Support for such temporal segmentation comes from observations of the temporal dynamics of fluent speech. In many languages, such as English, German, Polish, and Chinese, the semantic processing occurs in intervals of a few seconds and phrases (i.e., logical verbal segments) are limited in time to this duration (Pöppel, 1994, 1997). Such temporal chunking reflects the existence of a specific binding mechanism that links successive events (e.g., syllables or words) into longer perceptual units limited in time up to a few seconds.

Further support for the existence of such a processing platform comes from a number of literature studies concerning motor behavior, duration discrimination, reproduction and production of temporal intervals, sensorimotor-synchronization, spontaneous rate of change of perception of ambiguous figures, short term memory, slow cortical potentials, and mismatch

---

Abbreviations: TIP, temporal information processing; ISI, inter-stimulus-interval; TOT, temporal-order threshold; MIIL, measured integration interval length; HTE, high temporal efficiency; LTE, low temporal efficiency.

negativity (for summary, see Pöppel, 1994, 1997, 2004, 2009; Szelag et al., 2002; Kowalska and Szelag, 2006; Matsuda et al., 2015; also Wang et al., 2015).

This body of evidence supports the thesis that the brain provides a temporal processing platform for our mental activity with a duration limited up to a few seconds. This platform may reflect the operation of the temporal integration mechanism—one of hypothetical mechanisms on the highest level of Pöppel's framework model of time perception (see above). This mechanism binds sequences of elementary events together into perceptual (or conceptual) units of approx. 3 s durations. The existence of a temporal limit of a few seconds has been discussed for a long time in the literature and referred as the impression of the "subjective present" or the feeling of "now" (e.g., James, 1890; Fraisse, 1984; Pöppel, 1997, 2004; Dorato and Wittmann, 2020).

The limits of temporal integration across a time window of a few seconds may be also examined with the subjective accentuation paradigm (Szelag et al., 1996, 1997, 1998, 2004b; Szelag, 1997). This is the phenomenon reflecting that identical sounds within isochronous sequences can be perceived as unequal. In this paradigm, a listener hears a sequence of identical beats at one rate and, during such listening, by placing subjective accents on every n-th beat, the listener imposes a new rate, creating an individual rhythmic pattern. The listener can impose a new subjective structure onto identical sounds, but with a specific restriction: if the beats follow each other with an inter-beat-interval of, for example, 1 s, it is easy to impose a subjective structure by giving a subjective accent to every second or third beat; however, if the inter-beat-interval becomes too long (e.g., 5 s), the listener can no longer impose such a subjective structure and reports only separate beats (Szelag et al., 1997, 2004b). In such a case, temporal binding is impossible because the successive beats cannot be grouped within the time window of a hypothetical "subjective present," as the integration would exceed the assumed 3 s duration.

Referring to the temporal segmentation of behavior mentioned above, the time period of around 3 s constitutes the fundamental unit related to the neuro-cognitive machinery in normal humans. Within such time frame information can be grasped as a unit, therefore, the longer integration indicates more effective processing because more events are linked together and processed as a *gestalt*.

Individual differences in the duration of such integration period were evidenced in our previous studies (Szelag et al., 2004b) indicating reduced binding in patients with receptive language problems in auditory comprehension. As in the normal sample the upper limit of integration corresponds to the typical duration of phrases lasting in the conversational speech a few seconds (Pöppel, 1994, 1997; Szelag et al., 2004b), the reduced integration reflects a situation where the listener's brain cannot hold the information until the phrase is completed by a speaker. As a consequence, the ending of listener's integration does not correspond to the ending of phrases produced by a speaker. The former seems too short to grasp the whole phrase as a unit causing comprehension problems. By analogy, reduced capacity of binding may accompany non-optimal cognitive functioning, for example, less efficient working memory in normal subjects.

Taking the above rationale into account, the subjective accentuation paradigm was also applied in the present study to measure efficiency in the supra-second domain (see below).

Cross-Domains Relations in Temporal Information Processing

The above framework model raises the question of relationships between the different timescales. Given the evidence that our brains can generate discrete time quanta in the aforementioned two domains, there is ongoing debate as to whether performance on sub- and supra-second processing levels is related or whether they work independently in controlling behavioral activity (Mauk and Buonomano, 2004). In other words, the question is: are these two levels controlled by one underlying mechanism or by independent processes?

One may assume that such cross-domain overlapping may be expected from the theoretical point of view. Referring to the above hierarchical model of time perception, the rules of each hierarchy assume that each higher level should include phenomena observed at the lower level, but on the higher level new properties should be added.

The majority of studies on cross-domain comparisons focus on processing physical standards of defined durations in the sub-second range (typically up to 1 s, sometimes even up to 2 s) and the supra-second range (typically above a few seconds). These studies indicate a dissociation between these two timescales, suggesting the involvement of different neuronal processes in these two domains which are controlled by different mechanisms (e.g., Lewis and Miall, 2003a,b; Ulbrich et al., 2007; Morillon et al., 2009; Bangert et al., 2011; Gilaie-Dotan et al., 2011, 2016). Specifically, the sub-second range is assumed to be associated with motor and sensory processes and is known as "automatic timing." In contrast, the supra-second range, known as "cognitive timing," is associated more with cognitive mechanisms allowing the perception of accumulating durations.

The dissociation between sub- and supra-second interval timing is assumed to be reflected in the activity of neuronal circuits and represented by different types of oscillators generating spikes at regular intervals, which are built inside various circuits of the brain, known as "neural temporal units" (Merchant et al., 2008; Gupta, 2014; Gupta and Chen, 2016; Gupta and Merchant, 2017). This hypothesis is supported by the observation that training in discrimination between two durations can be generalized to different modalities but not to different durations (Buonomano and Karmarkar, 2002). As separate active clocks (pacemaker neurons) are proposed for each neural circuit, multiple calibration mechanisms of the proposed modular clock mechanism would be necessary to coordinate sub- and supra-second interval timing for controlling stable temporality and the impression of a continuous world (Gupta, 2014; Gupta and Chen, 2016).

Nevertheless, the studies mentioned above concentrated mostly on representation of the physical time of the perceived duration and, then, the encoding of the presented durations by neural circuits. These studies have predominantly employed duration judgment paradigms (reproduction, estimation, discrimination, comparison, etc.). In this context, one may ask about cross-domains interactions in perceptual timing for paradigms free from any duration judgment and, thus, from any translation of the physical time into neural processes. We therefore used such tasks in our study.

On the other hand, several various complementary models of TIP were proposed by Buonomano and Karmarkar (2002) supporting the idea of a centralized timing mechanism (or a pacemaker) for different tasks in parallel to separate populations of neurons for different intervals. Accordingly, in such labeled-line models, different intervals are coded by activity in independent and discrete populations of neurons. On the contrary, in population clock models, time is coded by the population activity of a large group of neurons and timing requires dynamic interaction between neurons for the parallel processing of interval, duration, order, and sequence cues. According to these authors, population models seem more likely to be the basis of timing in the range of tens to hundreds of milliseconds.

Convincing evidence for the neural representation of a pacemaker comes from electrophysiological and neuroimaging studies. Using event-related potential technology, Zhang et al. (2021) revealed similar activity patterns for sub- and supra-second time perception. Furthermore, Nani et al. (2019) provided a more recent meta-analysis of 84 published articles for a total of 109 experiments employing motor and non-motor (perceptual) tasks. They showed that both sub- and supra-second conditions recruit cortical and subcortical areas, but subcortical ones are activated more in sub-second tasks than those in supra-second tasks, in which a greater contribution from cortical activation was evidenced. However, in all studied conditions, common activations were observed in the SMA (rostral and caudal parts) along with the striatum and claustrum. These areas are supposed to be an essential node in different networks engaged in time processing.

To summarize, in light of this evidence, the mechanisms underlying cross-domain overlapping remain unclear and the question of relations between the timescales is still unanswered.

Experimental Aim

Given the unclear relationships between TIP on sub- and supra-second levels, the present study asks whether we are equipped with a hypothetical core timing mechanism. Therefore we investigated whether better efficiency on the millisecond scale is accompanied by better processing on supra-second one. Specifically, we verified whether persons characterized by more efficient TIP in the sub-second domain considered as the basic level in the above framework hierarchical model are also more efficient in the supra-second domain. To avoid the above reservations regarding duration judgment paradigms which employ often the involvement of specific reference system for the use of conventional time units (e.g., seconds, Kowalska and Szelag, 2006), we used a novel approach—rarely studied before, as far as we are aware—employing for cross-domain comparisons the efficiency of temporal resolution in a temporal order judgment task (sub-second domain) and the limits of temporal integration in a subjective accentuation task (supra-second domain). These two paradigms incorporate the intrinsic TIP, free from any influences of the translation from physical time to neural processes.

MATERIALS AND METHODS

Participants

The initial sample consisted of 118 young healthy volunteers (61 female/57 male), aged between 20 and 27 years (M ± SD = 23 ± 2 years). They were recruited *via* social media in the Warsaw area. All participants were right-handed native Polish speakers. They reported no systemic diseases, neurological or psychiatric disorders, head injuries in the past, addictions, or the use of medication that affects the nervous system. Moreover, all participants reported a lack of any regular musical education. Participants were screened for normal levels of cognitive abilities with the Raven Standard Progressive Matrices, as well as for normal hearing levels using pure-tone audiometry (Audiometer MA33, MAICO).

The study was in line with the Declaration of Helsinki and was approved by the Bioethics Committee of Nicolaus Copernicus University (permission no. KB 289/2019). All participants provided written informed consent prior to the study.

Procedure

The experimental studies were conducted in a soundproof room in the Laboratory of Neuropsychology at the Nencki Institute of Experimental Biology. The methods included two parts: (1) screening for efficiency of sub-second timing using an auditory temporal-order judgment task and (2) assessment of efficiency of supra-second timing using a subjective accentuation task.

The experimental methods applied in both these paradigms were similar to those reported in our earlier papers; therefore they are only briefly summarized below. For the method applied in Part 1 see, for example, Szymaszek et al. (2009), Szelag et al. (2011, 2018), Bao et al. (2014), Nowak et al. (2016) for the method used in Part 2 see, for example, Szelag et al. (1996, 1997, 1998, 2004b), or Szelag (1997).

Part 1: Screening for Efficiency of Sub-Second Timing Using the Auditory Temporal-Order Judgment Paradigm

Two complementary tasks were applied using spatial and spectral stimulus presentation modes.

Stimuli

In both spatial and spectral tasks, paired acoustic stimuli were presented in rapid succession with various Inter-Stimulus Intervals (ISIs) separating the two stimuli in each pair. In the spatial presentation mode, paired clicks (square-wave pulses of 1 ms duration each) were presented monaurally in an alternating stimulation mode: one click was presented to one ear followed by another click to the other ear. In the spectral presentation mode, pairs of two 10 ms sinusoidal tones (i.e., a low tone of 400 Hz and a high tone of 3,000 Hz) were presented. These two paired tones were adjusted to equal loudness on the basis of isophones.

The binaural stimulus presentation mode was used and each tone pair was presented to both ears.

The stimuli were generated by a computer with a sound controller using Waves MaxxAudio Pro software and presented *via* headphones at a comfortable listening level. The ISIs within each pair reflected the time gap between the offset of the first stimulus and the onset of the second stimulus. The duration of the ISIs varied during the experiment according to a pre-defined adaptive algorithm (see below for detailed description).

Tasks

The participant's task was to report verbally the temporal order of two successive stimuli within each pair—their before—after relation. In the spatial mode, two alternative responses were possible: left–right or right–left. In the spectral mode, two alternative responses were possible: low–high or high–low. The experimental situation in these two tasks is displayed in **Figure 1**.

To control the duration of ISIs in consecutive trials, an adaptive algorithm based on maximum likelihood estimation was used. The implementation of this algorithm for testing young healthy listeners was based on the literature reports by Treutwein (1997), Fink et al. (2005, 2006), as well as on our previous studies (Wittmann and Szelag, 2003; Szymaszek et al., 2009; Szelag et al., 2011, 2018; Bao et al., 2013, 2014; Nowak et al., 2016).

The algorithm consisted of two steps (see Szelag et al., 2018, p. 4–5). In Step 1, the participant responded to 20 introductory trials comprising paired stimuli presented with fixed ISIs of varying durations in consecutive trials. They were presented first in decreasing (10 trials) and, subsequently, in increasing order (in the next 10 trials; i.e., down and up) according to pre-defined rules. In the spatial task, the ISI ranged from 160 to 1 ms (changing in 18 ms steps) and in the spectral task from 240 to 1 ms (changing in steps of 27 ms). The different testing ranges in the spatial and spectral tasks were based on our previous observations, indicating different performance in these two tasks in young participants. After completion of these 20 introductory trials, based on the correctness of the participant's responses, the program calculated the ISI value for the initial trial in Step 2 of testing at the 75% probability of correct responses according to maximum likelihood estimation (Treutwein, 1997).

In Step 2, 50 trials were presented. In each of these 50 trials, the ISI was adjusted adaptively: it decreased after each correct response and increased after each incorrect response. The exact values of decreased or increased ISIs were randomly selected from a pre-defined range which varied depending on the tested ISI. To ensure accurate assessment, decrement steps were 0.5–5% of the ISI value of the previous trial, while increments were 10–20% of the previous ISI value.

On the basis of 70 completed trials (i.e., 20 trials in Step 1 and 50 trials in Step 2), the auditory Temporal Order Threshold (TOT) value for each participant was obtained as the mean of the estimated likelihood, calculated at 75% probability of correct responses (Treutwein, 1997). The measured TOT was defined as the shortest ISI between two successive stimuli necessary for a participant to report their temporal order with at least 75% correctness.

To focus the participant's attention, each pair of stimuli was preceded by a warning signal delivered binaurally 1 s before the first stimulus in each pair. Then, the paired stimuli were presented monaurally (in the spatial task) or binaurally (spectral task). After each presentation, participants reported verbally the order of the two stimuli in the presented pair, i.e., left-right or right-left in the spatial task and high-low or low-high in the spectral task.

Prior to the collection of data, each participant was given a verbal instruction by the experimenter and, then, was presented with a few practice trials consisting of pairs with a relatively long ISI. In these practice trials, feedback on correctness achieved was given after each answer. All participants performed these practice trials satisfactorily. Next, the proper measurement started and no feedback on correctness was given.

The measurement was conducted with each participant individually in two separate sessions, separated by a break of a few days. In each session, both the spatial and spectral tasks were completed. The TOJ measurement lasted approximately 15 min for each task. The TOT values obtained in these two sessions were averaged and the mean TOT values in the spatial and spectral task were further analyzed.

Outcome Measure

The outcome measures were TOT values for the spatial and spectral tasks. These values reflected the participant's TIP efficiency in the millisecond domain (sub-second level) in these two tasks. Accordingly, lower TOT values reflected better performance (HTE), whereas higher TOT values corresponded to poorer performance (LTE).

Part 2: Assessment of Efficiency in Supra-Second Timing Using a Subjective Accentuation Paradigm

Stimuli

The auditory stimuli were metronome beats (square-wave clicks of 1 ms duration) generated by the Adobe Audition 3.0 program and presented *via* earphones binaurally at the nine following frequencies: 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, and 5 beats/s. This means that the inter-beat intervals in these sequences were: 1000, 667, 500, 400, 333, 286, 250, 222, and 200 ms, respectively. These frequencies of metronome beats were selected on the basis of our pretesting data, which showed that the majority of tested participants could not perform the experimental task when the beat rate was above or below the values enumerated above (see section "Introduction" for further explanation).

Task

Participants were asked to listen to these equally spaced sequences of beats and to mentally accentuate during such listening every n-th (e.g., second, third, fourth, or other) beat to create an individual rhythmic pattern, integrating as many beats as possible. Obviously, this subjective rhythm existed only in the participant's mind, but not objectively, as the presented sequences were monotonous in their nature and without any actual accentuation. After presentation, participants reported verbally the maximum number of beats they could unite into a rhythmic pattern for each presented beat sequence. This new group consisted of the accentuated beat and the unaccentuated

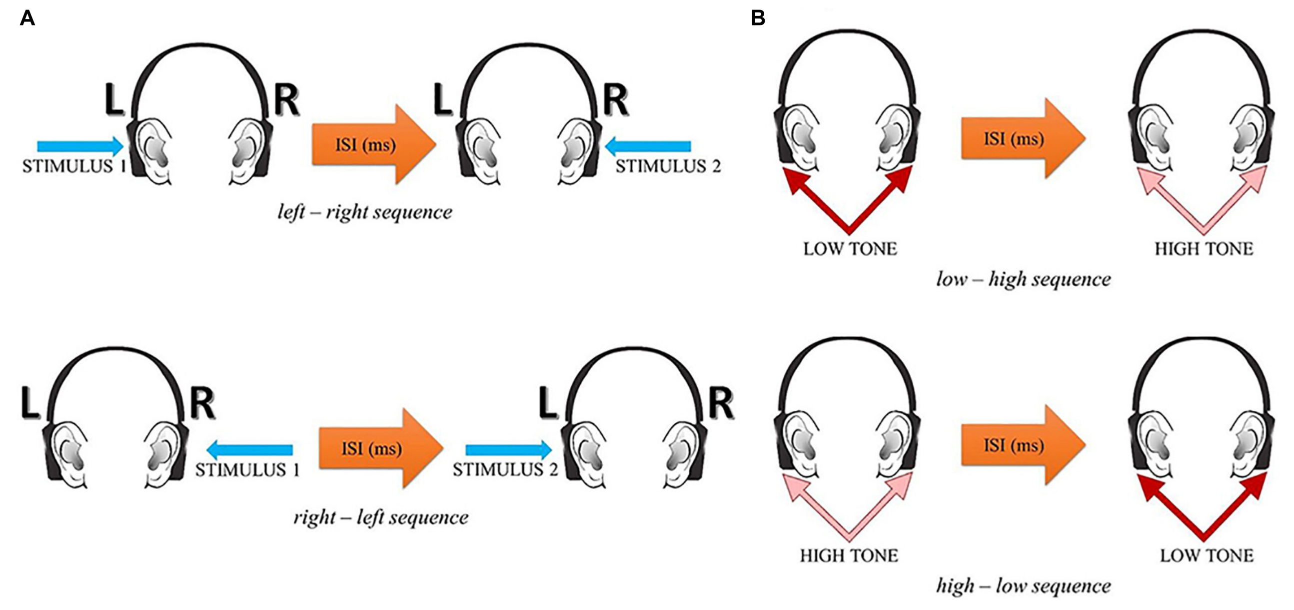


FIGURE 1 | The experimental situation in the spatial **(A)** and spectral **(B)** temporal-order judgment tasks (reprinted from Szelag et al., 2018, p. 4).

ones following it. The time interval in which the participant could integrate the information was reflected in measured integration interval length (MIIL) obtained by multiplying the number of reported beats being united by a time distance between two successive beats at a given metronome ratio. The duration of MIIL corresponded to the upper limit of integration capacity. The experimental situation for the subjective accentuation task is displayed in **Figure 2**.

Prior to collection of the data proper, each participant was given a verbal instruction by the experimenter and, then, was presented with a few practice trials in which some examples of metronome sequences were played. All participants could unite beats during these example sequences and completed the practice trials satisfactorily. Next, the measurement proper started.

During the measurement proper, nine metronome frequencies (enumerated above) were presented 10 times each in random order. The MIIL values obtained for these 10 presentations were averaged and the mean MIIL for a given metronome frequency was further analyzed for each participant. The study comprised 90 trials. The measurement was conducted with each participant individually and lasted approximately 30 min.

Outcome Measure

The MIIL (in ms) calculated for particular metronome frequencies, reflecting the duration of perceptual unit comprising subjectively grouped beats (**Figure 2**). This MIIL was defined as the extent of temporal integration in supra-second time domain.

RESULTS

All statistical analyses were conducted using IBM® SPSS® Statistics 28.

Study Design

The initial sample was screened for the level of temporal efficiency in the sub-second time domain using spatial and spectral temporal-order judgment tasks, which measured in each participant the temporal resolution in the auditory perception of temporal order. On the basis of the screening data obtained, two groups of participants were selected from the initial sample (*N* = 118): one group characterized by high temporal efficiency (HTE; *n* = 41) and the other group characterized by low temporal efficiency (LTE; *n* = 40). See below for the detailed procedure used to classify participants and the characteristics of these two groups. Next, the efficiency of performance of the HTE and LTE groups in the supra-second domain was compared using the subjective accentuation task. Finally, the relationships between the efficiency of TIP in sub- and supra-second ranges were tested in the HTE and LTE in the spatial and spectral task separately using the Spearman's rank correlation analysis.

Classification of Participants According to Their Efficiency in Sub-Second Timing

The median TOT values obtained in the initial participant sample (*N* = 118) in the spatial task was 40 ms and in the spectral task was 69 ms. On the basis of these two median values, two groups of participants (namely HTE and LTE) were selected from the initial sample (for the selection procedure, see **Figure 3**). These two far groups were selected to compare the performance between the defined efficiency of TIP, considering the clear-cut points between HTE and LTE. The remaining participants—those characterized by medium TOT values—were not considered in further analyses.

The HTE Group (*n* = 41, marked with red dots) was characterized by TOT values in both tasks below the median TOT. The LTE Group (*n* = 40, black dots) was characterized in

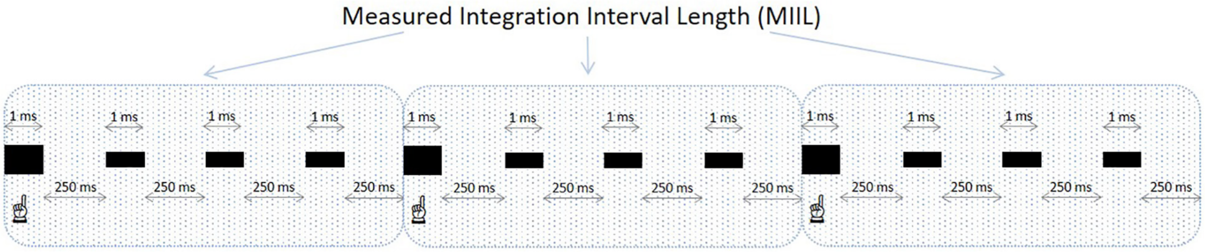


FIGURE 2 | An example of a grouping of beats in the subjective accentuation task. The frequency of presented beats is 4 beats/s and the participant accentuates every fourth beat (the accentuated beats are bolded and indicated). In this sequence, the inter-beat interval is 250 ms and the MIIL is 1,000 ms (4 ms × 250 ms). When calculating the MIIL, the duration of the integrated beats (4 ms × 1 ms) is not included.

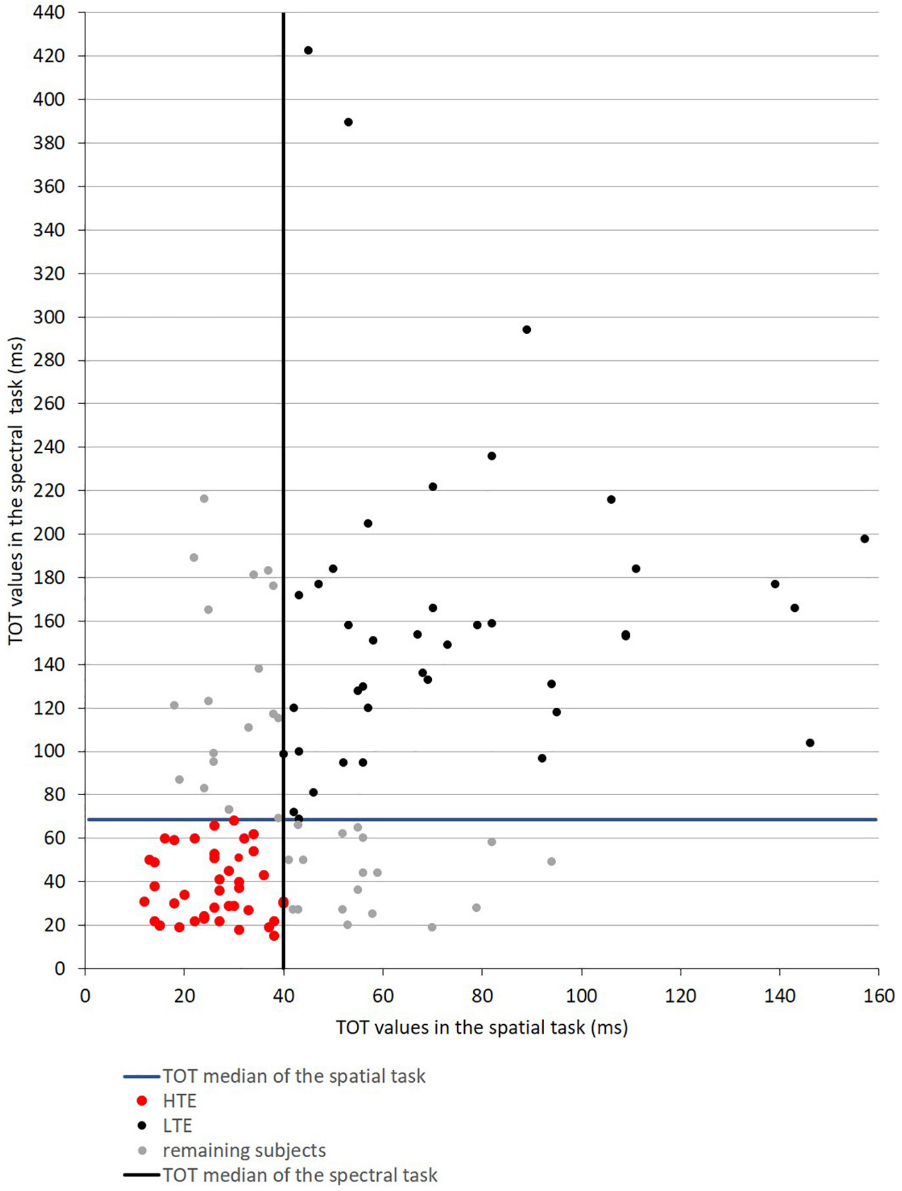


FIGURE 3 | Scatter plot data presenting TOT values obtained in the spatial and spectral temporal-order judgment tasks in the initial participant sample (*N* = 118). The vertical and horizontal solid lines reflect the median values of TOT in these two tasks, indicating four quartiles of obtained results depicted with different colors.

TABLE 1 | Characteristics of groups.

Group	*n*	Age (years)		Sex	TOT (mean)	
		Range	Mean (SD)	(Female/male)	Spatial	Spectral
HTE	41	From 20 to 27	23 (2)	16/25	26 ms	38 ms
LTE	40	From 20 to 26	23 (2)	23/17	75 ms	162 ms

both tasks by TOT values above the median TOT. Next, in the HTE and LTE groups the mean TOT in each task was calculated and further analyzed. The remaining participants (*n* = 37, gray dots) indicated mixed temporal efficiency—below the median TOT in one task but above this median TOT in the other, or vice versa. Therefore, they were not considered in further analyses. Detailed characteristics of the HTE and LTE groups are given in **Table 1**.

This classification of participants into HTE and LTE groups was further confirmed in two separate 2-way analyses of variance (ANOVAs) conducted on TOT values from the spatial (ANOVA 1) and spectral (ANOVA 2) tasks with "Group" (HTE vs. LTE) as between-subject variable. The data submitted to these analyses were transformed by natural logarithm, because the distribution of TOT values obtained in the spectral task deviated from the Gaussian distribution.

These two analyses revealed only a significant effect of "Group" [$F(1, 79) = 89.09$; $p < 0.001$; η2 = 0.53 and $F(1, 79) = 255.8$; $p < 0.001$; η2 = 0.76, for spatial and spectral tasks, respectively]. In the HTE group, the mean TOT value (26 ms for the spatial task and 38 ms for the spectral task) was lower than in the LTE group (75 ms for the spatial task and 162 ms for the spectral task), indicating better performance in the HTE group.

Supra-Second Timing Efficiency: Subjective Accentuation Task

The results of the present study are congruent, in general, with our previous observations using the subjective accentuation paradigm (Szelag et al., 1996, 1997, 2004b), indicating a specific integration process in supra-second intervals. The young healthy volunteers studied here could bind mentally temporally separated successive beats into larger perceptual units independently of their efficiency in sub-second timing (**Figure 4**). Thus, the separated beats were organized at the perceptual level into a higher-order structure that dominated their serial order.

Similarly to our previous studies, the MIIL in both groups strongly depended on the presented metronome frequency, but some regularities observed previously were also confirmed (**Figure 4**). Namely, for the lowest metronome frequency (1 beat/s) the MIIL did not exceed the 3 s time window which is typically assumed as the maximum limit of the temporal integration mechanism (Pöppel, 2004, 2009). On the other hand, for higher frequencies, the duration of MIIL was systematically shortened up to about 1 s.

Looking at the **Figure 4**, one can infer that at least three different strategies were used in the subjective accentuation task. First, if participants integrate information only by time (i.e., in a constant period of, e.g., 2, 3, or 4 s), the MIIL would be constant and independent of the presented frequency (horizontal continuous lines, **Figure 4**). Second, if the participants integrate information only by number (i.e., counting a constant number of beats, e.g., 2, 3, or 4), the MIIL would depend strongly on the presented frequency (dashed lines, **Figure 4**). Third, a combination of these two strategies would be possible in both groups, but integration by constant time dominates in the HTE group, whereas, in contrast, the LTE group supported the integration process with mental counting, especially for the higher metronome ratios.

To compare the performance of the HTE and LTE groups, the MIIL (transformed by square root extraction) values were submitted to a 2-way analysis of variance (ANOVA). The design included "Group" (HTE vs. LTE) as the between-subject variable and "Metronome Frequency" (nine ratios: 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, and 5 beats/s) as a within-subject variable. Greenhouse-Geisser correction was applied on the basis of results of Mauchly's test for sphericity. After the analysis of variance, to determine the sources of significance, the Bonferroni *post hoc* procedure was applied.

The analysis yielded a significant main effect of "Group" [$F(1, 79) = 8.58$, $p < 0.01$, η2 = 0.10], "Metronome Frequency," [$F(2.53, 199.48) = 102.93$, $p < 0.01$, η2 = 0.57], as well as the "Group × Metronome Frequency" interaction [$F(2.53, 199.48) = 3.42$, $p = 0.03$, η2 = 0.04].

In the HTE group, the mean MIIL (1,736 ms) was longer than that in the LTE group (1,411 ms). However, this main effect was modified by the presented metronome frequency. Pairwise comparison tests with Bonferroni correction showed that significant differences between groups were observed only for higher metronome frequencies (i.e., from 3 up to 5 beats/s), being non-significant for lower frequencies (i.e., from 1 up to 2.5 beats/s). These relationships are illustrated in **Figure 4**. For more descriptive statistics see **Table 2**.

These relationships between the efficiency of TIP in sub- and supra-second ranges in the HTE and LTE were further confirmed by the Spearman's rank correlations. We observed significant negative correlations between the TOT values achieved in the temporal-order judgment task (in spatial and spectral tasks, separately) and MIILs obtained in subjective accentuation task (**Table 3**). Better temporal resolution (lower TOT values) was accompanied by longer integration (longer MIILs) for all metronome tempos with the exception of 1.5 and 2 beats/s.

DISCUSSION

The results of the two experiments presented here provide convincing evidence for the existence of a close relationship between the efficiency of TIP in sub- and supra-second ranges. The clear relationships were evident for two different indices of participants' performance: TOT in the temporal-order judgment task (sub-second level) and MIIL in the subjective accentuation task (supra-second level). It is important to emphasize that these two tasks employed totally different experimental procedures based on intrinsic timing operations to measure performance on these two levels.

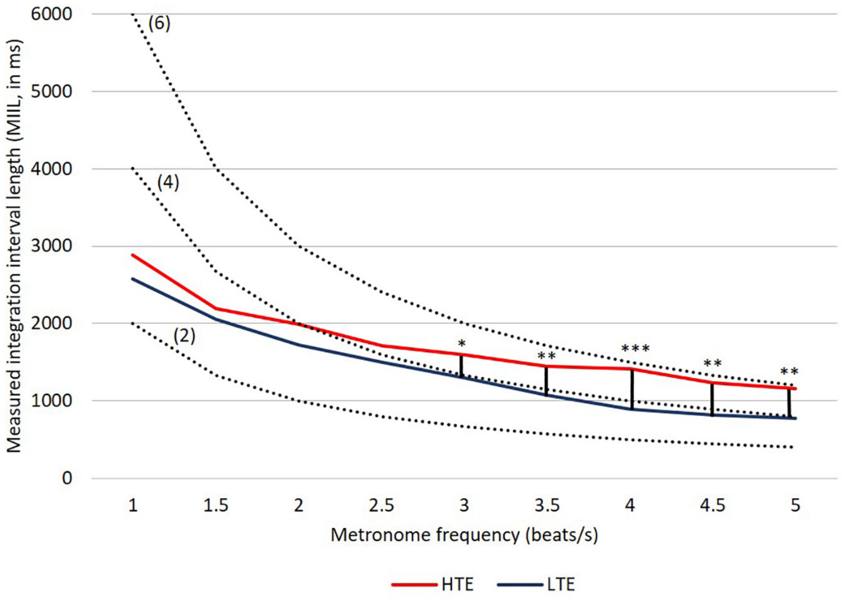


FIGURE 4 | MIILs for various metronome tempos in HTE and LTE groups. Gray lines reflect two hypothetical integration strategies: (A) integration in a constant time limited up to e.g., 3,000, 2,500, or 2,000 ms (solid horizontal lines) and (B) integration by mental counting ignoring the constant time (e.g., up to 2, 4 or 6 beats, dashed lines). Significant differences between groups are marked with asterisks: $*p < 0.05$; $**p < 0.01$; $***p < 0.001$. For more details see **Table 2** and text below.

It is interesting to note that listeners characterized by more efficient temporal resolution (i.e., HTE) could integrate information in the supra-second domain in longer units, reflecting more efficient temporal integration. In contrast, subjects less efficient in such resolution (i.e., LTE) indicated, in parallel, less efficient integration reflected in shorter MIILs than those observed in HTE (**Figure 4** and **Table 2**). Despite these differences, in both groups the upper limit of integration evidenced in the subjective accentuation task was within 3 s time "window."

Another important result was the use of different integration strategies in the LTE and HTE groups; however, this was evidenced only in situations in which the processed material required higher mental load—for higher metronome ratios (from 3 up to 5 beats/s; **Figure 4** and **Table 2**). In these situations, LTE relied more on the mental counting of consecutive beats (usually up to 4), but not so much on integration based on constant time related to the limits of temporal integration, as used by HTE in these situations (**Figure 4**). The application of such a counting strategy in LTE resulted in the creation of shorter rhythmic patterns (shorter MIIL), in contrast to the longer patterns in HTE, reflecting the ability to keep the information in longer units. The information processing in the latter group seems more efficient than the former one (see section "Introduction" for more explanations). The cross-domain relations in HTE and LTE reported here were further confirmed by the negative correlations, indicating that better temporal resolution (lower TOT values) was accompanied by longer binding (longer duration of MIILs) for nearly all metronome tempos (**Table 3**). To sum up, LTE displayed narrowed binding resources within a few seconds because of a shorter (i.e., less efficient) span of integration resources in tasks with higher mental load, as exemplified by the higher metronome frequencies presented here.

It should be stressed that, for such integration capacity, the difference between HTE and LTE was non-significant for tasks with lower mental load, exemplified by lower metronome ratios (i.e., from 1 up to 2.5 beats/s; **Figure 4** and **Table 2**). This means that the upper limit of temporal integration remained relatively stable, independently of the temporal resolution power. In both HTE and LTE, a similar integration strategy was employed for lower metronome ratios, as these two groups relied partially on constant time and partially on mental counting (**Figure 4**). Finally, our results support the thesis that the upper limit rarely exceeded 3 s intervals and was resistant to temporal resolution power, indexed by TOT.

The cross-domain relations reported in these two paradigms support the thesis that one "clock" (or neural mechanism) may be used for sub- and supra-second tasks. A solid conceptual background for understanding these relations may be provided by once again referring to Pöppel's taxonomy of TIP—in particular, to temporally discrete information processing within

TABLE 2 | The MIIL for particular metronome frequencies in HTE and LTE groups.

Group	Presented metronome frequencies (beats/s)								
	1	1.5	2	2.5	3 *	3.5 **	4 ***	4.5 **	5 **
HTE	2,880	2,193	1,990	1,713	**1,599**	**1,446**	**1,412**	**1,237**	**1,156**
LTE	2,573	2,055	1,719	1,500	**1,297**	**1,073**	**888**	**818**	**779**

Significant differences between groups (bolded) are marked by asterisks: $*p < 0.05$; $**p < 0.01$; $***p < 0.001$.

TABLE 3 | Spearman's rho correlation coefficients (and significance levels) between TOT values obtained in the spatial and spectral temporal-order judgment task and MIILs for various metronome tempos.

TOT task	Presented metronome frequencies (beats/s)								
	1	1.5	2	2.5	3	3.5	4	4.5	5
Spatial	−0.234*	−0.161	−0.208	−0.226*	−0.219*	−0.284*	−0.315**	−0.242*	−0.246*
Spectral	−0.242*	−0.128	−0.168	−0.217$^{\#}$	−0.239*	−0.322**	−0.370**	−0.266*	−0.271*

Asterisks indicate significant correlations: *$p < 0.05$; **$p < 0.01$; $^{\#}p < 0.051$.

a time window of approx. some tens of milliseconds (see section "Introduction"). Despite important individual differences, much experimental data has indicated that the perception of succession is controlled by the central timing mechanism, probably implemented in neuronal oscillations with a periodicity of about 25–40 Hz observed in electrophysiological activity (Van Rullen and Koch, 2003; Fink et al., 2006; Bao et al., 2013). Each period of such oscillations reflects an elementary processing unit (a system state) of a duration of approx. 30 ms. At a theoretical level, therefore, it was hypothesized that two events occurring within one such system state will be treated as co-temporal and fused into one unit. As a consequence, the before–after relationship cannot be established. In contrast, the before–after relation can be perceived if two stimuli occur in at least two successive oscillatory periods (Pöppel, 2009). Considerable data supporting the thesis of a common basic mechanism underlying the perception of temporal order can be provided from experiments on different sense modalities, indicating that TOT appears to have a similar numerical value of some tens of milliseconds for different sense modalities (visual, auditory, tactile; e.g., Hirsh and Sherrick, 1961; Pöppel, 1994, 1997; Kanabus et al., 2002).

The question remains as to what neural process could be a potential source of such oscillatory activity, providing the temporal constraints for sequencing ability? There is strong evidence that spontaneous (or stimulus triggered) gamma band oscillations corresponding in periodicity to the value of TOT play an important role in human cognition (Poeppel, 2003; Van Rullen and Koch, 2003; Heim et al., 2013). Furthermore, one may expect that the periodicity of neuronal oscillations might be modified by a hypothetical pacemaker, resulting in lower or higher TOT. For example, a higher pacemaker rate might lead to shorter periods of gamma oscillations and lower TOT, as evidenced in the HTE group. Conversely, a lower rate in such a hypothetical pacemaker might lead to longer periodicity of such oscillations and higher TOT values, reflected in our study in less efficient sequencing abilities in the LTE group (**Figure 3**). This could provide a theoretical explanation of the problem of individual differences in temporal resolution power.

Another electrophysiological candidate for a timekeeping mechanism could also be the beta rhythm, with a periodicity of 14–30 Hz (see van Wassenhove et al., 2019 for a recent review). Recently, evidence for the contribution of the beta rhythm to timing behavior was found in synchronization–continuation tasks in primates (e.g., Bartolo and Merchant, 2015) as well as in humans in tasks addressing predominantly supra-second timing. Specifically, Kononowicz and van Rijn (2015) reported the contribution of the beta power to the production of temporal intervals of 2.5 s duration, Kulashekhar et al. (2016) in duration judgment, and Wiener et al. (2018) in lengthening of duration experienced subjectively using transcranial alternating current stimulation (tACS). Moreover, Bernasconi et al. (2011) provided evidence for the contribution of beta range oscillations (18–23 Hz) to the perception of order in temporal order judgment tasks. To sum up, we cannot exclude the hypothesis that beta rhythm oscillations of a frequency of approx. 14–30 Hz (i.e., one period of ca. 30–70 ms duration) contribute to a time keeping mechanism in sub-second TIP.

The converging lines of evidence briefly summarized above consistently indicate that auditory perception of temporal order may represent a very basic mechanism of information processing rooted in electrophysiological indices. Further studies are needed to clarify which rhythm can be considered to be a basic oscillator that may actuate the pacemaker of the hypothetical internal clock. However, there is no doubt that temporal resolution is represented endogenously and based on intrinsic neural operations. They seem relevant for many cognitive functions, in both normal states and in pathological conditions (Nowak et al., 2016; Jablonska et al., 2020). Of course, these operations may be influenced by the nature of the presented material (spectral vs. spatial), as well as stimulus presentation modes (monaural vs. binaural). These influences were discussed in detail in our previous paper (Szelag et al., 2018).

The cross-domain relationships summarized in the Introduction point to a dissociation between neural circuits involved in TIP in different timescales. An important problem may be that these comparisons have often focused on duration judgment methods, using various paradigms involving the translation of classical time units (usually seconds or milliseconds) into neural processes. Such translation processes are not necessarily the same for different timescales. It would be difficult to accept that the same time frame operates in sub- and supra-second scales. To avoid such reservations, in the present study we tested endogenous mechanisms involved in the perception of temporal order and in temporal integration, which seem to be free from any bias from the classic time units and are rooted in intrinsic timing operations. It is a critical question whether such cross-domain dissociation may be also present in the case of the endogenous operations investigated in our study.

Final Conclusion

The results of the present study suggest that intrinsic timing operations on sub-second level may regulate TIP on the supra-second range. We can conclude that the temporal resolution in the tens of millisecond range reflected in the perception of temporal order is incorporated also at higher levels of TIP and may be essential for predicting individuals' efficiency in binding operations at the supra-second domain. Thus, neural entrainment in the sub-second range may help the brain to calibrate its timing for information processing in the supra-second range.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Bioethics Committee of Nicolaus Copernicus University (Permission No. KB 289/2019). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

ES conceptualized and designed the study, interpreted the data, wrote the manuscript, and is responsible for the final version of the manuscript. MS recruited the participants, acquired, analyzed and interpreted the data, and wrote the manuscript. AS interpreted the data and wrote the manuscript. All authors approved the final version of the manuscript.

FUNDING

This work was supported by the National Science Centre (Narodowe Centrum Nauki, NCN), Poland, grant no 2018/29/B/HS6/02038.

ACKNOWLEDGMENTS

We would like to thank Anna Bombinska for her technical assistance during the data collection phase.

REFERENCES

Bangert, A. S., Reuter-Lorenz, P. A., and Seidler, R. D. (2011). Dissecting the clock: understanding the mechanisms of timing across tasks and temporal intervals. *Acta Psychol.* 136, 20–34. doi: 10.1016/j.actpsy.2010.09.006

Bao, Y., Fang, Y., Yang, T., Wang, L., Szymaszek, A., and Szelag, E. (2014). Auditory perception of temporal order: a comparison between tonal language speakers with and without non-tonal language experience. *Acta Neurobiol. Exp.* 74, 98–103.

Bao, Y., Szymaszek, A., Wang, X., Oron, A., Pöppel, E., and Szelag, E. (2013). Temporal order perception of auditory stimuli is selectively modified by tonal and non-tonal language environments. *Cognition* 129, 579–585. doi: 10.1016/j.cognition.2013.08.019

Bartolo, R., and Merchant, H. (2015). β oscillations are linked to the initiation of sensory-cued movement sequences and the internal guidance of regular tapping in the monkey. *J. Neurosci.* 35, 4635–4640. doi: 10.1523/JNEUROSCI.4570-14.2015

Bernasconi, F., Manuel, A. L., Murray, M. M., and Spierer, L. (2011). Pre-stimulus beta oscillations within left posterior sylvian regions impact auditory temporal order judgment accuracy. *Int. J. Psychophysiol.* 79, 244–248. doi: 10.1016/j.ijpsycho.2010.10.017

Buhusi, C. V., and Meck, W. H. (2005). What makes us tick? Functional and neural mechanisms of interval timing. *Nat. Rev. Neurosci.* 6, 755–765. doi: 10.1038/nrn1764

Buhusi, C. V., Oprisan, S. A., and Buhusi, M. (2018). Biological and cognitive frameworks for a mental timeline. *Front. Neurosci.* 12:377. doi: 10.3389/fnins.2018.00377

Buonomano, D. V., and Karmarkar, U. R. (2002). How do we tell time? *Neuroscientist* 8, 42–51. doi: 10.1177/107385840200800109

Choinski, M., Szelag, E., Wolak, T., and Szymaszek, A. (2020). Working memory in aphasia: the role of temporal information processing. *Front. Hum. Neurosci.* 14:589802. doi: 10.3389/fnhum.2020.589802

Dorato, M., and Wittmann, M. (2020). The phenomenology and cognitive neuroscience of experienced temporality. *Phenomenol. Cogn. Sci.* 19, 747–771. doi: 10.1007/s11097-019-09651-4

Fink, M., Churan, J., and Wittmann, M. (2005). Assessment of auditory temporal-order thresholds–a comparison of different measurement procedures and the influences of age and gender. *Restor. Neurol. Neurosci.* 23, 281–296.

Fink, M., Ulbrich, P., Churan, J., and Wittmann, M. (2006). Stimulus-dependent processing of temporal order. *Behav. Process.* 71, 344–352. doi: 10.1016/j.beproc.2005.12.007

Fostick, L., and Babkoff, H. (2013). Different response patterns between auditory spectral and spatial temporal order judgment (TOJ). *Exp. Psychol.* 60, 432–443. doi: 10.1027/1618-3169/a000216

Fostick, L., and Babkoff, H. (2017). Auditory spectral versus spatial temporal order judgment: threshold distribution analysis. *J. Exp. Psychol. Hum. Percept. Perform.* 43, 1002–1012. doi: 10.1037/xhp0000359

Fraisse, P. (1984). Perception and estimation of time. *Annu. Rev. Psychol.* 35, 1–37. doi: 10.1146/annurev.ps.35.020184.000245

Gilaie-Dotan, S., Ashkenazi, H., and Dar, R. (2016). A possible link between supra-second open-ended timing sensitivity and obsessive-compulsive tendencies. *Front. Behav. Neurosci.* 10:127. doi: 10.3389/fnbeh.2016.00127

Gilaie-Dotan, S., Kanai, R., and Rees, G. (2011). Anatomy of human sensory cortices reflects inter-individual variability in time estimation. *Front. Integr. Neurosci.* 5:76. doi: 10.3389/fnint.2011.00076

Grondin, S., Dionne, G., Malenfant, N., Plourde, M., Cloutier, M. E., and Jean, C. (2007). Temporal processing skills of children with and without specific language impairment. *Can. J. Speech Lang. Pathol. Audiol.* 31, 38–46.

Gupta, D. S. (2014). Processing of sub-and supra-second intervals in the primate brain results from the calibration of neuronal oscillators via sensory, motor, and feedback processes. *Front. Psychol.* 5:816. doi: 10.3389/fpsyg.2014.00816

Gupta, D. S., and Chen, L. (2016). Brain oscillations in perception, timing and action. *Curr. Opin. Behav. Sci.* 8, 161–166. doi: 10.1016/j.cobeha.2016.02.021

Gupta, D. S., and Merchant, H. (2017). Understanding the role of the time dimension in the brain information processing. *Front. Psychol.* 8:240. doi: 10.3389/fpsyg.2017.00240

Heim, S., Keil, A., Choudhury, N., Friedman, J. T., and Benasich, A. A. (2013). Early gamma oscillations during rapid auditory processing in children with a language-learning impairment: changes in neural mass activity after training. *Neuropsychologia* 51, 990–1001. doi: 10.1016/j.neuropsychologia.2013.01.011

Heinrich, A., De la Rosa, S., and Schneider, B. A. (2014). The role of stimulus complexity, spectral overlap, and pitch for gap-detection thresholds in young and old listeners. *J. Acoust. Soc. Am* 136, 1797–1807. doi: 10.1121/1.4894788

Hirsh, I. J., and Sherrick, C. E. Jr. (1961). Perceived order in different sense modalities. *J. Exp. Psychol.* 62, 423–432. doi: 10.1037/h0045283

Jablonska, K., Piotrowska, M., Bednarek, H., Szymaszek, A., Marchewka, A., Wypych, M., et al. (2020). Maintenance vs manipulation in auditory verbal working memory in the elderly: new insights based on temporal dynamics of information processing in the millisecond time range. *Front. Aging Neurosci.* 12:194. doi: 10.3389/fnagi.2020.00194

James, W. (1890). "The consciousness of self," in *The Principles Of Psychology*, ed. W. James (New York, NY: Henry Holt and Co), 291–401. doi: 10.1037/10538-010

Kanabus, M., Szelag, E., Rojek, E., and Pöppel, E. (2002). Temporal order judgement for auditory and visual stimuli. *Acta Neurobiol. Exp.* 62, 263–270.

Kononowicz, T. W., and van Rijn, H. (2015). Single trial beta oscillations index time estimation. *Neuropsychologia* 75, 381–389. doi: 10.1016/j.neuropsychologia.2015.06.014

Kowalska, J., and Szelag, E. (2006). The effect of congenital deafness on duration judgment. *J. Child Psychol. Psychiatry* 47, 946–953.

Kulashekhar, S., Pekkola, J., Palva, J. M., and Palva, S. (2016). The role of cortical beta oscillations in time estimation. *Hum. Brain Mapp.* 37, 3262–3281. doi: 10.1002/hbm.23239

Lewis, P. A., and Miall, R. C. (2003a). Distinct systems for automatic and cognitively controlled time measurement: evidence from neuroimaging. *Curr. Opin. Neurobiol.* 13, 250–255. doi: 10.1016/s0959-4388(03)00036-9

Lewis, P. A., and Miall, R. C. (2003b). Brain activation patterns during measurement of sub-and supra-second intervals. *Neuropsychologia* 41, 1583–1592. doi: 10.1016/S0028-3932(03)00118-0

Matsuda, S., Matsumoto, H., Furubayashi, T., Hanajima, R., Tsuji, S., Ugawa, Y., et al. (2015). The 3-Second rule in hereditary pure cerebellar ataxia: a synchronized tapping study. *PLoS One* 10:e0118592. doi: 10.1371/journal.pone.0118592

Mauk, M. D., and Buonomano, D. V. (2004). The neural basis of temporal processing. *Annu. Rev. Neurosci.* 27, 307–340. doi: 10.1146/annurev.neuro.27.070203.144247

Merchant, H., Zarco, W., and Prado, L. (2008). Do we have a common mechanism for measuring time in the hundreds of millisecond range? Evidence from multiple-interval timing tasks. *J. Neurophysiol.* 99, 939–949. doi: 10.1152/jn.01225.2007

Mills, L., and Rollman, G. (1980). Hemispheric asymmetry for auditory perception of temporal order. *Neuropsychologia* 18, 41–47. doi: 10.1016/0028-3932(80)90082-2

Montemayor, C., and Wittmann, M. (2014). The varieties of presence: hierarchical levels of temporal integration. *Timing Time Percept.* 2, 325–338. doi: 10.1163/22134468-00002030

Morillon, B., Kell, C. A., and Giraud, A. L. (2009). Three stages and four neural systems in time estimation. *J. Neurosci.* 29, 14803–14811. doi: 10.1523/JNEUROSCI.3222-09.2009

Nani, A., Manuello, J., Liloia, D., Duca, S., Costa, T., and Cauda, F. (2019). The neural correlates of time: a meta-analysis of neuroimaging studies. *J. Cogn. Neurosci.* 31, 1796–1826. doi: 10.1162/jocn_a_01459

Nowak, K., Dacewicz, A., Broczek, K., Kupisz-Urbanska, M., Galkowski, T., and Szelag, E. (2016). Temporal information processing and its relation to executive functions in elderly individuals. *Front. Psychol.* 7:1599. doi: 10.3389/fpsyg.2016.01599

Oron, A., Szymaszek, A., and Szelag, E. (2015). Temporal information processing as a basis for auditory comprehension: clinical evidence from aphasic patients. *Int. J. Lang. Commun. Disord.* 50, 604–615. doi: 10.1111/1460-6984.12160

Poeppel, D. (2003). The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time'. *Speech Commun.* 41, 245–255. doi: 10.1016/s0167-6393(02)00107-3

Pöppel, E. (1994). Temporal mechanisms in perception. *Int. Rev. Neurobiol.* 37, 185–202. doi: 10.1016/S0074-7742(08)60246-9

Pöppel, E. (1997). A hierarchical model of temporal perception. *Trends Cogn. Sci.* 1, 56–61. doi: 10.1016/S1364-6613(97)01008-5

Pöppel, E. (2004). Lost in time: a historical frame, elementary processing units and the 3-second window. *Acta Neurobiol. Exp.* 64, 295–302.

Pöppel, E. (2009). Pre-semantically defined temporal windows for cognitive processing. *Philos. Trans. R. Soc. B Biol. Sci.* 364, 1887–1896. doi: 10.1098/rstb.2009.0015

Szelag, E. (1997). "Temporal integration of the brain as studied with the metronome paradigm," in *Time, Temporality, Now*, eds H. Atmanspacher and E. Ruhnau (Berlin: Springer Verlag), 121–131.

Szelag, E. (2018). Commentary: effects of video game training on measures of selective attention and working memory in older adults: results from a randomized controlled trial. *Front. Aging Neurosci.* 9:442. doi: 10.3389/fnagi.2017.00442

Szelag, E., Dreszer, J., Lewandowska, M., and Szymaszek, A. (2008). "Cortical representation of time and timing processes," in *Neuronal Correlates Of Thinking*, eds E. Kraft, B. Guylas, and E. Pöppel (Berlin: Springer Verlag), 185–196.

Szelag, E., Dreszer, J., Lewandowska, M., Medygral, J., Osinski, G., and Szymaszek, A. (2010). "Time and cognition from the aging brain perspective. Indiviual differences," in *Personality From Biological, Cognitive And Social Perspectives*, eds T. Maruszewski, M. W. Eysenck, and M. Fajkowska (New York, NY: Eliot Werner Publications INC), 87–114.

Szelag, E., Jablonska, K., Piotrowska, M., Szymaszek, A., and Bednarek, H. (2018). Spatial and spectral auditory temporal-order judgment (TOJ) tasks in elderly people are performed using different perceptual strategies. *Front. Psychol.* 9:2557. doi: 10.3389/fpsyg.2018.02557

Szelag, E., Kanabus, M., Kolodziejczyk, I., Kowalska, J., and Szuchnik, J. (2004a). Individual differences in temporal information processing in humans. *Acta Neurobiol. Exp.* 64, 349–366.

Szelag, E., Kolodziejczyk, I., Kanabus, M., Szuchnik, J., and Senderski, A. (2004b). Deficits of nonverbal auditory perception in postlingually deaf humans using cochlear implants. *Neurosci. Lett.* 355, 49–52. doi: 10.1016/j.neulet.2003.10.025

Szelag, E., Kowalska, J., Rymarczyk, K., and Pöppel, E. (1998). Temporal integration in a subjective accentuation task as a function of child cognitive development. *Neurosci. Lett.* 257, 69–72. doi: 10.1016/s0304-3940(98)00809-x

Szelag, E., Kowalska, J., Rymarczyk, K., and Pöppel, E. (2002). Duration processing in children as determined by time reproduction: implications for a few second temporal window. *Acta Psychol. (Amst)* 110, 1–19. doi: 10.1016/S0001-6918(01)00067-1

Szelag, E., Lewandowska, M., Wolak, T., Seniow, J., Poniatowska, R., Pöppel, E., et al. (2014). Training in rapid auditory processing ameliorates auditory comprehension in aphasic patients: a randomized controlled pilot study. *J. Neurol. Sci.* 338, 77–86. doi: 10.1016/j.jns.2013.12.020

Szelag, E., Szymaszek, A., Aksamit-Ramotowska, A., Fink, M., Ulbrich, P., Wittmann, M., et al. (2011). Temporal processing as a base for language universals: cross-linguistic comparisons on sequencing abilities with some implications for language therapy. *Restor. Neurol. Neurosci.* 29, 35–45. doi: 10.3233/RNN-2011-0574

Szelag, E., Von Seinbüchel, S., Raiser, M., de Langen, E. G., and Pöppel, E. (1996). Temporal constraints in processing of nonverbal rhythmic patterns. *Acta Neurobiol. Exp.* 56, 215–225.

Szelag, E., von Steinbüchel, N., and Pöppel, E. (1997). Temporal processing disorders in patients with Broca's aphasia. *Neurosci. Lett.* 235, 33–36. doi: 10.1016/s0304-3940(97)00703-9

Szymaszek, A., Dacewicz, A., Urban, P., and Szelag, E. (2018). Training in temporal information processing ameliorates phonetic identification. *Front. Hum. Neurosci.* 12:213. doi: 10.3389/fnhum.2018.00213

Szymaszek, A., Sereda, M., Pöppel, E., and Szelag, E. (2009). Individual differences in the perception of temporal order: the effect of age and cognition. *Cogn. Neuropsychol.* 26, 135–147. doi: 10.1080/02643290802504742

Szymaszek, A., Wolak, T., and Szelag, E. (2017). The treatment based on temporal information processing reduces speech comprehension deficits in aphasic subjects. *Front. Aging Neurosci.* 9:98. doi: 10.3389/fnagi.2017.00098

Teixeira, S., Machado, S., Flávia, P., Velasques, B., Silva, J. G., Sanfim, A. L., et al. (2013). Time perception distortion in neuropsychiatric and neurological disorders. *CNS Neurol. Disord. Drug Targets* 12, 567–582. doi: 10.2174/18715273113129990080

Treutwein, B. (1997). YAAP: yet another adaptive procedure. *Spat. Vis.* 11, 129–134.

Ulbrich, P., Churan, J., Fink, M., and Wittmann, M. (2007). Temporal reproduction: further evidence for two processes. *Acta Psychol.* 125, 51–65. doi: 10.1016/j.actpsy.2006.06.004

Ulbrich, P., Churan, J., Fink, M., and Wittmann, M. (2009). Perception of temporal order: the effects of age, sex, and cognitive factors. *Neuropsychol. Dev. Cogn. B Aging Neuropsychol. Cogn.* 16, 183–202. doi: 10.1080/13825580802411758

Van Rullen, R., and Koch, C. (2003). Is perception discrete or continuous? *Trends Cogn. Sci.* 7, 207–213. doi: 10.1016/s1364-6613(03)00095-0

van Wassenhove, V., Herbst, S. K., and Kononowicz, T. W. (2019). "Timing the brain to time the mind: Critical contributions of time-resolved neuroimaging for temporal cognition," in *Magnetoencephalography: From Signals To Dynamic Cortical Networks*, eds S. Supek and C. J. Aine (Cham: Springer), 855–905. doi: 10.1007/978-3-030-00087-5_67

von Steinbüchel, N., Wittmann, M., and Szelag, E. (1999a). Temporal constraints of perceiving, generating and integrating information: clinical evidence. *Restor. Neurol. Neurosci.* 14, 167–182.

von Steinbüchel, N., Wittmann, M., Strasburger, H., and Szelag, E. (1999b). Auditory temporal-order judgement is impaired in patients with cortical lesions in posterior regions of the left hemisphere. *Neurosci. Lett.* 264, 168–171. doi: 10.1016/s0304-3940(99)00204-9

Wang, L., Lin, X., Zhou, B., Pöppel, E., and Bao, Y. (2015). Subjective present: a window of temporal integration indexed by mismatch negativity. *Cogn. Process.* 16, 131–135. doi: 10.1007/s10339-015-0687-8

Wiener, M., Parikh, A., Krakow, A., and Coslett, H. B. (2018). An intrinsic role of beta oscillations in memory for time estimation. *Sci. Rep.* 8:7992. doi: 10.1038/s41598-018-26385-6

Wittmann, M. (2009). The inner experience of time. *Philos. Trans. R. Soc. Lond B Biol. Sci.* 364, 1955–1967. doi: 10.1098/rstb.2009.0003

Wittmann, M. (2013). The inner sense of time: how the brain creates a representation of duration. *Nat. Rev. Neurosci.* 14, 217–223.

Wittmann, M., and Szelag, E. (2003). Sex differences in perception of temporal order. *Percept. Mot. Skills* 96, 105–112. doi: 10.2466/pms.2003.96.1.105

Zhang, M., Zhang, K., Zhou, X., Zhan, B., He, W., and Luo, W. (2021). Similar CNV neurodynamic patterns between sub-and supra-second time perception. *Brain Sci.* 11:1362. doi: 10.3390/brainsci11101362

Conflict of Interest: 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 Psychology

ORIGINAL RESEARCH
published: 28 January 2022
doi: 10.3389/fpsyg.2021.772867

Detection and Recognition of Asynchronous Auditory/Visual Speech: Effects of Age, Hearing Loss, and Talker Accent

Sandra Gordon-Salant*, Maya S. Schwartz, Kelsey A. Oppler and Grace H. Yeni-Komshian

Department of Hearing and Speech Sciences, University of Maryland, College Park, MD, United States

OPEN ACCESS

Edited by:
Leah Fostick,
Ariel University, Israel

Reviewed by:
Kristin Van Engen,
Washington University in St. Louis, United States
Larry E. Humes,
Indiana University Bloomington, United States

****Correspondence:***
Sandra Gordon-Salant
sgsalant@umd.edu

Specialty section:
This article was submitted to Perception Science, a section of the journal Frontiers in Psychology

Received: *08 September 2021*
Accepted: *21 December 2021*
Published: *28 January 2022*

Citation:
Gordon-Salant S, Schwartz MS, Oppler KA and Yeni-Komshian GH (2022) Detection and Recognition of Asynchronous Auditory/Visual Speech: Effects of Age, Hearing Loss, and Talker Accent. Front. Psychol. 12:772867. doi: 10.3389/fpsyg.2021.772867

This investigation examined age-related differences in auditory-visual (AV) integration as reflected on perceptual judgments of temporally misaligned AV English sentences spoken by native English and native Spanish talkers. In the detection task, it was expected that slowed auditory temporal processing of older participants, relative to younger participants, would be manifest as a shift in the range over which participants would judge asynchronous stimuli as synchronous (referred to as the "AV simultaneity window"). The older participants were also expected to exhibit greater declines in speech recognition for asynchronous AV stimuli than younger participants. Talker accent was hypothesized to influence listener performance, with older listeners exhibiting a greater narrowing of the AV simultaneity window and much poorer recognition of asynchronous AV foreign-accented speech compared to younger listeners. Participant groups included younger and older participants with normal hearing and older participants with hearing loss. Stimuli were video recordings of sentences produced by native English and native Spanish talkers. The video recordings were altered in 50 ms steps by delaying either the audio or video onset. Participants performed a detection task in which they judged whether the sentences were synchronous or asynchronous, and performed a recognition task for multiple synchronous and asynchronous conditions. Both the detection and recognition tasks were conducted at the individualized signal-to-noise ratio (SNR) corresponding to approximately 70% correct speech recognition performance for synchronous AV sentences. Older listeners with and without hearing loss generally showed wider AV simultaneity windows than younger listeners, possibly reflecting slowed auditory temporal processing in auditory lead conditions and reduced sensitivity to asynchrony in auditory lag conditions. However, older and younger listeners were affected similarly by misalignment of auditory and visual signal onsets on the speech recognition task. This suggests that older listeners are negatively impacted by temporal misalignments for speech recognition, even when they do not notice that the stimuli are asynchronous. Overall, the findings show that when listener performance is equated for simultaneous AV speech signals, age effects are apparent in detection judgments but not in recognition of asynchronous speech.

Keywords: auditory-visual speech perception, aging, hearing loss, foreign-accented speech, detection of asynchronous auditory-visual speech, recognition of asynchronous auditory-visual speech

INTRODUCTION

Everyday speech recognition tasks stimulate both audition and vision. Successful processing in both modalities requires accurate detection and resolution of auditory and visual cues at an early stage of processing, and binding of these separate streams of processed auditory and visual stimuli into a unified percept at one or more later stages of integration.]See AV integration model of Grant and Bernstein (2019), shown in **Figure 1**]. Auditory and visual features of speech stimuli are complementary to each other, and also provide some redundancy, both of which enhance a listener's understanding of the speech signal and underscore the importance of accurate integration. Additionally, auditory-visual (AV) integration for speech signals is aided at multiple stages of processing by the listener's knowledge of the language, as well as by the availability of contextual cues. Finally, the listener's cognitive abilities contribute to the process of AV integration for speech. Specifically, working memory aids prediction about the spoken message as it unfolds over time, attention enables the listener to focus on the target message and ignore irrelevant information, and processing speed assists the listener in rapidly integrating, recognizing, and responding to a spoken message. [The reader is referred to Peelle and Sommers (2015), which proposes a dynamic process of AV integration consisting of early and later integration mechanisms in auditory cortex and posterior superior temporal sulcus, based on neurophysiological evidence].

One critical property for efficient integration of multisensory information is the temporal coherence between auditory and visual stimuli, which occurs naturally when these signals derive from the same source and have the same onset (Brooks et al., 2015). For naturally occurring speech signals, the relative onset of auditory and visual signals may not be perfectly aligned in time when it is received by the listener. For example, due to differences in the transmission speed of sound and light, the auditory signal arrives later than the visual signal when the talker is more than 10 m away from the receiver (Navarra et al., 2009). Visible speech information also arrives sooner than auditory information because preparatory movements of the jaw often precede speech production (Chandrasekaran et al., 2009; Schwartz and Savariaux, 2014). However, video signals transmitted through high-fidelity transmission (e.g., video presentation via television, streaming to a monitor or real-time remote face-to-face communication), may be prone to a lag in optical cues relative to acoustic cues (e.g., Grant et al., 2004). These examples of auditory-visual asynchrony are tolerated well by young listeners with normal hearing, who detect a range of asynchronies in auditory and visual speech signals as synchronous. Specifically, young normal-hearing listeners are relatively insensitive to asynchronies between about −50 ms (auditory lead/visual lag) to +150 ms (auditory lag/visual lead), such that there is a temporal window of approximately 200 ms over which asynchronous AV stimuli are detected as simultaneous. This window is referred to as the "AV simultaneity window" (Richards et al., 2017). The AV simultaneity window is remarkably robust, and has been observed for isolated nonsense syllables as well as for sentence-length materials (Grant et al., 2004). Additionally, the range of AV asynchronies over which young, normal-hearing adults maintain the same level of speech recognition performance, referred to as the "AV speech integration window," is comparable to the 200 ms-wide AV simultaneity window, as measured with detection judgments (Grant and Seitz, 1998; Grant et al., 2004).

Advanced age may affect the efficiency of AV integration, particularly for asynchronous AV signals, because of age-related changes in auditory temporal processing. Older listeners exhibit slowed auditory temporal processing on simple measures of temporal acuity and duration discrimination (Fitzgibbons and Gordon-Salant, 1994; Snell, 1997), more complex tasks of duration discrimination in tonal sequences (Fitzgibbons and Gordon-Salant, 1994, 2001), and recognition of time-compressed speech (Gordon-Salant and Fitzgibbons, 1993). In contrast, advanced age does not appear to have a consistent effect on processing rate for visual information, with some studies reporting age-related delays on visual gap detection and temporal order judgment tasks (Humes et al., 2009; Busey et al., 2010) and others reporting a minimal effect of age on temporal processing of visual signals, depending on signal and task complexity (Brooks et al., 2015; Guest et al., 2015). Given that older listeners consistently show slowed auditory temporal processing but may not experience slowed visual processing, it might be expected that the auditory signal arrives later than the visual signal at the central integrator, resulting in a shift in the AV simultaneity window, possibly in the negative direction, during the AV synchrony/asynchrony detection task. To illustrate with a hypothetical example, an AV stimulus presented at −100 ms AV asynchrony indicates that the auditory signal is presented 100 ms before the visual signal (i.e., auditory lead) and may be perceived as out of sync by younger listeners. However, if there is slowed processing of that auditory signal by an older listener, then it may be perceived as synchronous with the visual stimulus; the simultaneous judgment at −100 ms would be seen as a shift in the AV simultaneity window in the negative direction, relative to that observed for younger listeners. It is noted that individuals with hearing impairment, either young or old, do not show deficits in auditory temporal processing beyond those attributed

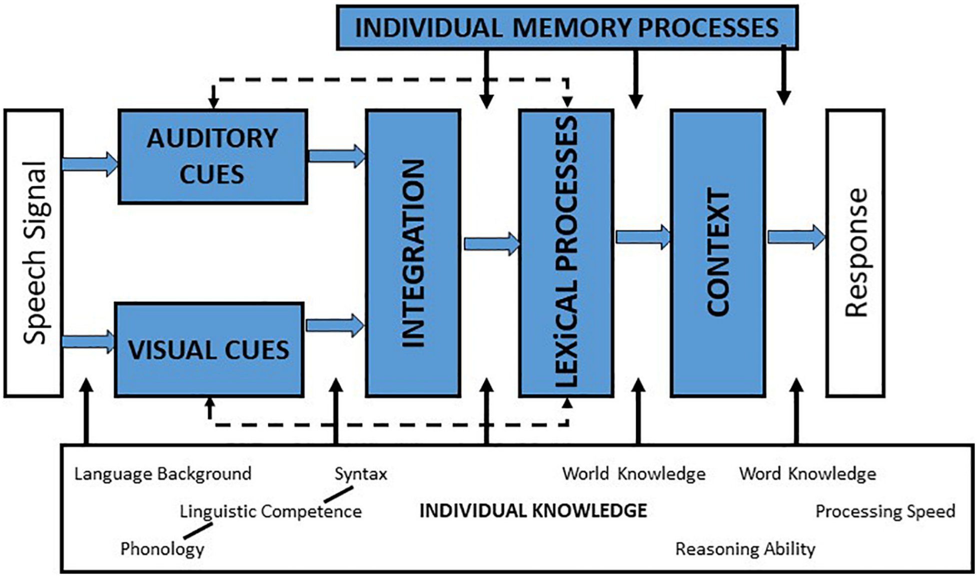


FIGURE 1 | Schematic of a simple model of auditory-visual speech processing, adapted from Grant and Bernstein (2019) (Reproduced with permission from Springer publishers via the Copyright Clearance Center).

to age (Fitzgibbons and Gordon-Salant, 1994), suggesting that individuals with hearing impairment should exhibit similar patterns of AV integration (AV simultaneity windows and AV speech integration) as individuals with normal hearing when they are matched in age.

The effects of age and/or hearing loss on the detection of AV asynchrony are somewhat mixed. Hay-McCutcheon et al. (2009) reported that older listeners with normal hearing or who used cochlear implants exhibited more negative thresholds of asynchrony in the auditory lead/visual lag conditions than middle-aged listeners, but no threshold differences in the auditory lag/visual lead conditions. In contrast, Başkent and Bazo (2011) reported comparable AV simultaneity windows by young listeners with normal hearing and older listeners with hearing loss. Neither of these previous studies compared performance on the AV asynchrony detection task between younger and older listeners who were matched for hearing sensitivity, nor between listeners with normal hearing and hearing loss who were matched in age. The present study seeks to overcome these limitations by evaluating the performance of three listener groups: young listeners with normal hearing, older listeners with normal hearing, and older listeners with hearing loss, in an effort to tease out possible effects due to age separately from those attributed to hearing loss.

The model of AV integration efficiency proposed by Grant and Bernstein (2019) incorporates cognitive abilities that influence AV speech recognition performance at multiple stages of the integration process. Because advanced age is characterized by declines in working memory (Park et al., 2002), processing speed (e.g., Salthouse, 2009; Lipnicki et al., 2017), and attentional control (Carlson et al., 1995; Milham et al., 2002), possible deterioration of AV integration by older adults may be associated with declines in cognitive abilities. For auditory-only signals, recognition of noisy, accented, or fast speech by older listeners correlates with cognitive abilities, including attention/inhibition (Janse, 2012), processing speed (Füllgrabe et al., 2015; Gordon-Salant et al., 2016), and working memory (Rönnberg et al., 2008, 2013; Gordon-Salant and Cole, 2016). For AV speech integration tasks, it may be predicted that older people will require more time to perform a higher-level task, such as recognizing misaligned auditory and visual stimuli. Thus, age-related decline in processing speed may result in poorer speech recognition performance by older than younger listeners in asynchronous conditions. It is therefore hypothesized that the AV speech integration window of older listeners will be narrower than that observed for younger listeners. This prediction is supported, in part, by previous findings that older listeners (both with and without hearing loss) demonstrated significant declines in speech recognition (relative to maximum performance) in most auditory lead/visual lag conditions, but younger listeners rarely showed a decrement in these conditions (Gordon-Salant et al., 2017). In that study, processing speed was identified as the principal cognitive factor contributing to the variance in AV speech recognition scores. Two limitations of this prior study were that the range of auditory lag/visual lead asynchronies was quite limited, and that all listeners were tested at the same fixed SNR, resulting in different levels of overall performance by the three listener groups. The current study addressed these limitations by (1) presenting a broad range of AV asynchronies from −450 ms to +450 ms; and (2) testing each listener at an individually adjusted SNR to yield 70.7% correct performance in the speech recognition task (synchronous condition).

Foreign-accented speech is ubiquitous in contemporary society and is often characterized by differences in timing information compared to native-English speech, including alterations in vowel and sentence duration (Guion et al., 2000;

Gordon-Salant et al., 2010a), lexical and suprasegmental stress patterns (Flege and Bohn, 1989; Trofimovich and Baker, 2006; Zhang et al., 2008; Gordon-Salant et al., 2015), and onsets of voicing in fricatives and affricates (Gordon-Salant et al., 2010a). In addition to these auditory-based changes with foreign-accented English, visible speech information may also be altered as a result of differences in speech production (Summers et al., 2010). There are few studies of AV integration with foreign-accented speech. At least one study has reported a reduced benefit of visual cues for recognition of foreign-accented speech relative to native English speech by younger listeners (Yi et al., 2013). In the auditory-only mode, older listeners exhibit considerable difficulty recognizing foreign-accented speech, which appears to be associated with the temporal modifications in foreign-accented English coupled with older listeners' deficits in auditory temporal processing (Gordon-Salant et al., 2010b, 2013, 2015). Thus, it is possible that the integration of auditory and visual information by older listeners is more challenging when recognizing foreign-accented speech than native English speech in conditions with auditory or visual delays, because older listeners will be less able to take advantage of visual and auditory cues that are misaligned to aid in resolving this type of speech signal. In other words, recognition of foreign-accented speech may be quite low in auditory lead and auditory lag conditions; the net effect is predicted to be a narrower AV speech integration window for foreign-accented speech than for native English. Further, it may be expected that older listeners with and without hearing loss will recognize foreign-accented speech more poorly in asynchronous AV conditions than younger listeners. Older listeners with hearing loss are expected to exhibit even narrower AV speech integration windows than older normal-hearing listeners, given the excessive difficulties of these listeners in recognizing foreign-accented speech (Gordon-Salant et al., 2013).

The overall objective of this investigation was to examine the extent to which slowed auditory temporal processing associated with advanced age is a source of altered AV integration, as assessed on tasks of detection and recognition of asynchronous AV speech. The influences of talker accent, listener hearing sensitivity, and cognitive abilities were also examined. The main experimental questions were: (1) do age and hearing sensitivity affect detection of AV asynchrony across a broad range of asynchronies? (2) Do age and hearing sensitivity affect recognition of AV asynchronous speech across a broad range of asynchronies? (3) Is there an effect of talker native language on listeners' detection and recognition of asynchronous speech? (4) Do cognitive abilities affect the speech integration window? It was predicted that older listeners with and without hearing loss would exhibit negative shifts in the AV simultaneity window (as measured on the detection task) and narrower AV speech integration windows (as measured on the speech recognition task) relative to younger listeners. It was also expected that foreign-accented speech would result in a narrowing of both the AV simultaneity window and the AV speech integration window, particularly by older listeners. Finally, it was expected that processing speed and working memory would be the most important cognitive domains associated with recognition of asynchronous AV signals, consistent with previous research (Gordon-Salant et al., 2017). The results are expected to shed light on the impact of age and hearing loss on the ability to perceive AV signals, particularly when they are misaligned in time and spoken with a foreign accent, as is now commonplace.

MATERIALS AND METHODS

Participants

Listeners were recruited primarily on the basis of age and hearing sensitivity and were assigned to one of three groups of 17 listeners per group. A power calculation was conducted to determine the sample size with 80% power, significance level of 0.05, and an effect size (Cohen's *d*) of 0.5, using mean and standard deviation data from a prior investigation of AV asynchrony (Gordon-Salant et al., 2017). The calculated sample size of 16 was increased by 1 to account for possible attrition. The young listeners with normal hearing (YNH; females = 11) were between 18 and 26 years of age (*Mean* = 20.8 years, *s.d.* = 50) and exhibited pure tone thresholds <25 dB HL (re: ANSI S3.6-2018, American National Standard Specification for Audiometers, 2018) between 250 and 4000 Hz. The older listeners with normal hearing (ONH; females = 15) fulfilled the same hearing criteria as the YNH listeners and were between 65 and 76 years of age (*Mean* = 70.1 years, *s.d.* = 0.87). The older listeners with hearing impairment (OHI; females = 3) were between 67 and 77 years (*Mean* = 72.0 years, *s.d. = 1.0*) and had a mild-to-moderate gradually sloping sensorineural hearing loss. The mean audiometric thresholds of the three listener groups are shown in **Figure 2**. Additional hearing criteria for all participants were monosyllabic word recognition scores of 80% or higher on Northwestern University Test No. 6 (Tillman and Carhart, 1966), normal tympanograms, and acoustic reflex thresholds present at levels consistent with data reported by Gelfand et al. (1990), indicative of normal hearing or a cochlear lesion (for the OHI listeners). Mean word recognition scores were 99.41, 99.29, and 94.6% for the YNH, ONH, and OHI listeners, respectively. All participants were native speakers of English and were required to pass a cognitive screening test (Montreal Cognitive Assessment, MoCA; Nasreddine et al., 2005) with a standard passing score of 26 or higher. They also were required to demonstrate normal visual acuity (20/40 or better), with or without correction.

Stimuli

The stimuli were 720 IEEE sentences (Rothauser et al., 1969). Video recordings of all 720 sentences were made by three male native speakers of English (NE) and three male native speakers of Spanish (NS) at a professional recording studio (National Foreign Language Center, University of Maryland) using green-screen technology. Details of the recording procedures are reported in Waddington et al. (2020). Multiple speakers (rather than a single speaker) were used to increase the generalizability of the results. The speakers were all graduate students at the University of Maryland and ranged in age from 28–39 years. The native speakers of English had a general American dialect. The native speakers of Spanish came from South American countries (Peru,

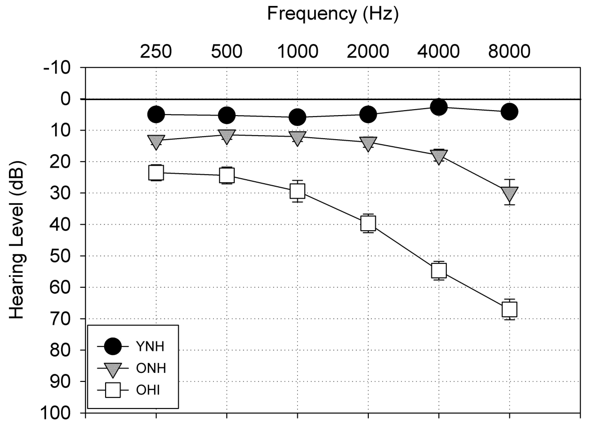


FIGURE 2 | Mean audiometric thresholds from 250 to 8000 Hz in dB Hearing Level (re: ANSI 2018) for the young normal-hearing listeners (YNH), older normal-hearing listeners (ONH), and older hearing-impaired listeners (OHI). Error bars represent one standard error of the mean.

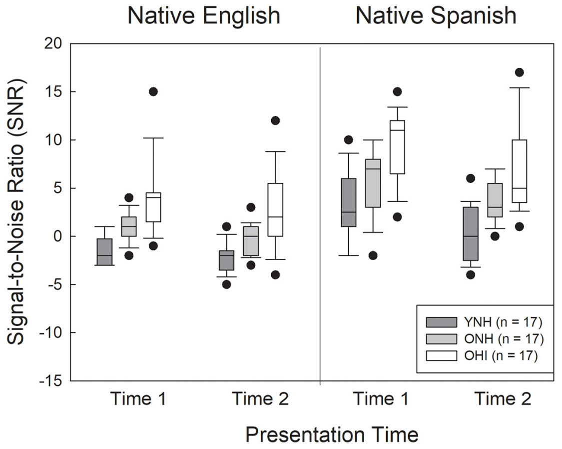


FIGURE 3 | Box plots for signal-to-noise ratios (SNRs) corresponding to approximately 70.7% correct recognition for the young normal-hearing listeners (YNH), older normal-hearing listeners (ONH), and older hearing-impaired listeners (OHI) listeners for native English and native Spanish talkers. Medians: Inside box lines. Upper and lower quartiles: top and bottom edges of the box, respectively. The endpoints of the whiskers represent the range of values without the outliers.

Argentina, and Chile) and moved to the United States after the age of 12 years. Ratings of their degree of accentedness by 10 YNH listeners indicated that they were all perceived as having a moderate Spanish accent (scores ranging from 4.95 to 5.90 on a scale of 1–9, with 1 indicating no accent and 9 indicating a heavy accent). The recordings were equated in root-mean-square (RMS) level across all speakers, and a calibration tone was created to be equivalent to this RMS level.

A six-talker babble consisting of spoken passages in English produced by three NE and three NS male talkers was used as the background noise. A description of the creation of this babble has been reported previously (Gordon-Salant et al., 2013). A calibration tone equivalent in RMS to the babble was also created.

Different lists of sentences were created for the three tasks administered in the experiment; these tasks are described in detail in the Procedures section: (a) the preliminary adaptive procedure, (b) the AV detection task, and (c) the AV recognition task. For the adaptive procedure, four sentence lists of 21 sentences each were created: two lists each of the NE talkers and two lists each of the NS talkers. These lists were used to determine the individual's signal-to-noise ratio (SNR) prior to the detection task and prior to the recognition task.

For the detection task, there were three lists spoken by the three NE talkers (19 sentences/talker × 3 talkers = 57 sentences/list), and similarly, there were three lists spoken by the three NS talkers. Each of the 19 sentences spoken by each talker on a list was presented at a unique AV asynchrony, ranging from −450 ms (auditory lead) to +450 ms (auditory lag) in 50 ms steps (i.e., 19 asynchronies).

For the AV speech recognition task, 30 sentence lists were created with 15 NE lists and 15 NS lists. Each list consisted of four sentences spoken by each NE or NS talker, for a total of 12 sentences on a list, and featured a single AV asynchrony, ranging from −300 ms to +400 ms in 50 ms steps. None of the sentences were repeated between lists.

A custom-designed AV editing software application (Scenario Designer©, v. 1.5.3, created at the University of Maryland, College Park, MD, United States) was used to import the video files, scale and position the talker on the video monitor, and insert background babble. The media files for this application included the video recordings of the 720 sentences by each of the NE and NS talkers, as well as asynchronous versions of each of these videos. In the asynchronous version, the entire visual image (V) was manipulated to occur either before or after the onset of the audio signal. The talker was positioned in the center of the monitor and scaled for a full head and shoulders shot, with a solid blue screen inserted in the background. The six-talker babble was uploaded into the Scenario Designer software and used as the audio background noise.

Procedures

Testing was conducted in a double-walled sound-attenuating booth at the University of Maryland. The Scenario Designer©software installed on a Mac computer controlled stimulus presentation and data collection. The speech and noise channels of the computer's audio output were directed to separate channels of an audiometer (Interacoustics AC40, Eden Prairie, MN, United States). The levels of the speech and noise were controlled through the audiometer, with the speech level fixed at 85 dB SPL for all testing and the noise level varied individually, as described below. Calibration tones associated with the speech and noise were used to calibrate signal levels daily (Larson Davis 824 sound level meter with 2-cm^3 coupler, Provo, UT, United States). Speech and noise signals were presented monaurally to the listener's better ear through an Etymotic insert earphone (ER-3A). The video output of the Mac computer was displayed on a television monitor (32-inch Samsung television). The listener was seated 1-m from the television screen.

Three tasks were conducted multiple times over the course of the experiment: (1) the adaptive procedure; (2) AV asynchrony detection; and (3) recognition of AV asynchronous speech. In the adaptive procedure, synchronous AV sentences spoken by either the NE or NS talkers were presented in a background of 6-talker babble to the listener. The participants were asked to repeat the sentence. A two-down, one-up adaptive rule was applied, based on keyword accuracy (3 or more of 5 words correct → correct response), in which the babble level was adjusted to yield the SNR corresponding to 70.7% correct recognition (Levitt, 1971). The initial step size was 4 dB, which was reduced to 2 dB for sentences 5–21. The SNR corresponding to 70.7% correct recognition was determined following the procedures described for the Hearing in Noise test (HINT; Nilsson et al., 1994). The adaptive procedure was presented four times over the course of the experiment: once/each prior to the AV asynchrony detection task for the NE talkers and the NS talkers, and once/each prior to the AV recognition task with NE talkers and NS talkers. For each administration, lists developed for the adaptive procedure featuring the NE talkers were used prior to the detection and recognition tasks with NE talkers, and a comparable procedure was used for the tasks featuring the NS talkers. The adaptive procedure was repeated prior to the presentation of each experimental measure (detection or recognition) to ensure that the SNR was adjusted to yield 70.7% correct performance in the synchronous condition, immediately prior to the presentation of a new experimental task.

In the AV asynchrony detection task, lists of mixed synchronous and asynchronous AV sentences spoken by either the NE or NS talkers were presented in the babble adjusted to the SNR corresponding to the individual's 70.7% correct recognition performance. After each sentence presentation, the listener was asked to respond "yes" if the auditory and visual presentation of the sentence was perceived as synchronous (in sync) and "no" if the auditory and visual presentation of the sentence was perceived as out of sync. The experimenter recorded each response. Each participant was presented with all AV asynchrony detection lists over the course of the experiment, resulting in nine judgments at each AV asynchrony for the NE talkers and nine judgments at each AV asynchrony for the NS talkers.

In the AV recognition task, 15 lists of sentences, each featuring a single AV asynchrony and spoken by either the NE or NS talkers (total of 30 conditions), were presented to listeners at the SNR corresponding to their 70.7% performance level for simultaneous AV signals. Recognition scores were derived as the percent of 60 keywords repeated correctly per list at each AV asynchrony. For each of the experimental tasks (detection and recognition), lists were blocked by talker accent and presented in randomized order across subjects.

Experimental testing was conducted over two visits, usually completed within 1 week. Each visit included the adaptive procedure, the detection task, a repeat of the adaptive procedure, and the recognition task. All tasks for the NE talkers were conducted during one visit, and all tasks for the NS talkers were conducted during the other visit, with the order of these visits randomized across subjects.

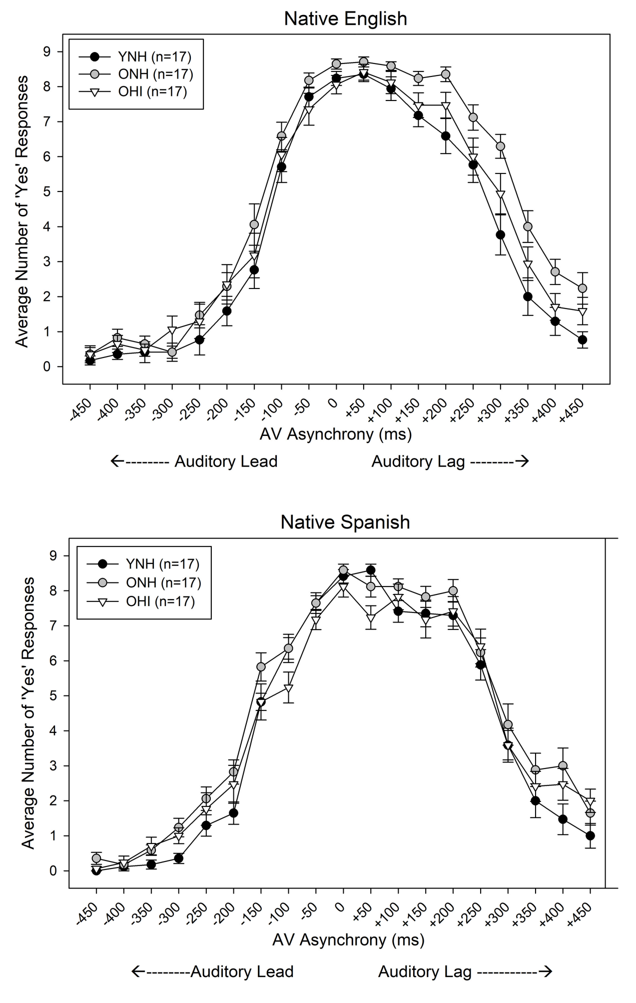


FIGURE 4 | Mean 'yes' responses (total possible = 9) for each degree of AV asynchrony, ranging from –450 ms (auditory lead) to +450 ms (auditory lag), obtained from YNH, ONH, and OHI listeners for native English talkers **(Top)** and native Spanish talkers **(Bottom)**. Error bars represent one standard error of the mean.

Listeners also were tested on a battery of cognitive measures. The cognitive measures assessed working memory [Listening SPAN (LSPAN), Daneman and Carpenter, 1980], processing speed [Digit Symbol Coding and Symbol Search from the Wechsler Adult Intelligence Scale (WAIS- III); Wechsler, 1997], inhibition (Flanker test from the NIH Toolbox, Eriksen and Eriksen, 1974), and executive function (Trail-making task, forms A and B, Reitan, 1958). The Flanker task and L-SPAN were administered via a tablet and PC, respectively, while all other cognitive measures were administered in a paper and pencil format.

The entire procedure was completed in approximately 4 h. Participants were compensated for their time in the experiment. This study involving human participants was reviewed and approved by the University of Maryland Institutional Review Board for Human Research. The participants provided their written informed consent to participate in this study.

RESULTS

Signal-to-Noise Ratio Thresholds

Initial data analysis examined the SNR values corresponding to approximately 70% correct recognition, obtained prior to the detection and recognition tasks. Box plots showing the SNR results (medians, upper and lower quartiles, upper and lower extremes) for the three listener groups in the two test administrations for both the NE and NS talkers are shown in **Figure 3**. The figure shows that the three listener groups performed differently from each other, and the SNRs were lower (better) for the NE than the NS talkers. In addition, SNR values tended to decrease from the first test administration to the second. An analysis of variance (ANOVA) was conducted on listener SNR values in a split-plot factorial design with two within-subjects factors (test time, talker accent) and one between-subjects factor (group). Results revealed significant main effects of listener group [$F(2,47) = 24.53$, $\eta_p^2 = 0.51$, $p < 0.001$], talker accent [$F(1,47) = 119.79$, $\eta_p^2 = 0.48$, $p < 0.001$], and test time [$F(1,47) = 43.50$, $\eta_p^2 = 0.72$, $p < 0.001$]. There was also a significant accent by time interaction [$F(1,47) = 5.41$, $\eta_p^2 = 0.10$, $p = 0.024$]. None of the other interactions were significant. *Post hoc* analysis (Bonferroni) of the group effect showed that the YNH listeners had lower SNRs than the ONH listeners ($p = 0.01$) and the OHI listeners ($p < 0.001$), and that the ONH listeners had lower SNRs than the OHI listeners ($p < 0.001$). It is likely that the slight differences in hearing threshold between the YNH and ONH listeners, and the substantial threshold differences between the ONH and OHI listeners, accounted for this pattern of group effects. The source of the interaction between talker accent and test time appears to be a greater change in performance from test time one to test time two for the NS talkers (mean SNR difference = 2.36, $t = 5.9$, *Cohen's* $d = 0.83$, $p < 0.001$) than for the NE talkers (mean difference in SNR = 1.26, $t = 4.3$, *Cohen's* $d = 0.61$, $p < 0.001$). Listener performance for both NE and NS talkers improved (SNRs lower) in test administration two compared to test administration one.

Auditory-Visual Asynchrony Detection

The mean AV asynchrony detection judgments of the three listener groups are shown in **Figure 4**. The data are plotted as number of "yes" responses, indicating the AV stimulus was perceived as synchronous, out of a total of nine presentations for each AV asynchrony. As expected, listeners of all three groups generally perceived stimuli in the 0 ms AV condition as synchronous, for both unaccented (NE) and accented (NS) talkers. Additionally, the mean AV simultaneity windows are asymmetric around the synchronous (0 ms AV) condition, with listeners showing greater sensitivity (perceiving asynchronies) for auditory lead/visual lag stimuli compared to auditory lag/visual lead stimuli, as reported by others (Grant et al., 2004).

The approach to data analysis for both the detection and recognition judgments was guided by the goal of comparing data for each type of judgment to data reported previously. To facilitate these comparisons, analyses were selected that would enable determination of the lead and lag conditions in which performance was significantly different from the simultaneous (0 ms) condition. Subsequently, the AV simultaneity windows for detection and the AV speech integration windows for recognition could be determined. To that end, the AV asynchrony detection judgments were analyzed with a model building approach (Hox et al., 2017) using generalized linear mixed effects regression analysis (glmer) in the lme4 package with R studio software (Bates et al., 2015). The dependent variable was the binary response (synchronous, coded as 0, and asynchronous, coded as 1) for each trial of AV stimulus presentation. Initial full model testing included all fixed factors of talker accent (dichotomous variable, coded as 0 = NE talker and 1 = NS talker), group (categorical variable, coded as 0 = YNH, 1 = ONH, 2 = OHI), and AV asynchrony conditions [(19 AV conditions, ranging from −450 ms to +450 ms, each tested dichotomously with 0 = synchronous condition (0 ms), 1 = specific negative or positive asynchronous condition], as well as all interactions between these main effects. The random effects of participant and sentence, as well as random slopes of asynchrony by participant, also were included in the model. The full model that converged was referenced to YNH listeners, NE talkers, and the 0 ms (synchronous) AV stimulus presentation. The model included the random effect of participant and significant fixed effects of AV asynchrony between −450 and −50 ms, and between +150 and +450 ms (based on the Wald ratio z-statistic in the model output, which compares the coefficient's estimated value with the standard error for the coefficient when data are normally distributed). The fixed effects of listener group and talker native language were not significant ($z > 0.05$). However, there were significant two-way interactions between talker native language and listener group at seven asynchronies (−450, −300, −250, −200, +300, +350, and +450 ms), and several three-way interactions between listener group, talker native language, and AV asynchrony. Results of the model output are shown in the **Supplementary Material** (interactions that were not significant removed to save space).

Because the variation in the AV simultaneity window for the two types of talkers for each listener group was of primary interest, the three-way interactions were explored further. To that end, subsequent general linear mixed effects analyses were conducted in which the reference listener group and talker's native language were re-leveled. A significance level of $z < 0.01$ from the model output was applied to determine which AV asynchronies were detected as significantly different from simultaneity (0 ms). This strategy permitted an assessment of the range of AV asynchronies over which detection performance was not significantly different from maximal performance (at simultaneity) separately for each listener group and talker type. **Table 1** shows the results of these analyses, including the minimum auditory lead/visual lag condition (most negative asynchrony) at which detection performance was not significantly different from simultaneity, the maximum auditory lag/visual lead condition (most positive asynchrony) at which performance also did not differ significantly from synchrony, and the difference between these two values (i.e., the AV simultaneity window). Three findings are apparent: (1) the AV simultaneity window of the YNH listeners did not differ for the NS and NE

TABLE 1 | Minimum auditory lead and maximum auditory lag asynchronies (in ms) at which detection of AV asynchrony of three listener groups was not significantly different from detection of simultaneous AV stimuli.

Group	Talker	Auditory Lead	Auditory Lag	AV Simult. Window
YNH	NE	0 ms	100 ms	100 ms
	NS	−50 ms	50 ms	100 ms
ONH	NE	−50 ms	200 ms	250 ms
	NS	0 ms	200 ms	200 ms
OHI	NE	−50 ms	200 ms	250 ms
	NS	0 ms	100 ms	100 ms

Also shown is the AV simultaneity window, in ms. YNH, young normal hearing; ONH, older normal hearing; OHI, older hearing-impaired. NE, native English talker; NS, native Spanish taker.

talkers; (2) the AV simultaneity window of the two older groups was narrower for the NS talkers than for the NE talkers; and (3) for the NE talker, the width of the AV simultaneity window was wider for the two older groups than for the younger group.

Auditory-Visual Recognition

Recognition scores for the three listener groups in the 15 synchrony/asynchrony conditions for the NE and NS talkers are shown in **Figure 5**. The SNR adaptive procedure was successful in equating the three listener groups in the 0 ms AV (synchronous) condition at approximately 70% correct level of performance, as confirmed by one-way ANOVAs indicating no significant performance differences between the three groups for either the NE talkers [$F(2,50) = 0.169$, $p = 0.845$] or the NS talkers [$F(2,50) = 0.773$, $p = 0.467$]. Mean speech recognition scores in the synchronous condition were between 66 and 70% across groups and talkers, indicating a close approximation to the target 70.7% recognition score.

A statistical model was fit to the sentence recognition data using the glmer analysis, following the model building approach described above (Hox et al., 2017). The first iteration of model building included random effects of participant and sentence, and random slopes of asynchrony by participant (to examine variation in participants by level of AV asynchrony), as well as fixed effects of listener group, talker native language and degree of asynchrony and all interactions between these effects. The dependent variable was the trial-by-trial number of keywords correct out of five possible, for each sentence presented. Initial model testing included all fixed factors of interest: talker accent (dichotomous variable, coded as 0 = NE talker and 1 = NS talker), group (categorical variable, coded as 0 = YNH, 1 = ONH, 2 = OHI), and AV asynchrony conditions [15 AV conditions, ranging from −300 ms to +400 ms; each tested dichotomously with 0 = synchronous condition (0 ms), 1 = specific negative or positive asynchronous condition], as well as their interactions. The referent talker accent was NE, referent group was YNH, and referent AV asynchrony condition was 0 ms. Model testing proceeded with iteratively removing the highest-order fixed effects and interactions that were not significant (z-statistic > 0.05 in the model output), and re-running the model. Improvement in model fit between the full model and subsequent models was assessed with the ANOVA test.

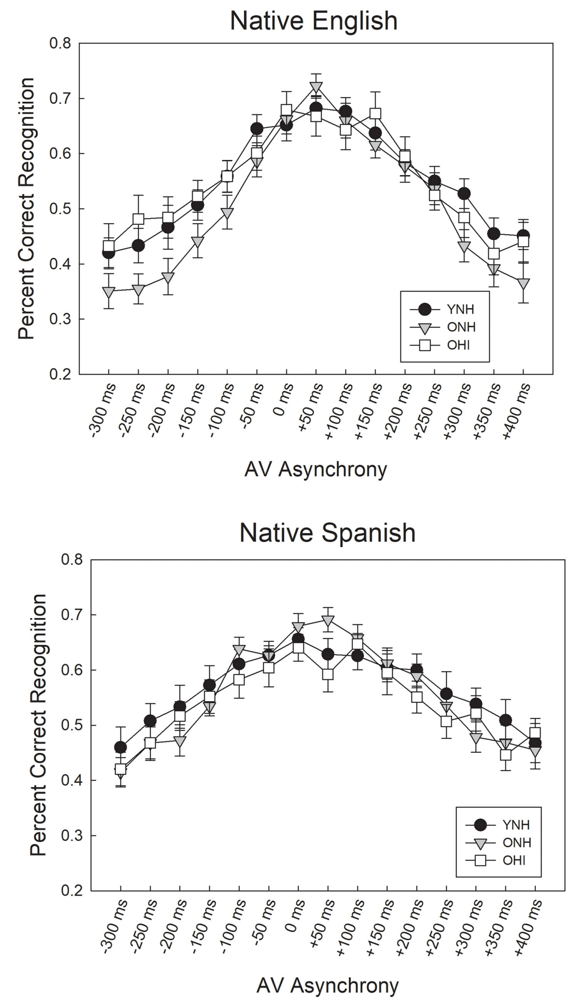


FIGURE 5 | Mean speech recognition scores for asynchronous AV IEEE sentences, ranging from –300 ms (auditory lead) to +400 ms (auditory lag), obtained from YNH, ONH, and OHI listeners for native English talkers **(Top)** and native Spanish talkers **(Bottom)**. Error bars represent one standard error of the mean.

The best-fitting model derived from these fixed and random effects, shown in **Table 2**, included the random effects of participant and sentence item, and the fixed effect of asynchrony condition. The fixed effects of listener group and talker native language and all interactions were not significant and subsequently were removed from the final model. With reference to the 0 ms (synchronous) condition, each fixed asynchrony condition was significantly different ($z < 0.001$), with the exception of the +50 and +100 ms AV asynchronies, as shown in the table. That is, performance in each negative AV asynchrony condition (−50 ms through −300 ms) was significantly different from the synchronous condition (0 ms), and performance for the positive AV asynchronies between +150 ms through +400 ms was also significantly different from the synchronous condition. Overall, the results suggest that the AV speech integration window for sentence recognition, based on the range between the

TABLE 2 | Final model of YNH, ONH, and OHI listener speech recognition performance.

	Coefficient	*SE*	*z*	*p*
Intercept	0.84	0.17	5.06	<0.001
AV asynchrony − 300	−1.71	0.21	−8.03	<0.001
AV asynchrony − 250	−1.51	0.21	−7.20	<0.001
AV asynchrony − 200	−1.44	0.21	−6.91	<0.001
AV asynchrony − 150	−1.17	0.21	−5.68	<0.001
AV asynchrony − 100	−0.99	0.20	−4.82	<0.001
AV asynchrony − 50	−0.46	0.21	−2.23	<0.001
AV asynchrony + 50	0.41	0.22	1.84	>0.05
AV asynchrony + 100	−0.12	0.21	−0.55	>0.05
AV asynchrony + 150	−0.42	0.21	−2.03	<0.05
AV asynchrony + 200	−0.42	0.21	−2.03	<0.05
AV asynchrony + 250	−0.66	0.20	−3.24	<0.01
AV asynchrony + 300	−1.23	0.21	−5.96	<0.001
AV asynchrony + 350	−1.40	0.21	−6.71	<0.001
AV asynchrony + 400	−1.51	0.21	−7.19	<0.001
Talker NS × Asynch − 100	0.84	0.29	2.86	<0.01
Talker NS × Asynch − 100 × YNH	−0.89	0.42	−2.13	<0.05

minimum negative and minimum positive asynchronies where performance was not significantly different from 0 ms, was between 0 ms to +100 ms, or 100 ms wide, and was similar for all three listener groups and for both NE and NS talkers.

Predictors of Recognition Performance

Mean scores (and standard errors) on the six cognitive measures for the three listener groups are shown in **Table 3**. ANOVAs were conducted separately for each of these measures and revealed a significant effect of listener group on each test [Digit Symbol: $F(2) = 37.05$, $\eta_p^2 = 0.61$, $p < 0.001$; Symbol Search: $F(2) = 22.13$, $\eta_p^2 = 0.48$, $p < 0.001$; LSPAN: $F(2) = 14.95$, $\eta_p^2 = 0.384$; Trail Making A: $F(2) = 10.97$, $\eta_p^2 = 0.314$, $p < 0.001$; Trail Making B: $F(2) = 11.78$, $\eta_p^2 = 0.329$, $p < 0.001$; Flanker (uncorrected): $F(2) = 28.59$, $\eta_p^2 = 0.55$, $p < 0.001$]. *Post hoc* multiple comparison tests using the Bonferroni correction revealed that the YNH listeners had significantly higher scores than the two older listener groups on the Digit Symbol, Symbol Search, LSPAN, and Flanker tests, and significantly lower scores than the two older groups on the Trail Making A and B tests. However, there were no differences in the performance between the two older groups on any measure ($p > 0.05$, each measure).

TABLE 3 | Mean scores (and standard deviations) of the three listener groups on the six cognitive measures.

	YNH		ONH		OHI	
	Mean	*s.d.*	Mean	*s.d.*	Mean	*s.d.*
Digit symbol	90.11	11.90	62.48	13.10	55.35	12.31
Symbol search	41.94	7.09	28.29	6.91	28.23	6.04
LSPAN	4.21	1.25	2.79	0.81	2.68	0.50
Trail making A	18.77	5.02	26.32	7.54	29.24	7.33
Trail making B	35.90	9.60	60.01	19.49	71.92	31.39
Flanker	112.81	5.18	97.35	7.91	98.59	6.01

TABLE 4 | Results of linear multiple regression analyses with three predictor variables retrieved (HF-PTA, Trail Making A, and LSPAN).

Variables retrieved	**SNR condition**			
	NE speech		**NS speech**	
	Time 1	**Time 2**	**Time 1**	**Time 2**
HF-PTA (1)	0.559	0.546	0.697	0.757
Trail making A (2)	0.635	–		
LSPAN (2)			0.741	0.791

Cumulative variance (r^2) accounted for by significant predictor variables, in the order retrieved by stepwise multiple linear regression [first (1), second (2)], is shown for SNRs measured at two test intervals for native English (NE) and native Spanish (NS) talkers. Criteria for significance of each retrieved variable in the table is $p < 0.05$, and the significance of each regression model associated with each retrieved variable is $p < 0.001$.

The best-fitting model for the asynchronous AV sentence recognition scores, described above, was next probed to determine which predictor variables of cognition and hearing sensitivity improved the model fit (based on the ANOVA test). Model testing proceeded from the reduced model described above to subsequently include, in separate iterations, each of the predictor variables (all continuous variables): working memory (L-SPAN), speed of processing (Digit Symbol Coding, Symbol Search), attention/inhibition (Flanker score), executive function (Trail Making A and B), pure-tone hearing thresholds [quantified as pure-tone average of thresholds at 0.5, 1, and 2 kHz (the PTA), and as high-frequency pure-tone average of thresholds at 1 k, 2 k, and 4 kHz (the HF-PTA)]. Scores were converted to *z*-scores prior to entering each variable sequentially into the model. None of the predictor variables improved model fit.

Finally, an analysis was conducted to determine the cognitive and hearing sensitivity variables that best predicted the SNRs at which listeners achieved 70% correct performance for NE and NS speech in the simultaneous AV condition. Because significant differences in these SNR values were observed between the first and second administrations, separate analyses were conducted for each of the four dependent measures (NE and NS speech, time 1 and time 2). Linear multiple regression analyses were conducted using a reduced set of predictor variables to minimize the effects of multicollinearity. The predictor variables were the Digit Symbol Coding, LSPAN, Flanker-uncorrected, and Trail Making A tests, and HF-PTA. Results of the linear regression analyses with the step-wise method are shown in **Table 4**, and revealed that for each SNR measure, the predictor variable that accounted for the most variance was HF-PTA. The Trail Making A test of executive function accounted for additional variance in recognition of NE speech in the first administration, and the LSPAN test of working memory accounted for additional variance for recognition of NS speech in both the first and second test administrations.

DISCUSSION

This study evaluated whether older listeners with and without hearing loss exhibit different patterns of AV integration for asynchronous auditory-visual sentences compared to younger listeners with normal hearing, and whether such patterns were influenced by talker accent and task. Results generally showed that detection of AV asynchrony varied between younger and older listeners, with different patterns observed for talkers of different native language backgrounds. However, for the speech recognition task, all three listener groups showed comparable effects of AV asynchrony for both NE and NS talkers. These findings and their interpretation are explained in more detail below.

Auditory-Visual Asynchrony Detection

Younger and older listeners showed different AV simultaneity windows across the range of AV asynchronies assessed, and these patterns varied with talker native language. For the native English talker, the minimum auditory lead condition perceived as synchronous was more negative for the ONH and OHI listeners compared to the YNH listeners. The negative shift, on average, was 50 ms for both the ONH and OHI listeners, consistent with the hypothesis that slowed auditory temporal processing, but not visual processing, by older listeners may have delayed perception of the auditory stimulus. That is, an auditory stimulus presented prior to a visual stimulus was perceived as more closely aligned in time to the visual stimulus by older listeners than younger listeners, imposing an overall negative shift in the AV simultaneity window. The findings also showed that the derived AV simultaneity window for NE talkers was at least twice as wide for older listeners than younger listeners. It appears that older listeners were less sensitive to AV asynchronies in the auditory lag/visual lead conditions than the younger listeners, which contributed to the differences in the width of the AV simultaneity window. The negative shift in the asynchronous AV stimuli in auditory lead conditions that are perceived as simultaneous by older listeners is consistent with findings reported by Hay-McCutcheon et al. (2009). While both the current study and that of Hay-McCutcheon et al. (2009) showed a wider range of AV asynchronies as judged as simultaneous for older listeners than younger listeners, the source of the increased width was different in the two studies. Hay-McCutcheon et al. (2009) attributed the increased range to more negative auditory lead thresholds exclusively, whereas the source of the wider windows of older listeners in the current study is associated with both more negative auditory lead asynchronies and more positive auditory lag/visual lead asynchronies yielding comparable detection performance to that observed for simultaneous AV stimuli. The increase in the positive side of the AV simultaneity window among older listeners compared to younger listeners may reflect less sensitivity to visually leading AV stimuli. This decreased sensitivity for visual leading asynchronous AV speech may serve older listeners well for situations where there is a delay in electronic transmission of auditory signals relative to visual signals, such as with hearing aids or with internet communication. The present results, however, are in contrast to findings of Başkent and Bazo (2011), who reported no differences in the AV simultaneity window between younger listeners with normal hearing and older listeners with hearing loss. One possible source of this discrepancy may be the method used to determine the range of AV asynchronies that are judged to be comparable to maximum performance (observed for simultaneous AV stimuli). Başkent and Bazo (2011) measured the AV simultaneity window encompassed by the 50% threshold points for auditory-leading and visual-leading stimuli, whereas the current study measured the window encompassed by AV asynchronies that were judged as not significantly different from simultaneous AV stimuli. Thus, different methods for calculating the AV simultaneity window may yield discrepant findings; the current method was chosen to facilitate comparison between detection and recognition judgments.

The AV simultaneity windows were narrower for sentences spoken by NS talkers compared to those spoken by NE talkers for the older listeners, but not for the younger listeners. This difference was attributed to a change in the asynchrony detection threshold for auditory leading signals for both older listener groups, which shifted in a more positive direction for NS talkers relative to NE talkers, suggesting that older listeners became more sensitive (i.e., perceived asynchrony) in auditory lead conditions for the more challenging NS talkers. Additionally, the OHI listeners' judgments for visual leading signals (positive AV asynchronies) shifted in a more negative direction, indicating that these listeners also were highly sensitive to asynchronies for NS talkers in visual lead conditions. While the source of the narrower AV simultaneity windows by the older listeners is not known, one possible explanation is that these listeners were so challenged by the Spanish-accented speech that they paid more attention to the visual stimuli to recognize the sentence, and as a result, disparities in the relative onset of auditory and visual information became more obvious. It should be noted that the AV simultaneity window for the native English talkers was somewhat narrower than the typical 200 ms integration window reported in other studies for younger listeners (e.g., Grant et al., 2004). Differences in method across the different studies likely accounted for the variation in window size, including the use of an adaptive procedure to equate listener performance prior to measuring the detection thresholds for asynchronous stimuli in the current study, as well as the use of multiple NE and NS talkers and a babble background.

Auditory-Visual Recognition

Recognition of synchronous and asynchronous AV sentences was examined to determine whether or not listeners' recognition performance is affected by asynchronous presentation of AV stimuli, and whether possible differences in AV integration between younger and older listeners impact performance on this task The results generally show that all listener groups exhibited significant declines in recognition performance for asynchronous presentation of AV sentences, when recognition performance was equated for synchronous speech. However, contrary to expectation, statistical modeling of the sentence recognition scores failed to reveal effects of listener age or hearing loss. It was expected that older listeners would show significantly greater declines in performance in the auditory lead conditions compared to younger listeners, as was shown in a previous study (Gordon-Salant et al., 2017). In auditory lead conditions, the lips are clearly misaligned with the talker's voice, requiring listeners to inhibit the distracting effect of poor bi-sensory signal alignment. It was expected that older listeners, who often have a compromised ability to inhibit irrelevant or distracting stimuli (e.g., Hasher and Zacks, 1988; Alain and Woods, 1999; Presacco et al., 2016) would be more impacted by such temporal onset misalignments that are perpetuated through the duration of the sentence. In a previous experiment (Gordon-Salant et al., 2017), the younger listeners performed near ceiling for the synchronous sentence stimuli and maintained a high level of performance for all asynchronies, indicating that they were minimally impacted by the temporal misalignments. However, the ONH and OHI listeners' recognition performance for synchronous AV speech was considerably poorer than that of the YNH listeners, and these two older groups showed significant declines in recognition in auditory lead conditions. It appears, then, that the current technique of equating listener performance to the same level in the 0 ms AV synchronous condition is critically important for evaluating the extent to which age and hearing loss, *per sé*, affect the AV speech integration window. This is reinforced by the observation that performance of YNH, ONH, and OHI listeners at the same fixed SNR will be inherently different, with one group or another performing at or near the ceiling or floor, making it difficult to observe the differential impact of the asynchronous distortions on recognition performance by the different listener groups. The finding that older listeners did not exhibit significantly greater declines in speech recognition than younger listeners in auditory lead conditions in the current study may also reflect the effects of slowed auditory processing among older listeners. That is, if processing of the auditory information in auditory lead conditions is delayed among older listeners, then the auditory signal may appear more synchronous with the visual signal, and relatively high recognition performance is maintained. Further investigation of this possible mechanism is warranted, using more discrete steps of the asynchronous stimulus presentation.

The statistical model of asynchronous AV sentence recognition performance also failed to show an effect of talker native language. Thus, listeners showed the same pattern of decline in speech recognition scores with AV asynchronies for both the NE and NS talkers. This finding was also contrary to expectation, as recognition of the asynchronous sentences spoken by NS talkers was expected to be extremely challenging, especially for older listeners. The method of equating performance for simultaneous AV stimuli separately for the NS and NE talkers likely reduced the expected performance declines for asynchronous NS sentences. The final model revealed that for both talker accents, recognition performance in the auditory lead conditions between −50 and −300 ms was significantly different from performance in the synchronous condition. Similarly, recognition performance in the visual lead/auditory lag conditions between +150 and +400 ms was significantly different from the synchronous condition. Based on these results, the AV asynchronies over which listener performance was comparable to that observed in the simultaneous AV condition (i.e., the AV speech integration window) was 100 ms wide, for speech produced by both NE and NS talkers. This window is comparable to the AV simultaneity window identified in the AV detection task for NE and NS talkers for young normal-hearing listeners. However, for older listeners, the AV integration window observed for recognition judgments was narrower than the AV simultaneity window on the detection task, particularly for the NE talker. Taken together, these results suggest that even though older listeners may be relatively insensitive in detecting asynchronies in AV speech stimuli, the same stimuli have a deleterious impact on recognition performance. In contrast, younger listeners appear to have difficulty accurately recognizing asynchronous AV sentences at the same asynchronies where they detect the presence of asynchrony.

These findings have implications for everyday communication. For example, there are many face-to-face interactions in daily life where the auditory and visual speech information may be misaligned in time, including internet communication (i.e., Zoom meetings), television programming, excessive distance between talker and receiver, or electronic amplification of the talker's speech with additional signal processing (see Gordon-Salant et al., 2017 for a review). The current findings suggest that all listeners, regardless of age and hearing loss, may have considerable difficulty accurately recognizing such asynchronous signals. Although the older listeners did not exhibit significantly poorer recognition performance or different speech integration windows than the younger listeners, these findings may not reflect age-related performance patterns in everyday listening situations, where the SNR is not individually adapted, but rather is similar for all listeners, depending on their location in the auditory scene.

Cognitive measures and hearing sensitivity were not significant predictor variables for recognition of asynchronous AV sentences. Two possibilities may account for these findings. The first is that the specific measures used to quantify cognitive ability were not sufficiently sensitive to identify individual variation. The second is that there was not sufficient variation in speech recognition performance among the listeners in the asynchronous AV conditions, because listener performance was equated in the synchronous condition. A related issue is the method of setting SNR prior to testing perception of asynchronous speech. A previous study of age-related differences in recognition of asynchronous AV stimuli (Gordon-Salant et al., 2017) presented asynchronous AV stimuli at the same SNR to all listeners, with some groups performing near ceiling and other groups performing near floor for particular stimuli. In that study, the cognitive measure of speed of cognitive processing contributed to variation in recognition performance in asynchronous conditions. Thus, the results are very different when stimuli are presented at the same fixed SNR to all participants vs. when stimuli are presented at an individually adjusted SNR to equate performance level across participants. Comparing the present findings with those reported previously (Gordon-Salant et al., 2017) tentatively suggests that the method of setting SNR, and the resulting level of recognition performance for synchronous AV signals, are important factors in determining the impact of cognitive abilities on recognition of asynchronous AV speech signals presented in noise. That is, adjusting the SNR on an individual basis may have provided compensation for the effects of cognitive decline or hearing sensitivity, or both. It appears that these predictors may be relevant when the SNR is fixed, as in many everyday listening situations.

Signal-to-Noise Ratio Thresholds

The SNRs corresponding to 70.7% correct recognition performance were significantly better during the second administration compared to the first administration for both NE and NS talkers. The adaptive measure was conducted twice with the same talkers (NE or NS) on each test day in order to equate performance immediately prior to the administration of the detection task and the recognition task. The improvement in SNR threshold in the second administration on the same test day may reflect, in part, a simple effect of learning the task (as sentences were not repeated). However, performance improved significantly more for the NS talkers than the NE talkers (effect size was strong for NS talkers and moderate for NE talkers), and may be one manifestation of rapid adaptation to foreign-accented speech reported previously (Clarke and Garrett, 2004; Bradlow and Bent, 2008; Gordon-Salant et al., 2010c; Bieber and Gordon-Salant, 2021). The effect of listener group did not interact with test time nor talker accent, indicating that both groups showed the same improvement with the second administration of the adaptive procedure. The rapid improvement in recognition of foreign-accented speech by younger and older listeners is consistent with previous findings (Gordon-Salant et al., 2010c; Bieber and Gordon-Salant, 2017) and is underscored here as a variable to control in future studies that employ multiple presentations of foreign-accented speech. Listener high-frequency pure-tone average accounted for the most variance in SNR scores across the two test administrations and for NE and NS speech. These results are highly consistent with prior findings of the importance of hearing sensitivity for recognition of speech in noise (e.g., Humes and Dubno, 2010), as well as the importance of working memory for recognition of degraded speech (Rönnberg et al., 2013), when the speech signals are presented in the auditory mode. However, the current findings extend these principles to recognition of speech presented in the AV mode and to recognition of foreign-accented speech.

Summary and Conclusion

This experiment examined integration of asynchronous AV native English and foreign-accented sentences by younger and older listeners, as manifested on detection and recognition tasks. Compared to younger listeners, older listeners are less sensitive to auditory lead asynchronies and perceive wider ranges of AV asynchronous sentences as synchronous, especially for NE talkers. These findings reflect possible age-related slowing of auditory speech streams in auditory lead/visual lag conditions and reduced sensitivity to asynchrony in auditory lag/visual lead conditions. In contrast, younger and older listeners with normal hearing and older listeners with hearing loss showed comparable patterns of AV speech integration, indicating that listener age and hearing loss did not impact recognition of asynchronous AV sentences. Although the AV simultaneity window determined from detection judgments was wider for NE talkers than NS talkers by older listeners, there were no differences in recognition performance for NE and NS talkers across a broad range of AV asynchronies. Overall, these findings suggest that unlike younger listeners, older listeners' speech recognition may be negatively impacted by asynchronous AV speech stimuli that they judged as synchronous.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are publicly available. The data can be found here: http://hdl.handle.net/1903/28269.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by University of Maryland Institutional Review Board. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

SG-S designed the experiments, oversaw data collection, analyzed data, and wrote the manuscript. MS assisted with stimulus creation, implemented and performed the experiments, and conducted some of the data analyses. KO collected and analyzed some of the data. GY-K was involved in designing the experiments and manuscript preparation. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the National Institute on Aging of the NIH (Grant # R01AG009191) awarded to SG-S.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the contributions of Bryan Anderson and Mike Buckhout-White for creating the recordings used in the experiments, and to Bryan Anderson for developing the Scenario Designer© software.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.772867/full#supplementary-material

REFERENCES

Alain, C., and Woods, D. L. (1999). Age-related changes in processing auditory stimuli during visual attention: evidence for deficits in inhibitory control and sensory memory. *Psychol. Aging* 14, 507–519. doi: 10.1037//0882-7974.14.3.507

ANSI S3.6-2018, American National Standard Specification for Audiometers (2018). *(Revision of ANSI S3.6-1996, 2004, 2010)*. New York: American National Standards Institute.

Başkent, D., and Bazo, D. (2011). Audiovisual asynchrony detection and speech intelligibility in noise with moderate to severe sensorineural hearing impairment. *Ear. Hear.* 32, 582–592. doi: 10.1097/AUD.0b013e31820fca23

Bates, D., Maechler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Statist. Softw.* 67, 1–48. doi: 10.18637/jssv067.i01

Bieber, R. E., and Gordon-Salant, S. (2017). Adaptation to novel foreign-accented speech and retention of benefit following training: influence of aging and hearing loss. *J. Acoust. Soc. Am.* 141, 2800–2811. doi: 10.1121/1.4980063

Bieber, S. R., and Gordon-Salant, S. (2021). Improving older adults' understanding of challenging speech: auditory training, rapid adaptation and perceptual learning. *Hear. Res.* 402, 1–16. doi: 10.1016/j.heares.2020.108054

Bradlow, A. R., and Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition* 106, 709–729. doi: 10.1016/j.cognition.2007.04.005

Brooks, C. J., Anderson, A. J., Roach, N. W., McGraw, P. V., and McKendrick, A. M. (2015). Age-related changes in auditory and visual interactions in temporal rate perception. *J. Vis.* 15, 1–13. doi: 10.1167/15.16.2

Busey, R., Craig, J., Clark, C., and Humes, L. (2010). Age-related changes in visual temporal order judgment performance: relation to sensory and cognitive capacities. *Vis. Res.* 50, 1628–1640. doi: 10.1016/j.visres.2010.05.003

Carlson, M. C., Hasher, L., Zacks, R. T., and Connelly, S. L. (1995). Aging, distraction, and the benefits of predictable location. *Psychol. Aging* 10, 427–436.

Chandrasekaran, C., Trubanova, A., Stillittano, S., Caplier, A., and Ghazanfar, A. A. (2009). The natural statistics of audiovisual speech. *PLoS Comput. Biol.* 5:e1000436. doi: 10.1371/journal.pcbi.1000436

Clarke, C. M., and Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *J Acoust. Soc. Am.* 116, 3647–3658. doi: 10.1121/1.1815131

Daneman, M., and Carpenter, P. (1980). Individual differences with working memory and reading. *J. Verb. Learn. Verb. Behav.* 19, 450–466. doi: 10.1016/S0011-5371(80)90312-6

Eriksen, B. A., and Eriksen, C. W. (1974). Effects of noise letters upon identification of a target letter in a non-search task. *Percept. Psychophys.* 16, 143–140. doi: 10.3758/BF03203267

Fitzgibbons, P., and Gordon-Salant, S. (1994). Age effects on measures of auditory temporal sensitivity. *J. Speech Hear. Res.* 37, 662–670. doi: 10.1044/jshr.3703.662

Fitzgibbons, P., and Gordon-Salant, S. (2001). Aging and temporal discrimination in auditory sequences. *J. Acoust. Soc. Am.* 109, 2955–2963. doi: 10.1121/1.1371760

Flege, J. E., and Bohn, O.-S. (1989). An instrumental study of vowel reduction and stress placement in Spanish-accented English. *Stud. Sec. Lang. Acquis.* 11, 35–62. doi: 10.1017/S0272267100007828

Füllgrabe, C., Moore, B. C. J., and Stone, M. A. (2015). Age-group differences in speech identification despite matched audiometrically normal hearing: contributions from auditory temporal processing and cognition. *Front. Aging Neurosci.* 6:347. doi: 10.3389/fnagi.2014.00347

Gelfand, S. A., Schwander, T., and Silman, S. (1990). Acoustic reflex thresholds in normal and cochlear-impaired ears: effects of no-response rates on 90th percentiles in a large sample. *J. Speech Hear. Disor.* 55, 198–205. doi: 10.1044/jshd.5502.198

Gordon-Salant, S., and Cole, S. S. (2016). Effects of age and working memory capacity on speech recognition performance in noise among listeners with normal hearing. *Ear. Hear.* 37, 592–602. doi: 10.1097/AUD.0000000000000316

Gordon-Salant, S., and Fitzgibbons, P. (1993). Temporal factors and speech recognition performance in young and elderly listeners. *J. Speech Hear. Res.* 36, 1276–1285. doi: 10.1044/jshr.3606.1276

Gordon-Salant, S., Yeni-Komshian, G., Fitzgibbons, P., and Cohen, J. I. (2015). Effects of age and hearing loss on recognition of unaccented and accented multisyllabic words. *J. Acoust. Soc. Am.* 137, 884–897. doi: 10.1121/1.4906270

Gordon-Salant, S., Yeni-Komshian, G., Fitzgibbons, P., Cohen, J. I., and Waldroup, C. (2013). Recognition of accented and unaccented speech in different maskers by younger and older listeners. *J. Acoust. Soc. Am.* 134, 618–627. doi: 10.1121/1.4807817

Gordon-Salant, S., Yeni-Komshian, G., Pickett, E., and Fitzgibbons, P. J. (2016). Perception of contrastive bi-syllabic lexical stress in unaccented and accented words by younger and older listeners. *J. Acoust. Soc. Am.* 139, 1132–1148. doi: 10.1121/1.4943557

Gordon-Salant, S., Yeni-Komshian, G. H., and Fitzgibbons, P. J. (2010a). Recognition of accented English in quiet by younger normal-hearing listeners and older listeners with normal hearing and hearing loss. *J. Acoust. Soc. Am.* 128, 444–455. doi: 10.1121/1.3397409

Gordon-Salant, S., Yeni-Komshian, G. H., and Fitzgibbons, P. J. (2010b). Perception of accented English in quiet and noise by younger and older listeners. *J. Acoust. Soc. Am.* 128, 3152–3160. doi: 10.1121/1.3495940

Gordon-Salant, S., Yeni-Komshian, G. H., Fitzgibbons, P. J., and Schurman, J. (2010c). Short-term adaptation to accented English by younger and older listeners. *J. Acoust. Soc. Am.* 128, EL200–EL204. doi: 10.1121/1.3486199

Gordon-Salant, S., Yeni-Komshian, G. H., Fitzgibbons, P. J., Willison, H. M., and Freund, M. S. (2017). Recognition of asynchronous auditory-visual speech by younger and older listeners: a preliminary study. *J. Acoust. Soc. Am.* 142, 151–159. doi: 10.1121/1.4992026

Grant, K. W., and Bernstein, J. G. W. (2019). "Toward a Model of Auditory-Visual Speech Intelligibility" in *Multisensory Processes. Springer Handbook of Auditory Research*. eds A. Lee, M. Wallace, A. Coffin, A. Popper, and R. Fay (Germany: Springer). 33–58. doi: 10.1007/978-3-030-10461-0_3

Grant, K. W., Greenberg, S., Poeppel, D., and van Wassenhove, V. (2004). Effects of spectro-temporal asynchrony in auditory and auditory-visual speech processing. *Semin. Hear.* 25, 241–255.

Grant, K. W., and Seitz, P. F. (1998). Measures of auditory-visual integration in nonsense syllables and sentences. *J. Acoust. Soc. Am.* 104, 2438–2450. doi: 10.1121/1.423751

Guest, D., Howard, C. J., Brown, L. A., and Gleeson, H. (2015). Aging and the rate of visual information processing. *J. Vis.* 15, 1–25. doi: 10.1167/15.14.10

Guion, S. G., Flege, J. E., Liu, S. H., and Yeni-Komshian, G. Y. (2000). Age of learning effects on the duration of sentences produced in a second language. *Appl. Psychol.* 21, 205–228. doi: 10.1017/SO142716400002034

Hasher, L., and Zacks, R. T. (1988). "Working memory, comprehension, and aging: a review and a new view" in *The Psychology of Learning and Motivation: advances in Research and Theory*. ed. G. H. Bower (Cambridge: Academic Press). 193–225.

Hay-McCutcheon, M. J., Pisoni, D. B., and Hunt, K. K. (2009). Audiovisual asynchrony detection and speech perception in hearing-impaired listeners with cochlear implants: a preliminary analysis. *Int. J. Audiol.* 48, 321–333.

Hox, J. J., Moerbeek, M., and Van de Schoot, R. (2017). *Multilevel Analysis: techniques and Applications*, 2nd Edn. New York: Routledge/Taylor & Francis Group.

Humes, L. E., Busey, T. A., Craig, J. C., and Kewley-Port, D. (2009). The effects of age on sensory thresholds and temporal gap detection in hearing, vision, and touch. *Atten. Percept. Psychophys.* 71, 860–871. doi: 10.3758/APP.71.4.860

Humes, L. E., and Dubno, J. R. (2010). "Factors affecting speech understanding in older adults," in *The Aging Auditory System*, eds. S. Gordon-Salant, R. Frisina, A. Popper and R. Fay (New York, NY: Springer), 211–257.

Janse, E. (2012). A non-auditory measure of interference predicts distraction by competing speech in older adults. *Neuropsychol. Dev. Cogn. B Aging Neuropsychol. Cogn.* 19, 741–758. doi: 10.1080/13825585.2011.652590

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *J. Acoust. Soc. Am.* 49, 467–477.

Lipnicki, D. M., Crawford, J. D., Dutta, R., et al. (2017). Age-related cognitive decline and associations with sex, education and apolipopoprotein E. genotype across ethnocultural groups and geographic regions: a collaborative cohort study. *PLoS Med.* 14:e1002261. doi: 10.1371/journal.pmed.1002261

Milham, M. P., Erickson, I. I., Banich, M. T., Kramer, A. F., Webb, A., Wszalek, T., et al. (2002). Attentional control in the aging brain: insights from an fMRI study of the Stroop task. *Brain Cog.* 49, 277–296. doi: 10.1006/brcg.2001.1501

Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., et al. (2005). The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *J. Am. Geriatr. Soc.* 53, 695–699. doi: 10.1111/j.1532-5415.2005.53221.x

Navarra, J., Hartcher-O'Brian, J., Piazza, E., and Spence, C. (2009). Adaptation to audiovisual asynchrony modulates the speeded detection of sound. *Proc. Natl. Acad. Sci. U.S.A.* 106, 9169–9173. doi: 10.1073/pnas.0810486106

Nilsson, M., Soli, S. D., and Sullivan, J. A. (1994). Development of the Hearing In Noise Test for the measurement of speech reception thresholds in quiet and in noise. *J. Acoust. Soc. Am.* 95, 1085–1099. doi: 10.1121/1.408469

Park, D. C., Lautenschlager, G., Hedden, T., Davidson, N. S., Smith, A. D., Smith, P. K., et al. (2002). Models of visuospatial and verbal memory across the adult life span. *Psychol. Aging* 17, 299–320. doi: 10.1037/0882-7974.17.2.299

Peelle, J. E., and Sommers, M. S. (2015). Prediction and constraint in audiovisual speech perception. *Cortex* 68, 169–181. doi: 10.1016/j.cortex.2015.03.006

Presacco, A., Simon, J. Z., and Anderson, S. B. (2016). Evidence of degraded representation of speech in noise, in the aging midbrain and cortex. *J. Neurophysiol.* 116, 2346–2355. doi: 10.1152/jn.00372.2016

Reitan, R. M. (1958). Validity of the trail making test as an indicator of organic brain damage. *Percept. Mot. Skills* 8, 271–276. doi: 10.1038/nn.2331

Richards, M. D., Goltz, H. C., and Wong, A. M. F. (2017). Alterations in audiovisual simultaneity perception in amblyopia. *PLoS One* 12:e0179516. doi: 10.1371/journal.pone.0179516

Rönnberg, J., Lunner, T., Zekveld, A., Sörqvist, P., Danielsson, H., Lyxell, B., et al. (2013). The ease of language understanding (ELU) model: theoretical, empirical, and clinical advances. *Front. Syst. Neurosci.* 7:31. doi: 10.3389/fnsys.2013.00031

Rönnberg, J., Rudner, M., Foo, C., and Lunner, T. (2008). Cognition counts: a working memory system for ease of language understanding (ELU). *Int. J. Audiol.* 7, S99–S105. doi: 10.1080/14992020802301167

Rothauser, E. H., Chapman, W., Guttman, N., Nordby, K. S., Silbiger, H. R., and Urbanek, G. E. et al. (1969). IEEE recommended practice for speech quality measurements. *IEEE Transac. Acoust. Speech Sign. Proces.* 17, 225–246.

Salthouse, T. A. (2009). When does age-related cognitive decline begin? *Neurobiol. Aging* 30, 507–514. doi: 10.1016/j.neurobiolaging.2008.09.023

Schwartz, J.-L., and Savariaux, C. (2014). No, there is no 150 ms lead of visual speech on auditory speech, but a range of audiovisual asynchronies varying from small audio lead to large audio lag. *PLoS Comput. Biol.* 10:e1003743. doi: 10.1371/journal.pcbi.1003743

Snell, K. B. (1997). Age-related changes in temporal gap detection. *J. Acoust. Soc. Am.* 101, 2214–2220. doi: 10.1121/1.418205

Summers, C., Bohman, T. M., Gillam, R. B., Pena, E. D., and Bedore, L. M. (2010). Bilingual performance on nonword repetition in Spanish and English. *Int. J. Lang. Commun. Disord.* 45, 480–493. doi: 10.3109/13682820903198058

Tillman, T. W., and Carhart, R. W. (1966). *An Expanded Test for Speech Discrimination Utilizing CNC Monosyllabic Words. Northwestern University Auditory Test No. 6. SAM-TR-66–55*. Dayton: USAF School of Aerospace Medicine. 1–12. doi: 10.21236/ad0639638

Trofimovich, P., and Baker, W. (2006). Learning second language suprasegmentals: effect of L2 experience on prosody and fluency characteristics of l2 speech. *Stud. Second. Lang. Acquis.* 28, 1–30. doi: 10.1017/SO272263106060013

Waddington, E., Jaekel, B. N., Tinnemore, A., Gordon-Salant, S., and Goupell, M. J. (2020). Recognition of accented speech by cochlear-implant listeners: benefit of audiovisual cues. *Ear Hear.* 41, 1236–1250. doi: 10.1097/AUD000000000000842

Wechsler, D. (1997). *Wechsler Adult Intelligence Scale, Third Edition (WAIS-III)*. San Antonio: Pearson Assessment.

Yi, H. G., Phelps, J. E., Smiljanic, R., and Chandrasekaran, B. (2013). Reduced efficiency of audiovisual integration for nonnative speech. *J. Acoust. Soc. Am.* 134, EL387–EL393. doi: 10.1121/1.4822320

Zhang, Y., Nissen, S. L., and Francis, A. L. (2008). Acoustic characteristics of English lexical stress produced by native Mandarin speakers. *J. Acoust. Soc. Am.* 123, 4498–4513. doi: 10.1121/1.2902165

Conflict of Interest: 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 Psychology

REVIEW
published: 16 February 2022
doi: 10.3389/fpsyg.2022.831670

Age and Auditory Spatial Perception in Humans: Review of Behavioral Findings and Suggestions for Future Research

*Michael Keith Russell**

Department of Psychology, Bellevue University, Bellevue, NE, United States

It has been well documented, and fairly well known, that concomitant with an increase in chronological age is a corresponding increase in sensory impairment. As most people realize, our hearing suffers as we get older; hence, the increased need for hearing aids. The first portion of the present paper is how the change in age apparently affects auditory judgments of sound source position. A summary of the literature evaluating the changes in the perception of sound source location and the perception of sound source motion as a function of chronological age is presented. The review is limited to empirical studies with behavioral findings involving humans. It is the view of the author that we have an immensely limited understanding of how chronological age affects perception of space when based on sound. In the latter part of the paper, discussion is given to how auditory spatial perception is traditionally conducted in the laboratory. Theoretically, beneficial reasons exist for conducting research in the manner it has been. Nonetheless, from an ecological perspective, the vast majority of previous research can be considered unnatural and greatly lacking in ecological validity. Suggestions for an alternative and more ecologically valid approach to the investigation of auditory spatial perception are proposed. It is believed an ecological approach to auditory spatial perception will enhance our understanding of the extent to which individuals perceive sound source location and how those perceptual judgments change with an increase in chronological age.

Keywords: age, spatial, methodology, ecology, perception

Edited by:
Leah Fostick,
Ariel University, Israel

Reviewed by:
Diane Baier,
Austrian Academy of Sciences (OeAW), Austria
Thomas A. Stoffregen,
University of Minnesota Twin Cities, United States

****Correspondence:***
Michael Keith Russell
mirussell@bellevue.edu

Specialty section:
This article was submitted to Perception Science, a section of the journal Frontiers in Psychology

Received: *08 December 2021*
Accepted: *24 January 2022*
Published: *16 February 2022*

Citation:
Russell MK (2022) Age and Auditory Spatial Perception in Humans: Review of Behavioral Findings and Suggestions for Future Research. Front. Psychol. 13:831670. doi: 10.3389/fpsyg.2022.831670

INTRODUCTION

In the real world, events occur within our vicinity and empirical research suggests we are able to detect and correctly identify those events when relying on sound alone. Individuals are nearly perfect at identifying bouncing and breaking glass jars (Warren and Verbrugge, 1984). Individuals are also capable of using sound to determine the gender of the pedestrian (Li et al., 1991) and whether a pedestrian is approaching or withdrawing (e.g., Kozhevnikova and Zhukov, 1990). Using sound, individuals are highly capable at recognizing the filling of a vessel and whether a vessel was filled to the brim (Cabe and Pittenger, 2000). When relying solely on sound, individuals are capable of judging the size of unseen dropped rods (Carello et al., 1998) and wooden balls (Grassi, 2005; Grassi et al., 2013), the roughness of a surface

(Lederman, 1979), the material composition of a plate (Giordano and McAdams, 2006), the hardness of a mallet (Freed, 1990), the elasticity of a bouncing ball (Warren et al., 1987), and the shape of a struck object (Kunkler-Peck and Turvey, 2000).

Aside from judging the properties of an object or actions that occurred nearby, individuals commonly determine the spatial position of unseen sound sources. When an unexpected sound occurs (e.g., a glass dropped into a hard surface), individuals naturally look in the direction of the objects that created the event. Furthermore, individuals often need to detect the location of an unseen sound-producing object in order to avoid injury and potentially death. Being able to merely detect the existence of a viciously barking dog, a chain saw, or a motor vehicle, for example, is insufficient. Individual also needs to be keenly aware of the position of the potentially damaging object as it relates to their position. It would also be advantageous for individuals to be able to determine if and how the spatial position between object and perceiver is changing. In order to avoid collision, a pedestrian must be able to accurately determine that an automobile is approaching and when it will arrive at the individual's location, as has been examined in a number of studies (e.g., Yan et al., 2007; Pörschmann and Störig, 2009; Braly et al., 2021).

Using an egocentric frame of reference, the position of an object can be described in a variety of ways. An important task an observer could perform is lateralization. Here, an observer simply needs to determine whether a target is located to their right or left (relative to the midline of the observer's body). A more precise method of evaluating the ability of individuals to locate an unseen sound-producing object is to ask them to report the object's distance, azimuth, or elevation. Distance refers to the extent of space between observer and target. Azimuth refers to the left-right, lateral, or horizontal angle (measured in degrees) between a sound source and the median plane of the observer (i.e., the plane corresponding to the observer's midline). A target at an azimuth of 0, 90, 180, and 270° represents a target located precisely ahead, to the right, behind, and to the left of the individual, respectively. Elevation (also referred to as altitude) refers to the up-down or vertical angle (also measured in degrees) between a sound source and the horizontal plane of the observer. A target at an elevation of 0, 90, 180, and 270° represents a target located precisely ahead, above, behind, and below the individual, respectively. A target located at the origin (0° azimuth and 0° elevation) is located at ear level and directly in front of the individual. It is also possible for researchers to measure spatial acuity by calculating the minimum audible angle, i.e., the smallest perceptually detectable difference in position of two sound sources. As can be expected, the greater the physical separation between two sound sources, the greater the ability to discriminate the sound sources (e.g., Hartmann and Rakerd, 1989a; Perrott and Saberi, 1990; Brimijoin and Akeroyd, 2014).

It is well known that the ability of individuals to localize an unseen sound source is dependent on the difference in sound reaching the two ears (i.e., interaural differences) and the shape of the pinna (e.g., Middlebrooks and Green, 1991; Carlile, 1996; Blauert, 1997). When a sound source is located either directly ahead (0° azimuth) or directly behind an individual (180° azimuth), the sound contacts both ears at the same time and with equal intensity (interaural differences are null). When a sound-producing object is off center, the sound contacts the closer ear sooner (interaural time difference) and with greater intensity (interaural level difference) in comparison with the more distant ear. Interaural time and level differences are known to increase with an increase in deviation from 0° and reach maximal levels when a sound source is located at 90° (directly to the right) or 270° (directly to the left). It is also well known that interaural differences are dependent on signal frequency. In short, interaural time differences are limited to low-frequency sounds (below 1,500 Hz) while interaural level differences are limited to high-frequency sounds (above 1,500 Hz). The pinna is also known to influence the ability of individuals to localize sound sources. Perceived sound position is a function of a sound's spectral properties and those properties are influenced by the shape of the pinna and target azimuth and elevation. Not surprisingly, alterations of the pinna have been found to impact the ability of individuals to localize sounds in the horizontal (e.g., Musicant and Butler, 1984; Oldfield and Parker, 1984; Hofman et al., 1998) and vertical planes (e.g., Roffler and Butler, 1968; Oldfield and Parker, 1984; Hofman et al., 1998).

Possibly not surprising to the reader are the decrements in perceptual capabilities that are coincident with an increase in chronological age. With respect to vision, an increase in age is often accompanied by an increase in the hardening (presbyopia) and the opacity (cataracts) of the lens. Possibly less well known is the apparent negative impact of age on perceptual judgments involving other modalities. In brief, an increase in chronological age has been found to negatively impact olfactory perception (e.g., Doty et al., 2011; Zhang and Wang, 2017; Olofsson et al., 2021), haptic perception (e.g., Thompson et al., 1965; Kleinman and Brodzinsky, 1978; Norman et al., 2016), and gustatory perception (e.g., Kaneda et al., 2000; Murphy et al., 2002; Fukunaga et al., 2005).

A complex relationship exists between age and the perception of sound source location. As will become evident, an accurate understanding of the impact of chronological age on auditory spatial perception requires the consideration of numerous factors. One such factor is hearing loss. Hearing loss often coincides and becomes more severe with the advancement of age. For example, it is common for individuals to become increasingly less sensitive to high-frequency sounds with an increase in chronological age (e.g., Rodríguez Valiente et al., 2014), which are considered important for sound localization (e.g., Butler and Humanski, 1992; Best et al., 2005; Zonooz et al., 2019). However, it is not always the case that hearing loss occurs with an increase in age. A small number of studies have determined the correlation between age and hearing loss to be weak or insignificant (e.g., Abel and Hay, 1996; Neher et al., 2011; Buchholz and Best, 2020). In addition, health and environmental factors have the potential to impair an individual's hearing (for a review see Jayakody et al., 2018). Thus, it is possible hearing impairment is a direct result of those factors and not age.

Two purposes exist with regard to the present paper. Initially, a summary of the research relating chronological age and auditory spatial perception will be provided. The literature reviewed has been limited to empirical studies with behavioral findings in humans.[1] Discussion is further limited to empirical studies that treated age as an independent variable. Undoubtedly, hearing loss is common among older individuals. Nonetheless, it is not a certainty an individual's hearing will deteriorate with age. Chronological age and hearing impairment are, in fact, discrete variables. The intent of the present paper was to examine the degree to which chronological age affects auditory spatial perception. To accomplish that task, it seemed prudent to consider age independently of any confounding variables. Thus, studies that treated age and hearing loss as a single variable were excluded. To forewarn the reader, each of the studies will be described in more detail than is typically presented in an empirical or review article. The inclusion of a greater than normal amount of information is necessary for it relates to the latter part of the paper. In the latter part of the paper, comparisons will be drawn between traditional laboratory investigations and real-world settings. Despite the advantages of conducting research in a particular manner (e.g., simple sounds, anechoic settings, and stationary observers), the possibility exists that the traditional approach to auditory spatial perception fails provide insight into how chronological age affects judgments of sound source location when they occur in natural settings. It is believed an ecologically based approach will yield findings that relate directly to how individuals of varying age perceive the spatial position of an unseen sound source under everyday circumstances. It is further believed that an ecologically based approach will provide information that enhances the scientific communities' understanding of the abilities of aged individuals to locate sounds and that information can, in turn, be used to enhance the performance of aged individuals in real-world settings.

SIGNIFICANT EFFECTS OF AGE ON AUDITORY SPATIAL PERCEPTION

Table 1 contains a brief summary of the characteristics of participants, design, and analysis(es) performed relative to each of the studies subsequently discussed.

Lateralization Perception

As mentioned previously, an important auditory location task an individual can perform is that of lateralization. Individuals need only determine whether a sound source is located to the right or left of the individual's midline. In short, a decrease in the ability to lateralize sounds is associated with an increase in chronological age. Szymaszek et al. (2006) examined the ability of individuals to determine the order (left-right or right-left) of two sequentially presented clicks. The period of time between click presentations was systematically varied. Young ($M = 24$ yrs., 8 mos.) and elderly ($M = 64$ yrs., 6 mos.) participants with normal hearing were compared. Threshold for lateralization was defined as the minimal time period between stimulus presentations that permitted 75% correct order identification. The thresholds for elderly individuals were significantly greater than that for young individuals. The mean threshold for young individuals was 66 ms. For elderly individuals, it was 88 ms.

Fink et al. (2005) employed the same task and likewise compared young ($M = 25$ yrs.) and elderly ($M = 61.7$ yrs.) individuals. While an increase in age was again found to negatively affect lateralization perception, threshold differences between the two age groups were dependent on the stimulus (clicks or tones), the method used to calculate the threshold (staircase or maximum-likelihood performance), and test session (session 1, 2, and 3). On average, the thresholds of older individuals were significantly greater than those of younger individuals. For young participants, the threshold was approximately 50 ms for clicks and approximately 15 ms for tones. For older participants, the mean threshold was approximately 65 ms for both clicks and tones. The mean difference in thresholds between the two age groups was smaller for clicks than for tones: 14.9 and 48.45 ms, respectively. With regard to test session, when the stimulus was a click, older and younger participants differed only with regard to the third test session and only when the staircase method was used to calculate the threshold. When the stimulus was a tone, the threshold difference between the two age groups decreased notably with an increase in session. For tones, the threshold for older participants decreased with an increase in test session. Thresholds were approximately 80 ms for session 1, 60 ms for session 2, and 40 ms session 3. For young participants, threshold values were largely unaffected by test session and were less than 20 ms.

Kołodziejczyk and Szelag (2008) likewise presented participants with the task of determining the order of a pair of stimuli (square wave tones). In that study, the thresholds for three age groups were determined: young ($M = 22$ yrs.), elderly ($M = 66$ yrs.), and very old ($M = 101$ yrs., 1 mo.). Threshold differences between the three age groups were evident. The

---

[1] Humans and most nonhumans are capable of localizing unseen sound sources despite substantial differences between species. With regard to humans, the size and structure of the head are such that sound either travels around it or is blocked by it (thus, creating what is referred to as acoustic shadow). In several species (e.g., fish, amphibians, and reptiles) where tissue density is substantially lower, sound travels through the body, head, or mouth. In humans and other species, the ears are physically separated in space, which provides an opportunity to determine sound source position using interaural differences. In other species (e.g., birds), the ears are physically and internally coupled by an interaural canal thereby making interaural differences moot. Additional physical differences between species (e.g., form of the tympanic membrane, presence of the three bones of the middle ear, and presence and shape of the pinna) as well as significant differences in experimental methodology make comparisons between species difficult. For these reasons, the present paper will focus on the abilities of humans of various ages to determine the position of a sound-producing object. However, interspecies comparisons suggest the existence of lawful relationships. Extent of high-frequency hearing appears to be related to the distance between the ears. Sound localization acuity is inversely related to the breadth of an animal's field of best vision. A number of publications discuss interspecies and intraspecies differences and similarities with regard to auditory spatial acuity (e.g., Heffner and Heffner, 1998, 2014, 2016, 2018).

TABLE 1 | Summary of characteristics of participants, design, and analysis related to studies presented in the literature review.

References	Age groups (in years)	Sample size	Stimuli	Range of localization (in degrees)	Statistical measure (s)
Abel et al. (2000)	7 age groups: 10–19, 20–29, 30–39, 40–49, 50–59, 60–69, 70–81. Means and standard deviations were not provided.	16 per age group	Broadband noise. One-third-octave noise band centered on 0.5 kHz and 4 kHz. Duration = 300 ms	Horizontal plane: 15–165	ANOVA and Regression
Abel and Hay (1996)	3 age groups: Young-Normal (18–38), Old-Normal (41–58), Old-Hearing Impaired (42–73). Means and standard deviations were not provided.	*N* = 24 Young-Normal, *N* = 24 Old-Normal, *N* = 23 Old-Hearing Impaired	Broadband noise. One-third octave noise bands centered at 0.5 or 4 kHz. Duration = 300 ms.	Horizontal plane: 30 to 150	ANOVA
Addleman et al. (2019)	2 age groups: Younger (21–33), Older (58–78). Means and standard deviations were not provided.	*N* = 13 Younger, *N* = 12 Older	Pink noise bursts. Frequency range = 0.2–8 kHz. Duration = 200 ms.	Horizontal plane: 10 to 180	ANOVA
Briley and Summerfield (2014)	3 age groups: Young (*M* = 22.9, *SD* = 2.8), Younger-Old (*M* = 65.1, *SD* = 3.6), Older-Old (*M* = 76.8, *SD* = 2.5)	*N* = 6 Young, *N* = 6 Younger-Old, *N* = 5 Older-Old	Summed pure tone frequencies/ pink noise. Frequency range = 0.1–5 kHz. Duration = 1,510 ms.	Horizontal plane: 0 to 75	Descriptive statistics
Brungart et al. (2017)	2 are groups: Normal Hearing (*M* = 31.9), Hearing Impaired (*M* = 54.7). Standard deviations were not provided.	*N* = 16 Normal Hearing, *N* = 20 Hearing Impaired	7 periodic chirp signals. Frequency range = 0.1–15 kHz. Stimulus duration either 250, 100, or 4,000 ms.	Horizontal plane = −150 to +150 Vertical plane = −28 to +28	Correlation and Regression
Dobreva et al. (2011)	3 age groups per experiment: Experiment 1: Young (19–41), Middle Age (45–66), Elderly (70–81). Experiment 2: Young (19–37), Middle Age (51–66), Elderly (71–81). Means and standard deviations were not provided.	Experiment 1: *N* = 19 Young, *N* = 11 Middle Age, *N* = 12 Elderly. Experiment 2: *N* = 8 Young, *N* = 7 Middle Age, *N* = 6 Elderly	Band-limited, flat spectrum, Gaussian noise bursts. Duration = 150 ms.	Horizontal plane = −60 to +60 Vertical plane = −25 to +25	ANOVA
Fink et al. (2005)	2 age groups: Younger (*M* = 25), Elderly (*M* = 61.7). Standard deviations were not provided.	*N* = 20 per group	2 sound signals: clicks or pure tones. Clicks were noise, rectangular pulses. Duration = 1 ms. Sinusoidal tones 0.8 and 1.2 kHz. Duration = 10 ms.	Not applicable. Temporal order (lateralization) task. Sounds presented *via* headphones.	ANOVA
Freigang et al. (2014)	2 age groups: Young (*M* = 24.1, *SD* = 2.3), Older (*M* = 68.1, *SD* = 5.5)	*N* = 22 Young, *N* = 53 Older	Narrowband noise centered at 0.5 (0.375–0.75 kHz) and 3.0 kHz (2.25–4.5 kHz). Duration = 500 ms.	Horizontal plane: −98 to +98	ANOVA, T-test, and Correlation
Kolodziejczyk and Szelag (2008)	3 age groups: Young (*M* = 22, *SD* = 1.1), Elderly (*M* = 66, *SD* = 0.7), Very Old (*M* = 101.1, *SD* = 0.11)	*N* = 17 Young, *N* = 18 Elderly, *N* = 11 Very Old	300 Hz square tones. Duration = 15 ms.	Not applicable. Temporal order (lateralization) task. Sounds presented *via* headphones.	ANOVA
Otte et al. (2013)	3 age groups: Children (*M* = 9.5, *SD* = 1.3), Young Adults (*M* = 24.9, *SD* = 4.9), Older Adults (*M* = 68.4, *SD* = 4.7)	*N* = 18 Children, *N* = 10 Young Adults, *N* = 14 Older Adults	Gaussian white noise high-pass filtered at 0.5 kHz or low-pass filtered at 5, 7, 11, or 20 kHz. Duration = 150 ms.	Horizontal plane = −75 to +75 Vertical plane = −55 to +55	T-test and Correlation

(Continued)

TABLE 1 | Continued

References	Age groups (in years)	Sample size	Stimuli	Range of localization (in degrees)	Statistical measure (s)
Rønne et al. (2016)	2 age groups: Normal Hearing (*M* =39, *SD* =11), Hearing Impaired (*M* =64, *SD* =15)	*N* =13 Hearing Impaired, *N* =11 Normal Hearing	White noiseband pass filtered to 0.4–16 kHz. Duration of stimulus not provided.	Horizontal plane: 0 to 120	ANOVA
Savel (2009)	Age treated as a continuous variable. 48 individuals between 18 and 48 (*M* =31, *SD* =10), 2 individuals 61 and 62 (*M* =61.5, *SD* =0.5)	*N* =48 Young, *N* =2 Older	Filtered noise 0.25–2 kHz. Duration = 50 ms.	Horizontal plane: −77 to +77	Correlation
Szymaszek et al. (2006)	2 age groups: Young (*M* =24 yrs., 8 months), Elderly (*M* =64 yrs., 6 mos.) Standard deviations were not provided.	*N* =17 Young, *N* =16 Elderly	"Clicks." No further description provided. Duration = 1 ms.	Not applicable. Temporal order (lateralization) task. Sounds presented *via* headphones.	ANOVA and Newman–Keuls
Whitmer et al. (2014)	Age treated as a continuous variable. Median = 63. Means and standard deviations were not provided.	*N* =35	0.1 kHz click train. Duration = 500 ms.	Horizontal plane: −30 to +30	T-test and Correlation
Whitmer et al. (2021)	Age treated as a continuous variable. No specific age information was provided. Based on Figure 1, it appears participants ranged in age from 20 to 70.	N = 28	Speech segments with either 5- or 10-kHz cutoff frequency. Duration >5,000 ms.	Horizontal plane: −90 to +90	ANOVA and Correlation

mean threshold for the young, elderly, and very old participants was 37, 60, and 191 ms, respectively. Significance was limited to the differences between the young and very old, and between the elderly and very old.

Localization Perception

The ability to determine which side of the body a sound source resides can be considered a rudimentary perceptual skill. Individuals perceiving and acting in real-world settings need to accurately locate sound-producing objects. Thus, the accuracy with which individuals can determine the precise location of a sound source is more insightful. Generally speaking, the findings of numerous studies suggest sound localization ability decreases with an increase in chronological age. Abel et al. (2000) evaluated the ability of seven age groups (10–19, 20–29, 30–39, 40–49, 50–59, 60–69, and 70–81 yrs.) to identify which loudspeaker (out of 4 or 8) was the origin of a target stimulus. In every ANOVA conducted, age was a significant factor. As expected, localization performance decreased with an increase in age, particularly beyond the first three decades of life. Regression analyses revealed that age accounted for 12–26 percent of the variance in perceptual accuracy.

In a study by Dobreva et al. (2011), young (19–41 yrs.), middle aged (45–66 yrs.), and elderly (70–81 yrs. old) individuals were required to direct (*via* a joystick) a laser-LED beam at the perceived location of the sound stimulus. The ability of individuals to accurately perceive azimuth and elevation was determined. In terms of perceived azimuth, while the main effect of age was not significant, age was a significant variable when different sound stimuli were taken into consideration. The three age groups were highly similar when the sound was low-pass (0.1–1 kHz) or high-pass (3–20 kHz) noise. Differences between the age groups were striking when the sound was ultra-high-pass (10–20 kHz) noise. Young participants were significantly more accurate than both middle-aged and elderly participants. In terms of perceived elevation, the main effect of age was again found to be significant. An increase in age was associated with a decreased ability to judge sound source elevation. Generally speaking, the perceptual judgments of young individuals were the most accurate and the performance of middle-aged individuals tended to be more similar to the performance of elderly individuals than to young individuals.

Whitmer et al. (2014) positioned participants in the center of a cloth covering a 24-speaker array that ranged from −45 to +45°. From one of the speakers, a 100-Hz click train was presented. The task of the participant was to indicate on a monitor situated in front of them where they believe the sound originated. For each individual, localization precision (variability) and accuracy (absolute difference between mean perceived and actual location) were calculated. Localization precision was significantly correlated with age ($r = +0.68$) even when hearing loss was controlled ($r = +0.46$). Participants younger than 50 years of age were significantly more variable in their judgments than individuals 50 or more years old; 3.9 and 10.1°, respectively. When three levels of age (younger: 26–42; middle: 45–65; and older: 74–81) were compared, age was again found to be a significant variable. Perceptual variability for the younger, middle, and older individuals was found to be 3.7, 6.8, and 12.4°, respectively. Localization accuracy was also significantly correlated with age ($r = +0.39$). Regrettably, information about the accuracy level of different age groups was not provided.

Freigang et al. (2014) compared the ability of older ($M = 68.1$ yrs.) and young ($M = 24.1$ yrs.) adults to accurately point to the origin of an unseen sound. The accuracy of perceptual judgments was determined using a "torch" that emitted a light invisible to humans. The minimum audible angle was calculated. In each trial, three sounds were sequentially presented. Two of the stimuli can from the same location while the third originated from a different location. The task of the participant was to point at the perceived location of the deviant sound. The main effect of age was significant as well as the age*position (frontal vs. peripheral targets) interaction. Overall, young individuals were more accurate than older individuals. In addition, the minimum audible angle was approximately two times greater for older than younger individuals. This was true regardless of whether the stimulus (noise) was low-frequency (0.5 kHz) or high-frequency (3.0 kHz) centered. The correlation between age and minimum audible angle was between 0.2387 and 0.3160.

Rønne et al. (2016) likewise compared the minimum audible angles of young ($M = 39$) and old ($M = 64$) individuals. In a manner similar to that of Freigang et al. (2014), participants were exposed to three sounds on each trial and their task was to identify the origin of the deviant sound. The results are similar to those obtained by Freigang and colleagues. The minimum audible angle was notably smaller for young individuals (10°) than for older individuals (20–30°). It was also discovered that manipulation of pinna-related cues affected the accuracy and variability of responses made by older participants. Younger participants appeared unaffected by such manipulations.

INSIGNIFICANT EFFECTS OF AGE ON AUDITORY SPATIAL PERCEPTION

Based on the summary provided thus far, an increase in chronological age appears to adversely affect the ability of individuals to determine whether an unseen sound source is located to the right or left and to precisely locate a sound-producing object. However, as will shortly become evident, the findings of other studies as well as the findings of a number of the aforementioned studies suggest auditory spatial perception is unrelated to chronological age.

Lateralization Perception

As mentioned previously, Fink et al. (2005) discovered that older participants needed a significantly greater time period between two sequentially presented in order to accurately determine whether a sound was located to the right or left. While the results suggested age was a significant factor affecting perception, the extent of the effect of age was dependent on the acoustic properties of the target. When the target was a sinusoidal tone, perceptual differences between younger and older participants were clearly evident. However, when the sound stimulus was a click (noise), the two age groups performed in a highly similar manner. Thus, the argument that age negatively affects sound lateralization should be promoted with qualification.

The previously mentioned study by Kołodziejczyk and Szelag (2008) also seems to support the argument that chronological age negatively affects auditory lateralization ability. While individuals in their 60s required a greater separation in time between the two sequentially present stimuli than did individuals in their 20 s, the threshold difference was not statistically significant. The suggestion that age negatively influences lateralization was only supported by the performance of centenarians. Without additional investigation, it remains possible the deterioration in performance by those over 100 years of age was a function of factors, such as cognitive processing and not audition.

Localization Perception

With regard to sound localization, the findings of two of the earlier mentioned studies suggest a null effect of age. In addition to discovering significant differences between seven age groups, Abel et al. (2000) also discovered similarities (i.e., non-significant differences). Regardless of age, all participants were poor at locating a 0.5 kHz sound and all were highly accurate at locating broadband noise. The negative impact of age was primarily limited to the 4 kHz sound stimulus. The reported overall 7–23% decrease in accuracy that coincided with an increase in age was largely a reflection of the 4 kHz sound. If performance related to the 4 kHz sound is excluded, the overall mean difference in accuracy between the youngest (10–19 yrs. of age) individuals and those in their 60s was 6.5% for the 0.5 kHz sound and 4.0% for broadband noise. With regard to specific sounds, the mean difference in accuracy between the youngest and oldest (70–81 yrs. of age) individuals was 11.5% for the 0.5 kHz sound and 9.7% for broadband noise. One could argue the difference in error between young and old and between the youngest and oldest is not substantial. Moreover, while age is a linear and continuous variable, the observed decrease in sound localization accuracy did not decrease linearly or continuously with an increase in age. This observation suggests factors other than age or one or more moderating or mediating factors may have affected auditory localization judgments (e.g., Nambu et al., 2013; Trapeau and Schönwiesner, 2018; Bednar and Lalor, 2020).

The previously mentioned study by Dobreva et al. (2011) appears to provide incontrovertible support for the notion that an increase in age negatively affects the ability of individuals to localize sounds in terms of both azimuth and elevation. As mentioned previously, younger individuals were more accurate than middle-aged and elderly individuals in both dimensions. However, Dobreva and colleagues also discovered similarities in performance across age groups. All participants overestimated azimuth and underestimated elevation. In addition, all participants were less precise in their estimations of elevation than azimuth. Similar to Abel et al. (2000), the negative impact of age was stimulus dependent. Age was a significant factor for certain types of wideband noise (e.g., 10–20 kHz), but not for others (e.g., 3–10 kHz). Age was a null factor when the sound was low-frequency narrowband noise.

Addleman et al. (2019) compared younger (21–33 yrs.) and older (58–78 yrs.) individuals in terms of lateralization and localization performance. Comparisons were also made between centrally (10–30°) and peripherally (60–80°) located sound sources. Performance was evaluated in terms of precision (absolute error) and variability. In short, the two age groups did not significantly differ in any discernible manner. Both age groups were equally accurate (approximately 7° of error) and equally variable (again, approximately 7°). This was true regardless of whether the sound source was located centrally or peripherally.

Abel and Hay (1996) compared young-normal hearing (18–38 yrs.), old-normal hearing (41–58 yrs.), and old-hearing impaired (42–73 yrs.) individuals. Given the focus of the present paper is specifically with regard to the relationship between age and auditory spatial perception independent of hearing impairment, the performance of the old-hearing impaired group will not be discussed. The task of participants was to identify which one of an array of speakers was the origin of a target sound. The background was either quiet or continuously contained white noise. In brief, age was not found to be significant. The presence of background sound negatively affected the performance of both age groups. Variations in signal frequency (0.5 or 4 kHz) equally affected the performance of the two age groups. Age of participant was also found to be unrelated to either the accuracy of left-right judgments and front-back judgments. In a similar vein, both age groups made a comparable number of front-back and back-front reversals and both age groups were 20% more likely to make a back-to-front reversal than a front-to-back reversal.

While the focus of a study by Brungart et al. (2017) involved a comparison of young-normal hearing and older hearing-impaired individuals, the impact of age was treated, in certain circumstances, as a discrete independent variable. In that study, participants were instructed to point a handheld wand at the perceived location of an actual unseen sound source (experiment 1) or at a virtual sound source (experiment 2). Aside from the finding that an increase in angle of head movement resulted in a decrease in front-back confusions and azimuth error for both normal and hearing-impaired individuals, a stepwise regression revealed predicted localization perception was unaffected by participant age. This was true for both free-field and virtual sound localization conditions.

In a study by Otte et al. (2013), a laser pointer was affixed to the head of participants and participants were instructed to point to the laser dot at the sound source. The performance of three age groups was compared: children ($M = 9.5$ yrs.), young adults ($M = 24.9$ yrs.), and older adults ($M = 68.4$ yrs.). Judgments of target azimuth and elevation were evaluated independently. With regard to target azimuth, the three age groups performed in a highly similar manner. The correlation between actual and perceived azimuth was 0.94, 0.95, and 0.94 for children, young, and older adults, respectively. Mean absolute error was also highly similar among the different age groups. With regard to target elevation, judgments were less accurate and more variable for all three age groups. The results suggest that the accuracy of and variability in auditory spatial judgments are influenced more by concha height than by age. For all three age groups, concha height was positively related to perceptual accuracy and inversely related to perceptual variability.

Savel (2009) compared individuals varying in age on their ability to locate an unseen sound source. Loudspeakers were positioned in a 180° arc and concealed by a curtain. Participants viewed a computer screen that likewise contained a 180° arc and their task was to indicate on the computer screen the perceived location of the target sound. Perceptual accuracy was evaluated in three ways: root mean square error, standard deviation, and standard error. The results suggested perceptual abilities are unaffected by age. The correlation between age and each of the three dependent variables was not significant. Interestingly, correlations ranged from −0.32 to +0.37, which suggests an increase in age may result in either perceptual impairment as perceptual improvement. On an aside, the results further revealed left-right discrimination was unrelated to participant age.

Whitmer et al. (2021) likewise compared the ability of individuals of various ages to locate a sound. At the start of each trial, participants would hear a 5-s segment of speech from a talker located at 0° azimuth. Following a one-second delay, a speech segment from a different talker was played from a different location. The task of the participant was to turn their head and/or chair to directly face the new talker. Performance was evaluated in nine ways: trajectory start time, trajectory end time, trajectory duration, accuracy (absolute error), peak velocity, time of peak velocity, complexity, rate of misorientations, and rate of reversals. In brief, age was not significantly correlated with any of the dependent variables.

As mentioned earlier, Freigang et al. (2014) compared younger and older individuals in their ability to localize an unseen sound source. Both perceptual accuracy and minimum audible angle were calculated. With regard to perceptual accuracy, the results of various ANOVAs suggested the two age groups differed significantly. Interestingly, perceptual accuracy was not significantly correlated with participant age. This latter finding suggests perceptual judgments were influenced by a variable other than, but confounded with, age. With regard to the minimum audible angle, ANOVAs and correlation analysis suggested age was a significant factor. The authors concluded that age is not a relevant factor in terms of the ability to locate sound sources but is relevant in terms of the ability of individuals to discriminate between sound source locations.

Briley and Summerfield (2014) likewise determined the minimum audible angle of different age groups. Three age groups were compared: young ($M = 22.9$ yrs.), younger-old ($M = 65.1$ yrs.), and older-old ($M = 76.8$ yrs.). Similar to Freigang et al. (2014), participants were exposed to three pairs of sounds each separated by a period of silence. Two of the pairs originated from the same location and the third pair originated from a different location. The minimum audible angle was calculated using a 71% correct detection threshold. Although no inferential statistics were presented, the findings suggest young and younger-old adults have highly similar perceptions. When the target was located directly ahead (0° azimuth), mean MAA was 5.8° for the young participants and 6.1° for the younger-old participants. Interestingly, when the target was located in the periphery, young participants had *larger* minimum audible angles than younger-old participants. The younger-old group also seemed to be more homogenous than the young group for peripherally positioned targets. The impact of chronological age was largely limited to the performance of the older-old group. The older-old group had substantially larger minimum audible angle thresholds (8.3°) than either the young or the younger-old groups. Considering the older-old group had pronounced hearing loss, it is possible the poor performance displayed by the older-old group was a reflection of hearing loss and not age.

SUGGESTIONS FOR FUTURE RESEARCH EVALUATING THE IMPACT OF AGE ON AUDITORY SPATIAL PERCEPTION

The summary of the research provided previously should make it clear that the relationship between chronological age and auditory spatial perception is ambiguous. Obviously, arguments can be made that the differences in findings reflect differences in methodology. Instead, I prefer to make the argument that a more accurate understanding of the effect of age on spatial perception requires an approach notably different from the one traditionally taken. More specifically, the argument will be made that an ecological approach to auditory spatial perception is required and will yield a more accurate understanding of how age affects auditory spatial judgments. Gaver (1993) argued that auditory perception experiments in general should focus on everyday listening (i.e., perception of events) rather than musical listening (i.e., perception of acoustic properties). A similar approach to that proposed by Gaver will be taken here.[2] The subsequently offered approach will focus on the auditory perception of space, but it is equally applicable to other avenues of research. Recommendations for future research will be provided.

INDIVIDUAL FACTORS

Stationary Vs. Dynamic Point of Observation

In laboratory experiments, it is fairly common for observer motion to be constrained to some degree. Participants are typically seated and head movement is limited. To more effectively control changes in head position, researchers often employ a chin rest or bite bar. Limitations on head and body position are imposed for good reasons. During sound presentation, a change in head or body position can dramatically alter the sound entering the auditory canal that, in turn, can dramatically alter perception. The control of head and body movement provides an opportunity for researchers to determine the extent to which spatial judgments are influenced by factors, such as interaural level difference, interaural time differences, and various acoustic properties (e.g., frequency, intensity, and phase). Of the 16 studies previously reviewed, all 16 imposed limitations of the posture of the observers.

In real-world settings, organisms are active. We habitually change our body posture and often change our position within an environment. Rather than being passive recipients of stimulation, we intentionally seek out information and our motion makes information available when previously it was not. While humans are incapable of swiveling their ears as is commonly executed by a number of nonhumans (e.g., cats, dogs, and horses), the ears of humans are attached to a head which can be tilted and pivoted. The head is attached to a body that is capable of changing position in multiple ways (forward-backward, left-right, and up-down). By changing our position in space, we are able to detect information that could not have been detected otherwise. A number of studies have found that changes in head position have the potential to influence the judgments of sound source location (e.g., Ashmead et al., 1995; Perrett and Noble, 1997; Wightman and Kistler, 1999).

In real-world settings, individuals could become better aware of the position of a sound source by simply altering the head and/or body position. By altering the orientation of the head until interaural differences are eliminated, an observer is able to determine whether an unseen sound source is located to the right or left and they are aware of the object's precise location. Changing one's location within a setting also provides an opportunity for individuals to determine the change in distance of an unseen sound source. Movement that produces an increase in signal intensity at the point of observation suggests approach while movement that produces a decrease in intensity suggests withdrawal. If limitations placed on the natural response of observers potentially to yield faulty perceptual judgments, then it seems prudent to permit the individual to respond in a natural manner. Rather than limiting observer motion, researchers should permit participants to move in the manner they typically do in real-world settings. By doing so,

---

[2] The reader may note that the ecological approach discussed in the present paper resembles the ecological approach proposed by Gibson (1979) and wonders why the Gibsonian approach was not discussed in more depth. First, I intend to subsequently submit a paper that directly contrasts¸ with regard to auditory spatial perception, the traditional and Gibsonian, ecological approaches. That paper will delve more deeply into the Gibsonian approach and include discussion of concepts, such as affordances, effectivities, mutuality, and prospective control. Second, while the Gibsonian, ecological approach could have been included in the present paper, I do not believe the Gibsonian, ecological approach can be completely and accurately presented in a brief manner, as would have been required here. I believe the contrast between the traditional and Gibsonian approaches should be the sole focus of a separate paper and considered independently of the relationship between chronological age and auditory spatial perception.

the results obtained will better reflect the actual capabilities of individuals.

Verbal Vs. Action-Based Tasks

Researchers investigating auditory spatial perception have traditionally required participants to verbally report target location. Participants convey their perception of target azimuth and elevation in degrees. While not discussed thus far, participants are often required to report the distance of a target using feet/inches or meters/centimeters. A clear advantage to doing so is that it permits researchers the opportunity to precisely determine perceptual accuracy. Using signed error, researchers are capable of determining the exact degree to which participants either underestimate or overestimate sound source position. Using absolute error, researchers are capable of determining the overall size of perceptual error.

Despite the advantage of employing the aforementioned response metrics, the simple fact is that it is unnatural. Individuals in real-world settings rarely, if ever, report an auditory target's position in degrees, feet and inches, or meters and centimeters. Instead, observers typically respond to a sound. Individuals look in the direction of sudden, unexpected sounds. A telephone rings and the individual alters their gaze so that it is in the direction of the phone. A knock on the door frequently elicits the individual approaching and opening a door. As stated previously, individuals are active organisms. Empirically speaking, a number of studies have discovered that individuals are highly accurate at making action-based judgments using vision (e.g., Warren, 1984; Mark, 1987; Warren and Whang, 1987), haptics (e.g., Bingham et al., 1989; Malek and Wagman, 2008; Hajnal et al., 2020), and sound (e.g., Rosenblum et al., 1996; Russell and Turvey, 1999; Russell and Schneider, 2006).

In 1979, James J. Gibson coined the term affordance. Briefly, affordances refer to the possible actions that can be performed with an object and reflect the relationship between the perceiver and the object being perceived. Rather than estimating target azimuth in angles and distance using feet and inches, individuals performing an affordance task are required to report whether a particular action is or is not possible. Previous research supports the notion that individuals are highly capable of using sound to judge the affordances of objects (Rosenblum et al., 1996; Riehm et al., 2019) discovered that individuals are highly accurate at judging whether an object affords grasping. In fact, sound-based judgments were as accurate as those based on vision. Russell and Turvey (1999), Gordon and Rosenblum (2004), and Russell (2020) found that individuals are able to accurately determine whether a gap is large enough to afford passage. O'Neill and Russell (2017) determined that individuals are highly capable of judging whether the height of a surface affords stepping on or walking under.

In a similar manner, what one hears influences how one acts. Using only sound, individual is capable of determining when they need to act so that a moving target can be intercepted (Vernat and Gordon, 2010). What an individual hears influences their ability to maintain a stable posture (e.g., Soames and Raper, 1992; Stoffregen et al., 2009b, 2019) and the stability of their posture influences their ability to successfully use sound to judge whether a surface can be stepped upon or walked under O'Neill and Russell (2017). By acting, individuals are able to detect information otherwise unavailable. Speigle and Loomis (1993) and Ashmead et al. (1995) discovered that individuals are more accurate and less variable in their estimations of target distance when walking during the broadcast of the sound than if they were stationary. It is argued here that action responses are more natural (i.e., ecologically relevant) and incorporating such metrics in empirical investigations will yield a better understanding of how age impacts our perception of the world.

Verbal Vs. Action Tasks

As stated previously and as was evident in the literature review, laboratory investigations commonly require participants to verbally estimate the position of a sound source and those verbal responses rarely reflect the types of actions performed in the real world. It is worth noting that when individuals make action-based judgments, what the individual says they can do does not always reflect what they actually do. For example, Warren and Whang (1987) required participants to verbally report whether the width of an adjustable doorway was sufficiently large to permit passage. The results revealed that participants required a larger gap when they actually walked through the gap than when they made verbal judgments from a stationary point of observation. Significant differences between perceptual judgments and actual performance have been documented in terms of whether a stair is low enough to be stepped upon (Warren, 1984), an object is close enough to be grasped (e.g., Carello et al., 1989; Wagman and Morgan, 2010), an expanse can be stepped or leaped over (Cole et al., 2013; Day et al., 2015), the slope of a surface is small enough for it to be ascended (Kinsella-Shaw et al., 1992), and wheelchair users can pass through a gap (e.g., Higuchi et al., 2004; Stoffregen et al., 2009a; Rodrigues et al., 2014).

While the literature suggests a disparity between perception and action, the manner by which researchers determine individual's capabilities may be a poor reflection of what the individual can actually do. For example, measuring arm length may not accurately reflect whether an individual can grasp an object. Human observers can alter the maximum reach extent by bending forward or pivoting the torso at the waist. The findings of Butler et al. (2011), Zhong and Yost (2013), O'Neill and Russell (2017), and Wagman et al. (2019) suggest posture and postural stability affect the ability of individuals to judge accurately whether an object can be grasped. In a similar fashion, Konczak et al. (1992) determined that the maximum surface height an individual can step up on is a function not only of the individual's leg length, but also that individual's hip flexibility and ability to generate a sufficient amount of torque to lift the body. It should also be noted that a number of studies have discovered insignificant differences between perceptual judgments and actual performance (e.g., Mark et al., 1990; Rosenblum et al., 1996; Franchak et al., 2010).

Irrespective of whether differences exist between perception and action, and regardless of whether differences exist between an individual's actual action capabilities and how researchers ascertain those capabilities, it would seem prudent to evaluate how individuals actually behave in real-world settings. Requiring individuals to perform an action, rather than make a verbal response, can be expected to provide greater insight into exactly how they actually perceive sound source position. For example, participants could be asked to look in the direction of a sound source. In real-world settings, individuals often direct their gaze at sound-producing objects (e.g., cell phone, orator, and animal). Likewise, participants could be asked to intercept a moving auditory target as was required by Vernat and Gordon (2010). Doing so would provide insight into the ability of individuals of various ages to accurately determine the movement of sound-producing objects. Knowing when a sound source will arrive at a particular location is useful to pedestrians attempting to cross a street when automobiles are moving in both directions. In such instances, audition, not vision, would be more helpful. Requiring participants to perform an action-based task can notably enhance the ecological validity of a study. More importantly, the use of action-based tasks provides researchers with the opportunity to more precisely evaluate the ability of individuals to be aware of the world and how age impacts that ability.

Unidimensional Vs. Multidimensional Perception

With rare exception (e.g., Makous and Middlebrooks, 1990), laboratory investigations into auditory spatial perception involve the manipulation of sound source position along a single dimension. With respect to the perception of azimuth, targets are positioned horizontally in an arc (thereby maintaining a constant observer-source distance) and at a fixed elevation (typically ear height). With respect to the perception of elevation, targets are positioned vertically in an arc (again maintaining a constant observer-source distance) and at a fixed azimuth (the center of the arc is aligned with the observer's midline). With respect to the perception of distance, the span between observer and target is varied, but all locations have identical elevations and azimuth is typically 0° (i.e., extending outward from with the observer's midline). Given that researchers are often interested in the accuracy with which individuals can locate a target within a particular dimension or researchers seek to identify the information supporting perception within a particular dimension, it is advantageous to vary one dimension while other dimensions are invariant.

In real-world settings, sound sources commonly exist at different azimuths, elevations, and distance. An auditory target may be near or far, in front or behind, above or below, and at varying positions to the left or right. Rarely do targets vary along a single dimension. A change in the position of a sound source or the point of observation commonly involves sequential or simultaneous changes across multiple dimensions. Imagine the case of a stationary pedestrian whose midline is perpendicular to a road. Coincident with the approach (withdrawal) of an automobile is a decrease (increase) in distance and azimuth. An individual standing at the bottom of a flight of stairs listening to the footsteps of a pedestrian hears an object changing in distance and elevation. Rather than require individuals to judge the position of a sound source in only one dimension, future investigations in auditory spatial perception and those examining the impact of age on spatial perception may wish to explore the ability of individuals to judge the location of a sound-producing object in multiple dimensions. Independent examination of perception along different dimensions may not reflect the exact nature of how sound source position is perceived in real-world settings. Russell and Schneider (2006), for example, exposed participants to a sound source that varied in distance and azimuth and discovered that judgments were highly accurate when participants walked to the perceived location of the unseen sound and judgments were notably less accurate when participants independently reported target distance and azimuth. It is possible that, in a Gestaltian sense, the whole is different from the sum of its parts.

ENVIRONMENTAL FACTORS

Anechoic Vs. Echoic Settings

Laboratory studies are regularly conducted in anechoic or sound-altered settings. Sound absorbing panels cover most if not all of the interior of the experimental room (floor, ceiling, and walls) or sound absorbing material is contained within the room walls. Of the 16 studies reported previously in the literature review portion of the present paper, 14 used a setting that altered sound transmission in some manner. Of the two remaining studies, one was conducted in the home of older participants due to mobility issues and one study did not provide information about the setting possibly because stimuli were presented through headphones. Auditory researchers often minimize or eliminate reverberant sound since reverberations have the potential to influence auditory spatial judgments (e.g., Mershon and King, 1975; Mershon et al., 1989; Bronkhorst and Houtgast, 1999; Zahorik et al., 2005). In order for researchers to maintain control over the acoustic information participants are exposed to, it is beneficial to minimize or eliminate reverberant sound.

Though open or free-field environments exist in natural settings, they should be considered a limiting case (i.e., abnormal). Spaces that produce little or no reverberant sound are highly uncommon in the real world. Instead, individuals commonly inhabit locations whose surfaces (floor, ceiling, and walls) reflect sound. Moreover, natural settings often contain surfaces composed of different materials (e.g., concrete, tile, and carpet) each of which can uniquely alter the sound reaching the point of observation. Therefore, it is possible the sound stimulating the ear is notably more complex and different from the sound at the origin. Regardless, it is common for perception and action to occur within settings that contain both direct and reverberant sound and previous research suggests sound reverberations significantly affect auditory spatial perception. Reverberant sound has been shown to alter sound localization judgments (e.g., Hartmann, 1983; Rakerd and Hartmann, 1985; Giguère and Abel, 1993) and the ability to detect and perceive the distance of walls (e.g., Griffin, 1944; Supa et al., 1944; Worchel and Dallenbach, 1947; Kellogg, 1962; Rosenblum et al., 2000). Ashmead et al. (1998), for example, found that observers rely on reverberations in order to walk down a corridor without colliding with the walls.

Despite the difficulty of ascertaining the change in sound that results from the presence of sound reflecting surfaces, future investigations should consider the extent to which individuals can accurately detect the position of an unseen sound source in echoic settings. Knowing how individuals of different ages estimate the position of a sound-producing object in an anechoic space may provide little to no insight into how they perceive sound source position in natural settings. Conducting research in environments commonly inhabited by individuals will provide researchers with a more accurate understanding of the information observers employ when determining sound source position and the extent to which age influences spatial perception.

Uncluttered Vs. Cluttered Setting

The creation of a setting free from reverberation requires the elimination of clutter. Here, clutter refers to objects other than the object that is the target of perception. Laboratory settings often exclude objects not directly necessary for performing an experiment. Experimental settings seemingly contain nothing more than a participant, chair, loudspeaker(s), materials to position the loudspeaker(s), and equipment related to the broadcast of sound. The extent to which previous research has utilized a clutter-free setting is often impossible to determine from descriptions provided in published papers. Aside from describing the experimental setting as anechoic, semi-reverberant, or echoic, authors rarely include information that would permit an understanding of the extent to which the experimental setting is cluttered. Often, the reader assumes clutter has been minimized or eliminated. The minimization and elimination of clutter exist for theoretical reasons. Clutter has the potential of creating reverberant sound that, as mentioned earlier, could dramatically alter the sound reaching the ear. In short, clutter has the potential to alter observer perceptions of sound source location.

Despite the advantages associated with the absence of clutter, real-world settings are rarely clutter free. Everyday settings regularly contain numerous and various objects that are unrelated to the task. The inclusion of clutter allows researchers to investigate auditory spatial perception, and its relationship to chronological age, in a situation akin to that which individuals normally find themselves. The use of cluttered settings enhances the ecological validity of studies and provides insight into the extent to which individuals accurately locate objects in everyday settings. As stated earlier, the inclusion of clutter may provide participants with information (e.g., reverberant sound) that enhances perceptual judgments.

The use of a cluttered setting also provides new avenues of research. A single piece of clutter could either obstruct or occlude a sound-producing object. A sound source is deemed obstructed if the object is small enough that sound can travel around it. If the object is large enough to fully block sound transmission, the sound source is deemed occluded. In real-world settings, individuals are commonly exposed to obstructed or occluded sound-producing objects. Despite the presence of the walls of my office, I can easily use sound to detect the existence of a pedestrian and I am able to detect the direction of their motion (approaching or receding). Based solely on what I hear, I am keenly aware of whether a colleague who is speaking to me is located in my office doorway or their adjacent office. Recently, Russell and Brown (2019), Gordon and Rosenblum (2004), and Kolarik et al. (2016) reported that individuals are highly capable at both detecting and creating occlusion. Given that clutter is common in natural settings, future investigations may wish to include occlusion and/or obstruction, for example, in their experimental designs when investigating auditory spatial perception and when determining the extent to which chronological age affects spatial judgments.

ACOUSTIC FACTORS

Simple Vs. Complex Sounds

Laboratory experiments investigating the impact of age on judgments of sound source position typically involve simple sounds. Out of the 16 studies reviewed earlier in this paper, 15 used pure tones or noise stimuli. Only one study used a complex sound stimulus (a speech segment). Simple sounds are sounds whose acoustic properties (e.g., frequency and amplitude) remain unchanged over time. Examples include musical notes, pure tones, and noise (e.g., white noise, pink noise, and brown noise). The use of simple sounds in laboratory experiments is expected given that signal frequency affects interaural level differences and interaural temporal differences. It is well known that an increase in frequency results in greater interaural level (intensity) differences and that a decrease in frequency results in greater interaural temporal differences. Changes in signal frequency result in changes in interaural differences that, in turn, result in changes in azimuth perception. With respect to auditory distance perception, while signal intensity is clearly a relevant factor, so is signal frequency. Changes in signal frequency have been shown to influence distance judgments (e.g., Blauert, 1997; Brungart, 1999; Kopčo and Shinn-Cunningham, 2011) and the ability of individuals to detect changes in sound level (e.g., Riesz, 1933; Jesteadt et al., 1977), which occurs with a change in actual distance. More specifically, the loss of higher frequency components yields the perception of a more distant sound source. Knowing that signal frequency affects distance perception, it should come as no surprise that the alteration of signal frequency affects the perception of approach or withdrawal. Gordon et al. (2013), for example, determined that a decrease in the center frequency of noise bands resulted in an increase in the accuracy of estimations of sound source arrival. In sum, the use of simple sounds provides researchers with an opportunity to examine the extent to which particular acoustic properties affect spatial perception.

In the real world, simple sounds are rare and complex sounds are highly common. Complex sounds are sounds whose acoustic properties (e.g., frequency, amplitude, amplitude envelopes, and onset and/or offset) vary over time. The vast majority of everyday events (e.g., a door closing, footsteps, and rain hitting the ground) creates complex auditory events. While advantages exist for using simple sounds, the lack of naturally occurring changes in acoustic structure may deprive participants of information necessary to accurately judge sound source location. Schutz and Gillard (2020) reviewed 443 articles and discovered that only 11% of 1,017 of the reported experiments used a dynamically varying sound. While flat tones are more common in laboratory settings, percussive tones are more representative of the sounds individuals normally encounter in real-world settings and, thus, are more often the sounds individuals judge and respond to. The inclusion of a sound that changes in structure has the potential to affect one's perception of the world. Individuals who watched two disks moving toward one another were vastly more likely to report the disks bouncing off one another if a click sound occurred at or near time of collision (Sekuler et al., 1997). More relevant to our discussion of complex sounds, Grassi and Casco (2009) discovered that the disks were perceived as bouncing off one another if a ramped sound (increasing intensity) occurred at the moment of contact. Damped sounds (those decreasing in intensity) created the impression the disks passed through one another. With regard to auditory spatial perception, Rakerd and Hartmann (1986) and Hartmann and Rakerd (1989b) found that the extent of signal onset and offset notably affected the ability of individuals to accurately determine the origin of a sound. These studies and others suggest that future investigations into the influence of age on perception of sound source location should determine the extent to which variations in dynamic sounds affect perceptual reports. The use of dynamic stimuli will serve to enhance our understanding of how it is that we perceived the world.

Brief Vs. Extended Sounds

It is incredibly common for researchers to utilize brief sounds when determining the extent to which chronological age impacts auditory spatial judgments. Of the 16 studies presented earlier, mean stimulus duration was 630.8 ms and only two studies used a stimulus that lasted more than 1 s. If the two exceptional studies are excluded from consideration, mean stimulus duration was a very brief 227.1 ms. The use of brief stimuli can be expected to reduce participation time and therefore minimize or prevent participant fatigue. However, I am unaware of any published papers that advocate the advantage of using brief sounds.

Though researchers commonly use brief stimuli, auditory events in real-world settings commonly last for extended periods of time. The sound of a pedestrian, an automobile, and a conversation, for example, persist for several seconds. Akin to the notion that a Fourier analysis transforms a complex sound into a conglomeration of simple sounds, the argument can be made that auditory events are nothing more than a composite of brief and sometimes repetitive sounds. The sound of a pedestrian, for example, can be considered as nothing more than a collection of a number of discrete and individually identifiable footsteps. Nonetheless, the counterargument can be made that the information serving as the foundation of auditory spatial judgments is the entire event. It is possible even slight differences between repetitive sounds and/or repeated exposure to the same sound may affect spatial perception. As discussed by Gaver (1993), the real world is composed of auditory events that exist over time and it is the entirely of the event that individuals perceive. Empirically speaking, stimulus duration affects auditory spatial judgments. More specifically, a decrease in stimulus duration has been shown to result in a decrease in perceptual accuracy (e.g., Macpherson and Middlebrooks, 2000; Vliegen and Van Opstal, 2004; Yost and Pastore, 2019). For both theoretical and empirical reasons, it is suggested that future investigations into the influence of age on auditory spatial perception should involve prolonged sounds and, if applicable, complete auditory events. By taking such an approach, it is expected our findings will more accurately reflect individual perception as it occurs in real-world settings, have enhanced ecologically validity, and yield a better understanding of how we perceive the world and how age influences those perceptions.

Stationary Vs. Dynamic Sound-Producing Objects

In addition to brief stimuli, investigations into the relationship between age and spatial perception typically involve auditory targets that are stationary during sound presentation. Of the 16 studies presented earlier, all 16 involved stationary sound sources. The use of stationary sources is advantageous since it provides an opportunity for researchers to determine the accuracy with which individuals can determine the location of a target. For example, Mills (1958) and Voss et al. (2004) determined that, depending on the frequency of the stimulus, the minimum audible angle was as low as 1–3°. Oldfield and Parker (1984) determined that the mean absolute error in terms of azimuth was 9° while mean absolute error in elevation was 12°. Similarly, Makous and Middlebrooks (1990) found mean error to be 2° for azimuth and 3.5° for elevation. Use of stationary sound sources also provides researchers with the opportunity to determine the extent to which various acoustic features (e.g., frequency, onset/offset, and intensity) influence spatial judgments. Since a change in observer-source position could dramatically alter the sound contacting the ear, it is imperative to hold source position constant.

As in laboratory settings, real-world settings often contain stationary sound-producing objects. The distance, azimuth, and elevation of a ringing telephone are constant unless acted upon by the individual. The same can be said for a wide variety of everyday objects (e.g., television, computer, and dishwasher). Nonetheless, real-world settings also contain sound sources that are in motion (e.g., motor vehicles, pedestrians, and birds). Use of dynamic sound sources (i.e., sound sources whose position changes over time) allows researchers to investigate the types of spatial judgments individuals frequently make in everyday settings; the kind of judgments not possible with a stationary source. Rosenblum et al. (1987, 1993), Schiff and Oldak (1990), and Silva et al. (2017) found individuals often and somewhat poorly estimate when a sound-producing object will arrive at their location. Neuhoff (1998, 2001) discovered individuals are apparently more sensitive to approaching than receding auditory objects. Neuhoff (2016) further discovered approaching sounds are perceived as moving faster than receding sounds. Recently, Russell and Herl (2021) discovered individuals are capable of distinguishing between auditory events that involved hard (forceful) and soft (gentle) contact. In addition to being ecologically valid, the use of dynamic sound sources provides an opportunity to explore new avenues of research that, in turn, will expand our understanding of how chronological age affects auditory spatial judgments.

SUMMARY

Two purposes existed with respect to the present paper. First, the author wished to review the research investigating the relationship between chronological age and auditory spatial perception. What follows is a brief summary of the findings:

- When examined independently of hearing impairment, only a relatively small number of studies have investigated the degree to which age influences auditory spatial perception.
- Roughly, an equal number of findings suggest an increase in age either adversely affects spatial perception or has little or no influence.
- The adverse effect of age on the perceived sound source location appears to be dependent on a variety of factors (e.g., signal frequency, presence of background noise, and centrally or peripherally located sources).
- Notable differences exist between studies in terms of experimental design (e.g., stimuli, task, and definitions of "older") which may or may not account for differences in findings.
- While the relationships between age and azimuth perception and age and elevation perception have been examined, it appears no known study has examined the impact of observer age on auditory distance perception.

Based on the literature, it is not yet possible to render a firm conclusion about the relationship between age and spatial perception. Additional research is clearly needed.

A second purpose of the present paper was to compare how auditory spatial investigations, including but not limited to those examining the impact of age, are often conducted in laboratory settings and the manner with which individuals perceive sound source position in real-world settings. It is hoped the comparison between laboratory and everyday settings is not interpreted as a criticism of previous research. When possible, the author has presented a rationale as to why experimenters have utilized particular settings, sounds, and methods. As stated throughout the paper, solid theoretical reasons exist for why research has been conducted in the custom it has been. The sole intention of the author is to simply provide alternatives to the traditional approach; alternatives that are based on the environments commonly inhabited by observers, the sounds they are frequently exposed to, and the manner with which they normally respond. What follows is a summary of the notable differences that exist between the laboratory environment and everyday settings:

- Laboratory research commonly involves stationary individuals making verbal judgments of a stationary sound source. Judgments are made in an anechoic, uncluttered setting. With rare exception, verbal judgments reflect one dimension. Sound stimuli are typically brief and simple in nature.
- In everyday settings, individuals are commonly exposed to auditory events that exist for extended periods of time and are complex in acoustic structure. Those sounds occur in echoic, cluttered spaces and can originate at any point across multiple dimensions. In response to a sound, individuals often alter the position of the head and/or body and either make an action-based judgment or behaviorally react to a sound-producing object. The sound-producing object can be stationary or dynamic.

In brief, numerous and vast differences exist between laboratory and real-world settings.

It is the belief of the author that, despite solid theoretical reasons, laboratory research is largely lacking in ecological validity. Research conducted in an ecologically valid manner permits researchers to investigate auditory spatial perception as individuals do in everyday settings. It is believed that an increase in ecological validity will yield results that more accurately reflect the abilities of individuals to judge the position of a sound-producing object. Using a more natural design also provides researchers with an opportunity to better determine the extent to which chronological age influences spatial judgments. Furthermore, research conducted in an ecologically relevant manner provides new avenues for research (e.g., perception of occluded sound sources). Those avenues have direct bearing on what individuals actually do in everyday settings. Finally, it is likely that ecologically based research will have greater external validity and greater construct validity in comparison with research conducted in the traditional manner. Clearly, additional research is needed to confirm or discredit the arguments presented here.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

REFERENCES

Abel, S. M., Giguère, C., Consoli, A., and Papsin, B. C. (2000). The effect of aging on horizontal plane sound localization. *J. Acoust. Soc. Am.* 108, 743–752. doi: 10.1121/1.429607

Abel, S. M., and Hay, V. H. (1996). Sound localization. The interaction of aging, hearing loss and hearing protection. *Scand. Audiol.* 25, 3–12. doi: 10.3109/01050399609047549

Addleman, A. D., Xiong, Y. Z., Nelson, P., and Legge, G. E. (2019). Effects of age and target modality on spatial localization on the horizontal plane. *J. Vis.* 19:117b. doi: 10.1167/19.10.117b

Ashmead, D. H., Davis, D. L., and Northington, A. (1995). Contribution of listeners' approaching motion to auditory distance perception. *J. Exp. Psychol. Hum. Percept. Perform.* 21, 239–256. doi: 10.1037/0096-1523.21.2.239

Ashmead, D. H., Wall, R. S., Eaton, S. B., Ebinger, K. A., Snook-Hill, M.-M., Guth, D. A., et al. (1998). Echolocation reconsidered: using spatial variations in the ambient sound field to guide locomotion. *J. Visual Impairment Blindness* 92, 615–632. doi: 10.1177/0145482X9809200905

Bednar, A., and Lalor, E. C. (2020). Where is the cocktail party? Decoding locations of attended and unattended moving sound sources using EEG. *NeuroImage* 205:116283. doi: 10.1016/j.neuroimage.2019.116283

Best, V., Carlile, S., Jin, C., and van Schaik, A. (2005). The role of high frequencies in speech localization. *J. Acoust. Soc. Am.* 118, 353–363. doi: 10.1121/1.1926107

Bingham, G. P., Schmidt, R. C., and Rosenblum, L. D. (1989). Hefting for a maximum distance throw: a smart perceptual mechanism. *J. Exp. Psychol. Hum. Percept. Perform.* 15, 507–528. doi: 10.1037/0096-1523.15.3.507

Blauert, J. (1997). *Spatial Hearing: The Psychophysics of Human Sound Localization*. Cambridge, MA: MIT Press.

Braly, A. M., DeLucia, P., and Oberfeld, D. (2021). Does affective content of sounds affect auditory time-to-collision estimation? *Auditory Percept. Cognit.* 4, 1–23. doi: 10.1080/25742442.2021.1997064

Briley, P. M., and Summerfield, A. Q. (2014). Age-related deterioration of the representation of space in human auditory cortex. *Neurobiol. Aging* 35, 633–644. doi: 10.1016/j.neurobiolaging.2013.08.033

Brimijoin, W. O., and Akeroyd, M. A. (2014). The moving minimum audible angle is smaller during self motion than during source motion. *Front. Neurosci.* 8:273. doi: 10.3389/fnins.2014.00273

Bronkhorst, A. W., and Houtgast, T. (1999). Auditory distance perception in rooms. *Nature* 397, 517–520. doi: 10.1038/17374

Brungart, D. S. (1999). Auditory localization of nearby sources III. Stimulus effects. *J. Acoust. Soc. Am.* 106, 3589–3602. doi: 10.1121/1.428212

Brungart, D. S., Cohen, J. I., Zion, D. J., and Romigh, G. D. (2017). The localization of non-individualized virtual sounds by hearing impaired listeners. *J. Acoust. Soc. Am.* 141, 2870–2881. doi: 10.1121/1.4979462

Buchholz, J. M., and Best, V. (2020). Speech detection and localization in a reverberant multitalker environment by normal-hearing and hearing-impaired listeners. *J. Acoust. Soc. Am.* 147, 1469–1477. doi: 10.1121/10.0000844

Butler, R. A., and Humanski, R. A. (1992). Localization of sound in the vertical plane with and without high-frequency spectral cues. *Percept. Psychophys.* 51, 182–186. doi: 10.3758/BF03212242

Butler, A. A., Lord, S. R., and Fitzpatrick, R. C. (2011). Reach distance but not judgment error is associated with falls in older people. *J. Gerontol. A Biol. Sci. Med. Sci.* 66A, 896–903. doi: 10.1093/gerona/glr071

Cabe, P. A., and Pittenger, J. B. (2000). Human sensitivity to acoustic information from vessel filling. *J. Exp. Psychol. Hum. Percept. Perform.* 26, 313–324. doi: 10.1037//0096-1523.26.1.313

Carello, C., Anderson, K. L., and Kunkler-Peck, A. J. (1998). Perception of object length by sound. *Psychol. Sci.* 9, 211–214. doi: 10.1111/1467-9280.00040

Carello, C., Grosofsky, A., Reichel, F. D., Solomon, H. Y., and Turvey, M. T. (1989). Visually perceiving what is reachable. *Ecol. Psychol.* 1, 27–54. doi: 10.1207/s15326969eco0101_3

Carlile, S. (ed.) (1996). "Auditory space," in *Virtual Auditory Space: Generation and Applications. Neuroscience Intelligence Unit* (Berlin, Heidelberg: Springer).

Cole, W. G., Chan, G. L., Vereijken, B., and Adolph, K. E. (2013). Perceiving affordances for different motor skills. *Exp. Brain Res.* 225, 309–319. doi: 10.1007/s00221-012-3328-9

Day, B. M., Wagman, J. B., and Smith, P. J. (2015). Perception of maximum stepping and leaping distance: stepping affordances as a special case of leaping affordances. *Acta Psychol.* 158, 26–35. doi: 10.1016/j.actpsy.2015.03.010

Dobreva, M. S., O'Neill, W. E., and Paige, G. D. (2011). Influence of aging on human sound localization. *J. Neurophysiol.* 105, 2471–2486. doi: 10.1152/jn.00951.2010

Doty, R. L., Petersen, I., Mensah, N., and Christensen, K. (2011). Genetic and environmental influences on odor identification ability in the very old. *Psychol. Aging* 26, 864–871. doi: 10.1037/a0023263

Fink, M., Churan, J., and Wittmann, M. (2005). Assessment of auditory temporal-order thresholds – a comparison of different measurement procedures and the influences of age and gender. *Restor. Neurol. Neurosci.* 23, 281–296.

Franchak, J. M., van der Zalm, D. J., and Adolph, K. E. (2010). Learning by doing: action performance facilitates affordance perception. *Vis. Res.* 50, 2758–2765. doi: 10.1016/j.visres.2010.09.019

Freed, D. J. (1990). Auditory correlates of perceived mallet hardness for a set of recorded percussive sound events. *J. Acoust. Soc. Am.* 87, 311–322. doi: 10.1121/1.399298

Freigang, C., Schmiedchen, K., Nitsche, I., and Rübsamen, R. (2014). Free-field study on auditory localization and discrimination performance in older adults. *Exp. Brain Res.* 232, 1157–1172. doi: 10.1007/s00221-014-3825-0

Fukunaga, A., Uematsu, H., and Sugimoto, K. (2005). Influences of aging on taste perception and oral somatic sensation. *J. Gerontol. Ser. A Biol. Sci. Med. Sci.* 60, 109–113. doi: 10.1093/gerona/60.1.109

Gaver, W. W. (1993). What in the world do we hear? An ecological approach to auditory event perception. *Ecol. Psychol.* 5, 1–29. doi: 10.1207/s15326969eco0501_1

Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. Boston, MA: Houghton-Mifflin Company.

Giguère, C., and Abel, S. M. (1993). Sound localization: effects of reverberation time, speaker array, stimulus frequency, and stimulus rise/decay. *J. Acoust. Soc. Am.* 94, 769–776. doi: 10.1121/1.408206

Giordano, B. L., and McAdams, S. (2006). Material identification of real impact sounds: effects of size variation in steel, glass, wood, and plexiglass plates. *J. Acoust. Soc. Am.* 119, 1171–1181. doi: 10.1121/1.2149839

Gordon, M. S., and Rosenblum, L. D. (2004). Perception of sound-obstructing surfaces using body-scaled judgments. *Ecol. Psychol.* 16, 87–113. doi: 10.1207/s15326969eco1602_1

Gordon, M. S., Russo, F. A., and MacDonald, E. (2013). Spectral information for detection of acoustic time to arrival. *Atten. Percept. Psychophys.* 75, 738–750. doi: 10.3758/s13414-013-0424-2

Grassi, M. (2005). Do we hear size or sound? Balls dropped on plates. *Percept. Psychophys.* 67, 274–284. doi: 10.3758/bf03206491

Grassi, M., and Casco, C. (2009). Audiovisual bounce-inducing effect: attention alone does not explain why the discs are bouncing. *J. Exp. Psychol. Hum. Percept. Perform.* 35, 235–243. doi: 10.1037/a0013031

Grassi, M., Pastore, M., and Lemaitre, G. (2013). Looking at the world with your ears: how do we get the size of an object from its sound? *Acta Psychol.* 143, 96–104. doi: 10.1016/j.actpsy.2013.02.005

Griffin, D. R. (1944). Echolocation by blind men, bats, and radar. *Science* 100, 589–590. doi: 10.1126/science.100.2609.589

Hajnal, A., Olavarria, C. X., Surber, T., Clark, J. D., and Doyon, J. K. (2020). Comparison of two psychophysical methods across visual and haptic perception of stand-on-ability. *Psychol. Res.* 84, 602–610. doi: 10.1007/s00426-018-1076-6

Hartmann, W. M. (1983). Localization of sound in rooms. *J. Acoust. Soc. Am.* 74, 1380–1391. doi: 10.1121/1.390163

Hartmann, W. M., and Rakerd, B. (1989a). On the minimum audible angle: a decision theory approach. *J. Acoust. Soc. Am.* 85, 2031–2041. doi: 10.1121/1.397855

Hartmann, W. M., and Rakerd, B. (1989b). Localization of sound in rooms. IV: The Franssen effect. *J. Acoust. Soc. Am.* 86, 1366–1373. doi: 10.1121/1.398696

Heffner, H. E., and Heffner, R. S. (1998). "Hearing," in *Comparative Psychology: A Handbook. 1st Edn*. eds. G. Greenberg and M. M. Haraway (London: Routledge), 290–303.

Heffner, H. E., and Heffner, R. S. (2014). "The behavioral study of mammalian hearing," in *Perspectives on Auditory Research. 1st Edn*. eds. A. N. Popper and R. R. Fay (Berlin: Springer), 269–285.

Heffner, H. E., and Heffner, R. S. (2016). The evolution of mammalian sound localization. *Acoust. Today* 12, 20–27.

Heffner, H. E., and Heffner, R. S. (2018). "The evolution of mammalian hearing," in *To the Ear and back – Advances in Auditory Biophysics, American Institute of Physics Conference Proceedings*. eds. C. Bergevin and S. Puria (College Park, MD: American Institute of Physics Publishing), 30001–130008.

Higuchi, T., Takada, H., Matsuura, Y., and Imanaka, K. (2004). Visual estimation of spatial requirements for locomotion in novice wheelchair users. *J. Exp. Psychol. Appl.* 10, 55–66. doi: 10.1037/1076-898X.10.1.55

Hofman, P. M., Van Riswick, J. G. A., and Van Opstel, A. J. (1998). Relearning sound localization with new ears. *Nat. Neurosci.* 1, 417–421. doi: 10.1038/1633

Jayakody, D. M. P., Friedland, P. L., Martins, R. N., and Sohrabi, H. R. (2018). Impact of aging on the auditory system and related cognitive functions: a narrative review. *Front. Neurosci.* 12, 1–16. doi: 10.3389/fnins.2018.00125

Jesteadt, W., Wier, C. C., and Green, D. M. (1977). Intensity discrimination as a function of frequency and sensation level. *J. Acoust. Soc. Am.* 61, 169–177. doi: 10.1121/1.381278

Kaneda, H., Maeshima, K., Goto, N., Kobayakawa, T., Ayabe-Kanamura, S., and Saito, S. (2000). Decline in taste and odor discrimination abilities with age, and relationship between gustation and olfaction. *Chem. Senses* 25, 331–337. doi: 10.1093/chemse/25.3.331

Kellogg, W. N. (1962). Sonar system of the blind. *Science* 137, 399–404. doi: 10.1126/science.137.3528.399

Kinsella-Shaw, J. M., Shaw, B., and Turvey, M. T. (1992). Perceiving 'walk-on-able' slopes. *Ecol. Psychol.* 4, 223–239. doi: 10.1207/s15326969eco0404_2

Kleinman, J. M., and Brodzinsky, D. M. (1978). Haptic exploration in young, middle-aged, and elderly adults. *J. Gerontol.* 33, 521–527. doi: 10.1093/geronj/33.4.521

Kolarik, A. J., Scarfe, A. C., Moore, B. C. J., and Pardhan, S. (2016). An assessment of auditory-guided locomotion in an obstacle circumvention task. *Exp. Brain Res.* 234, 1725–1735. doi: 10.1007/s00221-016-4567-y

Kołodziejczyk, I., and Szelag, E. (2008). Auditory perception of temporal order in centenarians in comparison with young and elderly subjects. *Acta Neurobiol. Exp.* 68, 373–381

Konczak, J., Meeuwsen, H. J., and Cress, M. E. (1992). Changing affordances in stair climbing: the perception of maximum climbability in young and older adults. *J. Exp. Psychol. Hum. Percept. Perform.* 18, 691–697. doi: 10.1037//0096-1523.18.3.691

Kopčo, N., and Shinn-Cunningham, B. G. (2011). Effect of stimulus spectrum on distance perception for nearby sources. *J. Acoust. Soc. Am.* 130, 1530–1541. doi: 10.1121/1.3613705

Kozhevnikova, E. V., and Zhukov, S. Y. (1990). Features of auditory evaluation of an approaching and receding sound source. *Sens. Syst.* 3, 194–201.

Kunkler-Peck, A. J., and Turvey, M. T. (2000). Hearing shape. *J. Exp. Psychol. Hum. Percept. Perform.* 26, 279–294. doi: 10.1037/0096-1523.26.1.279

Lederman, S. J. (1979). Auditory texture perception. *Perception* 8, 93–103. doi: 10.1068/p080093

Li, X., Logan, R., and Pastore, R. (1991). Perception of acoustic source characteristics: walking sounds. *J. Acoust. Soc. Am.* 90, 3036–3049. doi: 10.1121/1.401778

Macpherson, E. A., and Middlebrooks, J. C. (2000). Localization of brief sounds: effects of level and background noise. *J. Acoust. Soc. Am.* 108, 1834–1849. doi: 10.1121/1.1310196

Makous, J. C., and Middlebrooks, J. C. (1990). Two-dimensional sound localization by human listeners. *J. Acoust. Soc. Am.* 87, 2188–2200. doi: 10.1121/1.399186

Malek, E. A., and Wagman, J. B. (2008). Kinetic potential influences visual and remote haptic perception of affordances for standing on an inclined surface. *Q. J. Exp. Psychol.* 61, 1813–1826. doi: 10.1080/17470210701712978

Mark, L. S. (1987). Eyeheight-scaled information about affordances: a study of sitting and stair climbing. *J. Exp. Psychol. Hum. Percept. Perform.* 13, 361–370. doi: 10.1037//0096-1523.13.3.361

Mark, L. S., Balliett, J. A., Craver, K. D., Douglas, S. D., and Fox, T. (1990). What an actor must do in order to perceive the affordance for sitting. *Ecol. Psychol.* 2, 325–366. doi: 10.1207/s15326969eco0204_2

Mershon, D. H., Ballenger, W. L., Little, A. D., McMurtry, P. L., and Buchanan, J. L. (1989). Effects of room reflectance and background noise on perceived auditory distance. *Perception* 18, 403–416. doi: 10.1068/p180403

Mershon, D. H., and King, L. E. (1975). Intensity and reverberation as factors in the auditory perception of egocentric distance. *Percept. Psychophys.* 18, 409–415. doi: 10.3758/BF03204113

Middlebrooks, J. C., and Green, D. M. (1991). Sound localization by human listeners. *Annu. Rev. Psychol.* 42, 135–159. doi: 10.1146/annurev.ps.42.020191.001031

Mills, A. W. (1958). On the minimum audible angle. *J. Acoust. Soc. Am.* 30, 237–246. doi: 10.1121/1.1909553

Murphy, C., Schubert, C. R., Cruickshanks, K. J., Klein, B. E. K., Klein, R., and Nondahl, D. M. (2002). Prevalence of olfactory impairment in older adults. *J. Am. Med. Assoc.* 288, 2307–2312. doi: 10.1001/jama.288.18.2307

Musicant, A. D., and Butler, R. A. (1984). The psychophysical basis of monaural localization. *Hear. Res.* 14, 185–190. doi: 10.1016/0378-5955(84)90017-0

Nambu, I., Ebisawa, M., Kogure, M., Yano, S., Hokari, H., and Wada, Y. (2013). Estimating the intended sound direction of the user: toward an auditory brain-computer interface using out-of-head sound localization. *PLoS One* 8:e57174. doi: 10.1371/journal.pone.0057174

Neher, T., Laugesen, S., Jensen, N. S., and Kragelund, L. (2011). Can basic auditory and cognitive measures predict hearing-impaired listeners' localization and spatial speech recognition abilities? *J. Acoust. Soc. Am.* 130, 1542–1558. doi: 10.1121/1.3608122

Neuhoff, J. G. (1998). Perceptual bias for rising tones. *Nature* 395, 123–124. doi: 10.1038/25862

Neuhoff, J. G. (2001). An adaptive bias in the perception of looming auditory motion. *Ecol. Psychol.* 13, 87–110. doi: 10.1207/S15326969ECO1302_2

Neuhoff, J. G. (2016). Looming sounds are perceived as faster than receding sounds. *Cognit. Res. Principles Implications* 1:15. doi: 10.1186/s41235-016-0017-4

Norman, J. F., Adkins, O. C., Hoyng, S. C., Dowell, C. J., Pedersen, L. E., and Gilliam, A. N. (2016). Aging and the haptic perception of material properties. *Perception* 45, 1387–1398. doi: 10.1177/0301006616659073

Oldfield, S. R., and Parker, S. P. (1984). Acuity of sound localisation: a topography of auditory space. I. Normal hearing conditions. *Perception* 13, 581–600. doi: 10.1068/p130581

Olofsson, J. K., Ekström, I., Larsson, M., and Nordin, S. (2021). Olfaction and aging: a review of the current state of research and future directions. *i-Perception* 12, 1–24. doi: 10.1177/20416695211020331

O'Neill, S. M., and Russell, M. K. (2017). Impact of postural stability and modality on the perception of passage and surface climbing. *Ecol. Psychol.* 29, 54–68. doi: 10.1080/10407413.2017.1270153

Otte, R. J., Agterberg, M. J., Van Wanrooij, M. M., Snik, A. F., and Van Opstal, A. J. (2013). Age-related hearing loss and ear morphology affect vertical but not horizontal sound-localization performance. *J. Assoc. Res. Otolaryngol.* 14, 261–273. doi: 10.1007/s10162-012-0367-7

Perrett, S., and Noble, W. (1997). The contribution of head motion cues to localization of low-pass noise. *Percept. Psychophys.* 59, 1018–1026. doi: 10.3758/bf03205517

Perrott, D. R., and Saberi, K. (1990). Minimum audible angle thresholds for sources varying in both elevation and azimuth. *J. Acoust. Soc. Am.* 87, 1728–1731. doi: 10.1121/1.399421

Pörschmann, C., and Störig, C. (2009). Investigations into the velocity and distance perception of moving sound sources. *Acta Acust. United Acustica* 95, 696–706. doi: 10.3813/AAA.918198

Rakerd, B., and Hartmann, W. M. (1985). Localization of sound in rooms, II: the effects of a single reflecting surface. *J. Acoust. Soc. Am.* 78, 524–533. doi: 10.1121/1.392474

Rakerd, B., and Hartmann, W. M. (1986). Localization of sound in rooms, III: onset and duration effects. *J. Acoust. Soc. Am.* 80, 1695–1706. doi: 10.1121/1.394282

Riehm, C., Chemero, A., Silva, P. L., and Shockley, K. (2019). Virtual auditory aperture passability. *Exp. Brain Res.* 237, 191–200. doi: 10.1007/s00221-018-5407-z

Riesz, R. (1933). The relationship between loudness and the minimum perceptible increment of intensity. *J. Acoust. Soc. Am.* 4, 211–216. doi: 10.1121/1.1915601

Rodrigues, S. T., Galvão, N. C., and Gotardi, G. C. (2014). Visual estimation of apertures for wheelchair locomotion in novices: perceptual judgment and motor practice. *Psychol. Neurosci.* 7, 331–340. doi: 10.3922/j.psns.2014.040

Rodríguez Valiente, A., Trinidad, A., García Berrocal, J. R., Górriz, C., and Ramírez Camacho, R. (2014). Extended high-frequency (9-20 kHz) audiometry reference thresholds in 645 healthy subjects. *Int. J. Audiol.* 53, 531–545. doi: 10.3109/14992027.2014.893375

Roffler, S. K., and Butler, R. A. (1968). Factors that influence the localization of sound in the vertical plane. *J. Acoust. Soc. Am.* 43, 1255–1259. doi: 10.1121/1.1910976

Rønne, F. M., Laugesen, S., Jensen, N. S., and Pedersen, J. H. (2016). Minimum audible angles measured with simulated normally-sized and oversized pinnas for normal-hearing and hearing-impaired test subjects. *Adv. Exp. Med. Biol.* 894, 207–217. doi: 10.1007/978-3-319-25474-6_22

Rosenblum, L. D., Carello, C., and Pastore, R. E. (1987). Relative effectiveness of three stimulus variables for locating a moving sound source. *Perception* 16, 175–186. doi: 10.1068/p160175

Rosenblum, L. D., Gordon, M. S., and Jarquin, L. (2000). Echolocating distance by moving and stationary listeners. *Ecol. Psychol.* 12, 181–206. doi: 10.1207/S15326969ECO1203_1

Rosenblum, L. D., Wuestefeld, A. P., and Anderson, K. L. (1996). Auditory reachability: an affordance approach to the perception of sound source distance. *Ecol. Psychol.* 8, 1–24. doi: 10.1207/s15326969eco0801_1

Rosenblum, L. D., Wuestefeld, A. P., and Saldaña, H. M. (1993). Auditory looming perception: influences on anticipatory judgments. *Perception* 22, 1467–1482. doi: 10.1068/p221467

Russell, M. K. (2020). Impact of qualitative and quantitative visual experience on the auditory perception of apertures. *Auditory Percept. Cognit.* 3, 124–157. doi: 10.1080/25742442.2021.1888371

Russell, M. K., and Brown, S. (2019). Using sound to create and detect occlusion of an unseen sound source. *Auditory Percept. Cognit.* 2, 207–229. doi: 10.1080/25742442.2020.1773731

Russell, M. K., and Herl, C. (2021). Influence of event duration and impact intensity on the auditory perception of contact severity. *Auditory Percept. Cognit.* 4, 49–59. doi: 10.1080/25742442.2021.1965853

Russell, M. K., and Schneider, A. L. (2006). Sound source perception in a two-dimensional setting: comparison of action and nonaction-based response tasks. *Ecol. Psychol.* 18, 223–237. doi: 10.1207/s15326969eco1803_4

Russell, M. K., and Turvey, M. T. (1999). Auditory perception of unimpeded passage. *Ecol. Psychol.* 11, 175–188. doi: 10.1207/s15326969eco1102_3

Savel, S. (2009). Individual differences and left/right asymmetries in auditory space perception. I. Localization of low-frequency sounds in free field. *Hear. Res.* 255, 142–154. doi: 10.1016/j.heares.2009.06.013

Schiff, W., and Oldak, R. (1990). Accuracy of judging time to arrival: effects of modality, trajectory, and gender. *J. Exp. Psychol. Hum. Percept. Perform.* 16, 303–316. doi: 10.1037//0096-1523.16.2.303

Schutz, M., and Gillard, J. (2020). On the generalization of tones: a detailed exploration of non-speech auditory perception stimuli. *Sci. Rep.* 10:9520. doi: 10.1038/s41598-020-63132-2

Sekuler, R., Sekuler, A., and Lau, R. (1997). Sound alters visual motion perception. *Nature* 385, 308. doi: 10.1038/385308a0

Silva, R. M., Lamas, J., Silva, C. C., Coello, Y., Mouta, S., and Santos, J. A. (2017). Judging time-to-passage of looming sounds: evidence for the use of distance-based information. *PLoS One* 12:e0177734. doi: 10.1371/journal.pone.0177734

Soames, R. W., and Raper, S. A. (1992). The influence of moving auditory fields on postural sway behavior in man. *Eur. J. Appl. Physiol. Occup. Physiol.* 65, 241–245. doi: 10.1007/BF00705088

Speigle, J. M., and Loomis, J. M. (1993). "Auditory distance perception by translating observers." in *Proceedings of 1993 IEEE Research Properties in Virtual Reality Symposium*; October 25, 1993; San Jose: CA, USA, 92–99.

Stoffregen, T. A., Ito, K., Hove, P., Yank, J. R., and Bardy, B. G. (2019). The postural responses to a moving environment of adults who are blind. *J. Visual Impairment Blindness* 104, 73–83. doi: 10.1177/0145482X1010400203

Stoffregen, T. A., Villard, S., Kim, C., Ito, K., and Bardy, B. G. (2009b). Coupling of head and body movement with motion of the audible environment. *J. Exp. Psychol. Hum. Percept. Perform.* 35, 1221–1231. doi: 10.1037/a0014251

Stoffregen, T. A., Yang, C.-M., Giveans, M. R., Flanagan, M., and Bardy, B. G. (2009a). Movement in the perception of an affordance for wheelchair locomotion. *Ecol. Psychol.* 21, 1–36. doi: 10.1080/10407410802626001

Supa, M., Cotzin, M., and Dallenbach, K. M. (1944). "Facial vision": the perception of obstacles by the blind. *Am. J. Psychol.* 57, 133–183. doi: 10.2307/1416946

Szymaszek, A., Szelag, E., and Sliwowska, M. (2006). Auditory perception of temporal order in humans: the effect of age, gender, listener practice and stimulus presentation mode. *Neurosci. Lett.* 403, 190–194. doi: 10.1016/j.neulet.2006.04.062

Thompson, L. W., Axelrod, S., and Cohen, L. D. (1965). Senescence and visual identification of tactual-kinesthetic forms. *J. Gerontol.* 20, 244–249. doi: 10.1093/geronj/20.2.244

Trapeau, R., and Schönwiesner, M. (2018). The encoding of sound source elevation in the human auditory cortex. *J. Neurosci.* 38, 3252–3264. doi: 10.1523/JNEUROSCI.2530-17.2018

Vernat, J.-P., and Gordon, M. S. (2010). Indirect interception actions by blind and visually impaired perceivers: echolocation for interceptive actions. *Scand. J. Psychol.* 51, 75–83. doi: 10.1111/j.1467-9450.2009.00722.x

Vliegen, J., and Van Opstal, A. J. (2004). The influence of duration and level on human sound localization. *J. Acoust. Soc. Am.* 115, 1705–1713. doi: 10.1121/1.1687423

Voss, P., Lassonde, M., Gougoux, F., Fortin, M., Guillemot, J. P., and Lepore, F. (2004). Early- and late-onset blind individuals show supra-normal auditory abilities in far-space. *Curr. Biol.* 14, 1734–1738. doi: 10.1016/j.cub.2004.09.051

Wagman, J. B., Cialdella, V. T., and Stoffregen, T. A. (2019). Higher order affordances for reaching: perception and performance. *Q. J. Exp. Psychol.* 72, 1200–1211. doi: 10.1177/1747021818784403

Wagman, J. B., and Morgan, L. L. (2010). Nested prospectivity in perception: perceived maximum reaching height reflects anticipated changes in reaching ability. *Psychon. Bull. Rev.* 17, 905–909. doi: 10.3758/PBR.17.6.905

Warren, W. (1984). Perceiving affordances: visual guidance of stair climbing. *J. Exp. Psychol. Hum. Percept. Perform.* 10, 683–703. doi: 10.1037//0096-1523.10.5.683

Warren, W. H. Jr., Kim, E. E., and Husney, R. (1987). The way the ball bounces: visual and auditory perception of elasticity and control of the bounce pass. *Perception* 16, 309–336. doi: 10.1068/p160309

Warren, W. H. Jr., and Verbrugge, R. R. (1984). Auditory perception of breaking and bouncing events: A case study in ecological acoustics. *J. Exp. Psychol. Hum. Percept. Perform.* 10, 704–712. doi: 10.1037//0096-1523.10.5.704

Warren, W. H. Jr., and Whang, S. (1987). Visual guidance of walking through apertures: body-scaled information for affordances. *J. Exp. Psychol. Hum. Percept. Perform.* 13, 371–383. doi: 10.1037//0096-1523.13.3.371

Whitmer, W. M., McShefferty, D., Levy, S. C., Naylor, G., and Edwards, B. (2021). Changes in orientation behavior due to extended high-frequency (5 to 10 kHz) spatial cues. *Ear Hear.* 9, 1–9. doi: 10.1097/AUD.0000000000001113

Whitmer, W. M., Seeber, B. U., and Akeroyd, M. A. (2014). The perception of apparent auditory source width in hearing-impaired adults. *J. Acoust. Soc. Am.* 135, 3548–3559. doi: 10.1121/1.4875575

Wightman, F. L., and Kistler, D. J. (1999). Resolution of front-back ambiguity in spatial hearing by listener and source movement. *J. Acoust. Soc. Am.* 105, 2841–2853. doi: 10.1121/1.426899

Worchel, P., and Dallenbach, K. M. (1947). "Facial vision": perception of obstacles by the deaf-blind. *Am. J. Psychol.* 60, 502–553. doi: 10.2307/1417725

Yan, J.-J., Zhou, L., Xie, C.-X., Campos, J., and Sun, H.-J. (2007). Visual and auditory processing of distance and the time-to-collision of an approaching object. *J. Vis.* 7:754. doi: 10.1167/7.9.754

Yost, W. A., and Pastore, M. T. (2019). Individual listener differences in azimuthal front-back reversals. *J. Acoust. Soc. Am.* 146, 2709–2715. doi: 10.1121/1.5129555

Zahorik, P., Brungart, D. S., and Bronkhorst, A. W. (2005). Auditory distance perception in humans: A summary of past and present research. *Acta Acust. United Acust.* 91, 409–420.

Zhang, C., and Wang, X. (2017). Initiation of the age-related decline of odor identification in humans: a meta-analysis. *Ageing Res. Rev.* 40, 45–50. doi: 10.1016/j.arr.2017.08.004

Zhong, X., and Yost, W. A. (2013). Relationship between postural stability and spatial hearing. *J. Am. Acad. Audiol.* 24, 782–788. doi: 10.3766/jaaa.24.9.3

Zonooz, B., Arani, E., Körding, K. P., Remco Aalbers, P. A. T., Celikel, T., and Van Opstal, J. A. (2019). Spectral weighting underlies perceived sound elevation. *Sci. Rep.* 9:1642. doi: 10.1038/s41598-018-37537-z

Conflict of Interest: 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.

frontiers
in Psychology

ORIGINAL RESEARCH
published: 17 February 2022
doi: 10.3389/fpsyg.2022.816864

Speech Perception in Older Adults: An Interplay of Hearing, Cognition, and Learning?

Liat Shechter Shvartzman, Limor Lavie* and Karen Banai**

The Auditory Cognition Lab, Department of Communication Sciences and Disorders, University of Haifa, Haifa, Israel

Older adults with age-related hearing loss exhibit substantial individual differences in speech perception in adverse listening conditions. We propose that the ability to rapidly adapt to changes in the auditory environment (i.e., perceptual learning) is among the processes contributing to these individual differences, in addition to the cognitive and sensory processes that were explored in the past. Seventy older adults with age-related hearing loss participated in this study. We assessed the relative contribution of hearing acuity, cognitive factors (working memory, vocabulary, and selective attention), rapid perceptual learning of time-compressed speech, and hearing aid use to the perception of speech presented at a natural fast rate (fast speech), speech embedded in babble noise (speech in noise), and competing speech (dichotic listening). Speech perception was modeled as a function of the other variables. For fast speech, age [odds ratio (OR) = 0.79], hearing acuity (OR = 0.62), pre-learning (baseline) perception of time-compressed speech (OR = 1.47), and rapid perceptual learning (OR = 1.36) were all significant predictors. For speech in noise, only hearing and pre-learning perception of time-compressed speech were significant predictors (OR = 0.51 and OR = 1.53, respectively). Consistent with previous findings, the severity of hearing loss and auditory processing (as captured by pre-learning perception of time-compressed speech) was strong contributors to individual differences in fast speech and speech in noise perception. Furthermore, older adults with good rapid perceptual learning can use this capacity to partially offset the effects of age and hearing loss on the perception of speech presented at fast conversational rates. Our results highlight the potential contribution of dynamic processes to speech perception.

Keywords: perceptual learning, degraded speech, hearing aids, aging, age-related hearing loss

OPEN ACCESS

Edited by:
Leah Fostick,
Ariel University, Israel

Reviewed by:
Yang Zhang,
University of Minnesota Health Twin Cities, United States
Carine Signoret,
Linköping University, Sweden

****Correspondence:***
Liat Shechter Shvartzman
liat.shechter.86@gmail.com
Limor Lavie
llavie@welfare.haifa.ac.il
Karen Banai
kbanai@research.haifa.ac.il

Specialty section:
This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Psychology

Received: *17 November 2021*
Accepted: *26 January 2022*
Published: *17 February 2022*

Citation:
Shechter Shvartzman L, Lavie L and Banai K (2022) Speech Perception in Older Adults: An Interplay of Hearing, Cognition, and Learning? Front. Psychol. 13:816864. doi: 10.3389/fpsyg.2022.816864

INTRODUCTION

Aging is often accompanied by sensorineural hearing loss (presbycusis) and poor speech perception in daily listening environments (Committee on Hearing, Bioacoustics and Biomechanics (CHABA), 1988; Humes, 1996; Morrell et al., 1996; Pichora-Fuller, 1997; Gordon-Salant and Fitzgibbons, 2001; Dubno et al., 2008), especially under adverse listening conditions (e.g., in the presence of fast speech or competing noise; Pichora-Fuller and Singh, 2006; Schneider et al., 2010). There is tremendous variability in degraded speech perception among older adults. This variability is associated with sensory and cognitive factors (Committee on Hearing, Bioacoustics and Biomechanics (CHABA), 1988; Souza, 2016), as well as with individual differences in perceptual learning for speech (Karawani et al., 2017; Manheim et al., 2018; Rotman et al., 2020b). Hearing aids are the most common rehabilitation for speech perception difficulties in older adults with age-related hearing loss (Souza, 2016). However, like their non-hearing aid using peers, older adults who use hearing aids also vary widely on measures of speech perception. We hypothesize that the same factors that account for individual differences in degraded speech processing in adults with presbycusis are likely responsible for some of the variability in speech perception performance observed among hearing aid users. Therefore, the overall aim of the current study is to assess the relative contribution of sensory (i.e., hearing acuity) and cognitive factors (working memory, vocabulary, and selective attention), rapid perceptual learning, and the use of hearing aids to the identification of different types of degraded speech among older adults. We used three speech tasks—fast speech, speech in babble noise, and competing speech—which represent different challenges that can be encountered in daily listening situations, and which are known to pose difficulties for older adults with hearing loss (for review see Humes et al., 2012). The effects of cognitive factors, learning, and hearing aids might differ across these different conditions. Whereas the challenges associated with fast speech result from source degradation (speaking rapidly changes the temporal and spectral characteristics of speech, Koreman, 2006), the challenges associated with speech in babble noise and competing speech are associated with the listening environment (transmission degradation according to the terminology proposed by Mattys et al., 2012).

Speech Perception in Age-related Hearing Loss

Age-related hearing loss is a primary contributor to speech perception difficulties in older adults (e.g., Humes and Roberts, 1990; Jerger et al., 1991; Humes, 2002). Individuals with age-related sensorineural hearing loss often require favorable signal to noise ratios to recognize speech due to elevated hearing thresholds (Killion, 1997). Reduced audibility (e.g., Gates and Mills, 2005), impaired temporal synchrony (e.g., Hopkins and Moore, 2007, 2011), and broadening of auditory filters (e.g., Peters and Moore, 1992; Leek and Summers, 1993) have also been suggested to account for the connection between age-related hearing loss and reduced perception of speech in noisy environments. Overall, it is estimated that these sensory factors account for 50–90% of individual differences in speech perception (for review, see Humes and Dubno, 2010).

When listening to connected speech (i.e., utterances longer than one word such as sentences or longer units of speech) older adults with age-related hearing loss often have perceptual difficulties with rapid speech rates (e.g., Tun, 1998; Gordon-Salant and Fitzgibbons, 2001; Wingfeld et al., 2006), in the presence of background noise (e.g., Pichora-Fuller et al., 1995; Helfer and Freyman, 2008) or in the presence of competing speech in dichotic listening situations (e.g., Jerger et al., 1994, 1995; Roup et al., 2006). However, auditory factors might be insufficient to explain individual differences in these situations because listeners with identical audiograms can have vastly different speech perception abilities (Luterman et al., 1966; Phillips et al., 2000; Schneider and Pichora-Fuller, 2001; Pichora-Fuller and Souza, 2003). Even when matched for audiological factors, older adults often find it more difficult than their young counterparts to perceive and comprehend speech in adverse listening situations (Gordon-Salant and Fitzgibbons, 1993; Needleman and Crandell, 1995).

The contribution of audiometric thresholds to speech perception tends to be larger in relatively easy conditions (e.g., identifying words in a quiet background) than in more challenging ones (e.g., with temporally distorted speech and speech in noise; Gordon-Salant and Fitzgibbons, 1993, 2001; Humes, 2007). Furthermore, whereas audiometric factors typically allow reasonably accurate predictions of speech in quiet, using auditory thresholds often leads to over estimation of performance of speech in noise (Dubno et al., 1984; Schum et al., 1991; Hargus and Gordon-Salant, 1995). Thus, once a task becomes more demanding, additional factors are needed to explain performance, as explained below.

Cognitive Abilities and Speech Perception in Older Adults

Current models of speech recognition like the Ease of Language Understanding (ELU) model (Rönnberg et al., 2013) highlight the significance of cognitive factors for the processing of speech in ecological listening. The ELU model suggests that when the speech signal is degraded, for example, by competing noise or due to hearing loss, the automatic encoding of incoming speech may fail to match long-held representations within an individual's mental lexicon. When such failure occurs, explicit processing of the signal becomes necessary to achieve speech understanding. This is done by utilization of previous experience and context, as well as recruitment of linguistic knowledge and more domain-general cognitive resources (e.g., working memory and attention) to support listening (Pichora-Fuller et al., 1995; Akeroyd, 2008). By this account, individual differences in cognitive or linguistic functions are expected to contribute to individual differences in the explicit processes required for recognition under adverse conditions.

Consistent with the ELU model, studies suggest that individual differences in cognition are associated with individual differences in the processing of speech under adverse listening conditions (e.g., Salthouse, 1994, 1996). For example, cognitive speed of processing contributes to the perception of both time-compressed speech (a form of rapid speech; Wingfield et al., 1985; Wingfield, 1996; Dias et al., 2019) and speech in noise (Tun and Wingfield, 1999). Working memory and attention (specifically, divided attention and selective listening) are also associated with perception of time-compressed speech (Tun et al., 1992; Vaughan et al., 2006) and speech in noise

(Pichora-Fuller et al., 1995; Tun and Wingfield, 1999; Schneider et al., 2002, 2007; Tun et al., 2002). For dichotic speech, declines in attention are also related to performance declines (McDowd and Shaw, 2000; Rogers, 2000). Linguistic context can also positively contribute to speech perception (for review, see Burke and Shafto, 2008). Larger vocabulary in older adults and improved ability to utilize contextual cues facilitate speech perception in adverse listening conditions (Pichora-Fuller et al., 1995; Verhaeghen, 2003; Sheldon et al., 2008; Ben-David et al., 2015; Signoret and Ruder, 2019).

However, it is probably the combination of sensory and cognitive factors that affect speech perception of older adults (e.g., Cherry, 1953; Humes et al., 2006; Bronkhorst, 2015). If listeners possess a finite amount of information-processing resources (Kahneman, 1973), and if hearing-impaired older adults have to divert some of them to the normally automatic process of auditory encoding, then fewer resources will be available for subsequent higher-level processing (Rabbitt, 1990; Pichora-Fuller, 2003a). In addition, the interplay between sensory and cognitive factors can change in different listening conditions, but studies on the contribution of cognition to individual differences in speech perception in older adults often focused on a single task, making it hard to determine if either the contribution of cognition or the cognitive/sensory interplay changes across speech tasks. Whether the use of hearing aids changes, this interplay is also unknown.

Rapid Perceptual Learning Accounts for Variance in Speech Perception in Older Adults

Rapid perceptual learning also relates to the variability in perception of speech under challenging conditions (Peelle and Wingfield, 2005; Golomb et al., 2007; Manheim et al., 2018; Banai and Lavie, 2020; Rotman et al., 2020b). Rapid perceptual learning, defined as the ability to rapidly adapt to changes in one's environment, occurs under many adverse or sub-optimal conditions (Samuel and Kraljic, 2009). Perceptual learning is observed in old age, but it appears to be slower or reduced (Schneider and Pichora-Fuller, 2001; Forstmann et al., 2011; Lu et al., 2011) and more specific (Peelle and Wingfield, 2005) than in young adults (for a recent review, see Bieber and Gordon-Salant, 2021). Age-related hearing loss might have a further negative effect on learning. For example, older adults with preserved hearing exhibit poorer rapid learning of time-compressed speech compared to young adults, but better rapid learning than older adults with age-related hearing loss (Manheim et al., 2018). Relevant to the current study, across a range of speech tasks, rapid learning was documented in older adults with different levels of hearing (Peelle and Wingfield, 2005; Karawani et al., 2017; Manheim et al., 2018).

Perceptual learning and speech perception are related in the sense that learning contributes to future perception (Ahissar et al., 2009; Samuel and Kraljic, 2009; Banai and Lavie, 2021). However, recent studies suggest that the links could go beyond what could be expected from associations across different speech tasks (Banai and Lavie, 2021). Recent studies on perceptual learning (with both visual and speech materials) suggest that a general learning factor across learning tasks could serve as an individual capacity that supports performance across a range of scenarios (Yang et al., 2020; Dale et al., 2021; Heffner and Myers, 2021). Consistent with this view, we observed that individual differences in rapid perceptual learning of one type of speech (e.g., time-compressed speech) are consistently related to individual differences in speech perception under different adverse conditions (speech in noise and fast speech; Karawani et al., 2017;Manheim et al., 2018 ; Rotman et al., 2020b). Speech perception and rapid learning have also been found to be associated even when learning is assessed under conditions designed to offset the effects of age and hearing loss on speech perception (Manheim et al., 2018; Rotman et al., 2020b). We hypothesize that individuals who retain good rapid perceptual learning despite aging and hearing loss, can offset some of their negative impacts through rapid online learning (Banai and Lavie, 2021). To further explore this hypothesis, we now focus on the unique contribution of rapid perceptual learning to other challenging listening conditions, after accounting for sensory and cognitive factors and for the use of hearing aids.

Hearing Aid Use

For older adults with hearing loss, hearing aids are the most widely used rehabilitation devices. While hearing aids are unlikely to fully compensate for the auditory processing deficits of individuals with hearing impairment, they amplify sounds to improve audibility (Souza, 2016) and incorporate multiple algorithms intended to improve communication in adverse listening conditions (Neher et al., 2014). However, the perceptual results of using hearing aids depend not only on hearing aid technology but also on the factors described above. Moreover, long-term acclimatization induced benefits may further improve speech perception in some individuals. The effects reported in the literature include improved speech perception in noise, reduced distractibility to background noise, and reduced listening effort (Gatehouse, 1992; Munro and Lutman, 2003; Lavie et al., 2015; Habicht et al., 2016; Dawes and Munro, 2017; Lavie et al., 2021). However, other studies have failed to demonstrate improved identification of degraded speech (speech in noise, Dawes et al., 2013, 2014a,b; fast speech, Rotman et al., 2020a) in new or experienced hearing aid users. Thus, even though there are some indications for perceptual gains after months or years of hearing aid use, the effects of hearing aids on higher-level language processes in complex listening conditions are not well understood.

Unsurprisingly, most studies seeking to explain individual differences in aided speech perception have identified differences in hearing thresholds as the main source of variance (e.g., Tun and Wingfield, 1999; Humes, 2007). However, after controlling for the effects of audibility and age, working memory span score was correlated with both aided and unaided perception of speech in noise (Lunner, 2003), and the benefit from hearing aid algorithms (i.e., fast acting compression) was positively associated with cognitive skill (Lunner, 2003; Gatehouse et al., 2006; Cox and Xu, 2010; Souza et al., 2015).

Research Questions and Hypotheses

According to the literature review above, the interplay between the perception of different types of degraded speech, multiple cognitive factors, and perceptual learning is not sufficiently understood. It is also unclear whether the use of hearing aids changes the interplay among the different factors or results in plastic changes in speech perception (see Lavie et al., 2021 for a systematic review). The present study was designed to address these issues by investigating the contribution of hearing, cognition, rapid perceptual learning, and the contribution of long-term hearing aid use to three indices of speech perception: fast speech, speech in babble noise, and dichotic speech.

If, as explained above, perceptual learning is a capacity that can support other processes, such as "online" speech perception, rapid perceptual learning should explain unique variance in the perception of different types of distorted speech in addition to the known contributions of other sensory and cognitive factors. To this end, we use rapid learning of time-compressed speech as an index of learning for two reasons. First, the work reviewed above suggests that with this task, rapid learning rates are maintained even in older adults with hearing loss. Second, most listeners have no experience with this form of accelerated speech, and initial performance can be quite poor, making it easy to observe learning. Additionally, we hypothesize that the same factors that account for individual differences in degraded speech processing in adults with presbycusis also play a role when it comes to the effects of hearing aids, but the current literature (see Kalluri et al., 2019; Lavie et al., 2021 for recent reviews) makes it hard to draw more specific hypotheses, and therefore, in this regard, this is an exploratory study.

MATERIALS AND METHODS

Participants

A total of 95 potential participants were recruited *via* hearing clinics, retirement communities, and community centers. Potential participants were screened based on the following inclusion criteria: (1) age 65 and above; (2) bilateral, adult-onset, symmetric, sensory hearing loss of 30–70 dB, with flat or moderately sloping audiograms, and suprathreshold word recognition scores of ≥60% and air-bone gaps ≤15 dB; (3) no known neurological or psychiatric diagnoses; (4) normal or corrected-to-normal vision; (5) high proficiency in Hebrew; (6) normal cognitive status [a score of 24 or higher on the Mini-Mental State Examination (MMSE; Folstein et al., 1975)]; and (7) hearing aid use: we targeted only non-users (no prior experience with hearing aids and no plans to acquire hearing aids during the period of the study) and experienced hearing aid users [at least 6 months of bilateral hearing aid use; hearing aids were digital, with at least 16 amplification channels, at least four compression channels, noise reduction and anti-feedback algorithms, and wireless (ear to ear) processing]. Participants received modest monetary compensation for their participation and signed written informed consent forms. All aspects of the study were approved by the ethics committee of the Faculty of Social Welfare and Health Sciences at the University of Haifa (permit 362/18).

Twenty-two recruits failed to meet inclusion criteria and were excluded from the study: 12 for having insufficient hearing loss, five for having more severe hearing loss or low suprathreshold word recognition scores, two for asymmetric hearing loss, two for having insufficient experience with hearing aids [in the experienced hearing aid group (see below)], and one for reporting additional motor issues that could have influenced their responses on some of the tasks (e.g., block design and flanker). Three additional participants completed the first experimental session only (see experimental design below), and their data were thus excluded from all analyses.

The final study sample included 70 participants (23 males and 47 females) who met all inclusion criteria: 35 older adults with hearing loss (OHI) who did not use hearing aids and 35 older hearing-impaired adults who were experienced hearing aid users (OHI-HA). The two groups had similar ages, MMSE and cognitive scores, but hearing aid users had poorer hearing, somewhat poorer suprathreshold word recognition scores and somewhat higher education (see **Figure 1**; **Table 1**). Hearing thresholds were considered in our statistical modeling; the differences in word recognition (corresponding to 1–2 words difference) and education were considered negligible. Based on a power analysis on the data of our previous study (Rotman et al., 2020b), no effect for hearing aid use was expected even if we increased our sample size to 400 (200 in each group) participants, which was unrealistic. In contrast, a sample of 40 participants (20 in each group) was deemed sufficient to replicate the perceptual, learning, and cognitive effects reported by Rotman et al. (2020a) with a statistical power of 0.8. Power calculation was performed using the *simr* package (Green et al., 2016) in R.

Procedure

Testing was comprised of two sessions conducted 7–14 days apart at a hearing clinic or at the participants' home, based on each participant's preference (see **Figure 2**). Except for the audiometric assessments in the clinic (see below), all other testing was conducted in a quiet room in the clinic or in the participants' homes. In session I, potential participants were screened based on the inclusion criteria. Participants who met the inclusion criteria underwent assessments of rapid perceptual learning of time-compressed speech, perception of fast speech, and speech in noise and dichotic word identification. Session II included cognitive assessments and another assessment of time-compressed speech learning, data from which are not reported here. All testing was conducted by two clinical audiologists experienced in working with hearing-impaired patients and therefore accustomed to speaking loudly and clearly.

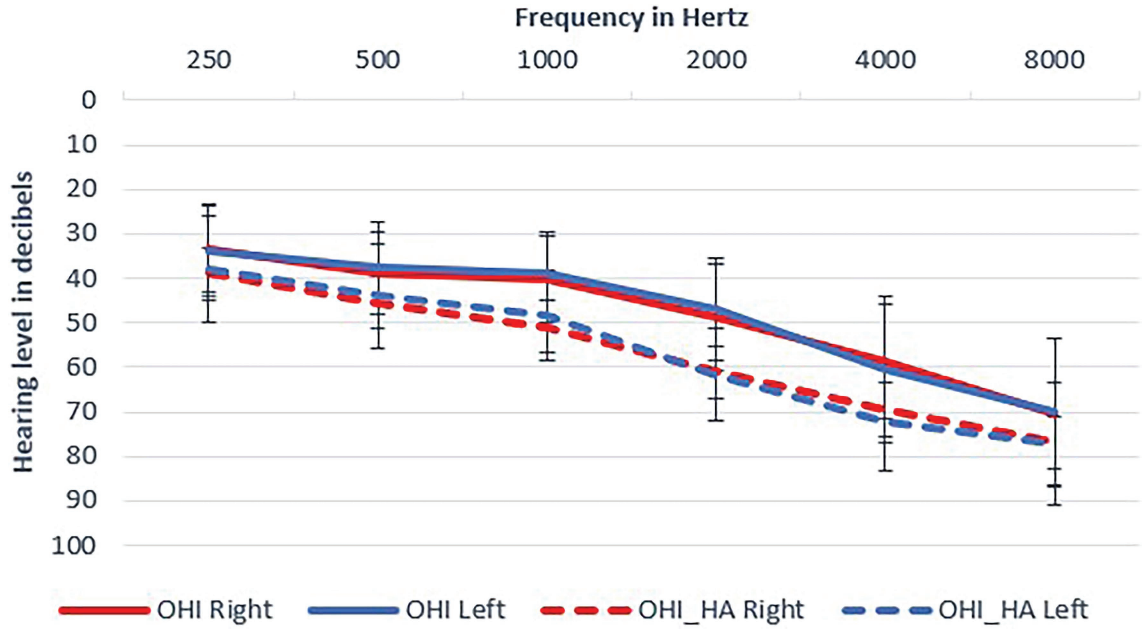


FIGURE 1 | Mean audiograms of participants. Mean thresholds and standard deviations are shown: older hearing-impaired adults (OHI) in full lines; older hearing-impaired adults who use hearing aids (OHI-HA) in dashed lines.

TABLE 1 | Age, hearing, word recognition, education, and cognitive screening.

	OHI	OHI-HA
Age (years)		
Mean (SD) [95% CI]	79 (7) [77–82]	81 (6) [78–83]
Median (IQR)	81 (73–84)	80 (76–85)
Hearing (PTA4, dB)		
Mean (SD) [95% CI]	46 (7) [43–49]	57 (8) [53–59]
Median (IQR)	44 (40–52)	56 (53–61)
Suprathreshold word recognition scores		
Mean (SD) [95% CI]	90 (9) [86–93]	84 (9) [81–87]
Median (IQR)	92 (86–95)	85 (77–90)
Years of education		
Mean (SD) [95% CI]	14 (3) [13–15]	16 (4) [14–17]
Median (IQR)	14 (12–15.5)	16 (12.5–18)
MMSE		
Mean (SD) [95% CI]	28 (1) [27–28]	28 (1) [28–29]
Median (IQR)	28 (27–29)	28 (28–29)

PTA, Pure-tone average; MMSE, Mini-mental state examination; CI, Confidence interval; IQR, Interquartile range; OHI, Hearing-impaired older adults; and OHI-HA, Hearing-impaired older adults who are experienced hearing aid users.

Task

Screening Assessments

Demographic Questionnaire

A questionnaire regarding education, handedness, lifestyle, and general health was used in the current study. The participants completed the questionnaire before completing further assessments.

Cognitive and Hebrew Screening

Participants were screened with a the MMSE (Folstein et al., 1975), with a cutoff score of 24 as an inclusion criterion. Proficiency in Hebrew was evaluated using a short screening with a series of questions and commands in Hebrew. To participate in the study, one had to complete this screening with a perfect score (see Lavie, 2011).

Audiological Assessments

A full pure-tone and speech audiometry (suprathreshold word recognition) was conducted in an acoustic booth, using the MAICO audiometer (model MA42) or at the participants' home with Inventis *Cello* and Piccolo portable audiometers and Silenta Supermax supra-aural headphones. Most comfortable levels (MCL) for speech were also assessed. The audiograms were classified based on Duthey (2013), with four frequencies pure-tone average (0.5, 1, 2, and 4 KHz) ≥ 30 dB as a criterion of hearing loss. Participants with up-to-date (≤6 months) audiograms were not evaluated again.

Speech Perception and Learning

Stimuli

Stimuli were 80 simple sentences in Hebrew, five to six words long, with a common subject-verb-object grammatical structure (adapted from Prior and Bentin, 2006). All sentences were recorded in a sound attenuating booth using a built-in MacBook Air microphone, sampled at 44 KHz and saved in WAV format. The root-mean-square levels of the recorded sentences were normalized using Audacity audio software version 2.2.0. Sentences were recorded by four native Hebrew speakers (three females and one male).

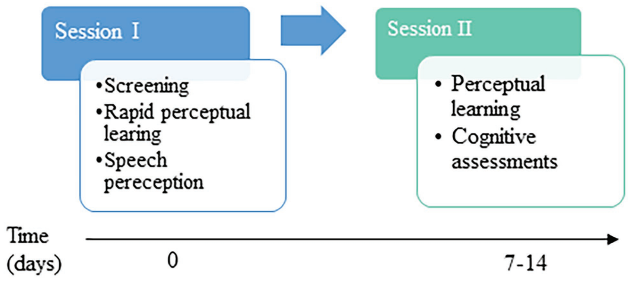


FIGURE 2 | Schematic illustration of study design.

Speaker 1 (female) recorded 10 different sentences at her natural fast rate ($M=183$ words/min, $SD=17$); speaker 2 (female) recorded 10 sentences at her natural fast rate ($M=210$ words/min, $SD=21$) and 10 sentences at her normal, unhurried rate ($M=111$ words/min, $SD=27$); speaker 3 (male) recorded 10 sentences at a normal rate of 88 words/min ($SD=10.30$); and speaker 4 (female) recorded 40 sentences at a normal rate of 102 words/min ($SD=12.68$). To minimize the effects of sentence familiarity on performance, there was no sentence repetition within or across conditions. In addition, Speaker 4 also recorded a list containing 25 pairs of monosyllabic words, adapted from the Hebrew PB-50 test (Lavie et al., 2015) for the dichotic word identification task (see below).

Presentation and Scoring

Speech materials were presented through Meze 99 classics headphones to both ears as follows: (1) unaided to the OHI group and (2) aided for the OHI-HA group (i.e., headphones were placed while participants wore their hearing aids). Stimuli were presented at each listener's preferred level. To determine this level, a pre-recorded short passage was played and listeners determined their preferred listening level. Because some of the participants were tested at home, and others in several rooms in the clinic, achieving constant acoustic settings for sound field presentation of the speech stimuli was impossible. Thus, we decided to test all participants with headphones and play the stimuli from the computers (in line with Rotman et al., 2020b). In the OHI-HA group, the testers verified that the hearing aids were working properly at the beginning of each session. After listening to each test stimulus (sentences or dichotic word pairs), participants were asked to repeat what they had heard, and the experimenter transcribed their replies. Each stimulus was presented only once, and no feedback was provided. Performance was scored off-line. For the rapid learning, fast speech and speech in noise tasks, all words, including function words, were counted for scoring. Scoring of the dichotic listening task is described below. Unless otherwise noted, the proportion of correctly recognized words/sentence was computed and used for statistical modeling, although for visualization proportion was averaged across sentences.

Rapid Perceptual Learning (Session I)

Ten sentences (recorded by Speaker 2) were presented as time-compressed speech. Time-compressed speech was chosen because learning with this form of speech was previously documented in older adults within and across sessions (e.g., Peelle and Wingfield, 2005; Golomb et al., 2007; Manheim et al., 2018; Rotman et al., 2020b). In addition, most older listeners have no experience with this type of artificially accelerated speech, making it useful in studying the correlations between learning and the recognition of other forms of degraded speech (e.g., naturally fast speech and speech in noise). Following earlier work (Rotman et al., 2020b), sentences were compressed to 45–50% of their original length (45% for participants with PTA of 26–47 dB and 50% for PTA ≥ 48 dB) in Matlab, using a pitch preserving algorithm (WSOLA, Verhelst and Roelands, 1993). Speech rates were adjusted based on hearing threshold to minimize the effects of hearing on the estimate of rapid learning.

Baseline recognition of time-compressed speech was defined as the proportion of correctly identified words in the two first sentences. Learning of time-compressed speech was defined as the rate of improvement in recognition over time. It was quantified as the linear slopes of the learning curves over an additional eight time-compressed speech sentences (for further details see Rotman et al., 2020b).

Speech Perception

Speech perception was evaluated using the following tasks:

Fast Speech

Twenty sentences (10 sentences recorded by Speaker 1 and 10 sentences recorded by Speaker 2).

Speech Recognition in Noise

Twenty sentences were presented (10 sentences recorded by Speaker 3 and 10 sentences recorded by Speaker 4). All sentences were embedded in a 4-talker babble noise with a fixed SNR level of +3 dB.

Dichotic Word Identification

Following previous research (Lavie et al., 2013, 2015), we used a list of 25 pairs of monosyllabic words, adapted from the Hebrew PB-50 test. One word of each pair was presented to the right ear while the other word was presented simultaneously to the left ear, and participants were required to repeat both words in whichever order they chose. For statistical analysis, the number of correctly repeated words in each ear was counted and two indices of dichotic listening were calculated as: the sum (= dominant ear score + non-dominant ear score) and the difference between the ears (= dominant ear - non-dominant ear).

Cognitive Assessments

A battery of cognitive assessments was used to evaluate cognitive status and identify characteristics that might influence participants' performance on the experimental tasks. This battery was administered at a comfortable auditory level that was defined by each participant to negate potential confounding effects of audibility on performance. The following subtests from the Wechsler Adult Intelligence Scale-Third Edition in Hebrew (WAIS-III; Wechsler, 1997) were used as: **vocabulary** (semantic knowledge), **digit span** (working memory), and **block design** (non-verbal reasoning).

All subtests were administrated and scored according to the test manual. Raw scores were converted to standardized scores.

TABLE 2 | Cognition and speech perception.

	OHI	OHI-HA
Cognition		
Vocabulary (scaled score)		
Mean (SD) [95% CI]	11.6 (2.4) [11–12]	13.9 (2.6) [13–15]
Median (IQR)	12 (10–13)	14 (12–16)
Working memory (scaled score)		
Mean (SD) [95% CI]	9.3 (2.2) [8–10]	10.5 (3.1) [9–11]
Median (IQR)	9 (7–10)	10 (8–12)
Block design (scaled score)		
Mean (SD) [95% CI]	10.9 (3.3) [10–12]	11.6 (4.1) [10–13]
Median (IQR)	11 (9–13)	10 (8–15)
Trail Making		
Mean (SD) [95% CI]	2.7 (1.0) [2.4–3.1]	2.4 (0.8) [2.1–2.7]
Median (IQR)	2.4 (2.0–3.7)	2.2 (1.8–2.8)
Flanker cost		
Mean (SD) [95% CI]	1.01 (0.01) [1.01–1.02]	1.02 (0.02) [1.01–1.02]
Median (IQR)	1.01 (1.01–1.02)	1.01 (1.01–1.02)
Speech perception		
FS (proportion correct)		
Mean (SD) [95% CI]	0.28 (0.19) [0.22–0.35]	0.22 (0.16) [0.16–0.27]
Median (IQR)	0.32 (0.04–0.62)	0.20 (0.02–0.60)
SIN (proportion correct)		
Mean (SD) [95% CI]	0.68 (0.20) [0.62–0.75]	0.50 (0.22) [0.42–0.58]
Median (IQR)	0.74 (0.03–0.96)	0.52 (0.07–0.96)
Dichotic listening (sum)		
Mean (SD) [95% CI]	0.63 (0.30) [0.53–0.73]	0.55 (0.22) [0.48–0.63]
Median (IQR)	0. 6 (0.16–1.32)	0.52 (0.24–1.08)
Dichotic listening (gap)		
Mean (SD) [95% CI]	0.21 (0.12) [0.17–0.26]	0.18 (0.12) [0.14–0.23]
Median (IQR)	0.24 (0–0.56)	0. 2 (0–0.4)
TCS baseline (proportion correct)		
Mean (SD) [95% CI]	0.158 (0.19) [0.11–0.20]	0.151 (0.13) [0.11–0.19]
Median (IQR)	0.09 (0–0.73)	0.18 (0–0.73)
TCS learning slope		
Mean (SD) [95% CI]	0.095 (0.07) [0.07–0.12]	0.094 (0.07) [0.07–0.12]
Median (IQR)	0.086 (–0.01–0.30)	0.090 (–0.002–0.22)

FS, Fast speech; SIN, Speech in noise; TCS, Time-compressed speech; PTA, Pure-tone average; CI, Confidence interval; IQR, Interquartile range; OHI, Older hearing-impaired adults; and OHI-HA, Older hearing-impaired adults who are experienced hearing aid users.

TABLE 3 | Correlations between speech perception, cognition, and learning among all participants.

	FS	SIN	Dichotic listening (sum)	Dichotic listening (gap)	Slope
Hearing	–0.50	–0.59	–0.35	–0.16	–0.38
Vocabulary	0.10	–0.03	–0.008	–0.11	0.11
Working memory	**0.36**	**0.33**	0.24	–0.004	**0.28**
Flanker cost	–0.05	–0.13	0.07	0.15	–0.18
TCS baseline	**0.56**	**0.49**	0.11	–0.05	**0.35**
Slope	**0.54**	**0.46**	0.23	0.06	–

Pearson correlations are shown. FS, Fast speech; SIN, Speech in noise; hearing = average PTA; and TCS baseline = average of first two sentences of time-compressed speech. Vocabulary and working memory = raw scores from corresponding tests; Slope = rapid perceptual learning slope, session I. Bold entries represent significant correlations (p < 0.05) after correcting for multiple testing with a Bonferroni correction.

Attention

Two tests were used as: (1) **Flanker test** (Eriksen and Eriksen, 1974). A computerized version of the well-validated Flanker test was used as a measure of inhibition and selective attention. The target stimulus was an arrow-head heading right or left, embedded in the middle of a row of five arrow-heads or other stimuli. Participants were asked to note the direction of a central arrow, which was flanked by arrows pointing in the same direction (congruent trials) or the opposite direction (incongruent trials) or non-arrow stimuli (neutral trials). Reaction time and accuracy were measured. The "flanker cost" for each participant was used for statistical analyses. The cost was calculated as the mean logRT.

(RT = reaction time in ms) of the correct responses in the incongruent trials divided by the mean log RT of the correct responses in the neutral trials. A higher flanker cost (>1) means poorer selective attention. (2) **Trail making test** (Reitan, 1958). Attention switching control was tested in two test conditions: in condition A, participants were asked to draw lines to connect circled numbers in a numerical sequence (i.e., 1-2-3) as rapidly as possible. In condition B, participants were asked to draw lines to connect circled numbers and letters in an alternating numeric and alphabetic sequence (i.e., 1-A-2-B) as rapidly as possible. Response speed was measured by a stopwatch.

RESULTS

Descriptive Statistics

As shown in **Table 2**, hearing aid users had somewhat higher vocabulary scores than the non-hearing aid group, and this was considered in the statistical analyses reported below. In both groups, there was large between-participant variance across all speech and learning tasks (the raw data and analysis code can be found at https://osf.io/sreq4).

As shown in **Table 3**, rapid perceptual learning of time-compressed speech was positively correlated with identification of fast speech and speech in noise, and negatively correlated with hearing thresholds. In addition, and as expected from the literature, speech perception was correlated with specific cognitive indices. Rapid learning of time-compressed speech also correlated with some of the cognitive measures.

Modeling Speech Perception As a Function of Age, Hearing, Cognition, Rapid Perceptual Learning, and Hearing Aid Use

The contribution of hearing and cognition to recognition accuracy in the speech tasks was studied in the past. Therefore, our modeling here focused on the unique additional contributions of perceptual learning and hearing aid use. To this end, modeling was performed in stages: hearing and cognition were modeled first, followed by learning, and then hearing aid use. With this approach, if a later model fits the data significantly better than a previous one (with a

TABLE 4 | Fast speech—model comparisons.

Model	Fixed effects	AIC	χ^2	*Df*	*p*
0	(Random effects)	3630.7	–	–	–
1	+ Background variables	3595.5	45.18	5	< 0.001
2	+ Baseline recognition of TCS	3581.7	15.76	1	< 0.001
3	+ Rapid learning slope	3576.4	7.32	1	0.007
4	+ Hearing aids	3578.4	0.03	1	0.868

As described in the main text, the initial model included age, hearing, vocabulary, working memory, and attention as predictors. Comparison models successively added the fixed effects of baseline recognition of time-compressed speech, rapid perceptual learning slope, and hearing aids. The random effects structure was identical across models.

model comparison), the predictor entered at the later stage has a unique contribution to speech recognition when all other included variables are considered. Within a given model, the coefficient of each predictor reflects its contribution while all other predictors in the model are kept constant. Since there were repeated measures for the fast speech and speech in noise, a series of generalized linear mixed models was run using the lme4 package in R (Bates et al., 2014). Single trial fast speech and speech in noise scores served as the dependent variables, and age, hearing, cognition, rapid perceptual learning, and hearing aid use served as the independent variables (i.e., the predictors). Given the number of predictors relative to sample size, and to avoid overloading the models, block design and trail making were excluded from the analysis; likewise, interactions were not modeled. The random effects structure consisted of random intercepts for both participant and sentence; predictors were standardized (z-scored) prior to modeling. Following earlier work, and due to dealing with proportion scores, binomial regressions with a logit link function (logistic regressions) were used (Rotman et al., 2020b).

Five models were constructed for fast speech and for speech in noise, starting with a model that included only the random effects (Model 0). Thereafter, each subsequent model added one additional predictor over the previous model(s), with the models building upon one another sequentially (e.g., model 1 = Model 0 + variable 1; Model 2 = Model 1 + variable 2; and Model 3 = Model 2 + variable 3). Model 1 included background variables of the participants as predictors, which included as: age, hearing, vocabulary, working memory, and attention. Model 2 included baseline recognition of time-compressed speech; Model 3 added the rapid perceptual learning slope; and Model 4 added hearing aid use (rated on a nominal scale—yes/no). To isolate the unique contribution of each additional variable, these four increasingly complex models were compared using likelihood ratio tests with the R ANOVA function.

Note that in general, correlations between the different predictors were not high (the highest Pearson correlations were $r = 0.43$ between vocabulary and working memory, $r = -0.38$ between hearing and learning, and $r = 0.35$ between learning and baseline recognition of TCS), suggesting that multicollinearity is not a serious concern. Likewise, all Variance Inflation Factors (VIF) were low (< 2), as reported below for the best fitting models.

Recognition of Fast Speech

The inclusion of the background variables in the model resulted in a better fit to the data than the model that included the random effects only. However, the addition of baseline recognition of time-compressed speech and rapid learning both improved the fits significantly, suggesting that rapid learning had a significantly unique contribution to the recognition of fast speech, beyond that of other variables. Hearing aids had no additional effect (see **Table 4**).

In the best fitting model (model 3), age, hearing, baseline recognition of time-compressed speech, and learning were all significant predictors of fast speech recognition (see **Table 5** which also includes model 1 with only background variables). Hearing was the strongest negative predictor (largest beta in absolute value, see **Table 5**) of fast speech recognition followed by age, indicating that fast speech recognition was poorer in individuals with more severe hearing loss and in older individuals. Baseline recognition of time-compressed speech and rapid learning were both positive predictors, suggesting that for a given age/hearing loss, listeners who maintained better perception and learning of time-compressed speech also maintained more accurate recognition of fast speech, regardless of hearing aid use (see **Figure 3**). Variance Inflation Factors for the best fitting model were 1.27 for age, 1.91 for hearing, 1.41 for vocabulary, 1.55 for working memory, 1.06 for attention, 1.31 for baseline recognition of TCS, and 1.46 for learning.

Recognition of Speech in Noise

The inclusion of the background variables in the model resulted in a better fit to the data than the model that included the random effects only (**Table 6**). However, baseline recognition of time-compressed speech improved the fits significantly (see **Table 6**), suggesting that time-compressed speech perception had a significant unique contribution to the recognition of speech in noise, beyond that of other variables. Hearing aids had no additional effect.

In the best fitting model (see **Table 7** which also includes model 1 with only background variables), hearing was the strongest predictor (i.e., largest beta in absolute value) of speech in noise recognition, followed by baseline recognition of time-compressed speech. Neither rapid learning nor hearing aid use further improved the fit (see **Figure 4**). Thus, lower hearing thresholds and more accurate time-compressed speech recognition were associated with better recognition of speech in noise. Variance Inflation Factors for the best fitting model

TABLE 5 | Results of generalized linear mixed-model for fast speech recognition as a function of the background variables (Model 1) and as a function of age, hearing, cognition, baseline recognition of time-compressed speech, and rapid perceptual learning as fixed effects (Model 3).

Fixed effect	Odds ratio	*β*	*SE*	95% CI	*Z*	*p*
Model 1						
Age	0.73	−0.32	0.12	[0.57, 0.92]	−2.63	0.009
Hearing (PTA4)	0.53	−0.64	0.12	[0.41, 0.67]	−5.28	<0.001
Vocabulary	1.21	0.22	0.13	[0.97, 1.59]	1.69	0.091
Working memory	1.32	0.28	0.13	[1.02, 1.70]	2.14	0.033
Attention	1.00	−0.002	0.12	[0.79, 1.26]	−0.02	0.984
Model 3						
Age	0.79	−0.24	0.10	[−0.44, −0.03]	−2.26	0.023
Hearing (PTA4)	0.62	−0.48	0.11	[−0.70, −0.27]	−4.35	<0.001
Vocabulary	1.15	0.14	0.11	[−0.07, 0.36]	1.30	0.219
Working memory	1.10	0.10	0.12	[−0.13, 0.32]	0.83	0.414
Attention	1.03	0.03	0.10	[−0.18, 0.23]	0.25	0.803
Baseline TCS	1.47	0.39	0.10	[0.18, 0.59]	3.68	<0.001
Learning	1.36	0.31	0.11	[0.09, 0.52]	2.77	0.008

PTA, Pure-tone average; TCS, Time-compressed speech; and CI, Confidence interval.

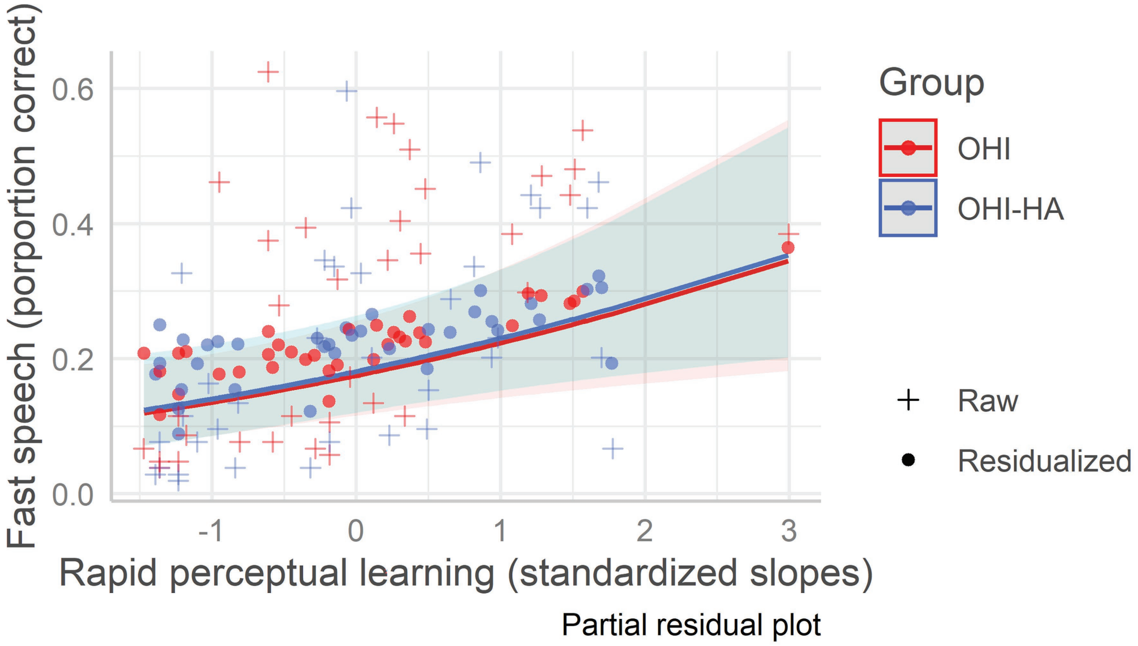


FIGURE 3 | Fast speech recognition as a function of rapid learning among older hearing-impaired adults and older hearing-impaired adults who use hearing aids. Older hearing-impaired adults (OHI) in red; older hearing-impaired adults who use hearing aids (OHI-HA) in blue. The *y*-axis indicates the correct perception percentage of fast speech, and the *x*-axis indicates the standardized rapid perceptual learning slope. The dots (residualized aggregate scores) mark the predicted scores; their deviation from the regression line indicates prediction error, while the pluses mark the raw/true scores. The shaded areas are the confidence intervals.

TABLE 6 | Speech in noise—model comparisons.

Model	Fixed effects	AIC	χ^2	*df*	*p*
0	(Random effects)	4454.3	–	–	–
1	+ Background variables	4417.5	46.79	5	0.000
2	+ Baseline recognition of TCS	4406.7	12.75	1	0.000
3	+ Rapid learning slope	4408.1	0.62	1	0.43
4	+ Hearing aids	4407.4	2.73	1	0.09

Model 1 included age, hearing, vocabulary, working memory, and attention as predictors. Comparison models successively added the fixed effects of baseline recognition of time-compressed speech, rapid perceptual learning slope, and hearing aids. The random effects structure was identical across models.

TABLE 7 | Results of generalizedlinear mixed-effects model for speech in noise recognition as a function of the background variables (Model 1) and as a function of age, hearing, cognition, and baseline recognition of time-compressed speech as fixed effects (Model 2).

Fixed effect	Odds ratio	*B*	*SE*	95% CI	*Z*	*p*
Model 1						
Age	0.79	−0.24	0.12	[−0.48, 0.01]	−1.90	0.057
Hearing (PTA4)	0.49	−0.71	0.12	[−0.95, −0.47]	−5.86	<0.001
Vocabulary	1.07	0.07	0.13	[−0.19, 0.32]	0.51	0.607
Working memory	1.41	0.35	0.13	[0.09, 0.61]	2.61	0.009
Attention	0.95	−0.05	0.11	[−0.27, 0.17]	−0.43	0.668
Model 2						
Age	0.85	−0.17	0.11	[−0.39, 0.06]	−1.45	0.147
Hearing (PTA4)	0.51	−0.67	0.11	[−0.89, −0.45]	−6.04	<0.001
Vocabulary	1.04	0.04	0.12	[−0.19, 0.27]	0.34	0.731
Working memory	1.23	0.21	0.13	[−0.04, 0.45]	1.65	0.099
Attention	0.94	−0.06	0.10	[−0.26, 0.14]	−0.62	0.533
Baseline TCS	1.53	0.42	0.11	[0.20, 0.65]	3.74	<0.001

PTA, Pure-tone average; TCS, Time-compressed speech; and CI, Confidence interval.

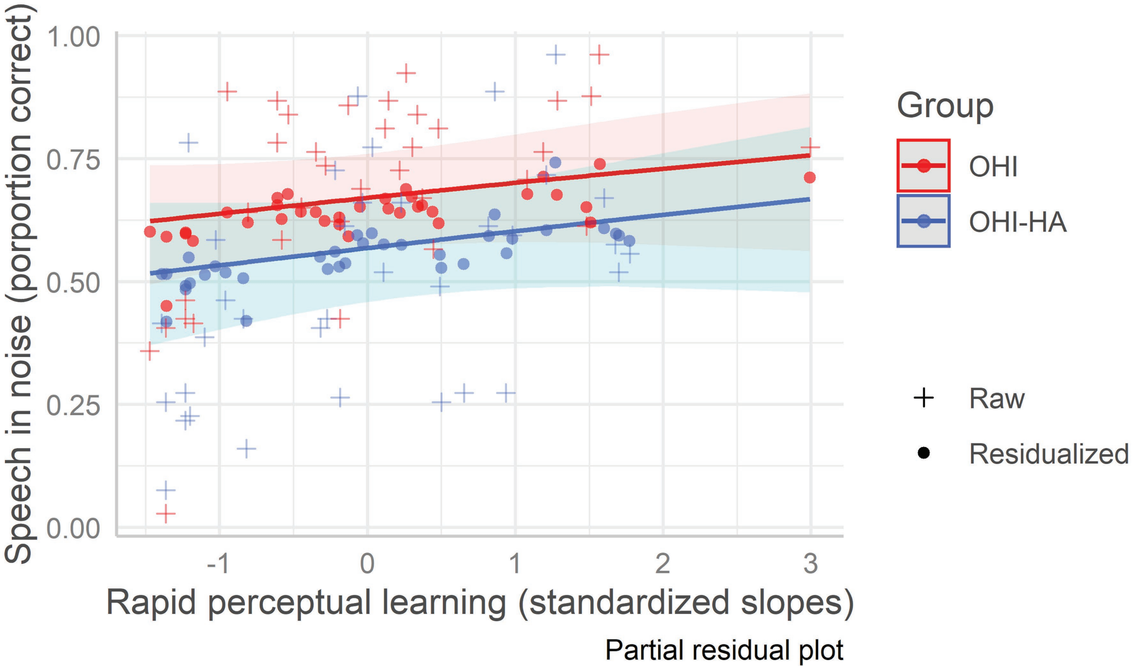


FIGURE 4 | Speech in noise recognition as a function of rapid learning among older hearing-impaired adults and older hearing-impaired adults who use hearing aids. Older hearing-impaired adults (OHI) in red; older hearing-impaired adults who use hearing aids (OHI-HA) in blue. The *y*-axis indicates the correct perception percentage of speech in noise, and the *x*-axis indicates the standardized rapid perceptual learning slope. The dots (residualized aggregate scores) mark the predicted scores; their deviation from the regression line indicates prediction error, while the pluses mark the raw/true scores. The shaded areas are the confidence intervals.

were 1.29 for age, 1.97 for hearing, 1.45 for vocabulary, 1.55 for working memory, 1.06 for attention, 1.33 for baseline recognition of TCS, and 1.53 for learning.

Dichotic Word Identification

Since there were no repeated measures (i.e., there was only one score for each participant), linear regression analyses were used. Four models were constructed for the dichotic listening task in two different ways: once with the dichotic sum serving as the dependent variable and once with the dichotic gap serving as the dependent variable. The models all included age, hearing, vocabulary, working memory, and attention as predictors. Thereafter, as with the models mentioned above, each subsequent model added one additional variable, with the models building upon one another sequentially. Model 2

TABLE 8 | Dichotic sum—results of the comparison of nested model tests.

Model	Fixed effects	AIC	Res. *df*	RSS	*df*	Sum of Sq.	*F*	*p*
1	Background variables	195.724	64	54.964				
2	+ Baseline recognition of TCS	197.715	63	54.957	1	0.001	0.01	0.93
3	+ Rapid perceptual learning slope	199.299	62	54.631	1	0.326	0.36	0.55
4	+ Hearing aids	201.293	61	54.627	1	0.004	0.00	0.94

As described in the main text, the initial model included age, hearing, vocabulary, working memory, and attention as predictors. Comparison models successively added the fixed effects of baseline recognition of time-compressed speech, rapid perceptual learning slope, and hearing aids. The random effects structure was identical across models.

TABLE 9 | Dichotic gap—results of the comparison of nested model tests.

Model	Fixed effects	AIC	Res. *df*	RSS	*df*	Sum of Sq.	*F*	*p*
1	Background variables	207.182	64	64.739				
2	+ Baseline recognition of TCS	208.607	63	64.210	1	0.530	0.51	0.48
3	+ Rapid perceptual learning slope	210.328	62	63.953	1	0.256	0.24	0.62
4	+ Hearing aids	212.189	61	63.827	1	0.127	0.12	0.73

As described in the main text, the initial model included age, hearing, vocabulary, working memory, and attention as predictors. Comparison models successively added the fixed effects of baseline recognition of time-compressed speech, rapid perceptual learning slope, and hearing aids. The random effects structure was identical across models.

included baseline recognition of time-compressed speech; Model 3 added the rapid learning slope; and Model 4 added hearing aids. To isolate the unique contribution of each additional variable, these four successively complex models were compared using an ANOVA Table for Comparison of Nested Model tests.

Dichotic Sum as the Dependent Variable. For the dichotic listening task, with dichotic sum serving as the dependent variable, model comparisons showed that there were no contributions of variables/effects that did not appear in the first model (see **Table 8**).

Dichotic Gap As the Dependent Variable

For the dichotic listening task, with dichotic gap serving as the dependent variable, model comparisons showed that there were no contributions of variables/effects that did not appear in the first model (see **Table 9**).

DISCUSSION

We assessed the relative contribution of hearing acuity, cognitive factors, and rapid perceptual learning to the identification of fast speech, speech in noise, and dichotic speech in older adults with hearing loss. Hearing acuity and time-compressed speech perception uniquely contributed to the perception of both fast speech and speech in noise. Rapid perceptual learning was a significant predictor of fast speech perception even after accounting for age, hearing, and cognition. Hearing aid use had no effect on any of the speech tasks. Our findings suggest that in older adults, good rapid perceptual learning can partially offset the effects of age and hearing loss on the perception of fast speech, but not on the perception of speech in noise or dichotic speech. Determining if this is due to inherent differences between the different speech tasks or due to other differences (e.g., overall level of difficulty) requires further investigation. Furthermore, the finding that time-compressed speech recognition is strongly associated with the perception of speech in noise suggests a potential link between the perception of these two types of challenging speech.

In the present study, hearing acuity was the strongest predictor of both fast speech and speech in noise perception. This finding is consistent with previous work on speech perception in older adults (e.g., Frisina and Frisina, 1997; Janse, 2009; Humes and Dubno, 2010). For example, Janse (2009) investigated the relative contributions of auditory and cognitive factors to fast speech perception in older adults. While hearing acuity, reading rate, and visual speed of processing were all significant predictors, hearing acuity was the strongest one. Similarly, for speech in noise among new and experienced hearing aid users, hearing loss was repeatedly identified as the primary and best predictor for unaided performance (Humes, 2002). Our study extends this finding to the perception of fast speech among hearing aid users.

An interesting outcome of the current study is that the initial performance of time-compressed speech remained the second strongest predictor of perception of both fast speech and speech in noise. These findings are in line with previous results regarding the perception of fast speech (Manheim et al., 2018; Rotman et al., 2020b) and extend them to speech in noise. Although fast speech is harder to recognize than time-compressed speech at similar rates, performance is correlated between these two tasks, and temporal processing is likely involved in the perception of both (Janse, 2004; Gordon-Salant et al., 2014). Indeed, the increased difficulties older adults have in processing distorted speech are thought to result in part from age-related declines in temporal processing (e.g., Pichora-Fuller and Singh, 2006; Anderson et al., 2011; Füllgrabe et al., 2015). Temporal cues within both the temporal envelope of the speech signal and its fine structure convey information that influences lexical, syntactic, and phonemic processing and these can support speech perception across a range of conditions (Kidd et al., 1984; Nelson and Freyman, 1987; Festen and Plomp, 1990; Rosen, 1992). Fast speech recognition can thus be affected by the temporal resolution of phonetic information and by linguistic context, suggesting that both low-level and high-level processes can independently contribute to the processing of temporally distorted speech (Gordon-Salant and Fitzgibbons, 2001; Pichora-Fuller, 2003b; Gordon-Salant et al., 2014).

As for the association between time-compressed speech and speech in noise recognition, loss of synchrony in aging auditory systems may disrupt the fine structure cues that important for recognizing speech in noise (Schneider and Pichora-Fuller, 2001). The fine structure of speech, in particular its harmonic structure, enables listeners to attend to a target speech source or to distinguish competing speech or noise sources, especially when they are spectrally similar to the target signal (Moore, 2008, 2011). Similarly, binaural advantage for detecting and identifying speech presented in a noisy background relies on the ability of the binaural system to process interaural, minimal timing differences (Levitt and Rabiner, 1967). If the perception of temporal fine structure affects both identification of speech in the presence of competing noise and fast speech, it is perhaps unsurprising that perception of time-compressed speech accounts for some of the individual differences in the perception of speech in noise. Indeed, speech reception threshold in fluctuating noise and susceptibility to time compression are highly correlated among normal-hearing and hearing-impaired older adults (Versfeld and Dreschler, 2002).

Our results indicate that the association across speech tasks is not limited to tasks that share obvious sensory characteristics. This suggests that common speech perception processes could underlie performance variability across a range of listening challenges in older adults with different levels of hearing. Consistent with this view, research on speech recognition under adverse listening conditions has shown relationships across different conditions (e.g., Borrie et al., 2017; Carbonell, 2017). For example, Carbonell (2017) found that performance was correlated across noise-vocoded, time-compressed, and speech in babble noise tasks, and regression models that predicted performance on one task based on performance of the other two also showed a strong relationship. Nevertheless, it is hard to determine whether these findings reflect common underlying processing. Furthermore, in some studies, correlations across speech conditions were more limited (Bent et al., 2016; McLaughlin et al., 2018). Bent et al. (2016) studied intelligibility under different types of signal adversity and showed that English-speaking listeners who were good at understanding non-native (Spanish) accent were also good at understanding a regional dialect (Irish English) and disordered speech (ataxic dysarthria). These results indicated that, rather than possessing a general speech skill, listeners may possess specific cue sensitivities and/or favor perceptual strategies that allow them to be successful with particular types of listening adversity. Therefore, at present, it is hard to determine whether differences between different speech conditions stem from differences in the requirements they pose on underlying auditory mechanisms, from differences in listening effort or from methodological issues. For example, in the current study and with similar tasks, recognition of fast speech was poorer than that of speech in babble, but using different fast talkers or a more challenging SNR could have changed this pattern. Further studies with conditions matched for accuracy might shed further light on this issue if listening effort is tracked and compared across conditions. As for older adults with hearing impairment, both general speech skills and specific cue sensitivities/perceptual strategies decline with aging. Further research is needed to understand individual differences in those declines, which could help shed light on the varying degrees of benefit from current rehabilitative strategies.

In contrast to previous work in older adults (e.g., Salthouse, 1994, 1996; Pichora-Fuller et al., 1995; Humes, 2007; Rotman et al., 2020b), in the present study, cognitive abilities (working memory, vocabulary, and selective attention) were not significant predictors of performance on any of the speech perception tasks. This suggests that the relationship between cognition and speech perception is not straightforward. Indeed, Akeroyd (2008) found inconsistencies across studies both when the speech and the cognitive tasks varied across studies, and also when the assessed cognitive domain (e.g., working memory) was constant and only the speech task differed. However, task and stimulus related factors do not provide a sufficient account for the discrepancies across studies, because in the current study, we used the same time-compressed, fast speech and cognitive tasks as in a previous study from our lab in which we did find an association between fast speech recognition and vocabulary (Rotman et al., 2020b). A recent review by Dryden et al. (2017) highlighted that not only do measures of speech in noise perception and cognitive tasks vary greatly across published studies, but research participant samples vary widely as well and can include any combination of young and old listeners with or without hearing loss, tested under aided or unaided listening conditions. Consistent with this view, in the current study, effect sizes (expressed in odd ratios) were similar to those observed in our previous study. Furthermore, based on our previous data (Rotman et al., 2020b), statistical power was adequate. On the other hand, hearing levels were more variable and this increased variability may have contributed to the lack of significant effects.

The current finding that hearing aid use had no effect on degraded speech perception is consistent with that of Rotman et al. (2020b). However, this finding contradicts previous research showing improved speech perception following hearing aid use (Gatehouse, 1992; Munro and Lutman, 2003; Lavie et al., 2015; Habicht et al., 2016; Dawes and Munro, 2017; Wright and Gagné, 2020). One potential explanation for this could be that in our study, the average hearing loss (PTA) in hearing aid users was approximately 10 dB more severe than in non-users (see **Table 1**). This greater severity of hearing loss could have masked a hearing aid induced effect despite the inclusion of PTAs in statistical modeling. Methodological differences, including: timing and duration of hearing aid use (e.g., Gatehouse, 1992; Munro and Lutman, 2003), variability of outcome measures (e.g., Larson et al., 2000; Humes et al., 2001), and lack of baseline tests before starting to use the hearing aids (e.g., Vogelzang et al., 2021), can also account for the discrepancy between studies. The above differences highlight the need for further research on speech processing among hearing aid users. For example, future studies should include an unaided condition for the group with hearing aids and an aided condition for the group without hearing aids. This could test differences between the effects of hearing aid use and the effects of amplification during testing, without using hearing aids between test sessions.

DATA AVAILABILITY STATEMENT

The data from this study is available at Open Science Foundation; https://osf.io/sreq4.

ETHICS STATEMENT

This study involves human participants. It was reviewed and approved by the ethics committee of the Faculty of Social Welfare and Health Sciences, University of Haifa. Protocol 362/18 participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

LL and KB designed the study, prepared the study materials, and edited the manuscript. LS recruited study participants, collected and analyzed the data, and wrote the manuscript with oversight and conceptual guidance from KB and LL. All authors approved the submitted version.

FUNDING

This study was supported by the Israel Science Foundation (grant number 206/18).

ACKNOWLEDGMENTS

The authors wish to thank Mor Ben-Dor for her help in the data collection.

REFERENCES

Ahissar, M., Nahum, M., Nelken, I., and Hochstein, S. (2009). Reverse hierarchies and sensory learning. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 285–299. doi: 10.1098/rstb.2008.0253

Akeroyd, M. A. (2008). Are individual differences in speech reception related to individual differences in cognitive ability? A survey of twenty experimental studies with normal and hearing-impaired adults. *Int. J. Audiol.* 47(Suppl. 2), S53–S71. doi: 10.1080/14992020802301142

Anderson, S., Parbery-Clark, A., Yi, H. G., and Kraus, N. (2011). A neural basis of speech-in-noise perception in older adults. *Ear Hear.* 32, 750–757. doi: 10.1097/AUD.0b013e31822229d3

Banai, K., and Lavie, L. (2020). Perceptual learning and speech perception: A new hypothesis. *Proceedings of the International Symposium on Auditory and Audiological Research*, 7, 53–60.

Banai, K., and Lavie, L. (2021). Rapid perceptual learning and individual differences in speech perception: The good, the bad, and the sad. *Audit. Percept. Cognition* 3, 201–211. doi: 10.1080/25742442.2021.1909400

Bates, D., Maechler, M., Bolker, B., and Walker, S. (2014). Lme4: linear mixed-effects models using Eigen and S4. R package version 1, 1–7.

Ben-David, B. M., Erel, H., Goy, H., and Schneider, B. A. (2015). "Older is always better": age-related differences in vocabulary scores across 16 years. *Psychol. Aging* 30, 856–862. doi: 10.1037/pag0000051

Bent, T., Baese-Berk, M., Borrie, S. A., and McKee, M. (2016). Individual differences in the perception of regional, nonnative, and disordered speech varieties. *J. Acoust. Soc. Am.* 140, 3775–3786. doi: 10.1121/1.4966677

Bieber, R. E., and Gordon-Salant, S. (2021). Improving older adults' understanding of challenging speech: auditory training, rapid adaptation and perceptual learning. *Hear. Res.* 402, 1–68. doi: 10.1016/j.heares.2020.108054

Borrie, S. A., Baese-Berk, M., Van Engen, K., and Bent, T. A. (2017). A relationship between processing speech in noise and dysarthric speech. *J. Acoust. Soc. Am.* 141, 4660–4667. doi: 10.1121/1.4986746

Bronkhorst, A. W. (2015). The cocktail-party phenomenon revisited: early processing and selection of multi-talker speech. *Atten. Percept. Psychophys.* 77, 1465–1487. doi: 10.3758/s13414-015-0882-9

Burke, D. M., and Shafto, M. A. (2008). "Language and aging," in *The Handbook of Aging and Cognition*. eds. F. I. M. Craik and T. A. Salthouse. *3rd Edn.* (New York: Psychology Press), 373–443.

Carbonell, K. M. (2017). Reliability of individual differences in degraded speech perception. *J. Acoust. Soc. Am.* 142, EL461–EL466. doi: 10.1121/1.5010148

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and two ears. *J. Acoust. Soc. Am.* 25, 975–979. doi: 10.1121/1.1907229

Committee on Hearing, Bioacoustics and Biomechanics (CHABA) (1988). Working group on speech understanding, Speech understanding and aging. *J. Acoust. Soc. Am.* 83, 859–895. doi: 10.1121/1.395965

Cox, R. M., and Xu, J. (2010). Short and long compression release times: speech understanding, real world preferences, and association with cognitive ability. *J. Am. Acad. Audiol.* 21, 121–138. doi: 10.3766/jaaa.21.2.6

Dale, G., Cochrane, A., and Green, C. S. (2021). Individual difference predictors of learning and generalization in perceptual learning. *Atten. Percept. Psychophys.* 83, 2241–2255. doi: 10.3758/s13414-021-02268-3

Dawes, P., and Munro, K. J. (2017). Auditory distraction and acclimatization to hearing aids. *Ear Hearing* 38, 174–183. doi: 10.1097/AUD.0000000000000366

Dawes, P., Munro, K. J., Kalluri, S., and Edwards, B. (2013). Unilateral and bilateral hearing aids, spatial release from masking and auditory acclimatization. *J. Acoust. Soc. Am.* 134, 596–606. doi: 10.1121/1.4807783

Dawes, P., Munro, K. J., Kalluri, S., and Edwards, B. (2014a). Auditory acclimatization and hearing aids: late auditory evoked potentials and speech recognition following unilateral and bilateral amplification. *J. Acoust. Soc. Am.* 135, 3560–3569. doi: 10.1121/1.4874629

Dawes, P., Munro, K. J., Kalluri, S., and Edwards, B. (2014b). Acclimatization to hearing aids. *Ear Hear.* 35, 203–212. doi: 10.1097/AUD.0b013e3182a8eda4

Dias, J. W., McClaskey, C. M., and Harris, K. C. (2019). Time-compressed speech identification is predicted by auditory neural processing, Perceptuomotor speed, and executive functioning in younger and older listeners. *J. Assoc. Res. Otolaryngol.* 20, 73–88. doi: 10.1007/s10162-018-00703-1

Dryden, A., Allen, H. A., Henshaw, H., and Heinrich, A. (2017). The association Between cognitive performance and speech-in-noise perception for adult listeners: A systematic literature review and meta-analysis. *Trends Hear.* 21:44675. doi: 10.1177/2331216517744675

Dubno, J. R., Ahlstrom, J. B., and Horwitz, A. R. (2008). Binaural advantage for younger and older adults with normal hearing. *J. Speech Lang. Hear. Res.* 51, 539–556. doi: 10.1044/1092-4388(2008/039)

Dubno, J. R., Dirks, D. D., and Morgan, D. E. (1984). Effects of age and mild hearing loss on speech recognition in noise. *J. Acoust. Soc. Am.* 76, 87–96. doi: 10.1121/1.391011

Duthey, B. (2013). Background paper 6.21 hearing loss. Geneva: WHO.

Eriksen, B. A., and Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Percept. Psychophys.* 16, 143–149. doi: 10.3758/BF03203267

Festen, J. M., and Plomp, R. (1990). Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing. *J. Acoustic. Soc. Am.* 88, 1725–1736. doi: 10.1121/1.400247

Folstein, M., Folstein, S., and McHugh, P. (1975). "mini-mental state": a practical method for grading the cognitive state of patients for the clinician. *J. Psychiatr. Res.* 12, 189–198. doi: 10.1016/0022-3956(75)90026-6

Forstmann, B. U., Tittgemeyer, M., Wagenmakers, E. J., Derrfuss, J., Imperati, D., and Brown, S. (2011). The speed-accuracy tradeoff in the elderly brain: A structural model-based approach. *J. Neurosci.* 31, 17242–17249. doi: 10.1523/JNEUROSCI.0309-11.2011

Frisina, D. R., and Frisina, R. D. (1997). Speech recognition in noise and presbycusis: relations to possible neural mechanisms. *Hear. Res.* 106, 95–104. doi: 10.1016/s0378-5955(97)00006-3

Füllgrabe, C., Moore, B. C. J., and Stone, M. A. (2015). Age-group differences in speech identification despite matched audiometrically normal hearing: contributions from auditory temporal processing and cognition. *Front. Aging Neurosci.* 6, 1–25. doi: 10.3389/fnagi.2014.00347

Gatehouse, S. (1992). The time course and magnitude of perceptual acclimatization to frequency responses: evidence from monaural fitting of hearing aids. *J. Acoust. Soc. Am.* 92, 1258–1268. doi: 10.1121/1.403921

Gatehouse, S., Naylor, G., and Elberling, C. (2006). Linear and nonlinear hearing aid fittings – 2. Patterns of candidature. *Int. J. Audiol.* 45, 153–171. doi: 10.1080/14992020500429484

Gates, G. A., and Mills, J. H. (2005). Presbycusis. *Lancet* 366, 1111–1120. doi: 10.1016/S0140-6736(05)67423-5

Golomb, J. D., Peelle, J. E., and Wingfield, A. (2007). Effects of stimulus variability and adult aging on adaptation to time-compressed speech. *J. Acoust. Soc. Am.* 121, 1701–1708. doi: 10.1121/1.2436635

Gordon-Salant, S., and Fitzgibbons, P. J. (1993). Temporal factors and speech recognition performance in young and elderly listeners. *J. Speech Hear. Res.* 36, 1276–1285. doi: 10.1121/1.2436635

Gordon-Salant, S., and Fitzgibbons, P. J. (2001). Sources of age-related recognition difficulty for time-compressed speech. *J. Speech Lang. Hear. Res.* 44, 709–719. doi: 10.1044/1092-4388(2001/056)

Gordon-Salant, S., Zion, D. J., and Espy-Wilson, C. (2014). Recognition of time-compressed speech does not predict recognition of natural fast-rate speech by older listeners. *J. Acoust. Soc. Am.* 136, EL268–EL274. doi: 10.1121/1.4895014

Green, P., MacLeod, C. J., and Nakagawa, S. (2016). SIMR: an R package for power analysis of generalized linear mixed models by simulation. *Methods Ecol. Evol.* 7, 493–498. doi: 10.1111/2041-210X.12504

Habicht, J., Kollmeier, B., and Neher, T. (2016). Are experienced hearing aid users faster at grasping the meaning of a sentence than inexperienced users? An eye-tracking study. *Trend. Hear.* 20, 1–13. doi: 10.1177/2331216516660966

Hargus, S. E., and Gordon-Salant, S. (1995). Accuracy of speech intelligibility index predictions for noise-masked young listeners with normal hearing and for elderly listeners with hearing impairment. *J. Speech Hear. Res.* 38, 234–243. doi: 10.1044/jshr.3801.234

Heffner, C. C., and Myers, E. B. (2021). Individual differences in phonetic plasticity Across native and nonnative contexts. *J. Speech Lang. Hear.* 64, 3720–3733. doi: 10.1044/2021_JSLHR-21-00004

Helfer, K. S., and Freyman, R. L. (2008). Aging and speech-on-speech masking. *Ear Hear.* 29, 87–98. doi: 10.1097/AUD.0b013e31815d638b

Hopkins, K., and Moore, B. C. (2007). Moderate cochlear hearing loss leads to a reduced ability to use temporal fine structure information. *J. Acoust. Soc. Am.* 122, 1055–1068. doi: 10.1121/1.2749457

Hopkins, K., and Moore, B. C. (2011). The effects of age and cochlear hearing loss on temporal fine structure sensitivity, frequency selectivity, and speech reception in noise. *J. Acoust. Soc. Am.* 130, 334–349. doi: 10.1121/1.3585848

Humes, L. E. (1996). Speech understanding in the elderly. *J. Am. Acad. Audiol.* 7, 161–167.

Humes, L. E. (2002). Factors underlying the speech-recognition performance of elderly hearing-aid wearers. *J. Acoust. Soc. Am.* 112, 1112–1132. doi: 10.1121/1.1499132

Humes, L. E. (2007). The contributions of audibility and cognitive factors to the benefit provided by amplified speech to older adults. *J. Am. Acad. Audiol.* 18, 590–603. doi: 10.3766/jaaa.18.7.6

Humes, L. E., and Dubno, J. R. (2010). "Factors affecting speech understanding in older adults," in *The Aging Auditory System. Vol. 34.* eds. S. Gordon-Salant, R. D. Frisina, A. N. Popper and R. R. Fay (New York: Springer), 167–210.

Humes, L. E., Dubno, J. R., Gordon-Salant, S., Lister, J. J., Cacace, A. T., Cruickshanks, K. J., et al. (2012). Central presbycusis: a review and evaluation of the evidence. *J. Am. Acad. Audiol.* 23, 635–666. doi: 10.3766/jaaa.23.8.5

Humes, L. E., Garner, C. B., Wilson, D. L., and Barlow, N. N. (2001). Hearing-aid outcome measured following one month of hearing aid use by the elderly. *J. Speech Lang. Hear. Res.* 44, 469–486. doi: 10.1044/1092-4388(2001/037)

Humes, L. E., Lee, J. H., and Coughlin, M. P. (2006). Auditory measures of selective and divided attention in young and older adults using single-talker competition. *J. Acoust. Soc. Am.* 120, 2926–2937. doi: 10.1121/1.2354070

Humes, L. E., and Roberts, L. (1990). Speech-recognition difficulties of the hearing-impaired elderly: the contributions of audibility. *J. Speech Hear. Res.* 33, 726–735. doi: 10.1044/jshr.3304.726

Janse, E. (2004). Word perception in fast speech: artificially time-compressed vs. naturally produced fast speech. *Speech Comm.* 42, 155–173. doi: 10.1016/j.specom.2003.07.001

Janse, E. (2009). Processing of fast speech by elderly listeners. *J. Acoust. Soc. Am.* 125, 2361–2373. doi: 10.1121/1.3082117

Jerger, J., Alford, B., Lew, H., Rivera, V., and Chmiel, R. (1995). Dichotic listening, event-related potentials, and interhemispheric transfer in the elderly. *Ear Hear.* 16, 482–498. doi: 10.1097/00003446-199510000-00005

Jerger, J., Chmiel, R., Allen, J., and Wilson, A. (1994). Effects of age and gender on dichotic sentence identification. *Ear Hear.* 15, 274–286. doi: 10.1097/00003446-199408000-00002

Jerger, J., Jerger, S., and Pirozzolo, F. (1991). Correlational analysis of speech audiometric scores, hearing loss, age, and cognitive abilities in the elderly. *Ear Hear.* 12, 103–109. doi: 10.1097/00003446-199104000-00004

Kahneman, D. (1973). Attention and effort. Prentice–Hall.

Kalluri, S., Ahmann, B., and Munro, K. J. (2019). A systematic narrative synthesis of acute amplification-induced improvements in cognitive ability in hearing-impaired adults. *Int. J. Audiol.* 58, 455–463. doi: 10.1080/14992027.2019.1594414

Karawani, H., Lavie, L., and Banai, K. (2017). Short-term auditory learning in older and younger adults. *Proceedings of the International Symposium on Auditory and Audiological Research*, 6, 1–8.

Kidd, G. Jr., Mason, C. R., and Feth, L. L. (1984). Temporal integration of forward masking in listeners having sensorineural hearing loss. *J. Acoust. Soc. Am.* 75, 937–944. doi: 10.1121/1.390558

Killion, M. C. (1997). The SIN report. *Hear. J.* 50, 28–30. doi: 10.1097/00025572-199710000-00002

Koreman, J. (2006). Perceived speech rate: the effects of articulation rate and speaking style in spontaneous speech. *J. Acoust. Soc. Am.* 119, 582–596. doi: 10.1121/1.2133436

Larson, V. D., Williams, D. W., Henderson, W. G., Luethke, L. E., Beck, L. B., Noffsinger, D., et al. (2000). Efficacy of 3 commonly used hearing aid circuits: A crossover trial. *JAMA* 284, 1806–1813. doi: 10.1001/jama.284.14.1806

Lavie, L. (2011). Plasticity of the Auditory System in the Elderly Following Hearing Aids Usage. Doctoral Dissertation, University of Haifa, Israel.

Lavie, L., Attias, J., and Karni, A. (2013). Semi-structured listening experience (listening training) in hearing aid fitting: influence on dichotic listening. *Am. J. Audiol.* 22, 347–350. doi: 10.1044/1059-0889(2013/12-0083)

Lavie, L., Banai, K., Karni, A., and Attias, J. (2015). Hearing aid-induced plasticity in the auditory system of older adults: evidence from speech perception. *J. Speech Lang. Hear. Res.* 58, 1601–1610. doi: 10.1044/2015_JSLHR-H-14-0225

Lavie, L., Banai, K., and Shechter Shvartzman, L. (2021). Plastic changes in speech perception in hearing-impaired older adults following hearing aid use: a systematic review. *Int. J. Audiol.*, 1–9. doi: 10.1080/14992027.2021.2014073

Leek, M. R., and Summers, V. (1993). Auditory filter shapes of normal- hearing and hearing- impaired listeners in continuous broadband noise. *J. Acoust. Soc. Am.* 94, 3127–3137. doi: 10.1121/1.407218

Levitt, H., and Rabiner, L. R. (1967). Binaural release from masking for speech and gain in intelligibility. *J. Acoust. Soc. Am.* 42, 601–608. doi: 10.1121/1.1910629

Lu, P. H., Lee, G. J., Raven, E. P., Tingus, K., Khoo, T., Thompson, P. M., et al. (2011). Age-related slowing in cognitive processing speed is associated with myelin integrity in a very healthy elderly sample. *J. Clin. Exp. Neuropsychol.* 33, 1059–1068. doi: 10.1080/13803395.2011.595397

Lunner, T. (2003). Cognitive function in relation to hearing aid use. *Int. J. Audiol.* 42, 49–58. doi: 10.3109/14992020309074624

Luterman, D. M., Welsh, O. L., and Melrose, J. (1966). Responses of aged males to time-altered speech stimuli. *J. Speech Hear. Res.* 9, 226–230. doi: 10.1044/jshr.0902.226

Manheim, M., Lavie, L., and Banai, K. (2018). Age, hearing, and the perceptual learning of rapid speech. *Trend. Hear.* 22, 1–18. doi: 10.1177/2331216518778651

Mattys, S. L., Davis, M. H., Bradlow, A. R., and Scott, S. K. (2012). Speech recognition in adverse conditions: A review. *Lang. Cogni. Proces.* 27, 953–978. doi: 10.1080/01690965.2012.705006

McDowd, J. M., and Shaw, R. J. (2000). "Attention and aging: A functional perspective," in *Handbook of Aging and Cognition*. eds. F. I. M. Craik and T. A. Salthouse. 2nd ed (New York: Erlbaum), 221–292.

McLaughlin, D. J., Baese-Berk, M. M., Bent, T., Borrie, S. A., and Van Engen, K. J. (2018). Coping with adversity: individual differences in the perception of noisy and accented speech. *Atten. Percept. Psychophys.* 80, 1559–1570. doi: 10.3758/s13414-018-1537-4

Moore, B. C. (2008). The role of temporal fine structure processing in pitch perception, masking, and speech perception for normal-hearing and hearing-impaired people. *J. Assoc. Res. Otolaryngol.* 9, 399–406. doi: 10.1007/s10162-008-0143-x

Moore, B. C. J. (2011). The importance of temporal fine structure for the intelligibility of speech in complex backgrounds. *Proceedings of the International symposium on Auditory and Audiological Research*, 3, 21–32.

Morrell, C. H., Gordon-Salant, S., Pearson, J. D., Brant, L. J., and Fozard, J. L. (1996). Age- and gender-specific reference ranges for hearing level and longitudinal changes in hearing level. *J. Acoust. Soc. Am.* 100, 1949–1967. doi: 10.1121/1.417906

Munro, K. J., and Lutman, M. E. (2003). The effect of speech presentation level on measurement of auditory acclimatization to amplified speech. *J. Acoust. Soc. Am.* 114, 484–495. doi: 10.1121/1.1577556

Needleman, A. R., and Crandell, C. C. (1995). Speech recognition in noise by hearing-impaired and noise-masked normal-hearing listeners. *J. Am. Acad. Audiol.* 6, 414–424

Neher, T., Grimm, G., and Hohmann, V. (2014). Perceptual consequences of different signal changes due to binaural noise reduction: do hearing loss and working memory capacity play a role? *Ear Hear.* 35, e213–e227. doi: 10.1097/AUD.0000000000000054

Nelson, D. A., and Freyman, R. L. (1987). Temporal resolution in sensorineural hearing-impaired listeners. *J. Acoust. Soc. Am.* 81, 709–720. doi: 10.1121/1.395131

Peelle, J. E., and Wingfield, A. (2005). Dissociations in perceptual learning revealed by adult age differences in adaptation to time-compressed speech. *J. Exp. Psychol. Hum. Percept. Perform.* 31, 1315–1330. doi: 10.1037/0096-1523.31.6.1315

Peters, R. W., and Moore, B. C. (1992). Auditory filter shapes at low center frequencies in young and elderly hearing- impaired subjects. *J. Acoust. Soc. Am.* 91, 256–266. doi: 10.1121/1.402769

Phillips, S. L., Gordon-Salant, S., Fitzgibbons, P. J., and Yeni-Komshian, G. (2000). Frequency and temporal resolution in elderly listeners with good and poor word recognition. *J. Speech Lang. Hear. Res.* 43, 217–228. doi: 10.1044/jslhr.4301.217

Pichora-Fuller, M. K. (1997). Language comprehension in older adults. *J. Speech Lan. Pathol. Audiol.* 21, 125–142.

Pichora-Fuller, M. K. (2003a). Cognitive aging and auditory information processing. *Int. J. Audiol.* 42(Suppl. 1):26. doi: 10.3109/14992020309074641

Pichora-Fuller, M. K. (2003b). Processing speed and timing in aging adults: psychoacoustics, speech perception, and comprehension. *Int. J. Audiol.* 42(Suppl. 2), S59–S67. doi: 10.3109/14992020309074625

Pichora-Fuller, M. K., Schneider, B. A., and Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *J. Acoust. Soc. Am.* 97, 593–608. doi: 10.1121/1.412282

Pichora-Fuller, M. K., and Singh, G. (2006). Effects of age on auditory and cognitive processing implications for hearing aid fitting and audiologic rehabilitation. *Trends Amplif.* 10, 29–59. doi: 10.1177/108471380601000103

Pichora-Fuller, M. K., and Souza, P. E. (2003). Effects of aging on auditory processing of speech. *Int. J. Audiol.* 42:(Suppl. 2), 11. doi: 10.3109/14992020309074638

Prior, A., and Bentin, S. (2006). Differential integration efforts of mandatory and optional sentence constituents. *Psychophysiology* 43, 440–449. doi: 10.1111/j.1469-8986.2006.00426.x

Rabbitt, P. M. A. (1990). Mild hearing loss can cause apparent memory failures which increase with age and reduce with IQ. *Acta Otolaryngologica (Suppl.* 111), 167–176. doi: 10.3109/00016489109127274

Reitan, R. M. (1958). Validity of the trail making test as an indicator of organic brain damage. *Percept. Mot. Skills* 8, 271–276. doi: 10.2466/pms.1958.8.3.271

Rogers, W. A. (2000). "Attention and aging," in *Cognitive Aging: A primer*. eds. D. C. Park and N. Schwarz (New York: Psychology Press), 57–73.

Rönnberg, J., Lunner, T., Zekveld, A., Sörqvist, P., Danielsson, H., Lyxell, B., et al. (2013). The ease of language understanding (ELU) model: theoretical, empirical, and clinical advances. *Front. Syst. Neurosci.* 7, 1–17. doi: 10.3389/fnsys.2013.00031

Rosen, S. (1992). Temporal information in speech: acoustic, auditory and linguistic aspects. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 336, 367–373. doi: 10.1098/rstb.1992.0070

Rotman, T., Lavie, L., and Banai, K. (2020a). Rapid perceptual learning of time-compressed speech and the perception of natural fast speech in older adults with presbycusis. *Proceedings of the International symposium on Auditory and Audiological Research*, 7, 93–100.

Rotman, T., Lavie, L., and Banai, K. (2020b). Rapid perceptual learning: a potential source of individual differences in speech perception under adverse conditions? *Trend. Hear.* 24, 1–16. doi: 10.1177/2331216520930541

Roup, C. M., Wiley, T. L., and Wilson, R. H. (2006). Dichotic word recognition in young and older adults. *J. Am. Acad. Audiol.* 54, 292–297. doi: 10.1044/1092-4388(2010/09-0230)

Salthouse, T. A. (1994). The aging of working memory. *Neuropsychology* 8, 535–543. doi: 10.1037/0894-4105.8.4.535

Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychol. Rev.* 103, 403–428. doi: 10.1037/0033-295X.103.3.403

Samuel, A. G., and Kraljic, T. (2009). Perceptual learning for speech. *Atten. Percept. Psychophys.* 71, 1207–1218. doi: 10.3758/APP.71.6.1207

Schneider, B. A., Daneman, M., and Pichora-Fuller, M. K. (2002). Listening in aging adults: from discourse comprehension to psychoacoustics. *Can. J. Exp. Psychol.* 56, 139–152. doi: 10.1037/h0087392

Schneider, B. A., Li, L., and Daneman, M. (2007). How competing speech interferes with speech comprehension in everyday listening situations. *J. Am. Acad. Audiol.* 18, 559–572. doi: 10.3766/jaaa.18.7.4

Schneider, B. A., and Pichora-Fuller, M. K. (2001). Age-related changes in temporal processing: implications for listening comprehension. *Semin. Hear.* 22, 227–240. doi: 10.1177/108471380601000103

Schneider, B. A., Pichora-Fuller, K., and Daneman, M. (2010). "Effects of senescent changes in audition and cognition on spoken language comprehension," in *The Aging Auditory System*. eds. S. Gordon-Salant, R. D. Frisina, R. R. Fay and A. N. Popper (New York: Springer International Publishing), 167–210.

Schum, D. J., Matthews, L. J., and Lee, F. S. (1991). Actual and predicted word-recognition performance of elderly hearing-impaired listeners. *J. Speech Lang. Hear. Res.* 34, 636–642. doi: 10.1044/jshr.3403.636

Sheldon, S., Pichora-Fuller, M. K., and Schneider, B. A. (2008). Priming and sentence context support listening to noise-vocoded speech by younger and older adults. *J. Acoust. Soc. Am.* 123, 489–499. doi: 10.1121/1.2783762

Signoret, C., and Ruder, M. (2019). Hearing impairment and perceived clarity of predictable speech. *Ear Hear.* 40, 1140–1148. doi: 10.1097/AUD.0000000000000689

Souza, P. (2016). "Speech perception and hearing aids," in *Hearing Aids*. eds. G. R. Popelka, B. C. J. Moore, R. R. Fay and A. N. Popper (New York: Springer International Publishing), 151–180.

Souza, P., Arehart, K., and Neher, T. (2015). Working memory and hearing aid processing: literature findings, future directions, and clinical applications. *Front. Psychol.* 6, 1–12. doi: 10.3389/fpsyg.2015.01894

Tun, P. A. (1998). Fast noisy speech: age differences in processing rapid speech with background noise. *Psychol. Aging* 13, 424–434. doi: 10.1037/0882-7974.13.3.424

Tun, P. A., O'Kane, G., and Wingfield, A. (2002). Distraction by competing speech in young and older adult listeners. *Psychol. Aging* 17, 453–467. doi: 10.1037/0882-7974.17.3.453

Tun, P. A., and Wingfield, A. (1999). One voice too many: adult age differences in language processing with different types of distracting sounds. *J. Gerontol. Psychol. Sci.* 54, P317–P327. doi: 10.1093/geronb/54B.5.P317

Tun, P. A., Wingfield, A., Stine, E. A., and Mecsas, C. (1992). Rapid speech processing and divided attention: processing rate versus processing resources as an explanation of age effects. *Psychol. Aging* 7, 546–550. doi: 10.1037//0882-7974.7.4.546

Vaughan, N., Storzbach, D., and Furukawa, I. (2006). Sequencing versus nonsequencing working memory in understanding of rapid speech by older listeners. *J. Am. Acad. Audiol.* 17, 506–518. doi: 10.3766/jaaa.17.7.6

Verhaeghen, P. (2003). Aging and vocabulary score: a meta-analysis. *Psychol. Aging* 18, 332–339. doi: 10.1037/0882-7974.18.2.332

Verhelst, W., and Roelands, M. (1993). An overlap-add technique based on waveform similarity (Wsola) for high quality time-scale modification of speech [paper presentation]. *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Minneapolis, MN, United States.

Versfeld, N. J., and Dreschler, W. A. (2002). The relationship between the intelligibility of time-compressed speech and speech in noise in young and elderly listeners. *J. Acoust. Soc. Am.* 111, 401–408. doi: 10.1121/1.1426376

Vogelzang, M., Thiel, C. M., Rosemann, S., Rieger, J. W., and Ruigendijk, E. (2021). Effects of age-related hearing loss and hearing aid experience on sentence processing. *Sci. Rep.* 11, 5994–5914. doi: 10.1038/s41598-021-85349-5

Wechsler, D. (1997). *WAIS-3: Wechsler Adult Intelligence Scale: Administration and Scoring Manual*. Psychological Corporation.

Wingfeld, A., McCoy, S. L., Peelle, J. E., Tun, P. A., and Cox, C. L. (2006). Effects of adult aging and hearing loss on comprehension of rapid speech varying in syntactic complexity. *J. Am. Acad. Audiol.* 17, 487–497. doi: 10.3766/jaaa.17.7.4

Wingfield, A. (1996). Cognitive factors in auditory performance: context, speed of processing, and constraints of memory. *J. Am. Acad. Audiol.* 7, 175–182.

Wingfield, A., Poon, L. W., Lombardi, L., and Lowe, D. (1985). Speed of processing in Normal aging: effects of speech rate, linguistic structure, and processing time. *J. Gerontol.* 40, 579–585. doi: 10.1093/geronj/40.5.579

Wright, D., and Gagné, J. P. (2020). Acclimatization to hearing aids by older adults. *Ear Hear.* 42, 193–205. doi: 10.1097/AUD.0000000000000913

Yang, J., Yan, F. F., Chen, L., Xi, J., Fan, S., Zhang, P., et al. (2020). General learning ability in perceptual learning. *Proc. Natl. Acad. Sci. U. S. A.* 117, 19092–19100. doi: 10.1073/pnas.2002903117

Conflict of Interest: 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 Psychology

ORIGINAL RESEARCH
published: 17 February 2022
doi: 10.3389/fpsyg.2022.820227

A Clinical Paradigm for Listening Effort Assessment in Middle-Aged Listeners

Ricky Kaplan Neeman[1,2]*, Ilan Roziner[1] and Chava Muchnik[1,2]

[1] Department of Communication Disorders, The Stanley Steyer School of Health Professions, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel, [2] Hearing, Speech and Language Center, Sheba Medical Cente, Ramat-Gan, Israel

Listening effort (LE) has been known to characterize speech recognition in noise regardless of hearing sensitivity and age. Whereas the behavioral measure of dual-task paradigm effectively manifests the cognitive cost that listeners exert when processing speech in background noise, there is no consensus as to a clinical procedure that might best express LE. In order to assess the cognitive load underlying speech recognition in noise and promote counselling for coping strategies, a feasible clinical paradigm is warranted. The ecological validity of such a paradigm might best be demonstrated in middle-aged adults, exhibiting intact hearing sensitivity on one hand, however, experiencing difficulties in degraded listening conditions, unaware of the implicated cognitive cost of speech recognition in noise. To this end, we constructed a dual-task paradigm that consists of a primary task of sentences-in-noise recognition and a secondary task of simple visual colored-shape matching. Research objective was to develop a clinical paradigm for the assessment of LE in middle-aged adults. Participants were 17 middle-aged adults (mean age of 52.81 years) and 23 young adults (mean age of 24.90 years). All participants had normal hearing according to age. Speech stimuli consisted of the Hebrew Matrix sentences in noise test. SRTn was obtained for 80% correct identification. Visual stimuli were colored geometric shapes. Outcome measures were obtained initially for each task separately, to establish performance ability, and then obtained simultaneously. Reaction time and accuracy in the secondary task were the defined metrics for LE. Results: LE was indicated for both groups, however, was more pronounced in the middle-aged, manifested in the visual accuracy and reaction time metrics. Both groups maintained the 80% correct recognition-in-noise in the dual-task, however, the middle-aged group necessitated a better SNR of 1.4dB than the normal hearing group. Moreover, the middle-aged group was taxed in a greater prolongation of reaction time, in order to uphold the correct recognition. Conclusion: a dual-task paradigm consisting of sentences-in-noise primary task combined with a simple secondary task successfully showed different manifestations of LE in middle-aged adults compared to young adults, thus approximating the use of such a paradigm in a clinical setting.

Keywords: listening effort, dual-task paradigm, middle- aged adults, cognitive cost, a clinical paradigm

OPEN ACCESS

Edited by:
Leah Fostick,
Ariel University, Israel

Reviewed by:
Erol J. Ozmeral,
University of South Florida,
United States
Viveka Lyberg Åhlander,
Åbo Akademi University, Finland

****Correspondence:***
Ricky Kaplan Neeman
kaplann@tauex.tau.ac.il

Specialty section:
This article was submitted to
Auditory Cognitive Neuroscience,
a section of the journal
Frontiers in Psychology

Received: *22 November 2021*
Accepted: *24 January 2022*
Published: *17 February 2022*

Citation:
Kaplan Neeman R, Roziner I and
Muchnik C (2022) A Clinical Paradigm
for Listening Effort Assessment
in Middle-Aged Listeners.
Front. Psychol. 13:820227.
doi: 10.3389/fpsyg.2022.820227

INTRODUCTION

Unraveling the difficulty of speech recognition in background noise has been a major challenge in hearing research for many years. The cause-effect relationship is still under investigation. In addition to speech stimuli attributes, masker noise types, and various characteristics of the listener, such as hearing sensitivity and age (e.g., Dubno, 2015), the cognitive component has been established as a key factor in the challenge (e.g., Gordon-Salant and Samuels Cole, 2016). The ability to suppress irrelevant, distracting context and focus on desired target information is essential for speech understanding in noise (Pichora-Fuller et al., 2016). Moreover, the listener sometimes is required to perform several tasks concurrently, while ignoring background noise (Gagné et al., 2017), and, therefore, is faced with a greater cognitive load (Peele, 2018). The cognitive cost that the listener is burdened with in such complex situations is termed: "listening effort" (LE). As stated by Pichora-Fuller et al. (2016), listening effort refers to "the deliberate allocation of mental resources to overcome obstacles in goal pursuit when carrying out a listening task". Demanding listening conditions on one hand, and increased motivation to overcome the distractions on the other hand, will affect the extent of cognitive resources allocated toward accomplishment of the target task.

Accumulating evidence shows increase in LE in the elderly (Tun et al., 2009; Gosselin and Gagné, 2011; Sommers and Phelps, 2016). As both cognitive ability and peripheral auditory function are known to decline with age (Tremblay and Backer, 2016), it is expected that older adults will exert more LE than young adults. In the middle-age (MA), on the other hand, it is more difficult to pre-establish expectations. Hearing acuity, as well as other auditory processing abilities, might not decline at the same manner. Whereas MA adults might not exhibit pure-tone thresholds elevation in the audiogram (Helfer et al., 2017), they were found to have complaints concerning their ability to understand speech in noise (Lee et al., 2015; Helfer et al., 2017). This finding was supported by research studies' evidence of deteriorated speech perception in noise (e.g., Lee et al., 2015; Goossens et al., 2017). In an attempt to explain these MA-related speech perception difficulties in the presence of normal hearing thresholds, it was assumed that temporal processing deficiencies might underlie some of these difficulties. Indeed, behavioral studies have found reduced supra-threshold temporal auditory processing capacities (Helfer and Vargo, 2009; Füllgrabe, 2013). Moreover, electrophysiological data demonstrated neural encoding deficits of temporal fine structure in participants aged 51-67 years (Clinard and Cotter, 2015). Nonetheless, the contribution of cognitive factors was argued to serve as a fundamental aspect in the decline of speech recognition in noise in the middle-aged (Helfer et al., 2017). Studies concerning LE in MA adults might shed more light on cognitive demands of speech recognition in noise. These studies, however, are scarce. Typically, LE in the MA group was studied as a part of a large age range of normal-hearing participants (e.g., Degeest et al., 2015), or in hearing-impaired participants (e.g., Desjardins and Doherty, 2013). Degeest et al. (2015) were among the first and few researchers that explored the effect of age on LE, in a group of 60 adults, aged 20-77 years. The primary task was recognition of digits-in-noise, and the secondary task required visual memory of the position of geometric figures on a screen. In order to rule out hearing sensitivity, the authors equated the experiment listening conditions, controlling for effects of differential speech intelligibility scores. Results showed that LE increased initially in the fourth decade of life and was related to the cognitive attribute of speech recognition in degraded listening conditions. Devesse et al. (2020) were among the few studies that focused specifically on participants in the age range of 45-60 years. The performance of 29 middle-aged adults was compared to that of 35 young adults in auditory-visual speech-in noise task that combined dual, triple, and quadruple secondary tasks, to approximate real life situations. Middle-aged adults were found to perform worse than the young adults in all tasks. Their findings highlighted the difficulties of speech in noise understanding of MA adults and their need to allocate cognitive resources in order to meet speech understanding in noise requirements.

Owing to the fact that speech recognition in noise partakes a fundamental role in audiological assessment, alongside with established data concerning age-dependent difficulties in speech in noise recognition, the need for integrating LE measures in the clinic emerges. A clinical measure of LE might demonstrate the listener's taxed cognitive capacity and provide means to identify the need for specific counseling and rehabilitation procedures (McGarrigle et al., 2014). Furthermore, such a measure might elucidate aspects of hearing disability, not yet manifested in hearing thresholds and correct recognition of speech stimuli (Lewis et al., 2016; Gagné et al., 2017; Alhanbali et al., 2019). A clinical measure of LE could be used when traditional speech perception tests result in ceiling effect (Houben et al., 2013), and might support hearing aids fitting by adequate adaptation of specific features that reduce LE (Hornsby, 2013) as well as help select a best-fit cochlear implant program (Pals et al., 2013).

Measures of LE vary among studies. Pupillometry was suggested as a sensitive measure reflecting the cognitive load encountered by the adult listener (Peele, 2018), however, dual-task measures might prove logistically more feasible for the clinical setting. In addition, as performing another task while processing speech is a ubiquitous situation, dual-task paradigms hold ecological validity (Gagné et al., 2017). Despite the great variability of dual-task experimental procedures described in the literature, there is no consensus as to a clinical procedure that might best express LE. The idea that LE is manifested in the secondary task measures led several researchers to characterize the appropriate secondary task that might best demonstrate the cognitive load inflicted upon the listener, in certain speech recognition in noise conditions. It has been suggested that a simple secondary task might not elicit the use of cognitive resources, but rather induce adaptation and habituation (Hasher and Zacks, 1979). For example, it has been shown that very little, or no change at all, was evident in LE while using a simple secondary task that required a button-press response when a red rectangle appeared on a screen. Conversely, a secondary task that demanded semantic judgment of noun recognition yielded increased sensitivity to LE (Picou et al., 2013; Picou and Ricketts, 2014). Alternatively, Ward et al. (2017) found that a visual monitoring task involving a key-press when a gray-scale image occurred twice (in a sequence of 206 images), demanded the use of cognitive processes, and LE was exhibited in the dual-task condition. Therefore, while task complexity might not solely indicate its compatibility for a secondary task, task modality also might influence dual-task performance and in turn, the allocation of cognitive resources appropriately. As denoted by Kahneman (1973), when both tasks, the primary and the secondary, draw resources from the same resource pool, performance in the primary task might be compromised. Kim et al. (2005) demonstrated increased interference in a Stroop meaning-comparison primary task, when the secondary task demanded recall of Korean verbal characters (letters). Accordingly, when both tasks engaged the phonological loop (Baddeley et al., 1998), the same limited resource pool interfered in the primary task performance. By contrast, a secondary task from a different modality, might prompt reallocation of unused resources with available reserve capacity. This idea is substantiated by studies using various visual secondary tasks that did not affect primary speech recognition in noise tasks (e.g., Hughes and Galvin, 2013; Ward et al., 2017), consistent with domain-specific attentional resources assumptions (e.g., Baddeley and Logie, 1999). Accordingly, primary and secondary tasks pertaining to different domains might better manifest LE, while preserving primary task performance (Grieco-Calub et al., 2017).

In face of the very few studies that investigated LE in the middle age specifically, and the need to incorporate LE in the audiology clinic, the purpose of the current study was to develop a clinical paradigm for the assessment of LE in middle-aged normal hearing (age-dependent) adults. In order for the paradigm to be well-suited to the clinical setting, and at the same time approximate real-life situations, the primary task consisted of sentences recognition-in-noise, and the secondary task was a simple, visual, basic shape-matching task.

MATERIALS AND METHODS

Participants

Twenty-three young female adults (range 21.33-28.34 years, mean = 24.90, *SD* = 1.86) and 17 middle-aged (seven males, ten females) adults (range 42.33-65.90 years, mean = 52.81, *SD* = 7.76) participated in the study. All participants self-reported no history of ear diseases, used Hebrew as their primary language, did not present attention disorders, and had no experience in hearing-in-noise experiments. Hearing thresholds in the young group did not exceed 15dBHL at octave frequencies from 0.25 through 8 kHz. In the middle-aged group, hearing thresholds were normal to age (in accordance with the 75th percentile: Engdahl et al., 2005) at the same frequencies. All participants were volunteers, and signed an informed consent form prior to data collection. The study was approved by the Institutional Review Board at Tel Aviv University.

Stimuli

Speech stimuli consisted of the Hebrew version of the Matrix sentences in noise test (Bugannim et al., 2019). Speech reception threshold in noise (SRTn) was obtained for the 80% of the words that were repeated correctly, using an adaptive procedure. Background noise was steady-state, test-specific, speech shaped noise, generated by superimposition of all sentences, presented at a fixed level of 60dBSPL. Sentences and noise were presented at initial SNR of 0dB, followed by increase or decrease of sentences level, depending on listeners correct word recognition.

Visual stimuli were three geometric shapes: squares, triangles and circles, in the colors of red, green and yellow (Hughes and Galvin, 2013). A colored shape was presented on a touch-screen for 0.5 second, followed by four colored shapes: the test shape and three foils. Participants had to touch the test shape they saw earlier.

Testing Apparatus

Testing was conducted in a sound-attenuating room. Participants sat on a chair, facing a loudspeaker located at a distance of one meter, 0° azimuth. Speech stimuli were presented from a Toshiba Satellite Pro laptop, routed through Auritec GmbH Earbox 3.0 sound card. Visual stimuli were displayed on a Sony S1 9.4" touchscreen tablet held by the participants, who indicated their response by touching the selected matched shape.

PROCEDURE

Dual-Task Paradigm

The dual-task paradigm consisted of a primary task: sentences recognition in noise, and a secondary task: visual shape-matching. Both primary and secondary tasks were performed initially as single tasks, and then simultaneously, as a dual-task.

Single task: A. At the beginning of the experiment, the shape-matching visual-motor task (secondary task) was performed for one minute, to familiarize the participants with the task. This time period allowed for presenting 25-36 shape-matching items. Participants were instructed to select and touch the matched shape as quickly and correctly as possible. Correct shape-matching and reaction time for each item were collected by the software. Following the practice trial, the shape-matching task was repeated for three minutes, allowing for presenting 70-105 items in order to equal the duration of each run of Matrix sentences. In keeping both primary and secondary tasks length identical, consistency across all test conditions was accomplished.

B. In the next stage, the Matrix sentences in noise was administered (primary task). Each Matrix sentences run consisted of 20 sentences, mixed with speech-shaped noise, presented at 60dBSPL, in initial SNR of 0dB. Participants were instructed to listen to each sentence and repeat aloud each word, as correctly as possible. Correct recognition of each word in a sentence led to a decrease in sentences intensity-level in relation to the noise intensity-level, thus decreasing SNR., whereas incorrect recognition led to an increase in sentences level, thus increasing the SNR. The first step-size was 3dB, followed by an exponential decrease in step-size, after each reversal of the presentation level. In the end, the speech reception threshold (SRTn) was calculated using the maximum likelihood method (Brand and Kollmeier, 2002). SRT 80% was obtained for each 20 sentence run. Participants performed three lists of 20 sentences due to the known training effect of the Matrix test (Kollmeier et al., 2015; Bugannim et al., 2019). As recommended by Kollmeier et al. (2015), each participant, being a naïve user of the test, performed two training lists of 20 sentences, and the speech reception threshold in noise (SRTn) was determined based on performance of the third list.

Dual-task: Subsequent to both single tasks performance, participants performed shape-matching and sentence recognition concurrently, instructed to give priority to the sentence's recognition task. Matrix sentences in noise were presented to each participant at the SRT 80% that was pre-determined at the single task trial. Stated differently, each participant performed the primary task in the dual-task condition at the SNR that yielded 80% recognition in the single task condition. Thus, listening conditions were fitted individually to participants ability of speech recognition in noise.

Data Analysis

All statistical analyzes were carried out using the IBM Statistical Package for the Social Sciences (SPSS) software version 27.0 for Windows. Descriptive statistics for the variables (Mean, SD) were calculated. Age-group characteristics of SRTn required to meet the 80% performance criterion, as well as correct-sentence-recognition in the dual-task, were compared using an independent-samples t-test. Next, a mixed model two-way ANOVA was performed with task (single vs. dual) as the within-subject variable and age-group (young vs. MA) as the between-subjects variable. Although hearing thresholds were normal for age for all participants (**Figure 1**), a comparison of the means of thresholds at 0.5, 1, 2, 4, and 8 kHz, yielded a significant difference, paired samples t-test, $t(38) = -8.43$, $p < 0.001$), $d = 4.72$. Consequently, we repeated each analysis including hearing threshold mean as a covariate in addition to the main and interaction effects of the research variables. Finally, following Salthouse and Somberg (1982), in order to control for individual differences in reaction time, as well as initial longer reaction times attributed to age (Meijer et al., 2009) already in the single task, we computed proportional dual-task cost (pDTC) using the following computation: RT pDTC = (RT single task - Rt dual-task)/RT single task X 100.

RESULTS

The auditory single task measure was the SRTn required to meet the 80% performance criterion, As can be seen in **Figure 2A**, middle-aged adults needed a better SNR (−4.4dB ± 0.43) compared to the young adults (−5.84dB ± 0.13). This result was found significant, in a paired-samples t-test, $t(38) = -3.49$, $p = 0.001$, consistent with previous research demonstrating the effect of age on SNR (Desjardins and Doherty, 2013; Degeest et al., 2015; Ward et al., 2017). On the other hand, when the individual SNR was provided in the dual-task to each participant, performance in the young group was, on average, 78.13% (± 0.9) and 76.06% (± 2.1) for the MA, as presented in **Figure 2B**. The difference between the groups was found insignificant, with $t(38) = 0.94$, $p = 0.35$. This finding suggests the efficiency of the study specific paradigm to manifest LE in the dual-task measures of visual accuracy and reaction time.

Figure 3 presents the means and standard errors for visual accuracy in the single and dual-tasks in the young and middle-aged groups. It can be seen that in both groups the accuracy decreased in the dual-task, from an average of 99.61% (± 0.14) to 91.38% (± 1.16) and from 97.56% (± 0.9) to 85.15% (± 2.4) in the young and middle-aged groups, respectively. ANOVA performed on these data revealed a significant main effect of task with F (1,38) = 91.91, $p < 0.001$, $\eta^2 = 0.71$, indicating the presence of LE in the sample. In addition, significant effect was obtained for age-groups, with $F(1, 38) = 7.57$, $p = 0.009$, $\eta^2 = 0.17$. The task X age interaction effect, however, was not significant $F(1, 38) = 3.80$, $p = 0.059$. Furthermore, adding to the analysis the variable of hearing thresholds as a covariate resulted in cancelation of the age-group main effect $F(1,37) = 0.55$, $p = 0.46$, whereas the task main effect persisted, $F(1,37) = 7.69$, $p = 0.009$, $\eta^2 = 0.17$. Thus, although a dual-task effect was obtained for visual accuracy, no age differences emerged for this effect.

Figure 4 depicts mean and standard errors for reaction time in the single and dual-tasks, in both groups. Prolongation in reaction time was evident for both groups, however, it was larger for the middle-aged. Whereas in the young group reaction time was prolonged from an average of 1,007.45msec (± 17.99) to an average of 1,391.04msec (± 57.01), in the middle-aged group the average for the single task was 1,742.77msec (± 217.33), while the average for the dual-task was 3,332.3msec (± 510.17). Statistical analysis indicated a significant main effect for task, with $F(1,38) = 32.82$, $p < 0.001$, $\eta^2 = 0.17$, underscoring the difficulty of dual vs single task. Furthermore, the greater prolongation that characterized the middle-aged group, as compared to the young group, was found significant as well, $F(1, 38) = 20.98$, $p < 0.001$, $\eta^2 = 0.17$. In addition, the task X age interaction effect was significant $F(1, 38) = 12.26$, $p = 0.001$, $\eta^2 = 0.17$. After the hearing threshold variable was added as a covariate to the ANOVA model, the main effect of task remained significant, $F(1, 37) = 4.77$, $p = 0.035$)], as well as the main effect of age-group, $F(1,37) = 10.89$, $p = 0.002$, $\eta^2 = 0.23$; and the task X age-group interaction, $F(1, 37) = 5.62$; $p = 0.023$, $\eta^2 = 0.13$. Thus, middle-aged adults exhibited a greater difficulty in the dual-task, irrespective of their hearing status. Notably, the calculation of the RT pDTC in both age-groups yielded a larger pDTC for the MA adults compared to the young adults: 0.95 ± 0.99, and 0.39 ± 0.30, respectively.

DISCUSSION

The present study aimed at setting a clinical paradigm for the assessment of LE in order to incorporate into the audiological evaluation an important marker of cognitive hearing. Whereas LE manifested by a dual-task paradigm has been a subject of ample research, no specific paradigm was suggested as suitable for the audiology clinic, despite the agreement upon the need of LE measure within the hearing evaluation and intervention

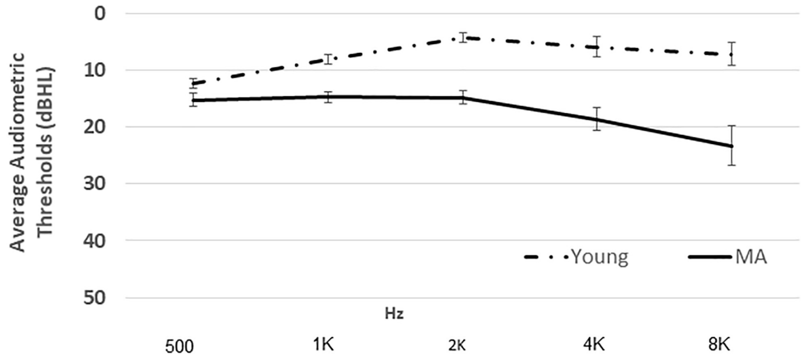


FIGURE 1 | Mean (± sd) air conduction thresholds for the young and middle-aged groups.

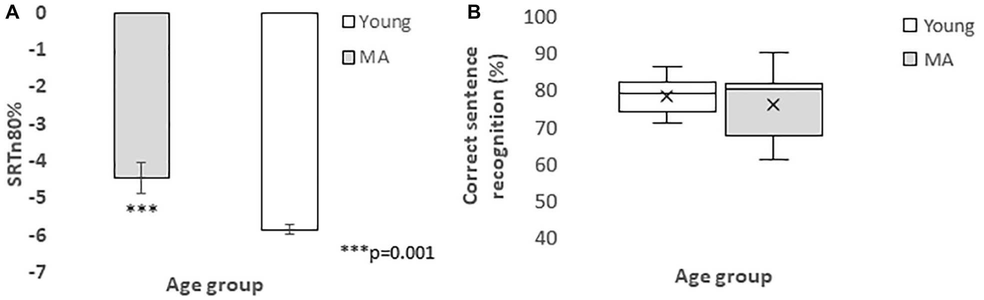


FIGURE 2 | in panel **(A)**, mean (± se) of SRTn 80% sentence recognition of young (white) and MA (gray) groups, in the *single* task. Note that the MA group needed a more positive SNR than the young group, even when the primary task was performed singly. In panel **(B)** distribution of correct sentence recognition for the young (white) and MA (gray), in the *dual-task*, performed at the individual SNR, obtained when tested singly. Lower and upper box boundaries represent the 25th-75th percentiles, lower and upper error bars represent 10th and 90th percentiles, respectively. The horizontal line inside the box represents the median and the X – the mean. Note the larger distribution of the MA results, albeit almost similar median and mean scores.

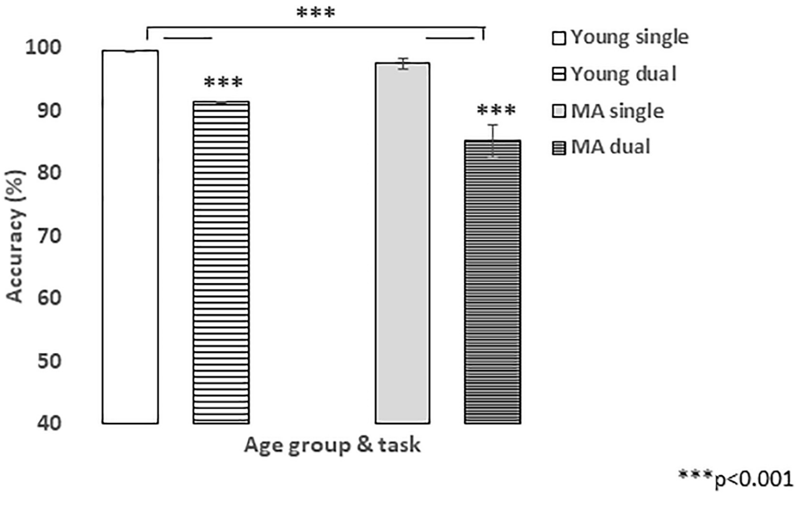


FIGURE 3 | Mean (± se) of visual task accuracy in the single (solid) and dual (pattern) tasks for Young (white) and MA (gray) groups.

framework (Bernarding et al., 2013; Houben et al., 2013; Pals et al., 2015; Alhanbali et al., 2019).

In order to meet the study criterion of 80% correct sentence recognition in the single task, MA adults needed a better SNR of 1.4dB compared to the young adults. In line with previous research (Helfer and Freyman, 2014; Degeest et al., 2015; Dubno, 2015; Helfer, 2015), this finding underscores the known difficulties of MA adults to process speech in

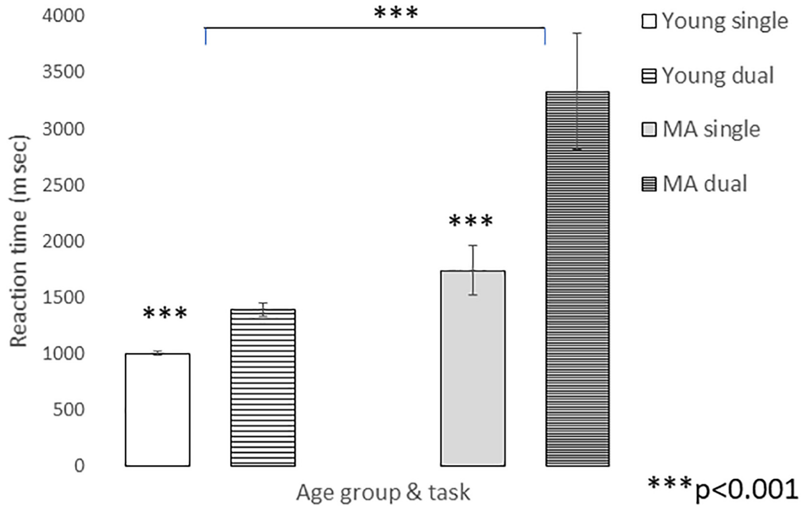


FIGURE 4 | Mean (± se) of visual reaction time in the single (solid) and dual (pattern) tasks for Young (white) and MA (gray) groups.

noise. Several explanations were offered in the literature to the reduced speech perception ability in the presence of normal hearing acuity. Auditory temporal capabilities were found to influence speech perception in noise capacity in the presence of normal hearing threshold (Helfer and Vargo, 2009; Füllgrabe, 2013), as well as the contribution of elevated extended high frequency thresholds (see review by Helfer and Jesse, 2021). Thus, even though hearing acuity of the participants in the current study was age-appropriate (Engdahl et al., 2005), when compared to the young participants, differences were found significant, and might have inflicted upon the correct sentence recognition of the MA adults. Notwithstanding the peripheral domain influence, Besser et al. (2015) stressed the integration between the auditory and the cognitive systems, suggesting that changes in one domain are associated with changes in the other domain. This notion of the auditory-cognitive association is manifested in our LE results. Despite better SNR, our results indicated that LE was expended by MA adults in the dual-task, more than by young adults, revealed by both measures of the secondary task. Our findings are in line with the few studies that explored LE specifically in MA adults. Degeest et al. (2015) reported increase in LE at the fifth decade of life (SNR dependent). Cramer and Donai (2019) found increased LE in 40-55 years old participants, compared to participants aged 18-25 years. In a related study, a measure of cognitive load was found to increase in normal hearing 51-61 years old participants by Xia et al. (2015). Taken together, our data show that LE increases already in the middle age. Furthermore, consistent with previous studies (Degeest et al., 2015; Cramer and Donai, 2019), our finding of different SNR needed for the 80% correct recognition, suggests that performance accuracy of speech recognition in noise does not fully manifest MA adults' efforts to maintain successful recognition. The need to allocate more resources than young adults, albeit better listening conditions, and unrelated to hearing thresholds, supports the idea that other auditory processing factors, supra threshold, or otherwise different, affect performance in noise.

Setting a clinical paradigm for the assessment of LE in MA adults might face some hurdles. First, the selection of the appropriate secondary task has been controversial in the literature. Our findings show that using a simple, non-auditory secondary task, resulted in the manifestation of LE in young as well as in MA adults. Both secondary task measures: performance accuracy and reaction time indexed LE. This finding is in line with previous research that used a different modality secondary task, such as tactile (Fraser et al., 2010), or visual (Hughes and Galvin, 2013). In these studies, decreased performance of secondary task measures was evident, regardless of task difficulty. On the other hand, our findings differ from that of Picou and Ricketts (2014), that argued in favor of depth of processing, in order to determine LE by secondary task measures. Trying to solve this contention, it has been suggested that engaging attentional resources across modalities instead of drawing on the same modality, might better reflect LE (Grieco-Calub et al., 2017). Taken together, the use of visual, secondary task in the current study allowed for resources allocation, and expressed LE effectively. Notably, MA adults that visit the audiology clinic are not pre-screened for cognition, thus LE assessment using a simple task might be more beneficial, meeting various cognitive capacities of the MA adults.

Another hurdle in the appropriate paradigm of LE assessment in the MA is the use of RT as a measure of LE. It has been previously established that aging, in general, is related to slower processing of information (Salthouse, 2000), thus RT might not manifest LE in MA adults, being already prolonged in the single task compared to the young adults. Instead, our data demonstrated RT dual-task effect in both groups: young and MA, more so for the MA. Furthermore, in order to overcome individual differences in baseline reaction time, that might be affected by age, we calculated pDTC, and found a RT pDTC, in both groups, more pronounced for the MA. This finding is consistent with Gosselin and Gagné (2011) that showed pDTC in both word and tactile accuracy in older adults. Taken together, it is suggested that MA adults, comparable to older adults, prolonged their responses, more than young adults, to maintain accuracy in the primary task. The cognitive load of speech recognition in noise while matching visual-colored shapes burdened their processing ability, and compelled them to slow their responses. On the whole, this finding proposes the compatibility of the paradigm to assess LE in young and MA adults. Additional studies will need to address other age-groups such as older adults, as well as hearing impaired listeners.

Finally, in an attempt to find a suitable measure for clinical evaluation of LE, physiological measures should be considered as well. One such measure, pupillometry, was found as a measure of cognitive processing load, sensitive to difference in noise types and intelligibility levels (Koelewijn et al., 2012). Moreover, Karatekin et al. (2004) proposed that pupillometry can present the magnitude of resource allocation, and not only the yielding of cognitive capacities. It should be noted, however, that such a measure necessitates appropriate and costly equipment, and might be complicated and inconvenient for the hearing clinics. The current study dual-task paradigm, on the other hand, while reflecting the different proportions of resource allocation by MA compared to young adults, does not require any special equipment other than that found already in the typical hearing clinics. The paradigm is easy to explain, understand, and use, with a time duration of approximately 20 min. Clinicians might find the paradigm helpful, specifically in cases of patients that are not fully aware of the effort they exert in order to understand speech in background noise. These patients sometimes are reluctant to use remote microphone systems or hearing aids. LE assessment might help to encourage them to use such means.

LIMITATIONS

The current study demonstrated the compatibility of a specific dual-task paradigm to manifest LE in MA adults. In order to further substantiate the clinical sensitivity of the paradigm, more participants in the MA, as well as in older adults, are needed. Furthermore, the young adults group consisted of female-only participants. Future studies might consider a mixed-gender group.

In addition, we did not incorporate cognitive tests in the study, as patients coming to the audiological clinic are not pre-screened for cognition. Future studies including cognitive tests may offer the possibility to identify specific aspects of cognitive capacity associated with LE, and further elucidate LE trajectories. Likewise, a self-report LE tool might shed light on strategies listeners use to meet different aspects of listening demands, delineating the cognitive load they are burdened with.

CONCLUSION

The current study proposed a clinical tool to assess LE. The dual-task paradigm, using a non-auditory secondary task was found compatible for the assessment of LE in normal hearing young, and more so in MA adults. Hearing thresholds, though significantly different between the two groups, did not account for the greater LE that was manifested in the MA group. Incorporating such a paradigm in the routine clinical setting will address MA adults' subjective reports, while taking into consideration that successful communication is more than audibility and speech intelligibility.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Institutional Review Board at Tel Aviv University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

RKN and CM designed the study. RKN supervised the data collection and organized the database. IR performed the statistical analysis. All authors contributed to manuscript revision, read, and approved the submitted version.

ACKNOWLEDGMENTS

We thank I. Nachum, D. Tabachnik, N. Yud, and M. Valder, undergraduate students at the Communication Disorders department, the Stanley Steyer School of Health Professions, Sackler Faculty of Medicine, Tel Aviv University, for assisting with data collection.

REFERENCES

Alhanbali, S., Dawes, P., Millman, R. E., and Munro, K. J. (2019). Measures of listening effort are multidimensional. *Ear Hear.* 40, 1084–1097. doi: 10.1097/AUD.0000000000000697

Baddeley, A., Gathercole, S., and Papagno, C. (1998). The phonological loop as a language learning device. *Psychol. Rev.* 105, 158–173. doi: 10.1037/0033-295x.105.1.158

Baddeley, A. D., and Logie, R. H. (1999). "Working memory: the multiple-component model," in *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control*, eds A. Miyake and P. Shah (Cambridge: Cambridge University Press), 28–61. doi: 10.1017/CBO9781139174909.005

Bernarding, C., Strauss, D. J., Hannemann, R., Seidler, H., and Corona-Strauss, F. I. (2013). Neural correlates of listening effort related factors: influence of age and hearing impairment. *Brain Res. Bull.* 91, 21–30. doi: 10.1016/j.brainresbull.2012.11.005

Besser, J., Festen, J. M., Goverts, S. T., Kramer, S. E., and Pichora-Fuller, M. K. (2015). Speech-in-speech listening on the LiSN-S test by older adults with good audiograms depends on cognitionand hearing acuity at high frequencies. *Ear Hear.* 36, 24–41. doi: 10.1097/aud.0000000000000096

Brand, T., and Kollmeier, R. B. (2002). Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests. *J. Acoust. Soc. Am.* 111, 2801–2810. doi: 10.1121/1.1479152

Bugannim, Y., Roth, D., Zechoval, D., and Kishon-Rabin, L. (2019). Training of speech perception in noise in pre-lingual hearing-impaired adults with cochlear implants compared with normal hearing adults. *Otol. Neuroto.* 40, e316–e325. doi: 10.1097/MAO.0000000000002128

Clinard, C. G., and Cotter, C. M. (2015). Neural representation of dynamic frequency is degraded in older adults. *Hear. Res.* 323, 91–98. doi: 10.1016/j.heares.2015.02.002

Cramer, J. L., and Donai, J. J. (2019). Effects of signal bandwidth on listening effort in young- and middle-aged adults. *Int. J. Audiol.* 58, 116–122. doi: 10.1080/14992027.2018.1533258

Degeest, S., Keppler, H., and Corthalsa, P. (2015). The effect of age on listening effort. *J. Speech Lang. Hear. Res.* 58, 1592–1600. doi: 10.1044/2015_JSLHR-H-14-0288

Desjardins, J. S., and Doherty, K. A. (2013). Age-related changes in listening effort for various types of masker noises. *Ear Hear.* 34, 261–272. doi: 10.1097/AUD.0b013e31826d0ba4

Devesse, A., Wouters, J., and van Wieringen, A. (2020). Age affects speech understanding and multitask costs. *Ear Hear.* 41, 1412–1415. doi: 10.1097/AUD.0000000000000848

Dubno, R. (2015). Speech recognition across the life span: longitudinal changes from middle-age to older adults. *Am. J. Audiol.* 24, 84–87. doi: 10.1044/2015_AJA-14-0052

Engdahl, B., Tambs, K., Borchgrevink, H. M., and Hoffman, H. J. (2005). Screened and unscreened hearing threshold levels for the adult population: results from the Nord-Trøndelag hearing loss study. *Int. J. Audiol.* 44, 213–230. doi: 10.1080/14992020500057731

Fraser, S., Gagné, J. P., Alepins, M., and Dubois, P. (2010). Evaluating the effort expended to understand speech in noise using a dual-task paradigm: the effects of providing visual speech cues. *J. Speech Lang. Hear. Res.* 53, 18–33. doi: 10.1044/1092-4388(2009/08-0140)

Füllgrabe, C. (2013). Age-dependent changes in temporal-fine-structure processing in the absence of peripheral hearing loss. *Am. J. Audiol.* 22, 313–315. doi: 10.1044/1059-0889(2013/12-0070)

Gagné, J. P., Besser, J., and Lemke, U. (2017). Behavioral assessment of listening effort using a dual-task paradigm: a review. *Trends Hear.* 21, 2331216516687287. doi: 10.1177/2331216516687287

Goossens, T., Vercammen, C., Wouters, J., and van Wieringen, A. (2017). Masked speech perception across the adult lifespan: impact of age and hearing impairment. *Hear. Res.* 344, 109–124. doi: 10.1016/j.heares.2016.11.004

Gordon-Salant, S., and Samuels Cole, S. (2016). Effects of age and working memory capacity on speech recognition performance in noise among listeners with normal hearing. *Ear Hear.* 37, 593–602. doi: 10.1097/AUD.0000000000000316

Gosselin, P., and Gagné, J. P. (2011). Older adults expend more listening effort than young adults recognizing audiovisual speech in noise. *Int. J. Audiol.* 50, 786–792. doi: 10.3109/14992027.2011.599870

Grieco-Calub, T. M., Ward, K. M., and Brehm, L. (2017). Multitasking during degraded speech recognition in school-age children. *Trends Hear.* 21, 2331216516686786. doi: 10.1177/2331216516686786

Hasher, l, and Zacks, R. (1979). Automatic and effortful processes in memory. *J. Exp. Psychol. Gen.* 108, 356–388.

Helfer, K. S. (2015). Competing speech perception in middle age. *Am. J. Audiol.* 24, 80–83. doi: 10.1044/2015_AJA-14-0056

Helfer, K. S., and Freyman, R. L. (2014). Stimulus and listener factors affecting age-related changes in competing speech perception. *J. Acoust. Soc. Am.* 136, 748–759. doi: 10.1121/1.4887463

Helfer, K. S., and Jesse, A. (2021). Hearing and speech processing in midlife. *Hear. Res.* 402, 1–8. doi: 10.1016/j.heares.2020.108097

Helfer, K. S., Merchant, G. R., and Wasiuk, P. A. (2017). Age-related changes in objective and subjective speech perception in complex listening environments. *J. Speech Lang. Hear. Res.* 60, 3009–3018. doi: 10.1044/2017_JSLHR-H-17-0030

Helfer, K. S., and Vargo, M. (2009). Speech recognition and temporal processing in middle-aged women. *J. Am. Acad. Audiol.* 20, 264–271. doi: 10.3766/jaaa.20.4.6

Hornsby, B. W. Y. (2013). The effects of hearing aid use on listening effort and mental fatigue associated with sustained speech processing demands. *Ear Hear.* 34, 523–534. doi: 10.1097/AUD.0b013e31828003d8

Houben, R., van Doorn-Bierman, M., and Dreschler, W. A. (2013). Using response time to speech as a measure for listening effort. *Int. J. Audiol.* 52, 753–761. doi: 10.3109/14992027.2013.832415

Hughes, K. C., and Galvin, K. L. (2013). Measuring listening effort expended by adolescents and young adults with unilateral or bilateral cochlear implants or normal hearing. *Cochlear Implants Int.* 14, 121–129. doi: 10.1179/1754762812Y.0000000009

Kahneman, D. (1973). *Attention and Effort*. New Jersey, NJ: Englewood Cliffs.

Karatekin, C., Couperus, J. W., and Marcus, D. J. (2004). Attention allocation in the dual-task paradigm as measured through behavioral and psychophysiological responses. *Psychophysiology* 41, 175–185. doi: 10.1111/j.1469-8986.2003.00147.x

Kim, S. Y., Kim, M. S., and Chun, M. M. (2005). Concurrent working memory load can reduce distraction. *Proc. Natl. Acad. Sci. U. S. A.* 102, 16524–16529. doi: 10.1073/pnas.0505454102

Koelewijn, T., Zekveld, A. A., Festen, J. M., and Kramer, S. E. (2012). Pupil dilation uncovers extra listening effort in the presence of a single-talker masker. *Ear Hear.* 33, 291–300. doi: 10.1097/AUD.0b013e3182310019

Kollmeier, B., Warzybok, A., Hochmuth, S., Zokoll, M. A., Uslar, V., and Brand, T. (2015). The multilingual matrix test: principles, applications, and comparison across languages: a review. *Int. J. Audiol.* 54, 3–16. doi: 10.3109/14992027.2015.1020971

Lee, J. Y., Lee, J. T., Heo, H. J., Choi, C. H., Choi, S. H., and Lee, K. (2015). Speech recognition in real-life background noise by young and middle-aged adults with normal hearing. *J. Audiol. Otol.* 19, 39–44. doi: 10.7874/jao.2015.19.1.39

Lewis, D., Schmid, K., O'Leary, S., Spalding, J., Heinrichs-Graham, E., and High, R. (2016). Effects of noise on speech recognition and listening effort in children with normal hearing and children with mild bilateral or unilateral hearing loss. *J. Speech Lang. Hear. Res.* 59, 1218–1232. doi: 10.1044/2016_JSLHR-H-15-0207

McGarrigle, R., Munro, K. J., Dawes, P., Stewart, A. J., Moore, D. R., Barry, J. G., et al. (2014). Listening effort and fatigue: What exactly are we measuring? A british society of audiology cognition in hearing special interest group 'white paper'. *Int. J. Audiol.* 53, 433–440. doi: 10.3109/14992027.2014.890296

Meijer, W. A., de Groot, R. H. M., van Gerven, P. W. M., van Boxtel, M. P. J., and Jolles, J. (2009). Level of processing and reaction time in young and middle-aged adults and the effect of education. *Eur. J. Cogn. Psychol.* 21, 216–234. doi: 10.1080/09541440802091780

Pals, C., Sarampalis, A., and Başkent, D. (2013). Listening effort with cochlear implant simulations. *J. Speech Lang. Hear. Res.* 56, 1075–1084. doi: 10.1044/1092-4388(2012/12-0074)

Pals, C., Sarampalis, A., van Rijn, H., and Başkent, D. (2015). Validation of a simple response-time measure of listening effort. *J. Acoust. Soc. Am.* 138, EL187–EL192. doi: 10.1121/1.4929614

Peele, J. (2018). Listening effort: How the cognitive consequences of acoustic challenge are reflected in brain and behavior. *Ear Hear.* 39, 204–214. doi: 10.1097/AUD.0000000000000494

Pichora-Fuller, M., Kramer, S., Eckert, M., Edwards, B., Hornsby, B., Humes, L., et al. (2016). Hearing impairment and cognitive energy: the framework for understanding effortful listening (FUEL). *Ear Hear.* 37, 5S–27S. doi: 10.1097/AUD.0000000000000312

Picou, E., and Ricketts, T. (2014). The effect of changing the secondary task in dual-task paradigms for measuring listening effort. *Ear Hear.* 35, 611–622. doi: 10.1097/AUD.0000000000000055

Picou, E., Ricketts, T., and Hornsby, B. (2013). How hearing aids, background noise, and visual cues influence objective listening effort. *Ear Hear.* 34, e52–e64. doi: 10.1097/AUD.0b013e31827f0431

Salthouse, T. A. (2000). Aging and measures of processing speed. *Biol. Psychol.* 54, 35–54. doi: 10.1016/s0301-0511(00)00052-1

Salthouse, T. A., and Somberg, B. L. (1982). Time-accuracy relationships in young and old adults. *J. Gerontol.* 37, 349–353. doi: 10.1093/geronj/37.3.349

Sommers, M. S., and Phelps, D. (2016). Listening effort in younger and older adults: a comparison of auditory-only and auditory-visual presentations. *Ear Hear.* 37, 62S–68S. doi: 10.1097/AUD.0000000000000322

Tremblay, K. L., and Backer, K. C. (2016). Listening and learning: cognitive contributions to the rehabilitation of older adults with and without audiometrically defined hearing loss. *Ear Hear.* 37, 155S–162S. doi: 10.1097/AUD.0000000000000307

Tun, P. A., McCoy, S., and Wingfield, A. (2009). Aging, hearing acuity, and the attentional costs of effortful listening. *Psychol. Aging* 24, 761–766. doi: 10.1037/a0014802

Ward, K. M., Shen, J., Souza, P. E., and Grieco-Calub, T. M. (2017). Age-related differences in listening effort during degraded speech recognition. *Ear Hear.* 38, 74–84. doi: 10.1097/AUD.0000000000000355

Xia, J., Nooraei, N., Kalliri, S., and Edwards, B. (2015). Spatial release of cognitive load measured in a dual-task paradigm in normal-hearing and hearing-impaired listeners. *J. Acoust. Soc. Am.* 137, 1888–1898. doi: 10.1121/1.4916599

Conflict of Interest: 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 Psychology

ORIGINAL RESEARCH
published: 03 March 2022
doi: 10.3389/fpsyg.2022.797422

Difficulties Experienced by Older Listeners in Utilizing Voice Cues for Speaker Discrimination

Yael Zaltz* and Liat Kishon-Rabin

Department of Communication Disorders, The Stanley Steyer School of Health Professions, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel

Human listeners are assumed to apply different strategies to improve speech recognition in background noise. Young listeners with normal hearing (NH), e.g., have been shown to follow the voice of a particular speaker based on the fundamental (F0) and formant frequencies, which are both influenced by the gender, age, and size of the speaker. However, the auditory and cognitive processes that underlie the extraction and discrimination of these voice cues across speakers may be subject to age-related decline. The present study aimed to examine the utilization of F0 and formant cues for voice discrimination (VD) in older adults with hearing expected for their age. Difference limens (DLs) for VD were estimated in 15 healthy older adults (65–78 years old) and 35 young adults (18–35 years old) using only F0 cues, only formant frequency cues, and a combination of F0 + formant frequencies. A three-alternative forced-choice paradigm with an adaptive-tracking threshold-seeking procedure was used. Wechsler backward digit span test was used as a measure of auditory working memory. Trail Making Test (TMT) was used to provide cognitive information reflecting a combined effect of processing speed, mental flexibility, and executive control abilities. The results showed that (a) the mean VD thresholds of the older adults were poorer than those of the young adults for all voice cues, although larger variability was observed among the older listeners; (b) both age groups found the formant cues more beneficial for VD, compared to the F0 cues, and the combined (F0 + formant) cues resulted in better thresholds, compared to each cue separately; (c) significant associations were found for the older adults in the combined F0 + formant condition between VD and TMT scores, and between VD and hearing sensitivity, supporting the notion that a decline with age in both top-down and bottom-up mechanisms may hamper the ability of older adults to discriminate between voices. The present findings suggest that older listeners may have difficulty following the voice of a specific speaker and thus implementing doing so as a strategy for listening amid noise. This may contribute to understanding their reported difficulty listening in adverse conditions.

Keywords: speaker discrimination, older adults, fundamental frequency, formant frequencies, voice discrimination

Edited by:
Leah Fostick,
Ariel University, Israel

Reviewed by:
Antje Heinrich,
The University of Manchester,
United Kingdom
Sandra Gordon-Salant,
University of Maryland,
College Park, United States

***Correspondence:**
Yael Zaltz
yaelzaltz@gmail.com

Specialty section:
This article was submitted to
Perception Science,
a section of the journal
Frontiers in Psychology

Received: *18 October 2021*
Accepted: *24 January 2022*
Published: *03 March 2022*

Citation:
Zaltz Y and Kishon-Rabin L (2022)
Difficulties Experienced by Older
Listeners in Utilizing Voice Cues for
Speaker Discrimination.
Front. Psychol. 13:797422.
doi: 10.3389/fpsyg.2022.797422

INTRODUCTION

Older people often find it extremely difficult to understand speech and converse in noisy environments, particularly when the noise includes several speakers (e.g., Pichora-Fuller, 1997). Such difficulties may limit their ability to participate in social, occupational, and educational activities, isolating them from their families and friends (Cacciatore et al., 1999; Arlinger, 2003; Pronk et al., 2011; Gopinath et al., 2012). Many studies have attempted to assess the contribution of different factors to the difficulties older adults experience in speech perception amid noise and found that they explain the results only in part. These include elevated hearing thresholds (e.g., Jayakody et al., 2018), reduced frequency and temporal resolution (e.g., Anderson and Karawani, 2020), and declining cognitive functioning (e.g., Pichora-Fuller and Singh, 2006; Schneider et al., 2010). One approach to further our understanding of the difficulties experienced by older adults in noisy conditions is to assess their ability to apply listening strategies that are known to assist younger adults in listening amid noisy backgrounds. One such strategy includes identifying and following the acoustic voice cues of a target speaker, such as the fundamental frequency (F0) and formant frequencies of their voice (e.g., Bronkhorst, 2015). Young adults have been shown to efficiently implement this strategy to segregate the relevant and irrelevant speakers (e.g., Bronkhorst, 2015), relying on efficient spectral (formants) and temporal (F0) processing of the speech signal (e.g., Fant, 1960; Lieberman and Blumstein, 1988; Carlyon and Shackleton, 1994; Fu et al., 2004; Oxenham, 2008; Xu and Pfingst, 2008). Given that spectro-temporal processes are known to degrade with age (e.g., Vongpaisal and Pichora-Fuller, 2007; Souza et al., 2011; Schvartz-Leyzac and Chatterjee, 2015; Chintanpalli et al., 2016; Goupell et al., 2017; Anderson et al., 2021) it may be difficult for older adults to take advantage of differences in F0 and/or formant information for talker segregation. Age-related cognitive decline in executive functions, including attention, inhibition, and working memory (e.g., Mitchell et al., 2000; Salthouse, 2000; Harada et al., 2013), may add to the difficulty in utilizing these acoustic cues for voice discrimination. Thus, the goal of the present study was to examine the use of F0 and formant cues for voice discrimination and to assess the contribution of sensory and cognitive factors to this discrimination in older adults with normal hearing for their age (NHA), compared to young adults.

Multi-talker situations are particularly challenging for speech understanding because they force the listener to cope with both energetic (Brungart, 2001; Ezzatian et al., 2012) and informational masking (Durlach et al., 2003; Hoen et al., 2007; Ezzatian et al., 2012). Energetic masking occurs when the energy of the frequencies of the competing voices overlap those of the target voice, activating similar areas along the basilar membrane. Informational masking occurs because the competing speech may invoke related linguistic activity and/or divert attention from the target speech, interfering with the processing of the speech signal at higher linguistic or cognitive levels and making it difficult for the listener to focus on the auditory stream of interest and ignore non-relevant sounds (also known as the "cocktail party" effect, e.g., Kattner and Ellermeier, 2020). Therefore, to function well in multi-talker situations, the listener has to efficiently utilize both bottom-up (e.g., spectral separation) and top-down (e.g., focused attention) mechanisms. However, spectral and temporal processing have been shown to degrade with age, even in listeners with audiograms within normal hearing (Moore and Peters, 1992; Vongpaisal and Pichora-Fuller, 2007; Schvartz-Leyzac and Chatterjee, 2015; Chintanpalli et al., 2016; Goupell et al., 2017). A negative age effect has also been reported for higher cognitive abilities, such as executive functions, including memory, attention, and inhibition (e.g., Mitchell et al., 2000; Salthouse, 2000; Harada et al., 2013). The poor speech-in-noise understanding of older NH adults may, therefore, be the result of varying degrees of decline in their peripheral, linguistic, and/or central and cognitive processing (e.g., Working Group on Speech Understanding and Aging, 1988; Humes et al., 2012; Tremblay et al., 2021).

One listening strategy that is assumed to assist in segregating the target voice from competing non-relevant sounds is to identify and track the acoustic characteristics of the voice of the speaker of interest (e.g., Bronkhorst, 2015). These characteristics include fundamental frequency (F0), which is influenced by the length, mass, and rate of vibration of the vocal cords, and formant frequencies (i.e., resonant frequencies of the vocal tract), which are influenced by the vocal tract length (VTL; e.g., Darwin et al., 2003; Vestergaard et al., 2009, 2011; Mackersie et al., 2011; Schvartz-Leyzac and Chatterjee, 2015; Başkent and Gaudrain, 2016). Both of these cues provide robust information regarding the speaker, such as age and gender, as well as idiosyncratic characteristics that are unique to that speaker and his/her personality, nearly as unique as our fingerprints according to some researchers (Shultz, 2015). Studies in young adults and children have shown that these listeners rely heavily on both types of cues (F0 and formant frequencies) to discriminate among (Zaltz et al., 2020) and segregate talkers (e.g., Darwin et al., 2003) as well as to identify the gender of a specific speaker (e.g., Smith and Patterson, 2005; Smith et al., 2007; Skuk and Schweinberger, 2014; Başkent and Gaudrain, 2016). Moreover, listeners who have difficulty perceiving differences in F0 and formant frequencies, such as cochlear implant users, showed reduced voice discrimination, which may explain, at least in part, their poor performance when listening amid noise (e.g., Gaudrain and Başkent, 2018; Zaltz et al., 2018).

Studies suggest that F0 coding relies primarily on efficient processing of the temporal envelope and/or of the temporal fine-structure cues of the signal, whereas formant coding primarily involves place coding of spectral energy peaks (e.g., Fant, 1960; Lieberman and Blumstein, 1988; Carlyon and Shackleton, 1994; Fu et al., 2004; Chatterjee and Peng, 2008; Oxenham, 2008; Xu and Pfingst, 2008). Previous studies have shown the effect of age on F0 discrimination (e.g., Moore and Peters, 1992; Vongpaisal and Pichora-Fuller, 2007; Souza et al., 2011; Anderson et al., 2021). It has been shown, for example, that older adults require twice the difference between F0s to reach similar accuracy in concurrent vowel identification with harmonic complexes and synthetic vowels, compared to their younger peers (Moore and Peters, 1992; Vongpaisal and

Pichora-Fuller, 2007). Electrophysiological studies have demonstrated pronounced reductions in phase locking in older adults, suggesting reduced neural synchrony among this population (e.g., Anderson et al., 2021). These findings were interpreted to reflect impaired periodicity coding in older listeners (Schvartz-Leyzac and Chatterjee, 2015), which, in turn, may negatively influence the utilization of F0 cues for talker discrimination in this population. Other studies have showed the effect of age on the utilization of formant changes for vowel identification (Vongpaisal and Pichora-Fuller, 2007; Chintanpalli et al., 2016; Goupell et al., 2017). However, no study, to our knowledge, has specifically investigated the ability to use changes in F0, formants and their combination for speaker discrimination.

Efficient utilization of the relevant acoustic cues for talker discrimination may also require complex cognitive processing, such as, attending to F0 and formant information of the different talkers, and storing this information in memory for decision making and for future reference. Therefore, in older adults, the reported age-related deterioration in the ability to focus attention on the relevant features of the stimulus while inhibiting the processing of non-relevant features (McDowd and Shaw, 2000; Schneider et al., 2007; Harada et al., 2013), may negatively affect their ability to discriminate between speakers based on specific voice cues. Similarly, decline in working memory processes with age, including poor short-term maintenance and manipulation of information during encoding (e.g., Mitchell et al., 2000), and/or general slowing of cognitive processes (Salthouse, 2000), may add to difficulties in utilizing F0 and formant cues for talker identification and stream segregation. Support for this hypothesis can be found in a recent study where the authors argued that poor talker identification amid noise in older adults may have been related to their difficulty to learn and store in memory the voice information associated with a particular speaker (Best et al., 2018). It is possible that a simpler task that examines the perception of F0 and formant frequencies *via* discrimination rather than *via* identification may better assess the utilization of voice cues in older adults. Thus, the present study aimed to assess the use of F0 cues alone, formant cues alone, and the combination of F0 and formants in a VD task in older adults with NHA, and to compare their performance to that of young adults with NH. In addition, because even a simple discrimination task may require attention and working memory capabilities, our second aim was to assess the contribution of higher cognitive abilities, specifically, executive control abilities and working memory, to VD performance.

MATERIALS AND METHODS

Participants

A total of 50 participants were recruited for the present study: 15 older adults (65–78 years, mean = 68.93 ± 3.63 years; median = 68) and 35 young adults (18–35 years, mean = 22.29 ± 3.16 years; median = 22). The VD results of 15 participants from the young-adult group were previously reported (Zaltz et al., 2020). For the current study, we tested an additional 20 young adults to obtain a larger dataset for comparison. As no significant difference in age or test results was observed between the two groups of young adults ($p > 0.05$), their data were combined for all further analyses. The young adults had hearing sensitivity within the normal range in both ears, with pure-tone air conduction thresholds <20 dB HL at octave frequencies of 500–4,000 Hz (Ansi, 2018). For the older adults, eight participants had thresholds less or equal to ≤25 dB HL, five participants had thresholds less or equal to ≤40 dB HL, one had thresholds less or equal to ≤55 dB HL, and one had thresholds less or equal to ≤70 dB HL at octave frequencies of 500–4,000 Hz (**Figure 1**). Overall, pure-tone air conduction thresholds for the older adults were within the normal range for their age (Engdahl et al., 2005), with a pure-tone average across ears and four frequencies (PTA4) of less than 33 dB HL. None of the participants had any previous psychoacoustic experience in similar tasks, they had no known history of ear disease, and they had completed at least 12 years of formal education. All participants were fluent in Hebrew. Six of the older participants were native Hebrew speakers. The other nine older adults immigrated to Israel at a mean age of 24 (±11) years (range: 3–43 years), and thus were exposed to Hebrew for an average of 45 (±11) years (range of exposure: 29–72 years). Their mother tongues were French ($n = 4$), Arabic ($n = 3$), English ($n = 1$), and Romanian ($n = 1$). All older adults had cognitive ability levels within the normal range (Mini Mental State Examination score ≥ 27; based on the English version; Folstein et al., 1975), lived independently, and led an active life based on self-report. Informed consent was obtained from all participants. The study was approved by the Institutional Review Board of Ethics at Tel Aviv University.

Stimuli

The VD test included three shortened (3-word) sentences from the Hebrew version of the Matrix sentence test. All the sentences had a simple grammatical structure (noun, verb, adjective) from a vocabulary that is appropriate for 5 year olds, and were recorded by a native Hebrew female speaker (Bugannim et al., 2019), similar to Zaltz et al. (2020). Sentences were manipulated using a 13-point stimulus continuum, exponentially ranging in $\sqrt{2}$ steps from a change of −0.18 semitone to a change of −8 semitones. This manipulation was conducted in three separate dimensions: (1) F0, (2) formant frequencies, and (3) combined F0 + formants (Zaltz et al., 2020). For a detailed explanation of the stimuli for the VD test, see **Appendix A**.

Voice Discrimination Test

A three-interval, three-alternative forced-choice procedure was used to estimate VD based on difference limen (DL) for F0 cues, formant cues, and combined F0 + formant cues. A two-down one-up adaptive tracking procedure yielded DLs corresponding to a 70.7% detection threshold on the psychometric function (Levitt, 1971). Each trial consisted of two reference sentences and a comparison sentence, specified at a random interval. Inter-stimulus interval was 300 milliseconds. When a sentence

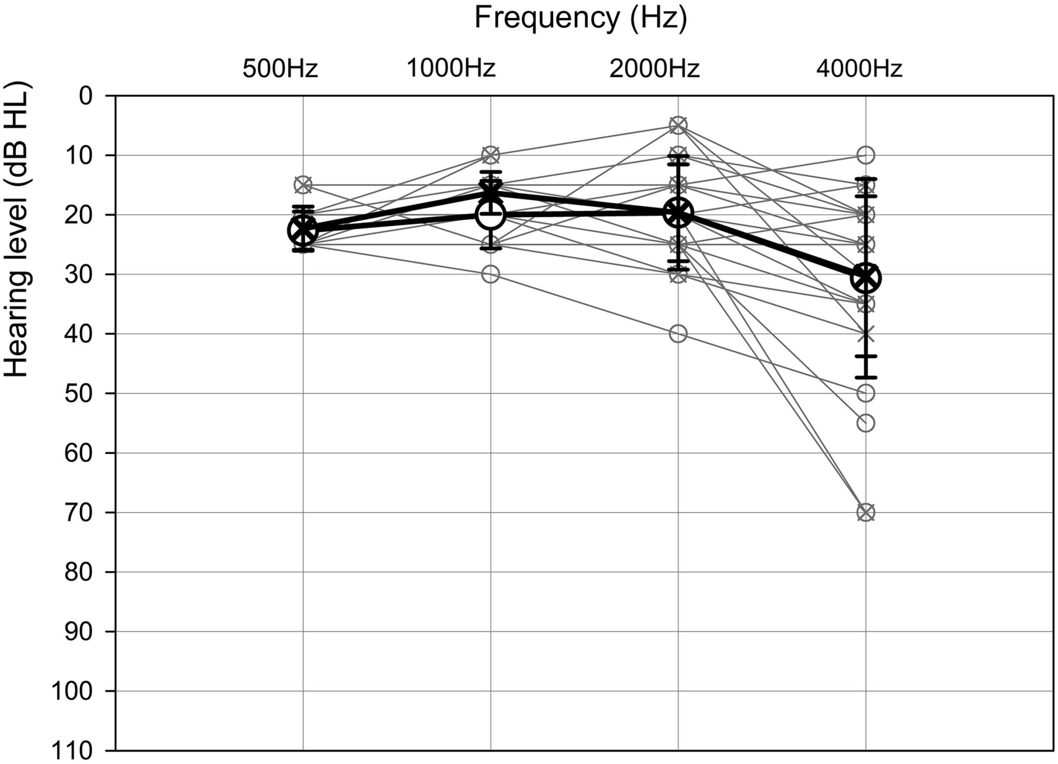


FIGURE 1 | Mean (thick black lines and symbols; ± 1 standard deviation) and individual (thin gray lines and symbols) hearing thresholds at 500 Hz, 1,000 Hz, 2000 Hz, and 4,000 Hz for the right (circle) and left (cross) ears for the older adults ($n = 15$).

was presented, one of the three squares on the PC monitor was highlighted to correspond to that sentence. The participant was instructed to select the sentence that "sounded different" by using the mouse to click on the corresponding square or telling the tester which sentence (1, 2, or 3) "sounded different." The same sentence was used for each VD threshold, resulting in a change only in the tested acoustic cue (F0, formants, or F0 + formants). No feedback was provided, and there was no time limit for responding. The difference between the stimuli was reduced by a factor of two until the first reversal and then reduced or increased by $\sqrt{2}$ until the sixth reversal. DLs were calculated as the arithmetic mean of the last four reversals.

Cognitive Tests

Cognitive capability was assessed only for older adults using the Hebrew version of the Mini Mental state examination test (based on the English version; Folstein et al., 1975). This test examined several mental functions, including orientation, memory, attention, naming, understanding of oral and written instructions, drawing, and writing. Overall, the test included 11 questions and lasted approximately 10 min. A score of 27 points or higher indicated cognitive ability within the normal range (Folstein et al., 1975).

Auditory working memory was assessed for all participants using a recorded version of the backward digit span subtest of the Wechsler Intelligence Scale (Wechsler, 1991). In the digit span test, the participants heard sequences of numbers (e.g., 2, 6, 4, and 3) and were asked to repeat them verbally in the reverse order. The passing criterion to proceed to the next longer sequence was one successful repetition of a sequence of a specific length. The score represented the number of correctly repeated sequences.

A combined effect of executive control abilities, mental flexibility, and the perceptual speed of processing was assessed for all participants using the Trail Making test (TMT) part A and part B (e.g., Bowie and Harvey, 2006; Sánchez-Cubillo et al., 2009). In part A of the test, the participants were instructed to manually connect consecutive numbers from 1–24 by drawing a line as quickly and accurately as possible. In part B of the test, the participants were instructed to manually connect a set of 24 consecutive numbers and letters (e.g., 1A, 2B, 3C…) in sequential order as quickly as possible while maintaining accuracy. If a participant made an error, the tester corrected the response before moving on to the next dot. The scores for the TMT parts A and B represent the time taken for the participant to complete the test accurately (in seconds).

Study Design

All participants took part in a single testing session. At the beginning of the session, each participant from the older adult group was tested using the Mini Mental test. All participants performed three VD thresholds, one with each voice cue (F0, formants, F0 + formants), with a different sentence presented for each cue. Voice cues and sentences were randomized across participants. Before formal testing, each participant performed a short familiarization task with each of the voice cues to ensure that the task was understood. After the completion of VD testing, cognitive tests were conducted. Overall, the testing lasted approximately 45–50 min, including two short breaks of 5–8 min. The participants were not compensated for their time.

Apparatus

The stimuli were delivered using a laptop personal computer through an external sound card and a GSI-61 audiometer to both ears *via* THD-50 headphones at approximately 35–40 dB SL above individual PTA4. The testing session took place in a sound-treated single-walled room.

Data Analysis

All of the VD data were log-transformed for ANOVA to normalize the distribution of the residuals (Kolmogorov–Smirnov test: $p > 0.05$) and allow for parametric statistics. *Post hoc* analyses were conducted using Bonferroni corrections for multiple comparisons. Pearson's coefficient correlations were conducted on the raw data. Corrections for multiple testing (ANOVA and correlations) were applied using the False Discovery Rate method. Statistical analyses were conducted using the SPSS-20 software.

RESULTS

Box whisker plots of the VD thresholds based on F0, formant, and F0 + formant cues for the young and older adult participants are shown in **Figure 2**. The older adults' thresholds were higher (i.e., worse) than those of the young adults for all voice cues. Two-way repeated measures ANOVA with Age as a between-subject variable and Cue (F0, formants, F0 + formants) as a within-subject variable revealed a significant effect of Age [$F(1,48) = 27.060$, $p < 0.001$, $\eta^2 = 0.361$], with the young adults showing better VD thresholds ($M = 0.69 \pm 0.38$), compared to the older adults ($M = 1.40 \pm 0.85$). There was a significant effect of Cue [$F(2,48) = 46.056$, $p < 0.001$, $\eta^2 = 0.490$] with no significant Age*Cue interaction [$F(2,48) = 0.439$, $p = 0.646$, $\eta^2 = 0.009$]. A pairwise comparison revealed better thresholds across groups with the formant cues ($M = 0.90 \pm 0.63$), compared to the F0 cues ($M = 1.18 \pm 0.78$; $p = 0.011$), with the best thresholds achieved with the combined F0 + formant cues ($M = 0.62 \pm 0.35$; F0 > F0 + formants, $p < 0.001$; Formants > F0 + formants, $p = 0.001$). Contrast analysis showed that both groups benefited similarly from the combined cues, compared to a single cue, with no significant Age*Cue interactions for F0, compared to F0 + formants [$F(1,48) = 0.656$, $p = 0.422$], or for formants, compared to F0 + formants [$F(1,48) = 0.01$, $p = 0.978$].

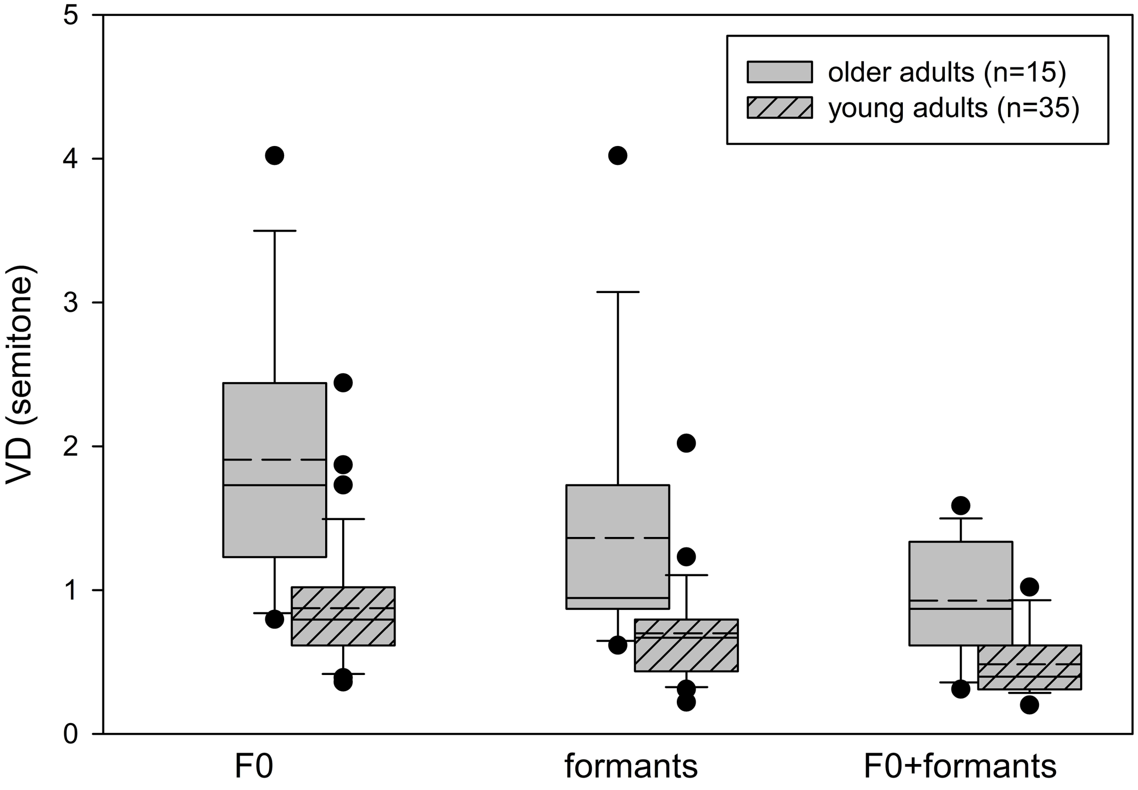


FIGURE 2 | Box plots for voice discrimination thresholds with the F0, formant, and combined F0 + formant cues for the young adults ($n = 35$) and the older adults ($n = 15$). Box limits include the 25th to 75th percentile data, the solid line within the box represents the median, and the dashed line represents the mean. Bars extend to the 10th and 90th percentiles. Black dots represent outliers.

Mauchly's test of sphericity revealed larger between-subject variability in the VD results for older adults than for young adults ($p = 0.019$). Thus, further analysis of the results was conducted at the individual level. Individual VD thresholds for F0 in relation to the formant cues for both young and older adults are shown in **Figure 3**. Also shown is the mean VD ± 1.5 standard deviation of the young adults (gray areas). The majority of the participants demonstrated better discrimination for formant cues, compared to F0 cues, with a higher proportion of this preference in older adults ($n = 12$, 80%) compared to younger adults ($n = 20$, 57%). However, this difference was not significant (Fisher's exact test: $p = 0.199$). Furthermore, although group analysis showed significantly worse thresholds for the older adults, individual analysis revealed that nine (60%), seven (47%), and five (33%) of the 15 older adults performed within the range of the young adults for the formant, F0, and formant+F0 conditions, respectively.

A comparison between the combined F0 + formant and single F0 or formant cues (**Figures 4A,B**) revealed that the majority of the participants benefited from the integration of two cues, compared to a single cue for VD, as is shown in **Figure 4**, by more data points below the diagonal line than above it. Specifically, for the older adults, 13 (87%) and 11 (73%) participants benefited from the combined cues over F0 only or formants only, respectively. For the young adults, 31 (89%) and 24 (69%) participants benefited from the combined cues over F0 only and formants only, respectively. There was no significant difference in proportions between older and younger participants (Fisher's exact test: $p > 0.05$). However, only six (40%) older adults performed within the range of the young adults with combined cues (**Figures 4A,B**).

Cognitive Abilities, Hearing Sensitivity, and Voice Discrimination

All the older adults exhibited cognitive performance in the normal range on each cognitive measure. The mean results of the cognitive tests for the young and older adults are shown in **Table 1**, along with the results of a one-way ANOVA comparing the two groups. Significantly better scores were achieved by the younger group in all cognitive tests. Pearson coefficient correlations were conducted separately for each group to test associations between the cognitive scores and age with the VD performance. For the older adults an additional Pearson coefficient correlation was conducted between hearing sensitivity (PTA4 averaged across both ears) and the VD performance.

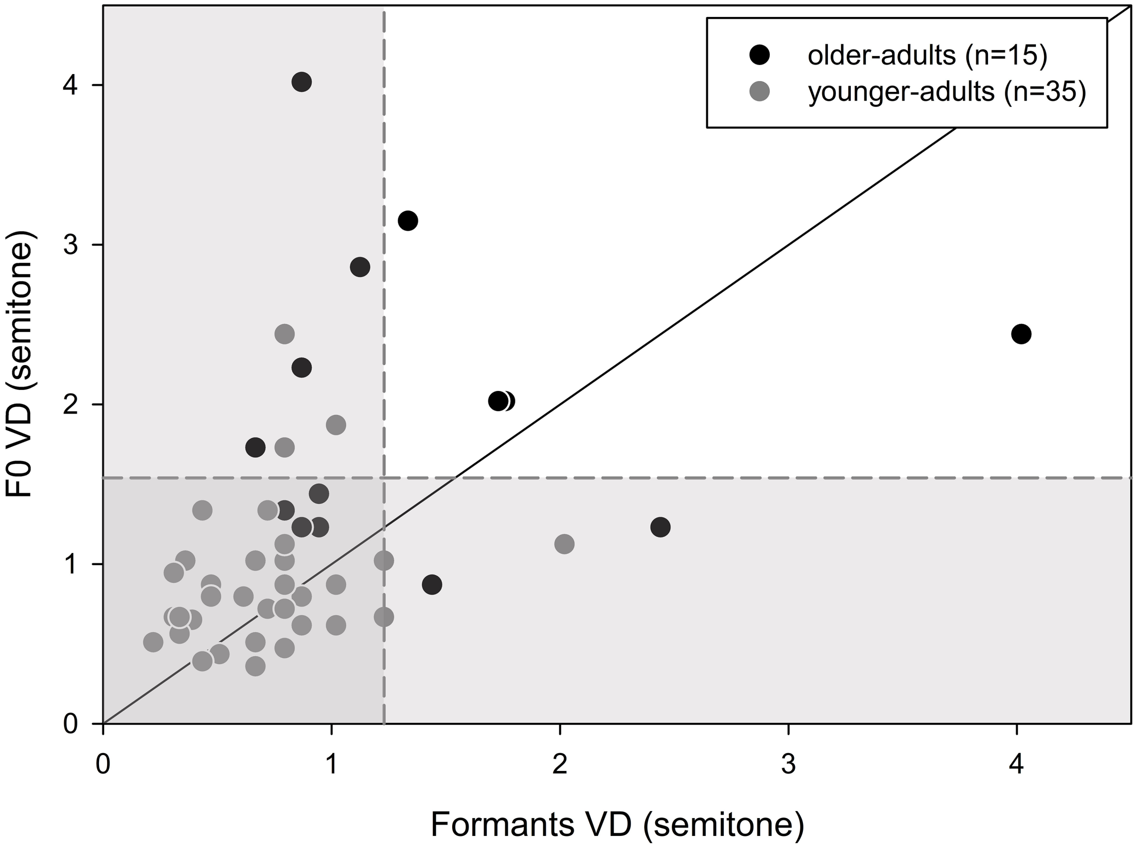


FIGURE 3 | Individual discrimination thresholds (DLs) for formants vs. F0 for the young ($n = 35$) and older adults ($n = 15$) The diagonal line is positioned at x = y. Data above the line show better thresholds for the formant cues, compared to the F0 cues, whereas data below the line show better thresholds for the F0 cues. The gray area represents the mean (±1.5 standard deviation) of the young adults' performance.

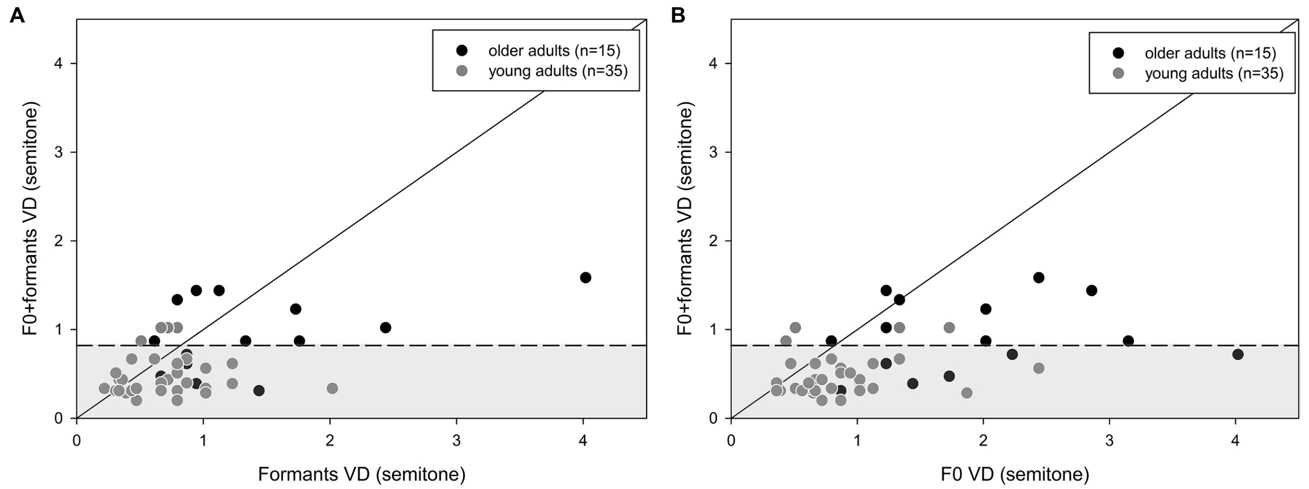


FIGURE 4 | Individual discrimination thresholds (DLs) for young ($n = 35$) and older adults ($n = 15$): **(A)** shows VD for F0 vs. F0 + formants, and **(B)** shows VD for formants vs. F0 + formants. The diagonal line is positioned at x = y. Points that are located on the line show that combining the cues did not change (improve or worsen) discrimination thresholds compared to only one cue. Data above the line show the advantage of the combined cues, compared to only one acoustic cue. Data below the line demonstrate a degrading effect of the combined cues, compared to one cue. The gray area represents the mean (±1.5 standard deviation) of the young adults' performance with the combined cues.

TABLE 1 | Mean performance for each of the cognitive tests for the young adults ($n = 35$) and older adults ($n = 15$).

	Young adults	Older adults	One-way ANOVA
Wechsler digit-span (#)	5.56 (±2.02)	4.4 (±1.12)	$F(1,48) = 4.32$ $p = 0.043$
TMT A (seconds)	22.46 (±7.87)	38.8 (±15.38)	$F(1,49) = 23.92$ $p < 0.01$
TMT B (seconds)	44.16 (12.10)	70.2 (±24.38)	$F(1,48) = 25.22$ $p < 0.01$

In parentheses are ± one standard deviation. Also shown are the results of the statistical analysis comparing the two groups of participants.

The full correlation results for these tests are shown in **Table 2**. For older adults, significant correlations were found in the combined F0 + formant testing condition between VD thresholds and TMT B scores [$r(13) = 0.592$, $r^2 = 0.35$, $p = 0.02$] (**Figure 5A**), and between VD thresholds and hearing sensitivity [$r(13) = 0.594$, $r^2 = 0.35$, $p = 0.02$] (**Figure 5B**). That is, shorter (i.e., better) TMT B times and better hearing sensitivity were associated with lower (i.e., better) VD thresholds when using F0 + formant cues. No significant association was found between TMT B scores and PTA4 ($p > 0.05$) for this group. No significant associations were found between cognitive abilities and VD results for the young adults ($p > 0.05$). The magnitudes of associations between F0 + formant VD and TMT B in the two groups (TMT_B X group interaction) was found statistically significant ($p = 0.026$), adding $R^2 = 0.063$ to the proportion of explained variance in F0 + formant VD.

DISCUSSION

In the present study, we examined the utilization of F0 and formant cues for VD in 15 older adults (65–78 years old) as compared to 35 young adults (18–35 years old). The results support the following findings: (a) Despite being generally high functioning, including hearing sensitivity expected for their age and normal cognitive function, the older adults as a group showed reduced ability to utilize voice cues (F0, formant frequencies, and the combined F0 + formants) for VD, compared to young adults; (b) However, individual analysis revealed large between-subject variability for the older adults, with 47–60% of them reaching VD performance that was within the range of young adults (mean ± 1.5 SD) with at least one acoustic cue; (c) An advantage was observed for formant cues, compared to F0 cues, for both young and older adults, in keeping with previous findings for children and young adults (Zaltz et al., 2020) and confirming that formants remain the reliable cue for VD throughout life; (d) Both the young and older participants benefitted more from the provision of the combined F0 + formant cues than from the provision of a single cue, supporting the hypothesis that older adults are capable of integrating spectral and temporal cues; (e) For the older adults, a combined effect of executive control abilities and speed of processing (as assume to be reflected by TMT B), and hearing sensitivity (PTA4) contributed significantly to the variance of VD in the combined F0 + formant condition, emphasizing the importance of basic requirements of bottom-up and top-down capabilities to allow for more advanced processing such as integration of acoustic cues.

Our major finding that as a group, older adults achieved larger (i.e., poorer) VD thresholds using F0 and formant frequency cues compared to young adults can be partly explained by age-related decline in cognitive abilities including a combined effect of processing speed, mental flexibility and executive control abilities. Normal aging is expected to include neurocognitive changes in working memory, attention, inhibition, and the speed of processing (e.g., Mitchell et al., 2000; Salthouse, 2000; Harada et al., 2013). These may be critical for focusing attention on the relevant acoustic

TABLE 2 | Pearson coefficient correlations between the cognitive scores, age, and hearing sensitivity (PTA4 averaged across both ears, available only for the older adults) and VD performance, separately for the young adults ($n=35$), and older adults ($n=15$).

	Young adults			Older adults		
	F0	**Formants**	**F0+formants**	**F0**	**Formants**	**F0+formants**
TMTA	−0.176	0.235	−0.137	0.197	−0.016	0.327
	(0.311)	(0.174)	(0.432)	(0.481)	(0.955)	(0.235)
TMTB	−0.315	−0.043	−0.069	−0.092	0.170	0.592*
	(0.070)	(0.811)	(0.700)	(0.745)	(0.546)	(0.020)
Wechsler digit-span	0.193	0.240	0.224	0.095	0.129	0.054
	(0.274)	0.172	(0.202)	(0.736)	(0.647)	(0.848)
Age	0.060	0.165	−0.085	−0.359	−0.017	0.081
	(0.731)	(0.345)	(0.626)	(0.189)	(0.953)	(0.775)
PTA4	–	–	–	0.149	0.413	0.594*
				(0.596)	(0.126)	(0.020)

*Significance level is shown in parentheses. *Correlation is significant at the 0.05 level. TMT, Trail Making Test.*

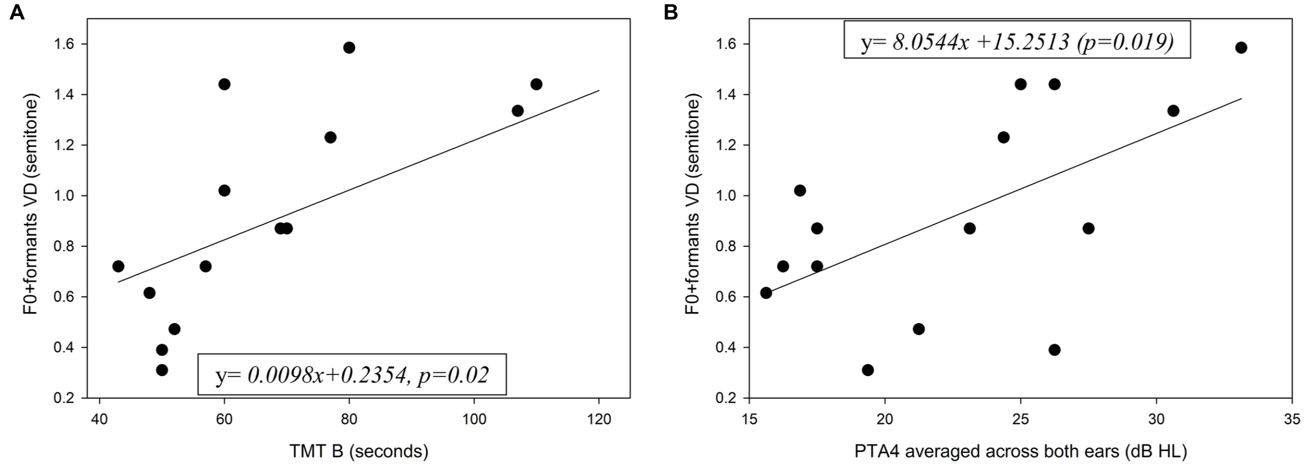


FIGURE 5 | Individual F0+formants VD for the older adults as a function of **(A)** performance on the TMT B test and **(B)** hearing sensitivity. PTA4=mean pure-tone thresholds at 500–4000Hz, averaged across both ears. Also shown in solid line for each graph is the function best fitting the data.

cues for VD and storing this information in memory long enough for decision making. The significant positive association found for the older adults between the F0+formant VD and TMT B results may suggest that attention focusing, inhibition and perceptual speed of processing played an important role in VD.

Another explanation for the poorer VD performance of the older adults may stem from the significant association that was found between hearing sensitivity (PTA4) and DL for the F0+formant condition in the older group. This association suggests that regardless of our attempts to compensate for loss in audibility by presenting the stimuli at 35–40dB above the average individual hearing thresholds (PTA4), older adults with poorer audiograms showed inferior VD performance. This finding may support the notion that audibility is necessary but not sufficient for good auditory processing and that resolving capabilities in the spectral and temporal domains are needed (e.g., Schneider et al., 2010). Indeed, spectro-temporal processing has been suggested to decline with age as a result of numerous deficits, including a subclinical loss of outer hair cells, broadened auditory filters, strial dysfunction, cochlear synaptopathy, and loss of neural synchrony (for a review, see Anderson and Karawani, 2020). Although temporal and spectral processing capabilities were not directly assessed in the current study, previous data suggest that F0 and formant frequency coding relies on efficient utilization of both temporal and spectral information (e.g., Fant, 1960; Lieberman and Blumstein, 1988; Carlyon and Shackleton, 1994; Fu et al., 2004; Chatterjee and Peng, 2008; Oxenham, 2008; Xu and Pfingst, 2008). Hence, deficits in the spectro-temporal processing, such as impairments in periodicity and fine-structure perception (Souza et al., 2011) may explain the poor performance in VD of older adults with greater loss of hearing sensitivity despite attempts to compensate for this loss of audibility.

One can argue that our finding of poorer VD for the older participants may have also been related to the fact that the VD test was conducted in Hebrew, which was the mother tongue for six of the 15 older participants, but for all the young participants. This hypothesis is based on the notion that testing in a second language (L2) may have increased working memory demands, and in our case, influencing VD performance for those older participants. We believe, however, that this explanation is less likely, because the stimuli for the VD task in the current study included only three sentences, one for each VD assessment. Moreover, for each threshold estimation, the same sentence was used. The listeners were, therefore, required to identify the odd sentence based only on psychoacoustic perception, with no linguistic decisions to be made. Furthermore, all three sentences were taken from the Hebrew version of the Matrix sentence test which comprises words that are suitable for 5-year-old Hebrew speakers (Bugannim et al., 2019), and included three words with a simple grammatical structure (noun, verb, adjective). Given that the nine older adults in our study whose mother tongue was not Hebrew were all fluent Hebrew speakers who were exposed to Hebrew for at least 29 years, with an average of 45 years, linguistic knowledge was probably not a contributing factor to the working memory demands of the VD task.

A second outcome of the present study is that the majority of older adults showed better discrimination thresholds with formant cues than F0 cues. This is similar to what was found in our young adults and in line with a recent study of school-age children (Zaltz et al., 2020). The advantage of formant cues for VD (in comparison to F0 cues) may be related to the amount of variation that each acoustic cue has in natural speech. While F0 varies significantly in terms of time for conveying prosodic information (standard deviation is approximately 3.7 semitones), formant frequencies are relatively constant over the duration of the vowel (standard deviation of approximately one semitone; Kania et al., 2006; Chuenwattanapranithi et al., 2008). The finding that formant frequencies remain a reliable cue for VD across the lifespan may also suggest faster degradation of temporal (F0) processing compared to spectral (formants) processing with age (Vongpaisal and Pichora-Fuller, 2007; Chintanpalli et al., 2016; Goupell et al., 2017).

Our finding that older adults benefit from combined cues to a similar degree as young adults supports the notion that older adults are able to integrate information from separately coded processes and use it to their advantage. Thus, despite the fact that formant frequencies were more easily accessed by older listeners, compared to F0 cues, this latter information was not ignored and was used to support the dominant channel of information for VD as long as their combined executive control and perceptual processing abilities were efficient (supported by the significant association found with TMT B; Bowie and Harvey, 2006; Sánchez-Cubillo et al., 2009).

Finally, our study showed a large between-subject variability in VD performance in the older group, with approximately half of the older adults showing a performance comparable to that of the young adults with at least one acoustic cue and 40% in the combined-cue condition. This variability was likely related to the considerable inter-individual variability reported for older adults in a range of auditory processing abilities and cognitive processes (e.g., Tun et al., 2012), as reflected by the significant associations found between VD performance and cognitive (TMT B) and sensory (PTA4) abilities. Given that the efficient discrimination and integration of F0 and formants are essential for speech segregation (e.g., Darwin et al., 2003; Vestergaard et al., 2009), age-related declines in cognitive and sensory processing that negatively affect these abilities may lead to impaired speech perception in noise.

LIMITATIONS OF THE STUDY AND SUGGESTIONS FOR FUTURE STUDIES

Although we tested a relatively homogeneous group of older adults (all with normal audiograms for their age, normal mental capabilities, relatively high functioning, and similar audibility of the stimuli) and assessed several cognitive abilities in addition to VD performance, our findings provided only a partial explanation for the poor VD performance found in older adults. Psychoacoustic tests and more sensitive working memory tests may be included in future studies to further assess the relationship between temporal and spectral abilities and VD in older adults, and better reflect the correlations with age-related declines in cognitive abilities. Also, although the differences in audibility between the young and older adults were compensated by presenting the stimuli at approximately 35–40 dB SL above the individual PTA4, one cannot rule out the possibility that the inferior VD performance of the older adults was related to their poor hearing sensitivity. We believe that this issue did not have a major effect on the VD results because the acoustic cues for F0 and formants are mainly in the low-mid frequency range. However, a different or additional compensation method, such as frequency-shaped amplification on the stimuli (e.g., Souza et al., 2011) may have better accommodated the high-frequency threshold elevations of some of the older listeners. Finally, future studies may want to control the mother tongue of the participants in order to assure that there is no effect of the mono versus multilingualism status of the participants on VD performance.

CONCLUSION

The present study is the first to test the contribution of F0, formant, and the combination of F0 and formant cues to VD in older adults with no hearing loss other than loss of hearing sensitivity as expected to decline with age. The findings indicate that many of the older adults found the VD task more difficult than NH young adults, presenting poorer VD thresholds across conditions. The VD thresholds in the F0 + formant condition for the older adults were associated with their audiograms, likely reflecting the importance of good bottom-up input and processing for efficient utilization of the acoustic voice cues of the talker. VD thresholds were also associated with TMT B scores, which are assumed to provide a general measure of executive control abilities, and perceptual processing, possibly reflecting their need to resort to cognitive resources in the presence of inefficient spectro-temporal processing (Schneider, 2011). These findings may explain the difficulties that older adults have in segregating talkers. The findings may also provide a possible explanation for the difficulty older adults face when listening to speech in a multi-talker environment. However, these assumptions need to be confirmed in future studies.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by The Institutional Review Board of Ethics at Tel Aviv University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

All authors contributed to this work to a significant extent, have read the article and agreed to submit it for publication after discussing the results and implications and commented on the article at all stages, and are, therefore, responsible for the reported research and have approved the final article as submitted.

ACKNOWLEDGMENTS

The authors wish to acknowledge the contributions of the following undergraduate students from the Department of Communication Disorders at Tel Aviv University for assisting in data collection: Elisheva Tibi, Sara Aviv, Alona Kandelei Shamor, Ester Reuveni, Daniel Lex Rabinovitch, and Orpaz Shariki.

REFERENCES

Anderson, S., Bieber, R., and Schloss, A. (2021). Peripheral deficits and phase-locking declines in aging adults. *Hear. Res.* 403:108188. doi: 10.1016/j.heares.2021.108188

Anderson, S., and Karawani, H. (2020). Objective evidence of temporal processing deficits in older adults. *Hear. Res.* 397:108053. doi: 10.1016/j.heares.2020.108053

Ansi, S. (2018). American national standard specification for audiometers. New York: ANSI S3.6.

Arlinger, S. (2003). Negative consequences of uncorrected hearing loss - A review. *Int. J. Audiol.* 42, 2S17–2S20. doi: 10.3109/14992020309074639

Başkent, D., and Gaudrain, E. (2016). Musician advantage for speech-on-speech perception. *J. Acoust. Soc. Am.* 139, EL51–EL56. doi: 10.1121/1.4942628

Best, V., Ahlstrom, J. B., Mason, C. R., Roverud, E., Perrachione, T. K., Kidd, G. Jr., et al. (2018). Talker identification: effects of masking, hearing loss, and age. *J. Acoust. Soc. Am.* 143, 1085–1092. doi: 10.1121/1.5024333

Bowie, C. R., and Harvey, P. D. (2006). Administration and interpretation of the trail making test. *Nat. Protoc.* 1, 2277–2281. doi: 10.1038/nprot.2006.390

Bronkhorst, A. W. (2015). The cocktail-party problem revisited: early processing and selection of multi-talker speech. *Atten. Percept. Psychophys.* 77, 1465–1487. doi: 10.3758/s13414-015-0882-9

Brungart, D. S. (2001). Informational and energetic masking effects in the perception of two simultaneous talkers. *J. Acoust. Soc. Am.* 109, 1101–1109. doi: 10.1121/1.1345696

Bugannim, Y., Roth, D. A., Zechoval, D., and Kishon-Rabin, L. (2019). Training of speech perception in noise in pre-lingual hearing impaired adults with cochlear implants compared with normal hearing adults. *Otol. Neurotol.* 40, e316–e325. doi: 10.1097/MAO.0000000000002128

Cacciatore, F., Napoli, C., Abete, P., Marciano, E., Triassi, M., and Rengo, F. (1999). Quality of life determinants and hearing function in an elderly population: Osservatorio Geriatrico Campano study group. *Gerontology* 45, 323–328. doi: 10.1159/000022113

Carlyon, R. P., and Shackleton, T. M. (1994). Comparing the fundamental frequencies of resolved and unresolved harmonics: evidence for two pitch mechanisms? *J. Acoust. Soc. Am.* 95, 3541–3554. doi: 10.1121/1.409971

Chatterjee, M., and Peng, S. C. (2008). Processing F0 with cochlear implants: modulation frequency discrimination and speech intonation recognition. *Hear. Res.* 235, 143–156. doi: 10.1016/j.heares.2007.11.004

Chintanpalli, A., Ahlstrom, J. B., and Dubno, J. R. (2016). Effects of age and hearing loss on concurrent vowel identification. *J. Acoust. Soc. Am.* 140, 4142–4153. doi: 10.1121/1.4968781

Chuenwattanapranithi, S., Xu, Y., Thipakorn, B., and Maneewongvatana, S. (2008). Encoding emotions in speech with the size code. *Phonetica* 65, 210–230. doi: 10.1159/000192793

Darwin, C. J., Brungart, D. S., and Simpson, B. D. (2003). Effects of fundamental frequency and vocal-tract length changes on attention to one of two simultaneous talkers. *J. Acoust. Soc. Am.* 114, 2913–2922. doi: 10.1121/1.1616924

Durlach, N. I., Mason, C. R., Shinn-Cunningham, B. G., Arbogast, T. L., Colburn, H. S., and Kidd, G. (2003). Informational masking: counteracting the effects of stimulus uncertainty by decreasing target-masker similarity. *J. Acoust. Soc. Am.* 114, 368–379. doi: 10.1121/1.1577562

Engdahl, B., Tambs, K., Borchgrevink, H. M., and Hoffman, H. J. (2005). Screened and unscreened hearing threshold levels for the adult population: results from the Nord-Trøndelag hearing loss study. *Int. J. Audiol.* 44, 213–230. doi: 10.1080/14992020500057731

Ezzatian, P., Li, L., Pichora-Fuller, M. K., and Schneider, B. A. (2012). The effect of energetic and informational masking on the time-course of stream segregation: evidence that streaming depends on vocal fine structure cues. *Lang. Cogn. Process.* 27, 1056–1088. doi: 10.1080/01690965.2011.591934

Fant, G. (1960). *Acoustic Theory of Speech Perception*. The Hague: Mouton.

Fitch, W. T., and Giedd, J. (1999). Morphology and development of the human vocal tract: a study using magnetic resonance imaging. *J. Acoust. Soc. Am.* 106, 1511–1522.

Folstein, M. F., Folstein, S. E., and McHugh, P. R. (1975). "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J. Psychiatr. Res.* 12, 189–198. doi: 10.1016/0022-3956(75)90026-6

Fu, Q. J., Chinchilla, S., and Galvin, J. J. (2004). The role of spectral and temporal cues in voice gender discrimination by normal-hearing listeners and cochlear implant users. *J. Assoc. Res. Otolaryngol.* 5, 253–260. doi: 10.1007/s10162-004-4046-1

Gaudrain, E., and Başkent, D. (2018). Discrimination of voice pitch and vocal-tract length in Cochlear implant users. *Ear Hear.* 39, 226–237. doi: 10.1097/AUD.0000000000000480

Gopinath, B., Schneider, J., McMahon, C. M., Teber, E., Leeder, S. R., and Mitchell, P. (2012). Severity of age-related hearing loss is associated with impaired activities of daily living. *Age Ageing* 41, 195–200. doi: 10.1093/ageing/afr155

Goupell, M. J., Gaskins, C. R., Shader, M. J., Walter, E. P., Anderson, S., and Gordon-Salant, S. (2017). Age-related differences in the processing of temporal envelope and spectral cues in a speech segment. *Ear Hear.* 38, e335–e342. doi: 10.1097/AUD.0000000000000447

Harada, C. N., Love, M. C. N., and Triebel, K. L. (2013). Normal cognitive aging. *Clin. Geriatr. Med.* 29, 737–752. doi: 10.1016/j.cger.2013.07.002

Hoen, M., Meunier, F., Grataloup, C. L., Pellegrino, F., Grimault, N., Perrin, F., et al. (2007). Phonetic and lexical interferences in informational masking during speech-in-speech comprehension. *Speech Comm.* 49, 905–916. doi: 10.1016/j.specom.2007.05.008

Humes, L. E., Dubno, J. R., Gordon-Salant, S., Lister, J. J., Cacace, A. T., Cruickshanks, K. J., et al. (2012). Central presbycusis: a review and evaluation of the evidence. *J. Am. Acad. Audiol.* 23, 635–666. doi: 10.3766/jaaa.23.8.5

Jayakody, D., Friedland, P. L., Martins, R. N., and Sohrabi, H. R. (2018). Impact of aging on the auditory system and related cognitive functions: a narrative review. *Front. Neurosci.* 12:125. doi: 10.3389/fnins.2018.00125

Kania, R. E., Hartl, D. M., Hans, S., Maeda, S., Vaissiere, J., and Brasnu, D. F. (2006). Fundamental frequency histograms measured by electroglottography during speech: a pilot study for standardization. *J. Voice* 20, 18–24. doi: 10.1016/j.jvoice.2005.01.004

Kattner, F., and Ellermeier, W. (2020). Distraction at the cocktail party: attenuation of the irrelevant speech effect after a training of auditory selective attention. *J. Exp. Psychol. Hum. Percept. Perform.* 46, 10–20. doi: 10.1037/xhp0000695

Lammert, A. C., and Narayanan, S. S. (2015). On short-time estimation of vocal tract length from formant frequencies. *PLoS One* 10:e0132193.

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *J. Acoust. Soc. Am.* 49, 467–477. doi: 10.1121/1.1912375

Lieberman, P., and Blumstein, S. E. (eds.) (1988). "Source-filter theory of speech production," in *Speech Physiology, Speech Perception, and Acoustic Phonetics* (Cambridge: Cambridge University Press), 34–50.

Mackersie, C. L., Dewey, J., and Guthrie, L. A. (2011). Effects of fundamental frequency and vocal-tract length cues on sentence segregation by listeners with hearing loss. *J. Acoust. Soc. Am.* 130, 1006–1019. doi: 10.1121/1.3605548

McDowd, J., and Shaw, R. (2000). "Attention and aging: a functional perspective," in *The Handbook of Aging and Cognition. 2nd Edn.* eds. F. Craik and T. Salthouse (NJ: Erlbaum), 221–292.

Mitchell, K. J., Johnson, M. K., Raye, C. L., Mather, M., and D'Esposito, M. (2000). Aging and reflective processes of working memory: binding and test load deficits. *Psychol. Aging* 15, 527–541. doi: 10.1037//0882-7974.15.3.527

Moore, B. C. J., and Peters, R. W. (1992). Pitch discrimination and phase sensitivity in young and elderly subjects and its relationship to frequency selectivity. *J. Acoust. Soc. Am.* 91, 2881–2893. doi: 10.1121/1.402925

Moulines, E., and Charpentier, F. (1990). Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech Commun.* 9, 453–467.

Oxenham, A. J. (2008). Pitch perception and auditory stream segregation: implications for hearing loss and cochlear implants. *Trends Amplif.* 12, 316–331. doi: 10.1177/1084713808325881

Pichora-Fuller, M. K. (1997). Language comprehension in older adults. *J. Speech. Lang. Path. Audiol.* 2, 125–142.

Pichora-Fuller, M. K., and Singh, G. (2006). Effects of age on auditory and cognitive processing: implications for hearing aid fitting and audiological rehabilitation. *Trends Amplif.* 10, 29–59. doi: 10.1177/108471380601000103

Pronk, M., Deeg, D. J., Smits, C., Van Tilburg, T. G., Kuik, D. J., Festen, J. M., et al. (2011). Prospective effects of hearing status on loneliness and depression in older persons: identification of subgroups. *Int. J. Audiol.* 50, 887–896. doi: 10.3109/14992027.2011.599871

Salthouse, T. A. (2000). Aging and measures of processing speed. *Biol. Psychol.* 54, 35–54. doi: 10.1016/S0301-0511(00)00052-1

Sánchez-Cubillo, I., Periáñez, J. A., Adrover-Roig, D., Rodríguez-Sánchez, J. M., Ríos-Lago, M., Tirapu, J., et al. (2009). Construct validity of the trail making test: role of task-switching, working memory, inhibition/interference control, and visuomotor abilities. *J. Int. Neuropsychol. Soc.* 15, 438–450. doi: 10.1017/S1355617709090626

Schneider, B. A. (2011). How age affects auditory-cognitive interactions in speech omprehension. *Audiol. Res.* 1:e10. doi: 10.4081/audiores.2011.e10

Schneider, B. A., Li, L., and Daneman, M. (2007). How competing speech interferes with speech comprehension in everyday listening situations. *J. Am. Acad. Audiol.* 18, 559–572. doi: 10.3766/jaaa.18.7.4

Schneider, B. A., Pichora-Fuller, K., and Daneman, M. (2010). "Effects of senescent changes in audition and cognition on spoken language comprehension," in *The Aging Auditory System: Springer Handbook of Auditory Research*. eds. S. Gordon-Salant, R. D. Frisina and A. N. Popper (New York, NY: Springer), 167–210.

Schvartz-Leyzac, K. C., and Chatterjee, M. (2015). Fundamental-frequency discrimination using noise-band-vocoded harmonic complexes in older listeners with normal hearing. *J. Acoust. Soc. Am.* 138, 1687–1695. doi: 10.1121/1.4929938

Shultz, D. (2015). The privacy arms race. When your voice betrays you. *Science* 347:494. doi: 10.1126/science.347.6221.494

Skuk, V. G., and Schweinberger, S. R. (2014). Influences of fundamental frequency, formant frequencies, aperiodicity, and spectrum level on the perception of voice gender. *J. Speech Lang. Hear. Res.* 57, 285–296. doi: 10.1044/1092-4388(2013/12-0314)

Smith, D. R., and Patterson, R. D. (2005). The interaction of glottal-pulse rate and vocal-tract length in judgements of speaker size, sex, and age. *J. Acoust. Soc. Am.* 118, 3177–3186. doi: 10.1121/1.2047107

Smith, D. R., Walters, T. C., and Patterson, R. D. (2007). Discrimination of speaker sex and size when glottal-pulse rate and vocal-tract length are controlled. *J. Acoust. Soc. Am.* 122, 3628–3639. doi: 10.1121/1.2799507

Souza, P., Arehart, K., Miller, C. W., and Muralimanohar, R. K. (2011). Effects of age on F0 discrimination and intonation perception in simulated electric and electroacoustic hearing. *Ear Hear.* 32, 75–83. doi: 10.1097/AUD.0b013e3181eccfe9

Tremblay, P., Brisson, V., and Deschamps, I. (2021). Brain aging and speech perception: effects of background noise and talker variability. *NeuroImage* 227:117675. doi: 10.1016/j.neuroimage.2020.117675

Tun, P. A., Williams, V. A., Small, B. J., and Hafter, E. R. (2012). The effects of aging on auditory processing and cognition. *Am. J. Audiol.* 21, 344–350. doi: 10.1044/1059-0889(2012/12-0030)

Vestergaard, M. D., Fyson, N. R., and Patterson, R. D. (2009). The interaction of vocal characteristics and audibility in the recognition of concurrent syllables. *J. Acoust. Soc. Am.* 125, 1114–1124. doi: 10.1121/1.3050321

Vestergaard, M. D., Fyson, N. R., and Patterson, R. D. (2011). The mutual roles of temporal glimpsing and vocal characteristics in cocktail-party listening. *J. Acoust. Soc. Am.* 130, 429–439. doi: 10.1121/1.3596462

Vongpaisal, T., and Pichora-Fuller, M. K. (2007). Effect of age on F0 difference limen and concurrent vowel identification. *J. Speech. Lang. Hear. Res.* 50, 1139–1156. doi: 10.1044/1092-4388(2007/079)

Wechsler, D. (1991). *Wechsler Intelligence Scale for Children—III*. San Antonio, TX: The Psychological Corporation.

Working Group on Speech Understanding and Aging (1988). Speech understanding and aging. *J. Acoust. Soc. Am.* 83, 859–895. doi: 10.1121/1.395965

Xu, L., and Pfingst, B. E. (2008). Spectral and temporal cues for speech recognition: implications for auditory prostheses. *Hear. Res.* 242, 132–140. doi: 10.1016/j.heares.2007.12.010

Zaltz, Y., Goldsworthy, R. L., Eisenberg, L. S., and Kishon-Rabin, L. (2020). Children with normal hearing are efficient users of fundamental frequency and vocal tract length cues for voice discrimination. *Ear Hear.* 41, 182–193. doi: 10.1097/AUD.0000000000000743

Zaltz, Y., Goldsworthy, R. L., Kishon-Rabin, L., and Eisenberg, L. S. (2018). Voice discrimination by adults with cochlear implants: the benefits of early implantation for vocal-tract length perception. *J. Assoc. Res. Otolaryngol.* 19, 193–209. doi: 10.1007/s10162-017-0653-5

Conflict of Interest: 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.

APPENDIX

Appendix A

Three-word sentences including a subject, a predicate, and an object (mean duration = 104.67 ± 6.11 msec) were used for the VD test. The sentences were shortened from the original 5-word sentences from the Hebrew version of the Matrix test to three words in order to minimize working memory demands. Sentences were manipulated using a 13-point stimulus continuum, exponentially ranging in $\sqrt{2}$ steps from a change of −0.18 semitone to a change of −8 semitones. This manipulation resulted in the following three separate dimensions: (1) F0, (2) formant frequencies, in which all formants were shifted down by an equal ratio, according to the 13-point stimulus continuum, and (3) combined, F0 + formants, in which both F0 and formants were shifted down in a similar manner. For example, the mean F0 for the F0-manipulated sentences varied by 0, −0.18, −0.26, −0.36, −0.51, −0.72, −1.02, −1.44, −2.02, −2.86, −4.02, −5.67, and − 8 semitones from the original sentence mean F0. Consequently, for the first sentence, the mean F0 was 175.62 Hz and the comparison sentences changed exponentially in $\sqrt{2}$ steps from 174 to 110.35 Hz, using the PSOLA algorithm (Moulines and Charpentier, 1990) for pitch extraction and manipulation. The formant frequencies for this sentence were manipulated exponentially, ranging in $\sqrt{2}$ steps from 0.99 (the smallest ratio between the original formant frequencies and the manipulated formant frequencies) to 0.63 (the highest ratio). This manipulation required resampling the stimulus to compress the frequency axis by a range of factors similar in ratio to the F0 change. The PSOLA algorithm was then applied to regain the original pitch and duration. Given that the average VTL of an adult female is approximately 140 mm (Fitch and Giedd, 1999), and formant frequencies are inversely proportional to VTL (Lammert and Narayanan, 2015), the formants range corresponded to a change from 2 to 88 mm. Note that this is a wider VTL range than expected from humans and was originally chosen to avoid floor effects for CI participants who had difficulty in perceiving normal difference limen (DL) for VTL (Zaltz et al., 2018). All manipulations were implemented using the PRAAT software version 5.4.17 (copyright 1992–2015 by Boersma and Weenink). Spectrographic displays for one of the sentences that was manipulated in F0, formants and F0 + formants for the VD test can be found in Zaltz et al. (2020).

frontiers in Psychology

ORIGINAL RESEARCH
published: 15 March 2022
doi: 10.3389/fpsyg.2022.838576

Do Age and Linguistic Status Alter the Effect of Sound Source Diffuseness on Speech Recognition in Noise?

Meital Avivi-Reich[1]*, Rupinder Kaur Sran[2,3] and Bruce A. Schneider[2]

[1] Department of Communication Arts, Sciences and Disorders, Brooklyn College, City University of New York, Brooklyn, NY, United States, [2] Human Communication Lab, Department of Psychology, University of Toronto Mississauga, Toronto, ON, Canada, [3] Department of Speech-Language Pathology, University of Toronto, Toronto, ON, Canada

One aspect of auditory scenes that has received very little attention is the level of diffuseness of sound sources. This aspect has increasing importance due to growing use of amplification systems. When an auditory stimulus is amplified and presented over multiple, spatially-separated loudspeakers, the signal's timbre is altered due to comb filtering. In a previous study we examined how increasing the diffuseness of the sound sources might affect listeners' ability to recognize speech presented in different types of background noise. Listeners performed similarly when both the target and the masker were presented via a similar number of loudspeakers. However, performance improved when the target was presented using a single speaker (compact) and the masker from three spatially separate speakers (diffuse) but worsened when the target was diffuse, and the masker was compact. In the current study, we extended our research to examine whether the effects of timbre changes with age and linguistic experience. Twenty-four older adults whose first language was English (Old-EFLs) and 24 younger adults whose second language was English (Young-ESLs) were asked to repeat non-sense sentences masked by either Noise, Babble, or Speech and their results were compared with those of the Young-EFLs previously tested. Participants were divided into two experimental groups: (1) A Compact-Target group where the target sentences were presented over a single loudspeaker, while the masker was either presented over three loudspeakers or over a single loudspeaker; (2) A Diffuse-Target group, where the target sentences were diffuse while the masker was either compact or diffuse. The results indicate that the Target Timbre has a negligible effect on thresholds when the timbre of the target matches the timbre of the masker in all three groups. When there is a timbre contrast between target and masker, thresholds are significantly lower when the target is compact than when it is diffuse for all three listening groups in a Noise background. However, while this difference is maintained for the Young and Old-EFLs when the masker is Babble or Speech, speech reception thresholds in the Young-ESL group tend to be equivalent for all four combinations of target and masker timbre.

Keywords: masking, aging, bilingualism, speech perception, diffuseness, amplification

Edited by:
Julian Keil,
University of Kiel, Germany

Reviewed by:
Eliane Schochat,
University of São Paulo, Brazil
Erol J. Ozmeral,
University of South Florida,
United States

***Correspondence:**
Meital Avivi-Reich
Meital.Avivi-Reich@brooklyn.cuny.edu

Specialty section:
This article was submitted to
Perception Science,
a section of the journal
Frontiers in Psychology

Received: *18 December 2021*
Accepted: *14 February 2022*
Published: *15 March 2022*

Citation:
Avivi-Reich M, Sran RK and Schneider BA (2022) Do Age and Linguistic Status Alter the Effect of Sound Source Diffuseness on Speech Recognition in Noise? Front. Psychol. 13:838576. doi: 10.3389/fpsyg.2022.838576

INTRODUCTION

Daily communication takes place in a variety of complex auditory settings that often contain several sound sources, some natural and some amplified. These competing sound sources make it difficult to extract a speech target masked by one or more competing sounds. A number of studies have examined different aspects of auditory scenes to be able to better understand how they may affect speech perception and comprehension. For example, researchers have examined how listening to and processing a speech target is affected by the number of auditory sound sources (e.g., Rosen et al., 2013), their intensity (e.g., Dos Santos Sequeira et al., 2010), spectral composition (e.g., Li and Fu, 2010; Roberts and Summers, 2020), and spatial location (e.g., Ezzatian et al., 2010; Avivi-Reich et al., 2014; Gygi and Shafiro, 2014; Bednar and Lalor, 2020). These studies contributed to our understanding of how the auditory scene and the acoustic input may affect the ways in which listeners detect, process, and encode acoustic signals and verbal information. One aspect of auditory scenes that has received very little attention is how the level of diffuseness of the sound sources affect speech recognition. This topic is becoming increasingly important given the increasing use of surround sound systems in our everyday lives.

Often, when amplification is used, a natural sound source (typically with a compact and defined location) is amplified and presented over more than a single loudspeaker. When an auditory stimulus (e.g., a human voice) is amplified and presented over multiple, spatial-separated loudspeakers, the signal's timbre is altered due to comb filtering, and the sound source is perceived to be more diffuse and with a broader auditory source width (Avivi-Reich et al., 2020). With the growing use of electric amplification and surround-sound systems, it would be useful to determine how the relative diffuseness and compactness of different sound sources affect speech recognition.

In a previous study (Avivi-Reich et al., 2020) we systematically examined how manipulating the diffuseness of the sound sources might affect the ability of young people with normal hearing to correctly identify target speech presented in different types of background noise. Twenty-four young adults were asked to repeat nonsense sentences that were presented in either Noise, Babble, or competing Speech. Participants were divided into two groups: (1) A Compact-Target group where the target sentences were presented over a single loudspeaker (compact target), while the masker was either presented over three spatially separated loudspeakers (diffuse masker) or over a single loudspeaker (compact); (2) A Diffuse-Target group, where the target sentences were diffuse while the masker was either compact or diffuse. The results of this study showed no significant Timbre effect in the absence of a timbre contrast (compact vs. diffuse) between target and masker. However, when there was a timbre contrast, the signal-to-noise ratios (SNRs) needed for 50% correct recognition of the target speech were higher when the masker was compact, and the target was diffuse, and lower when the target was compact, and the masker was diffuse. These results were consistent with the expected effects from comb filtering (for additional information and illustrations see Avivi-Reich et al., 2020), and also could reflect a tendency for attention to be drawn toward compact sound sources that may be perceived as closer in order to avoid dangerous situations or objects even without seeing them (Scharf, 1998; Farnè and Làdavas, 2002; Canzoneri et al., 2012). In vision, the tendency of closer items to have higher ecological salience is referred to as the behavior urgency hypothesis (Franconeri and Simons, 2003). These findings emphasize the importance of considering the level of diffuseness when designing and using amplification systems, especially when using amplification in order to enhance speech perception.

Speech perception in noise (SPIN) can be a demanding task both at peripheral and more central processing levels. Any competing sources in the auditory scene that temporally and spectrally overlaps the target speech signal creates overlapping excitation patterns in the cochlea and in the auditory nerve. This overlap might interfere with the perception and processing of the target at the auditory periphery, which often is referred to as energetic masking or peripheral masking (Durlach et al., 2003). In addition, when the masker contains meaningful speech, it is likely to initiate lexical processing of the masker, potentially allowing the content of irrelevant streams to intrude into working memory and interfere with the processing of the target message. This type of interference often is referred to as informational masking (Freyman et al., 1999; Durlach et al., 2003; Schneider et al., 2007, 2010; Kidd et al., 2008). While energetic masking seems to affect the early stages of sound perception and processing, informational masking is likely to affect later processes (Arbogast et al., 2002; Freyman et al., 2004; Ihlefeld and Shinn-Cunningham, 2008; Szalárdy et al., 2019).

Listeners can alleviate the effects of informational masking if they are able to segregate the different incoming auditory streams so that attention can be focused on processing the target stream. The ability to successfully segregate the streams largely depends on the perceptual similarities and dissimilarities between the target signal and other competing sound sources. Any differences among the sound sources could assist the listener in perceptually segregating the target stream from the competing sound sources, thereby providing a release from masking (Bregman, 1990). A large number of acoustic cues that could assist auditory stream segregation have been previously investigated in order to assess their potential to release the target signal from masking (e.g., Brungart et al., 2001; Humes et al., 2006; Vongpaisal and Pichora-Fuller, 2007). In the current study, we intend to continue investigating the possible role that timbre differences might play in auditory stream segregation (Bregman, 1990). This cue has received limited attention in the literature (see, for example, Freyman et al., 1999), and as far as we know our previous study was the first to systematically investigate its effect on speech recognition.

The current study aims to extend the previous study (Avivi-Reich et al., 2020) to populations other than young native-English listeners (Young-EFLs) to those who are known to experience greater difficulties when listening in complex auditory environment and may be affected differently by the diffuseness level of the different sound sources. Two such groups, whose ability to perceive speech in noise have been extensively studied, are older adult listeners for whom English is a first language

(Old-EFLs) as well as young adults for whom English is their second language (Young-ESLs). These two groups have been found to require more preferable listening conditions in order to achieve correct speech perception compared with young-EFL listeners (e.g., Rogers et al., 2006; Avivi-Reich et al., 2014, 2015; Francis et al., 2018). However, the reasons for their poorer SPIN are likely to be quite different and therefore the effect of sound source diffuseness on their SPIN may differ as well.

Aging and Speech Perception

Older adults often experience greater difficulties perceiving speech in noisy environments, even those who are considered to have normal hearing (Helfer and Freyman, 2008; Stevenson et al., 2015). Interestingly, not all types of maskers have a similar effect on younger and older listeners. One type of masker that seems particularly detrimental to older adults is competing speech (Tun and Wingfield, 1999; Helfer and Freyman, 2008; Goossens et al., 2017). It has also been suggested that older adults with normal hearing for their age benefit less than younger adults when the target voice and competing sound sources occupy different positions in space (Murphy et al., 2006; Marrone et al., 2008; Avivi-Reich et al., 2014), and when there are fluctuations in the masker signal (Stuart and Phillips, 1996; Dubno et al., 2003; Gifford et al., 2007). In addition, evidence suggests that older adults require a greater amount of time to establish stream segregation when listening in an environment that contains more than a single sound source compared to younger adults (Ben-David et al., 2012; Getzmann and Näätänen, 2015). Considering these age-related findings, it is important to examine if and how older adults' speech perception may be affected by changes in the diffuseness level of the sound sources in a noisy environment.

There are several possible reasons why older adults may be less able to use differences in diffuseness between target speech and competing sound sources to unmask the target speech. For example, when the masker is diffuse and the target is compact, older adults might not be able to fully use the troughs in the masker spectrum created by the comb filtering effect to improve speech perception (see Avivi-Reich et al., 2020 for more information regarding the effect of comb filtering under the different testing conditions). Other possible reasons may be related to age- related changes in the ability of listeners to form an auditory image of a diffuse vs. a compact sound, their ability to establish stream segregation between sound sources that are either presented over multiple loudspeakers or a single one, and/or their ability to focus their attention on the target stream.

Second Language and Speech Perception

When listening to a second language, listeners have lower performance than when listening to their first language on a number of speech perception measures (e.g., Ezzatian et al., 2010; Francis et al., 2018; Peng and Wang, 2019). This could be due, in part, to incomplete acquisition of the acoustic–phonetic characteristics in the second language. This incomplete knowledge might result in a reduced phoneme recognition in one's second or third language (Kroll and Steward, 1994). In addition, non-native listeners' second language semantic and linguistic processes may not be completely differentiated from their first language processes (FitzPatrick and Indefrey, 2009). This overlap between the two linguistic systems could result in greater competition as both systems are activated when listening. Hence, the degree and extent to which second language listeners might engage knowledge-driven processes (e.g., vocabulary and linguistic knowledge) to facilitate speech perception could differ from the pattern of engagement in the listeners' first language (Meador et al., 2000). In addition, this greater competition may require greater investment of attentional resources, leaving fewer resources available to attend to fine acoustic changes, such as those created by the presentation of a sound source over several loudspeakers rather than a single one.

MATERIALS AND METHODS

Participants

Twenty-four older listeners for whom English is their first language (Old-EFLs) and 24 younger listeners for whom English is their second language (Young-ESLs) participated in this study. Each group of participants was divided into two experimental groups. Twelve of the Old-EFLs (mean age: 73.08 years; *SD*: 4.60) and 12 of the Young-ESLs (mean age: 21.19 years; *SD*: 1.57) were tested using a compact target speech source (T_C); and of the other 12 Old-EFLs (mean age: 72.75 years; *SD*: 4.18) and 12 Young-ESLs (mean age: 21.02 years; *SD*: 1.95) were tested using a diffused target speech source (T_d). Listeners in the Old-EFL group were all born and raised in a country in which the primary language was English and were not fluent in any other language at the time of participation. Listeners in the Young-ESL were born and raised in a language other than English and did not attend an English or an American school before relocating to an English-speaking country at the age of 11 years old or later. The Young-ESL listeners were from a diverse linguistic background (1 Hindi, 1 Philipino, 1 Spanish, 1 Sinhalese, 1 Macedonian, 1 Indonesian, 1 Korean, 1 Russian, 4 Arabic, 2 Portuguese, 1 Malayalam, 1 Cantonese, 8 Mandarin). Their average age at the time of the relocation was 16.21 years (*SD* = 3.15). Participants were recruited from the University of Toronto Mississauga's Human Communication Lab database system. The database consists of younger adults who are students at the University of Toronto Mississauga and older adults who were individuals living independently in the community from the surrounding area (Mississauga, ON), who provided their own means of transportation to the laboratory. All participants completed a questionnaire regarding their general health, hearing, vision, and cognitive status. Only participants who reported that they were in good health and had no history of serious pathology were included. Participants had normal hearing for their age and no history of hearing disorders or previous use of hearing aids. The study reported here was approved by the Ethics Review Board of the University of Toronto.

Materials, Apparatus, and Procedure

All participants completed an Audiometric hearing test, the Nelson-Denny reading comprehension test (Brown et al., 1981), and the Mill Hill vocabulary test (Raven, 1965) during the first experimental session. The speech recognition task was administered during a second experimental session. Each of the two sessions was typically 1–1.5 h in duration. All participants provided their written informed consent to participate and were compensated monetarily for their participation.

Hearing Measures

Audiometric Testing

Pure-tone air-conduction thresholds were measured at nine frequencies (0.25–8 kHz) for both ears using an Interacoustics Model AC5 audiometer (Interacoustic, Assens, Denmark). All Young-ESL participants were required to have a pure tone threshold of 15 dB HL or lower from 0.25 to 8 kHz but were allowed to have one 20 dB HL threshold in one tested frequency in each ear. All Old-EFL participants were required to have a pure tone threshold of 25 dB HL or lower from 0.25 to 3 kHz. Older adults with hearing thresholds in the range described are usually considered to have normal hearing for their age (ISO 7029-2000). In addition, participants who demonstrated unbalanced hearing (more than 15 dB difference between ears at any tested frequency between 0.25 to 8 kHz) were excluded from participation. **Figure 1** plots the average audiometric thresholds for the left and right ears of the Old-EFLs and Young-ESLs in the present study along with the Young-EFLs in Avivi-Reich et al. (2020), separately for the two target groups (T_C vs. T_d).

Language Proficiency Measures

Vocabulary Knowledge

Participants were asked to complete the Mill Hill vocabulary test (Raven, 1965), which is a 20-item synonym test. In this task, participants were required to choose the closest synonym of each test item from a list of six alternatives. No time restraints were applied.

Reading Comprehension Skill

The Nelson-Denny test (Brown et al., 1981) was used to assess the reading comprehension skills of each participant. In this test, the participants had to read through eight independent passages and answer multiple-choice questions based on the content of the passages. This test includes a total of 36 questions and was limited to 20 min. Participants were instructed to answer as many questions as possible within the allotted time.

Semantically Anomalous Sentences-Recognition Task

The procedure for the sentence-recognition task was replicated from Avivi-Reich et al. (2020). In the experimental recognition task, listeners sat in a chair placed in the center of an Industrial Acoustic Company (IAC) sound-attenuated chamber. The internal dimensions of this chamber were 283 cm in length, 274 cm in width, and 197 cm in height. As described in Avivi-Reich et al. (2020), two loudspeakers were placed at 45° to the left and right of the listener, with a third placed directly in front of the listener. The distance between the center of the listener's head and each of the three loudspeakers was about 170 cm. The height of each loudspeaker was adjusted to match the ear level of a seated listener with an average body height. The acoustic stimuli for the present study were the same as those presented in Avivi-Reich et al. (2020), however the Signal to Noise Ratios (SNRs) used were adjusted to accommodate for age-related or language-related changes in speech recognition.

The target sentences used in the present study were the same as those reported in Avivi-Reich et al. (2020). Target sentences were 312 syntactically-correct-but-semantically-anomalous sentences spoken by a female talker and developed by Helfer (1997). Each sentence contained three target words in sentence frames such as "A *spider* will *drain* a *fork*," or "A *shop* can *frame* a *dog*" (target words italicized). The sentences were divided into 24 lists each comprising of 13 sentences. During the Compact-Target conditions, target sentences were presented over the front loudspeaker while the masker was either presented over all three loudspeakers to create a diffused image, or over the central loudspeaker only to create a compact image of the sound source. During the Diffuse-Target conditions, the target sentences were presented over all three loudspeakers to create a diffused target image while the masker was either presented from all three loudspeakers to create a diffused image, or over the central loudspeaker only to create a compact image of the masking sound source.

Target sentences were presented in one of three masking stimuli (Noise, Babble, Speech), as described in Avivi-Reich et al. (2020). The Noise masker was a steady-state speech-spectrum noise recorded from an audiometer (Interacoustic [Assens, Denmark] model AC5). The Babble was a 12-talker babble taken from the modified SPIN test (Bilger et al., 1984). The Speech masker was created using an additional set of semantically anomalous sentences spoken by two female talkers (315-s-long track presented in a continuous loop). The target sentences were presented at an average sound pressure of 55 dBA at the estimated center of a listener's head. The sound pressure level of the maskers was adjusted in order to produce 4 different SNRs depending on the listener Group, Masker Type, and the Timbre Condition tested. The sound pressure was measured using a Brüel and Kjær (Copenhagen, Denmark) KEMAR dummy-head to ensure that the voltages of the sounds presented in the three loudspeaker conditions were adjusted appropriately so that the sound pressure recorded at the KEMAR head in the three-loudspeaker conditions matched the sound pressure recorded at the KEMAR head in the single loudspeaker conditions. In addition, the sound level calibrations were confirmed using a Bruel and Kjaer sound level meter (Model 2260) at the location corresponding to the approximate center of a participant's head. However, these rigorous measuring procedures do not eliminate certain comb filtering effects which will be further addressed when discussing the results (for more details concerning comb filtering effects in these conditions, see Avivi-Reich et al., 2020).

The different SNRs used were initially chosen based on previous studies that used similar stimuli in noise (e.g., Avivi-Reich et al., 2018) and then altered according to the results of two rounds of preliminary pilot testing conducted under the present listening conditions. The SNRs used in the current study are presented in **Table 1**. A single list of 13 sentences was used for each of the SNR values that appear in the table.

Trials were blocked according to lists. All sentences in a list were presented at a constant SNR. In the two experimental groups

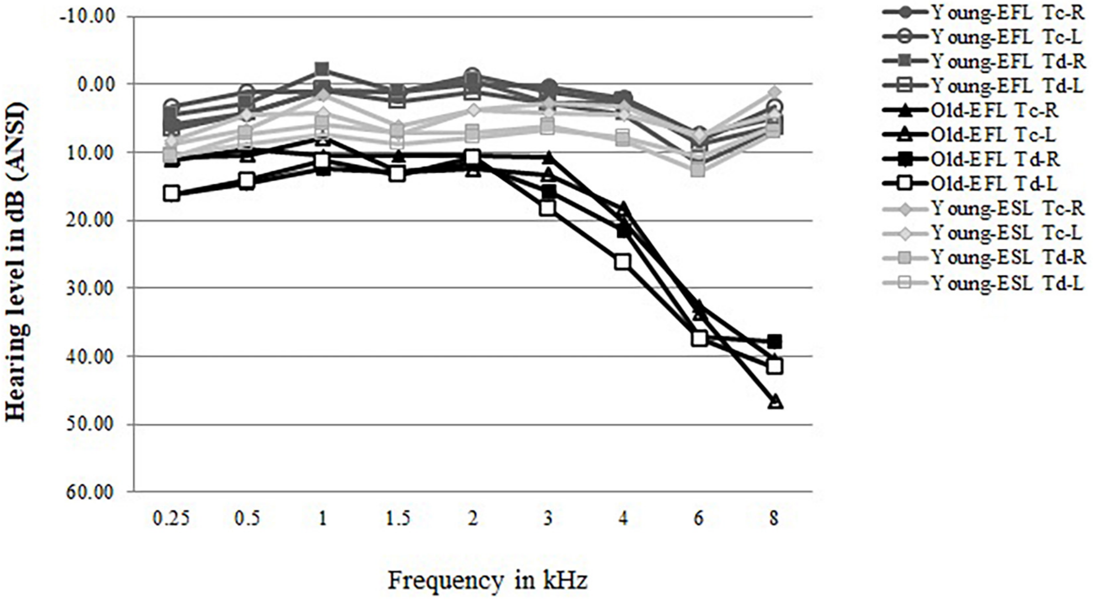


FIGURE 1 | Average audiograms for the two Old-EFL groups (T_C vs. T_d) and the two Young-ESL groups (T_d vs. T_C) as well as for the equivalent two Young-EFL groups from Avivi-Reich et al. (2020). Left and right ears are plotted separately.

(T_C, T_d), six participants were tested with a diffused masker (M_d) for the first 12 lists, and then with a compact masker (M_C) for the remaining 12. The reverse order was applied for the other six participants. Sentence lists and SNRs were counterbalanced across participants such that each list was presented at each of the 4 different SNRs an equal number of times within each group. Moreover, each list was presented in each of the four Timbre Conditions (T_CM_C, T_CM_d, T_dM_d, T_dM_C) and three Masker (Speech, Babble, Noise) combinations an equal number of times.

Before starting the experimental session, participants were given a brief explanation to become familiarized with the task. Participants were asked to repeat back the target sentence after each presentation and were scored for the correct repetition of any keyword. Performance was assessed in real-time while the session was taking place, and later by a second research assistant who listened to the participant's recorded responses. If there was a disagreement between the online assessment and the second listener's coding of the sentences, the two raters listened to the recording together, until they arrived at a consensus opinion. After each response by the participant, the researcher began the next presentation of the trial. Each trial began with the masker sound which was followed 1 s later by the target sentence. The masker remained on during the presentation of the target sentence, then the masker was turned off when the target sentence ended. After completing 12 lists, a short break was offered to the participants.

TABLE 1 | The values of the four Signal to Noise Ratios (SNRs) used under each condition: (1) compact targets and maskers (T_CM_C), 2) compact targets and diffuse maskers (T_CM_d), 3) diffuse targets and maskers (T_dM_d), and 4) diffuse targets and compact maskers (T_dM_C), for each of the three masker types (S, Speech; N, Noise; B, Babble), presented separately for the two experimental groups of the Young-ESL and Old-EFL participants.

Old-EFL						Young-ESL					
TcMc			TcMd			TcMc			TcMd		
S	N	B	S	N	B	S	N	B	S	N	B
10	8	−3	3	2	−10	11	6	−3	5	5	−7
4	3	−9	−3	−3	−16	5	1	−9	0	−1	−13
−2	−2	−15	−9	−8	−22	−1	−4	−15	−5	−7	−19
−8	−7	−21	−15	−13	−28	−7	−9	−21	−10	−13	−25
TdMc			TdMd			TdMc			TcMd		
S	N	B	S	N	B	S	N	B	S	N	B
14	11	4	10	8	1	11	9	−2	11	6	−3
8	6	−2	4	3	−5	5	4	−8	5	1	−9
2	1	−8	−2	−2	−11	−1	−1	−14	−1	−4	−15
−4	−4	−14	−8	−7	−17	−7	−6	−20	−7	−9	−21

RESULTS

Demographic Data

Table 2 presents the gender breakdown, mean age, Mill Hill test of vocabulary knowledge and Nelson-Denny test of reading comprehension results for the young English as first language Young-EFL participants (Young-EFL) in Avivi-Reich et al. (2020), and the older English as first language participants (Old-EFL), and the young English as a second language (Young-ESL) participants in this experiment. An Age Group (Young-Old) by Language Status (EFL-ESL) by Target Timbre Between-Subjects

TABLE 2 | The gender breakdown, mean age, Mill Hill vocabulary test and Nelson-Denny reading comprehension test results for the Young-EFL (taken from Avivi-Reich et al., 2020), and for the Old-EFL and the Young-ESL participants in this experiment. SE stands for Standard Error.

Group	Gender	Age in years	Mill Hill vocabulary	Nelson-Denny reading
Young EFLs compact target	4 Male 8 Females	Mean = 21.78 SE = 0.61	Mean = 14.50 SE = 0.36	Mean = 28.33 SE = 1.15
Young ESLs compact target	3 Males 9 Females	Mean = 21.19 SE = 0.45	Mean = 9.25 SE = 1.16	Mean = 18.08 SE = 1.77
Young EFLs diffuse target	1 Male 11 Females	Mean = 20.14 SE = 0.51	Mean = 13.00 SE = 0.77	Mean = 25.83 SE = 1.71
Young ESLs diffuse target	3 Males 9 Females	Mean = 21.02 SE = 0.56	Mean = 10.00 SE = 0.72	Mean = 20.67 SE = 1.77
Old EFLs compact target	1 Male 11 Females	Mean = 72.76 SE = 1.31	Mean = 15.45 SE = 0.68	Mean = 23.83 SE = 1.80
Old EFLs diffuse target	3 Males 9 Females	Mean = 72.75 SE = 1.21	Mean = 14.92 SE = 0.87	Mean = 22.67 SE = 1.65

ANOVA found a significant age difference between the younger and older groups [$F(1, 66) = 3{,}723.$, $p < 0.001$]. There were no differences in age between the EFL and ESL groups, and those participants in the Compact Target group and Diffuse Target Group. In addition, none of the interactions were significant (all F-values < 1).

An Age Group (Young-Old) by Language Status (EFL-ESL) by Target Timbre Between-subjects ANOVA on Mill Hill vocabulary scores found a highly significant effect of language status [EFLs had higher vocabulary scores than ESLs: $F(1, 66) = 26.905$, $p < 0.001$], and a nearly significant effect of Age-Group [$F(1, 66) = 3.258$, $p = 0.076$] where older adults had higher vocabulary scores than younger adults. The effect of Target Timbre failed to reach significance [$F(1, 66) < 1$], and there was no evidence of an interaction between Language Status and Target Timbre [$F(1, 66) = 2.001$, $p = 0.162$] and no evidence of an interaction between Age Group and Target Timbre [$F(1, 66) < 1$].

An Age Group (Young-Old) by Language Status (EFL-ESL) by Target Timbre Between-subjects ANOVA on Nelson Denny reading scores found a highly significant effect of language status [Young EFLs had better reading comprehension scores than Young ESLs: $F(1, 66) = 21.664$, $p < 0.001$], and a significant effect of Age-Group [$F(1, 66) = 5.358$, $p = 0.024$] where younger adults had higher reading scores than older adults. The effect of Target Timbre failed to reach significance [$F(1, 66) < 1$], and there was no evidence of an interaction between Age Group and Target Timbre [$F(1, 66) < 1$] or of an interaction between Language Status and Target Timbre [$F(1, 66) = 2.355$, $p = 0.130$].

Psychometric Functions

Figure 2 (Top Portion) shows the percentage of correctly identified keywords for the 24 young participants whose first language was English (Young-EFLs) as a function of SNR when the masker was speech spectrum noise (left panel), two-talker speech (center panel) or 12-talker babble (right panel). Twelve of these participants were presented with compact targets (Tc) only, while the other 12 participants were presented only with diffuse targets (T_d) (These data were adapted from Avivi-Reich et al., 2020). Psychometric functions are plotted separately for instances in which there is no contrast in timbre between the target and masker ($T_C M_C$ and $T_d M_d$), and those in which there is a timbre contrast between the target and masker ($T_C M_d$ and $T_d M_C$). Circles represent the data for the compact target (T_C) group with squares representing the data for the diffuse target (T_d) group. Logistic psychometric functions of the form

$$y = \frac{100 * a}{1 + e^{-\sigma(x-\mu)}} \tag{1}$$

were fit to these data points, where the parameter a is restricted to the range from 0 to 1, and $100^{*}a$ specifies the asymptotic value reached by the percent correct word recognition as the SNR, x, approaches infinity (i.e., when listening in quiet). The parameter μ denotes the value of x such that the percent correct word recognition reaches ½ of its asymptotic value, and σ controls the slope of the function (for a description of the fitting procedure see **Supplementary Appendix 1**). The 50% points on these fitted psychometric functions are indicated by the dashed vertical lines when the target speech was compact (T_C), and solid vertical lines for when the target speech was diffuse (T_d) and were used as estimates of the speech recognition threshold for that condition.

The center portion of **Figure 2** plots the equivalent data from the 24 older participants whose first language was English (Old-EFL), while the bottom portion shows the results from the 24 participants for whom English was a second language (Young-ESL). For all three groups, when there is no timbre contrast between target and masker ($T_C M_C$ or $T_d M_d$), the psychometric functions appear to be equivalent, independent of whether the target was compact (solid circles) or diffuse (solid squares). However, when there is a contrast in timbre between target and masker ($T_C M_d$ or $T_d M_C$), the psychometric functions for the conditions in which the target is diffuse (filled squares) are shifted to the right with respect to conditions in which target is compact (filled circles) in all three groups. There are, however, indications that Target Timbre, Masker Type and Language Status affects the 50% thresholds of the psychometric functions, as well as their slopes. First, **Figure 2** shows that thresholds are lowest for the Young-EFL group when compared to the other two groups. Second, when there is a timbre contrast between target and masker, the degree of separation between the psychometric functions for the $T_C M_d$ and the $T_d M_C$ conditions appears to depend on both their Linguistic Group, and the type of Masker (Noise, Babble, or Speech). It should also be noted that when there is no timbre contrast between target and masker, the effect of the signal-to-masker ratio appears to be the same independent of whether the target is compact or diffuse, as long as the masker timbre is the same as the target timbre.

To confirm these visual impressions, statistical analyses were conducted on individual participants with respect to the three parameters of the psychometric function. Specifically, psychometric functions were fit to all individuals in order to obtain individual estimates of the threshold, μ, the slope parameter, σ, and the asymptotic value (*a*) of the psychometric functions. We then conducted a 3 Group (Young-EFLs, Old-EFLs, Young-ESLs) × 2 Target Timbres (T_C vs. T_d) × 3 Masker Types (Noise, Babble, Speech) × 2 Masker Timbre conditions (M_C vs. M_d) ANCOVA with Participant Group, and Target Timber as between-subjects factors and Masker Type and Masker Timbre as within-subject factors, with vocabulary and reading comprehension as covariates, for thresholds and slopes, following the procedure recommended by Schneider et al. (2015). The results of this analysis of variance are shown in **Supplementary Table 1**. All four main effects (Masker Type, Masker Timbre, Target Timbre, and Group were highly significant ($p < 0.001$, for the main effects of all four factors). There were also 3 three-way interactions that were significant (MaskerType × TargetTimbre × Group, $p = 0.001$; MaskerType × MaskerTimbre × TargetTimbre, $p = 0.002$; MaskerType × MaskerTimbre × Group, $p = 0.01$), and 1 two-way interaction (MaskerType × Group, $p = 0.005$). None of the other interaction effects were significant. In addition, there was no evidence that the two covariates affected performance. Hence, none of the subsequent analyses involved the covariate measures.

Because **Figure 2** suggests that Target Timbre has a negligible effect on thresholds when the timbre of the target matches the timbre of the masker, we conducted two additional analyses to determine the sources of the interaction effects found in the omnibus ANOVA. First, we conducted a three Group (Young-EFLs, Old-EFLs, Young-ESLs) × two-target timbres (T_C & T_d) × three Masker Types (Noise, Speech, and Babble) ANOVA only for the conditions in which the timbre of the masker matched that of the target, with Group and Target Timbre as between-subjects factors, and Masker Type as a within-subject factor. **Supplementary Table 2** shows that when the target's timbre matches that of the masker, none of the effects involving the target's timbre are significant. Hence, the source of any of the interaction effects involving the target's timbre in the omnibus ANOVA are restricted to conditions in which there is a mismatch between the target's timbre and the masker's timbre.

A comparable analysis (see **Supplementary Table 3**) limited to when there was a mismatch between the target's timbre and the masker's timbre, however, found a significant three-way interaction between Target Timbre, Masker Timbre, and Group ($p < 0.001$). To identify the source of this three-way interaction, **Figure 3** plots how the thresholds for both T_CM_d and T_dM_C conditions change as a function of Group, separately for the Noise, Speech and Babble Maskers. Also shown are the average thresholds for the two conditions in which the target

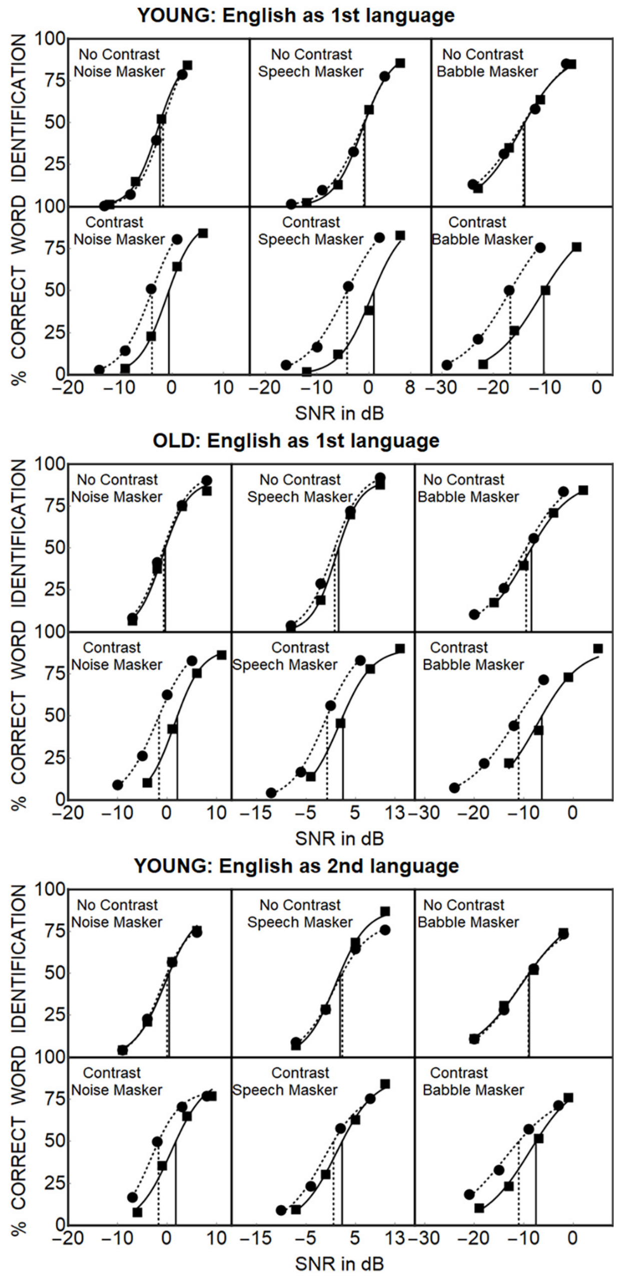


FIGURE 2 | Top. Young-EFL participants from Avivi-Reich et al. (2020). **Center.** Old-EFL participants, this experiment. **Bottom.** Young-ESL participants, this experiment. The percentage of words correctly identified is plotted as a function of SNR for each combination of the three maskers (Noise, Speech, and Babble) with the four target-masker combinations (TcMc, TdMd, TcMd, TdMc) for the three different groups of participants. The data for the No Contrast Conditions (TcMc and TdMd) are shown separately from the data when there is a timbre contrast between target and masker (TcMd and TdMc). Circles represent the average data for the Conditions where the target was compact (Tc), squares represent the data for diffuse targets (Td). Solid lines show the psychometric functions fit to that data when the target was diffuse (Td); dotted lines show the psychometric functions fit to the data when the target was compact (Tc). The SNRs corresponding to 50% correct identification when the target is compact are indicated by the dotted vertical lines. The solid vertical lines indicate the corresponding SNRs when the target is diffuse.

timbre matched the masker timbre (average of T_CM_C and T_dM_d thresholds). This figure indicates that for Noise maskers the separation between the T_CM_d and T_dM_C thresholds remains constant across the three Groups. However, for Speech and Babble Maskers, the advantage held by compact targets is severely diminished in the Young-ESL group compared to the Young-EFL group. Subsequent analyses in **Supplementary Appendix 2** shows that if the Young-ESL group is excluded from the analysis, there is no indication of an interaction between the two remaining EFL groups (Young-EFLs and Old-EFLs) and target timbre. However, when considering only young adults, there is a highly significant interaction between their linguistic status (EFL vs. ESL) and target timbre, highlighting the importance of the language status of people in a complex acoustic environment. An examination of **Figure 3** suggests, for young-ESL adults in both Babble and Speech Maskers, that the thresholds were essentially equivalent, for all combinations of target and masker timbre. Pairwise comparisons of the young-ESL thresholds among the four combinations of target and masker (T_CM_C, T_CM_d, T_dM_C, T_dM_d) failed to find any significant differences in threshold values when the masker was Babble for a Type 1 error of 0.05 (after applying a Bonferroni correction for the six comparisons). For the equivalent comparisons of Young-ESL thresholds in Speech, only one of the comparisons was significant (T_CM_C vs. T_CM_d). However, the difference in threshold between these two timbre conditions in the Young-ESL listeners (1.8 dB) was much smaller than the difference in the same two timbre conditions for the Young-EFL listeners (3.1 dB).

To determine the source of the two-way interaction in the omnibus ANOVA between Group and Masker Type when there is a mismatch between Target Timbre and Masker Timbre, in **Figure 4**, we plotted, for each of the Masker Types, the average thresholds for each of the Groups.

In **Figure 4**, the difference between Noise thresholds and Babble thresholds appears to be larger for Young-EFLs (11.6 dB) than it is for either Old-EFLs (9.6 dB) or Young-ESLs (10.6 dB). Similarly, the difference between Speech thresholds and Babble thresholds appears to be larger for Young-EFLs (12.4 dB) than it is for either Old-EFLs (10.7 dB) or Young-ESLs (11.8 dB). To confirm that the interaction between Masker Type and Group is due to the larger separation in the Young-EFL group between Noise and Babble, and between Speech and Babble than the comparable comparisons in the other two Groups, a separate ANOVA was conducted that excluded the Babble Masker condition. When the Babble Masking condition was excluded, there was no evidence of an interaction between Group and Masker Condition [$F(2, 66) = 0.270$, $p > 0.5$]. Hence, the two-way interaction between Masker Type and Group appears to be due to the very low threshold in Babble that is found in the Young-EFL participants.

Slopes of the Psychometric Functions

We also conducted an ANOVA on the slopes of the individual psychometric functions with Target Timbre and Group as between-subjects factors and Masker Type and Masker Timbre as within-subject factors. The only factor that significantly affected the slopes of the psychometric functions was the Masker Type [$F(2, 132) = 8.711$, $p < 0.001$]. As **Figure 2** suggests the slopes for Speech (Mean = 0.49) and for Noise (Mean = 0.41) are greater than those for Babble (Mean = 0.23). Pairwise T-test indicate that the difference in slopes between Noise and Speech were not significant [$T(71) = -1.08$, $p = 0.284$], but the differences in slopes between Noise and Babble [$T(71) = 8.87$, $p < 0.0001$], and Speech and Babble [$T(71) = 3.21$, $p = 0.002$] were significant (for more information see **Supplementary Table 4**).

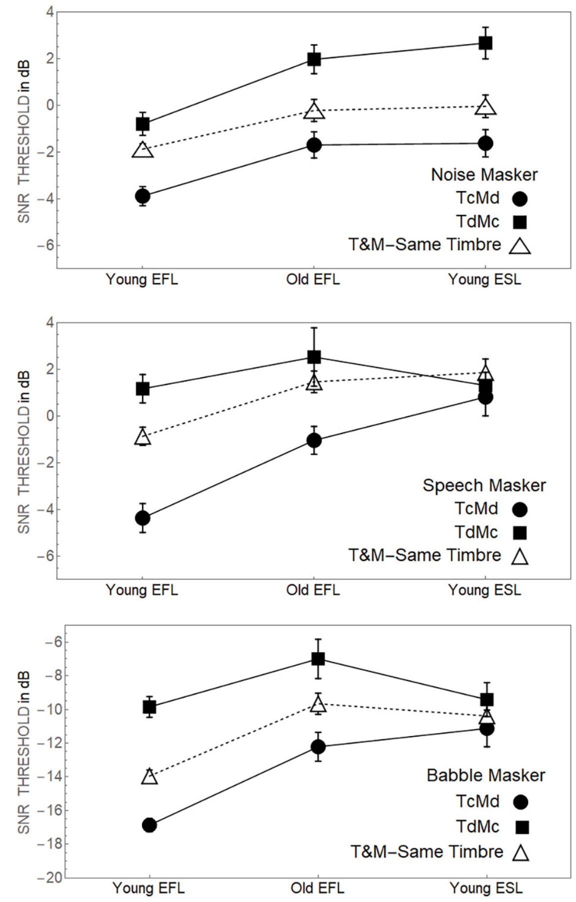


FIGURE 3 | Top panel. Fifty percent thresholds when the masker was speech spectrum noise for the three groups listening to sentences where the target was either compact or diffuse with the masker having the opposite timbre. **Middle Panel.** The equivalent data when the participants were listening to the sentences when the background is competing speech. **Bottom Panel.** The equivalent data when the participants were listening to the sentences when the background is babble. Standard error bars are shown.

Asymptotes of the Psychometric Functions

The mean asymptote (a) of the psychometric functions for the three linguistic groups were: (1) Young-EFLs (0.94); (2) Old-EFLs (0.92); and (3) Young-ESLs (0.84). A T-test of the difference between Young-EFL and Old-EFL asymptotes was not significant [$T(46) = -1.17$, $p = 0.25$]. A T-test of the difference between Young-EFL and Young-ESL asymptotes was significant

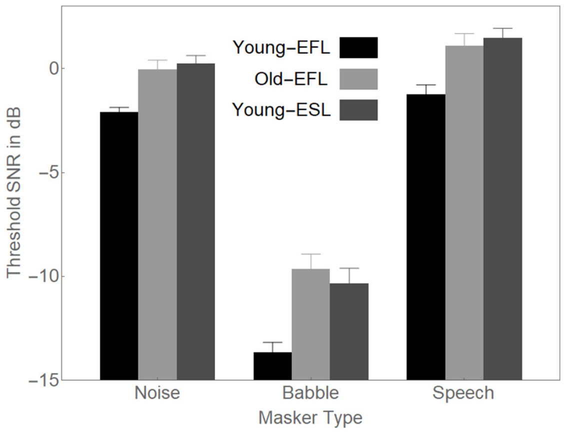


FIGURE 4 | Average 50% thresholds in Noise, Babble, and Speech for the three groups of participants. Error bars are standard errors of the means.

[$T(46) = -4.33$, $p < 0.0001$], as was the difference between Old-EFL and Young-ESL asymptotes [$T(46) = -3.14$, $p = 0.003$]. Hence, asymptotes for young and old native listeners were comparable, but both of these groups had significantly higher asymptotic values than did the Young-ESL group.

DISCUSSION

In the current study, three different masker types were used (Noise, Babble and Speech) to test the effect of sound source diffuseness on speech recognition in Young-ESL and Old-EFL listeners and compare their performance to that of the Young-EFL listeners previously tested (see Avivi-Reich et al., 2020). The results showed that for all three groups, when there is no timbre contrast between target and masker (T_CM_C or T_dM_d), the psychometric functions appear to be equivalent, independent of whether the target was compact or diffuse. In other words, the Target Timbre has a negligible effect on thresholds when the timbre of the target matches the timbre of the masker (T_CM_C or T_dM_d). These findings are similar to what was previously found in Young-EFL listeners (Avivi-Reich et al., 2020). However, when there is a contrast in timbre between target and masker (T_CM_d or T_dM_C), a significant separation between the T_CM_d and T_dM_C thresholds is evident in all three groups when the masker is Noise. Interestingly, for Speech and Babble Maskers, the advantage held by compact targets is severely diminished in the Young-ESL group compared to two EFL groups (young and old). Indeed thresholds for all four conditions (T_CM_C, T_CM_d, T_dM_d, T_d,M_C) appear to be quite similar (see **Figure 3**). This would suggest, that, in the presence of informational masking, Young-ESLs are unable to use timbre differences to attend to and process the target speech. These results indicate that listeners, whose linguistic status differs, respond to timbre differences differently depending on masker type. Young-EFLs and Old-EFLs appear to derive equivalent benefits from timbre differences between targets and maskers. Thus, it seems that while Old-EFLs generally need more favorable SNRs compared to Young-EFLs to correctly recognize speech in the presence of competing sounds, the different diffuseness levels between targets and maskers seem to affect both EFL age groups similarly.

In addition, a two-way interaction between Masker Type and Group was found, which appears to be due to the larger separation between Noise and Babble and between Speech and Babble thresholds in the Young-EFL group than in the other two Groups. In other words, the Young-EFL listeners, who overall had better (lower) speech recognition thresholds compared with the other two groups, did exceptionally better when the masker was Babble. Hence, when there is a babble of indistinguishable voices, Young-EFL listeners have exceptionally low thresholds compared to either Old-EFL listeners or young-ESL listeners.

Two possible reasons were previously suggested and discussed (Avivi-Reich et al., 2020) as to why listeners may find auditory scenes in which the target is compact and the masker is diffused more favorable than when there is no such timbre contrast between the sound sources, while they seem to find the opposite configuration (Target is diffuse and Masker is compact) less favorable than listening in an auditory scene with no timbre contrast. The first is that compact sound sources with a precise location may attract the listener's attention, giving the compact sound source a certain advantage, which could either serve speech recognition when the speech sound is compact, or potentially increase the interference when the irrelevant competing sound is the compact one. The second possible explanation is that the pattern of results found is consistent with what would be expected when taking into consideration the comb-filtering effects that occur when a sound source is played over multiple loudspeakers vs. when it is played over a single loudspeaker only. When the same sound is played over spatially separated loudspeakers, it will arrive at the ear of the listener at slightly different times. These delays result in some frequencies being enhanced, while others are canceled, producing peaks and troughs in the sound spectrum at the ears. Hence, when the masker is diffuse, there will be peaks and troughs in the spectrum of masker. If the listener can attend to and integrate the information in the speech target falling into the troughs of the masker, we might expect to find lower thresholds when the masker is diffuse and the target is compact. For a fuller explanation (see Avivi-Reich et al., 2020).

With these two possible explanations in mind, we would like to address the primary question raised by the current findings. First, why would all three groups (Young-EFLs, Old-EFLs, Young-ESLs) in the Noise condition, have lowest thresholds when the target is compact and the masker is diffuse (T_CM_d) and highest thresholds when the target is diffuse and the masker compact (T_dM_C) with the T_dM_d and T_CM_C conditions falling midway between the two? Second, why do the Young-EFL and Old-EFL listeners show this same pattern when the Masker is Babble or Speech, but not the young-ESL listeners, who perform equivalently in all four timbre conditions? To answer these questions, we will need to consider the ways in which the Noise masker is different than Babble and Speech, as well as the differences between EFL-listeners and ESL-listeners.

Noise, Babble and Speech maskers are all expected to cause interference resulting in a greater difficulty to recognize speech. However, the level of processing at which this interference occurs is likely to differ among masker types. All three masker types used in the current study (Noise, Babble, Speech) activated regions along the basilar membrane that undoubtedly overlap with those activated by the target speech. Such overlap energetically interferes with the encoding of the target speech signal causing peripheral or energetic masking (Pollack, 1975). When the masker used was speech from one or more talkers (Speech or Babble), it likely also interfered with the linguistic and semantic processing of the target speech causing informational masking as well as energetic masking (for a review, see Durlach et al., 2003; Freyman et al., 2004; Schneider et al., 2007, 2010; Kidd et al., 2008). Mattys et al. (2009) divided informational masking interference into three categories: (1) The effects of the masker competing for attention including the cost of inhibiting information coming from the competing speech; (2) interference from a known language when the masker itself is intelligible and meaningful, thereby leading to lexical-semantic interference; (3) additional cognitive load associated with the processing resources required when listeners need to divide their attention between the target and the masker. The three types of maskers used in the current study differ in the levels of energetic and informational making they cause. While the Noise masker generates relatively consistent energetic masking across a wide range of frequencies, it contains no verbal information and therefore is not expected to generate informational masking. Babble and Speech, however, lead to intensity fluctuations over time creating energetic peaks and troughs. In addition, it is reasonable to expect that due to the greater resemblance between the target speech and a speech masker (Speech or Babble), compared to that found between the target speech and a noise masker, stream segregation will be more difficult to obtain when the masker is speech or babble.

Several speech perception studies have included different types of maskers in order to study the effect type of masker may have on the extent to which listeners experience release from masking when provided with an assisting cue that could enhance speech perception (e.g., Freyman et al., 2004; Ezzatian et al., 2010; Mattys et al., 2010; Avivi-Reich et al., 2018). Their findings have shown that the amount of release provided by a particular manipulation differed depending on the type of masker that was presented. Interestingly, in several previous studies that examined spatial cues (such as location and spatial separation cues), the release from masking generally increases with the informational content of the masker (e.g., Arbogast et al., 2002; Ezzatian et al., 2010). For example, Ezzatian et al. (2010) asked young-EFL and young-ESL listeners to repeat sentences that were presented to them in the presence of either Noise, Babble or competing Speech, when the target and masker were co-located vs. when there was spatial separation between the two. Their results showed that the amount of release from masking due to spatial separation is larger when the masker is speech rather than noise. In addition, young-EFL and young-ESL listeners benefited equally from perceived spatial separation. This pattern of results resembles what was found for the Young-EFL listeners in the previous experiment, but somewhat contradicts the pattern found in the Young-ESL listeners.

Figure 3 suggests that for Young-ESL participants listening in the presence of a Babble or a Speech masker, thresholds for target speech recognition appear to be independent of the timbres of the target speech and the masker, and depend solely on the SNR (the one exception is the $T_C M_C$ vs. $T_C M_d$ comparison for the Speech Masker). We might expect such a result if the Young-ESL listeners were unable to take advantage of differences in timbre between target and masker. If that were the case, then thresholds would depend solely on the ratio of speech energy to masker energy.

Why might this be the case? The results from the conditions where the masker was Noise clearly indicates that speech recognition is sensitive to timbre differences between the target speech and masker for Young-ESL listeners. Hence, they can use these cues in some difficult listening situations. If that is the case, why do they not use these cues when the masker is Babble or Speech? One possibility is that in order to benefit from timbre differences, the listener has to allocate attentional resources to basic auditory processes in order to extract a benefit from timbre differences. In a previous paper, we pointed out that a diffuse masker produces troughs in the spectrum of the masker. If the listener is able to focus attentional resources in the frequency regions corresponding to the troughs and integrate the information from these troughs to extract the speech signal (Scharf et al., 1987), then we would expect lower speech recognition thresholds when the target is compact, and the masker is diffuse. The Young-ESL listeners can clearly do this when the masker is Noise, but not when the masker is Babble or Speech.

The reason for this difference may reside in the additional attentional resources that need to be deployed by second language listeners when the masker is either babble or speech. Second language listeners are found to have lower performance than listeners listening to their first language on a number of auditory speech-perception measures (Mayo et al., 1997; Bradlow and Pisoni, 1999; Meador et al., 2000; Bradlow and Bent, 2002; Cooke et al., 2008; Rogers and Lopez, 2008; Ezzatian et al., 2010; Avivi-Reich et al., 2014, 2015). Second language listeners tend to experience interference from their first language knowledge when listening to speech in their second language (Nábělek and Donahue, 1984; Bradlow and Pisoni, 1999; Cutler, 2001). The speech perception differences found between first and second language listeners could be due, in part, to incomplete acquisition of the acoustic–phonetic characteristics of the second language (e.g., Florentine, 1985; Mayo et al., 1997), which might lead to a reduced ability to correctly recognize the phonemes in one's second or third language (Bradlow and Pisoni, 1999; Meador et al., 2000). In addition, in second language listeners the semantic and linguistic processes in their second language may not be completely differentiated from those in their first (Kroll and Steward, 1994). Thus, this cross-linguistic interference could be a result of phonetic, phonemic and or phonotactic knowledge transfers (e.g., Polka, 1991, 1992). When both the target and the masker contain speech in their second language, second language listeners might find speech recognition to be especially difficult. The overlap between the two linguistic systems could result in greater competition as both systems are activated by more than a single incoming verbal stream. Hence, the degree and extent to which second language listeners must engage attentional and knowledge-driven processes (e.g., vocabulary and linguistic knowledge) to facilitate speech perception could differ from the pattern of engagement in first language listeners. This additional load may leave them with inadequate attentional resources to focus attention on particular regions along the basilar membrane.

If indeed the cause for the interaction found between the listeners' linguistic status and the effect of timbre contrast on speech recognition is due to greater draw on scarce attentional resources, it is reasonable to assume those could be captured by listening effort measurements. Thus, it is recommended that future studies use listening effort measures, such as pupilometry or dual-task, to further examine speech perception and the connection between linguistic experience and listening effort under different timbre conditions. The relationship between resource demand and listening-effort has been established by numerous studies (e.g., Koelewijn et al., 2012; Zekveld et al., 2014; Pichora-Fuller et al., 2016; Gagné et al., 2017; Tangkhpanya et al., 2019), incorporating a measure of effort would allow us to better understand the difficulties listeners might experience when listening to their second language in complex and acoustically amplified listening environment and contribute to the development of more accommodating sound amplification.

Why then are the Young-ESL listeners able to benefit as much from spatial separation as Young-EFL listeners? The reason might be that locating the azimuth positions of auditory objects is an automatic process, one that does not require attentional resources. The binaural system is exquisitely sensitive to time of arrival differences of a sound to the two ears, as well as differences in intensity. Time of arrival differences are coded at the level of the cochlear nucleus and are an intrinsic part of the auditory signal processed by higher-order brain structures. As such, they most likely do not require attentional resources to code and utilize these time of arrival differences. Timbre differences, however, most likely require attention to be focused on particular spectral areas. A number of studies have shown that when attention is focused on a particular region of the spectrum, the detection of a signal in that region is dramatically improved, suggesting that frequency-selective attention involves the operation of a "listening band," centered on the attended frequency (Scharf et al., 1987; Degerman et al., 2006; Riecke et al., 2017). Hence, if a listener could focus her or his attention on particular spectral regions, and integrate information across these regions, they could take advantage of the comb filtering provided by a diffuse masker. However, attentional selection has been characterized as a pool of attentional resources from which resources can be allocated to current tasks until the pool is exhausted (Kahneman, 1973; Lavie, 2005). Thus, if the attentional resources of the Young-ESL listeners were fully deployed at the lexical and semantic levels of processing, they might not have the resources to benefit from the increased signal-to-noise ratios that would be present in the troughs of the spectrum associated with a diffuse masker.

In summary, the results of the current study, which examines the effects of sound diffuseness levels on speech recognition in Young-ESL and Older-EFL listeners using three types of maskers (Noise, Babble, Speech) were compared to the results previously found in Young-EFLs. The comparison uncovered a significant difference in the timbre contrast effect found in the two EFL groups vs. the ESL group. While the two EFL groups demonstrated a benefit from such timbre contrast when the target was compact in the presence of all three masker types, the ESL group demonstrated improved speech recognition only when the diffused masker was Noise. A possible explanation as to why this three-way interaction was found statistically significant was suggested based on the listeners' linguistic experience, the interference caused by energetic vs. informational masking, and the explanations that were previously provided to explain the timbre contrast effects that were found (Avivi-Reich et al., 2020). The current study joins our previous study to form what we believe to be the only systematic investigation of sound diffuseness effect. The two studies together depict sound diffuseness level as an acoustic variable that could play a significant role in speech recognition, and its overall effect is dependent on variables such as the type of masker in which the target speech is presented and the linguistic experience of the listener. As the use of amplification becomes more common in both public and private listening environments, it is important to continue investigating the possible effects of using multiple loudspeakers on the speech perception of a variety of potential listeners.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Review Board of the University of Toronto. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

BS and MA-R conceived and planned the experiments and interpreted the results. MA-R and RKS carried out the experiments. BS took the lead in statistically analyzing the data. MA-R took the lead in writing the manuscript. All authors provided critical feedback and helped shape the research, analysis, and manuscript.

FUNDING

This research was supported by a grant from the Natural Sciences and Engineering Research Council of Canada (RGPIN 9952-13).

ACKNOWLEDGMENTS

We thank Jane Carey for assistance in recruiting participants and collecting data. We also would like to thank James Qi for his work in programming the experiments.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2022.838576/full#supplementary-material

REFERENCES

Arbogast, T. L., Mason, C. R., and Kidd, G. (2002). The effect of spatial separation on informational and energetic masking of speech. *J. Acous. Soc. Am.* 112, 2086–2098. doi: 10.1121/1.1510141

Avivi-Reich, M., Daneman, M., and Schneider, B. A. (2014). How age and linguistic competence alter the interplay of perceptual and cognitive factors when listening to conversations in a noisy environment. *Front. Syst. Neurosci.* 8:21. doi: 10.3389/fnsys.2014.00021

Avivi-Reich, M., Fifield, B., and Schneider, B. A. (2020). Can the diffuseness of sound sources in an auditory scene alter speech perception? *Atten. Percept. Psychophys.* 82, 1443–1458. doi: 10.3758/s13414-019-01808-2

Avivi-Reich, M., Jakubczyk, A., Daneman, M., and Schneider, B. A. (2015). How age, linguistic status, and the nature of the auditory scene alter the manner in which listening comprehension is achieved in multitalker conversations. *J. Speech Lang. Hear. Res.* 58, 1570–1591. doi: 10.1044/2015_JSLHR-H-14-0177

Avivi-Reich, M., Puka, K., and Schneider, B. A. (2018). Do age and linguistic background alter the audiovisual advantage when listening to speech in the presence of energetic and informational masking? *Atten. Percept. Psychophys.* 80, 242–261. doi: 10.3758/s13414-017-1423-5

Bednar, A., and Lalor, E. C. (2020). Where is the cocktail party? Decoding locations of attended and unattended moving sound sources using EEG. *NeuroImage* 205, 116283–116283. doi: 10.1016/j.neuroimage.2019.116283

Ben-David, B. M., Tse, V. Y., and Schneider, B. A. (2012). Does it take older adults longer than younger adults to perceptually segregate a speech target from a background masker? *Hear. Res.* 290, 55–63. doi: 10.1016/j.heares.2012.04.022

Bilger, R. C., Nuetzel, J. M., Rabinowitz, W. M., and Rzeczkowski, C. (1984). Standardization of a test of speech perception in noise. *J. Speech Hear. Res.* 27, 32–38. doi: 10.1044/jshr.2701.32

Bradlow, A. R., and Bent, T. (2002). The clear speech effect for non-native listeners. *J. Acoust. Soc. Am.* 112, 272–284. doi: 10.1121/1.1487837

Bradlow, A. R., and Pisoni, D. B. (1999). Recognition of spoken words by native and non-native listeners: talker-, listener-, and item-related factors. *J. Acoust. Soc. Am.* 106, 2074–2085. doi: 10.1121/1.427952

Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organization Of Sound*. Cambridge, MA: MIT Press.

Brown, J. I., Bennett, J. M., and Hanna, G. (1981). *The Nelson-Denny reading test*. Chicago, IL: Riverside.

Brungart, D. S., Simpson, B. D., Ericson, M. A., and Scott, K. R. (2001). Informational and energetic masking effects in the perception of multiple simultaneous talkers. *J. Acoust. Soc. Am.* 110, 2527–2538. doi: 10.1121/1.1408946

Canzoneri, E., Magosso, E., and Serino, A. (2012). Dynamic sounds capture the boundaries of peripersonal space representation in humans. *PLoS One* 7:e44306. doi: 10.1371/journal.pone.0044306

Cooke, M., Garcia Lecumberri, M. L., and Barker, J. (2008). The foreign language cocktail party problem: energetic and informational masking effects in non-native speech perception. *J. Acoust. Soc. Am.* 123, 414–427. doi: 10.1121/1.2804952

Cutler, A. (2001). Listening to a second language through the ears of a first. *Interpreting* 5, 1–23.

Degerman, A., Rinne, T., Salmi, J., Salonen, O., and Alho, K. (2006). Selective attention to sound location or pitch studied with fMRI. *Brain Res.* 1077, 123–134. doi: 10.1016/j.brainres.2006.01.025

Dos Santos Sequeira, S., Specht, K., Moosmann, M., Westerhausen, R., and Hugdahl, K. (2010). The effects of background noise on dichotic listening to consonant-vowel syllables: an fMRI study. *Laterality* 15, 577–596. doi: 10.1080/13576500903045082

Dubno, J. R., Horwitz, A. R., and Ahlstrom, J. B. (2003). Recovery from prior stimulation: masking of speech by interrupted noise for younger and older adults with normal hearing. *J. Acoust. Soc. Am.* 113, 2084–2094. doi: 10.1121/1.1555611

Durlach, N., Mason, C. R., Kidd, G., Arbogast, T. L., Colburn, H. S., and Shinn-Cunningham, B. G. (2003). Note on informational masking. *J. Acoust. Soc. Am.* 113, 2984–2987. doi: 10.1121/1.1570435

Ezzatian, P., Avivi, M., and Schneider, B. A. (2010). Do non-native listeners benefit as much as native listeners from spatial cues that release from speech masking? *Speech Commun.* 5, 919–929.

Farnè, A., and Làdavas, E. (2002). Auditory peripersonal space in humans. *J. Cogn. Neurosci.* 14, 1030–1043.

FitzPatrick, I., and Indefrey, P. (2009). Lexical competition in nonnative speech comprehension. *J. Cogn. Neurosci.* 22, 1165–1178. doi: 10.1162/jocn.2009.21301

Florentine, M. (1985). Non-native listeners' perception of American-English in noise. *Proc. Internoise* 85, 1021–1024.

Francis, A. L., Tigchelaar, L. J., Zhang, R., and Zekveld, A. (2018). Effects of second language proficiency and linguistic uncertainty on recognition of speech in native and nonnative competing speech. *J. Speech Lang. Hear. Res.* 61, 1815–1830. doi: 10.1044/2018_JSLHR-H-17-0254

Franconeri, S. L., and Simons, D. J. (2003). Moving and looming stimuli capture attention. *Percept. Psychophys.* 65, 999–1010. doi: 10.3758/bf03194829

Freyman, R. L., Balakrishnan, U., and Helfer, K. S. (2004). Effect of number of masking talkers and auditory priming on informational masking in speech recognition. *J. Acoust. Soc. Am.* 115, 2246–2256. doi: 10.1121/1.1689343

Freyman, R. L., Helfer, K. S., McCall, D. D., and Clifton, R. K. (1999). The role of perceived spatial separation in the unmasking of speech. *J. Acoust. Soc. Am.* 106, 3578–3588. doi: 10.1121/1.428211

Gagné, J. P., Besser, J., and Lemke, U. (2017). Behavioral assessment of listening effort using a dual-task paradigm: a review. *Trends Hear.* 21:233121651668728. doi: 10.1177/2331216516687287

Getzmann, S., and Näätänen, R. (2015). The mismatch negativity as a measure of auditory stream segregation in a simulated 'cocktail-party' scenario: effect of age.". *Neurobiol. Aging* 36, 3029–3037. doi: 10.1016/j.neurobiolaging.2015.07.017

Gifford, R. H., Bacon, S. P., and Williams, E. J. (2007). An examination of speech recognition in a modulated background and of forward masking in younger and older listeners. *J. Speech Lang. Hear. Res.* 50, 857–864. doi: 10.1044/1092-4388(2007/060)

Goossens, T., Vercammen, C., Wouters, J., and van Wieringen, A. (2017). Masked speech perception across the adult lifespan: impact of age and hearing impairment. *Hear. Res.* 344, 109–124. doi: 10.1016/j.heares.2016.11.004

Gygi, B., and Shafiro, V. (2014). Spatial and temporal modifications of multitalker speech can improve speech perception in older adults. *Hear. Res.* 310, 76–86. doi: 10.1016/j.heares.2014.01.009

Helfer, K. S. (1997). Auditory and auditory-visual perception of clear and conversational speech. *J. Speech Lang. Hear. Res.* 40, 432–443. doi: 10.1044/jslhr.4002.432

Helfer, K., and Freyman, R. (2008). Aging and speech-on-speech masking. *Ear Hear.* 29, 87–98. doi: 10.1097/AUD.0b013e31815d638b

Humes, L. E., Lee, J. H., and Coughlin, M. P. (2006). Auditory measures of selective and divided attention in young and older adults using single-talker competition. *J. Acoust. Soc. Am.* 120, 2926–2937. doi: 10.1121/1.2354070

Ihlefeld, A., and Shinn-Cunningham, B. (2008). Spatial release from energetic and informational masking in a selective speech identification task. *J. Acoust. Soc. Am.* 123, 4369–4379. doi: 10.1121/1.2904826

Kahneman, D. (1973). *Attention And Effort*. Englewood Cliffs, NJ: Prentice-Hall.

Kidd, G., Mason, C. R., Richards, V. M., Gallun, F. J., and Durlach, N. I. (2008). "Informational masking," in *Auditory Perception Of Sound Sources*, eds W. A. Yost, A. N. Popper, and R. R. Fay (New York, NY: Springer Handbook of Auditory Research), 143–190.

Koelewijn, T., Zekveld, A. A., Festen, J. M., and Kramer, S. E. (2012). Pupil dilation uncovers extra listening effort in the presence of a single-talker masker. *Ear Hear.* 33, 291–300. doi: 10.1097/AUD.0b013e3182310019

Kroll, J. F., and Steward, E. (1994). Category interference in translation and picture naming: evidence for asymmetric connections between bilingual memory representations. *J. Mem. Lang.* 33, 149–174. doi: 10.1016/j.jecp.2008.10.004

Lavie, N. (2005). Distracted and confused? Selective attention under load. *Trends Cogn. Sci.* 9, 75–82.

Li, T., and Fu, Q. J. (2010). Effects of spectral shifting on speech perception in noise. *Hear. Res.* 270, 81–88. doi: 10.1016/j.heares.2010.09.005

Marrone, N., Mason, C. R., and Kidd, G. (2008). The effects of hearing loss and age on the benefit of spatial separation between multiple talkers in reverberant rooms. *J. Acoust. Soc. Am.* 124, 3064–3075. doi: 10.1121/1.2980441

Mattys, S. L., Brooks, J., and Cooke, M. (2009). Recognizing speech under a processing load: dissociating energetic from informational factors. *Cogn. Psychol.* 59, 203–243. doi: 10.1016/j.cogpsych.2009.04.001

Mattys, S. L., Carroll, L. M., Li, C. K. W., and Chan, S. L. Y. (2010). Effects of energetic and informational masking on speech segmentation by native and non-native speakers. *Speech Commun.* 52, 887–899. doi: 10.1016/j.specom.2010.01.005

Mayo, L. H., Florentine, M., and Buus, S. (1997). Age of second-language acquisition and perception of speech in noise. *J. Speech Lang. Hear. Res.* 40, 686–693. doi: 10.1044/jslhr.4003.686

Meador, D., Flege, J. E., and Mackay, I. R. A. (2000). Factors affecting the recognition of words in a second language. *Bilingualism* 3, 55–67. doi: 10.1017/s1366728900000134

Murphy, D. R., Daneman, M., and Schneider, B. A. (2006). Why do older adults have difficulty following conversations? *Psychol. Aging* 21, 49–61. doi: 10.1037/0882-7974.21.1.49

Nábělek, A. K., and Donahue, A. M. (1984). Perception of consonants in reverberation by native and non-native listeners. *J. Acoust. Soc. Am.* 75, 632–634. doi: 10.1121/1.390495

Peng, Z. E., and Wang, L. M. (2019). Listening effort by native and nonnative listeners due to noise, reverberation, and talker foreign accent during english speech perception. *J. Speech Lang. Hear. Res.* 62, 1068–1081. doi: 10.1044/2018_JSLHR-H-17-0423

Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W. Y., Humes, L. E., et al. (2016). Hearing impairment and cognitive energy: the framework for understanding effortful listening (FUEL). *Ear Hear.* 37, 5S–27S. doi: 10.1097/AUD.0000000000000312

Polka, L. (1991). Cross-language speech perception in adults: phonemic, phonetic, and acoustic contributions. *J. Acoust. Soc. Am.* 89, 2961–2977. doi: 10.1121/1.400734

Polka, L. (1992). Characterizing the influence of native language experience on adult speech perception. *Percept. Psychophys.* 52, 37–52. doi: 10.3758/bf03206758

Pollack, I. (1975). Auditory informational masking. *J. Acoust. Soc. Am.* 57:S5.

Raven, J. C. (1965). *The Mill Hill Vocabulary Scale*. London: H.K. Lewis.

Riecke, L., Peters, J. C., Valente, G., Kemper, V. G., Formisano, E., and Sorger, B. (2017). Frequency-selective attention in auditory scenes recruits frequency representations throughout human superior temporal cortex. *Cereb. Cortex* 27, 3002–3014. doi: 10.1093/cercor/bhw160

Roberts, B., and Summers, R. J. (2020). Informational masking of speech depends on masker spectro-temporal variation but not on its coherencea. *J. Acoust. Soc. Am.* 148, 2416–2428. doi: 10.1121/10.0002359

Rogers, C. L., and Lopez, A. S. (2008). Perception of silent-center syllables by native and non-native english speakers. *J. Acoust. Soc. Am.* 124, 1278–1293. doi: 10.1121/1.2939127

Rogers, C. L., Lister, J., Febo, D. M., Besing, J. M., and Abrams, H. B. (2006). Effects of bilingualism, noise, and reverberation on speech perception by listeners with normal hearing. *Appl. Psycholinguist.* 27, 465–485. doi: 10.1017/s014271640606036x

Rosen, S., Souza, P., Ekelund, C., and Majeed, A. A. (2013). Listening to speech in a background of other talkers: effects of talker number and noise vocoding. *J. Acoust. Soc. Am.* 133, 2431–2443. doi: 10.1121/1.4794379

Scharf, B. (1998). "Auditory attention: the psychoacoustical approach," in *Attention*, ed. H. Pashler (Hove: Psychology Press), 75–117.

Scharf, B., Quigley, S., Aoki, C., Peachey, N., and Reeves, A. (1987). Focused auditory attention and frequency selectivity. *Percept. Psychophys.* 42, 215–223. doi: 10.3758/bf03203073

Schneider, B. A., Avivi-Reich, M., and Mozuraitis, M. (2015). A cautionary note on the use of the Analysis of Covariance (ANCOVA) in classification designs with and without within-subject factors. *Front. Psychol.* 6:474. doi: 10.3389/fpsyg.2015.00474

Schneider, B. A., Li, L., and Daneman, M. (2007). How competing speech interferes with speech comprehension in everyday listening situations. *J. Am. Acad. Audiol.* 18, 578–591. doi: 10.3766/jaaa.18.7.4

Schneider, B. A., Pichora-Fuller, M. K., and Daneman, M. (2010). "The effects of senescent changes in audition and cognition on spoken language comprehension," in *Springer Handbook of Auditory Research: The Aging Auditory System: Perceptual Characterization and Neural Bases of Presbycusis*, eds S. Gordon-Salant, R. D. Frisina, A. N. Popper, and R. R. Fay (New York, NY: Springer), 167–210.

Stevenson, R. A., Nelms, C. E., Baum, S. H., Zurkovsky, L., Barense, M. D., Newhouse, P. A., et al. (2015). Deficits in audiovisual speech perception in normal aging emerge at the level of whole-word recognition. *Neurobiol. Aging* 36, 283–291. doi: 10.1016/j.neurobiolaging.2014.08.003

Stuart, A., and Phillips, D. P. (1996). Word recognition in continuous and interrupted broadband noise by young normal-hearing, older normal-hearing, and presbyacusic listeners. *Ear Hear.* 17, 478–489. doi: 10.1097/00003446-199612000-00004

Szalárdy, O., Tóth, B., Farkas, D., György, E., and Winkler, I. (2019). Neuronal correlates of informational and energetic masking in the human brain in a multi-talker situation. *Front. Psychol.* 10:786. doi: 10.3389/fpsyg.2019.00786

Tangkhpanya, F., Carrour, M. L., Doucet, F., and Gagné, J. P. (2019). The effort required to comprehend a short documentary in noise: a comparison of younger and older francophones. *Am. J. Audiol.* 28, 756–761. doi: 10.1044/2019_AJA-HEAL18-18-0170

Tun, P., and Wingfield, A. (1999). One voice too many: adult age differences in language processing with different types of distracting sounds. *J. Gerontol.* 54B, 317–327. doi: 10.1093/geronb/54b.5.p317

Vongpaisal, T., and Pichora-Fuller, M. K. (2007). Effect of age on use of F0 to segregate concurrent vowels. *J. Speech Hear. Lang. Res.* 50, 1139–1156.

Yang, Z., Chen, J., Huang, Q., Wu, X., Wu, Y., Schneider, B. A., et al. (2007). The effect of voice cuing on releasing Chinese speech from informational masking. *Speech Commun.* 49, 892–904. doi: 10.1016/j.specom.2007.05.005

Zekveld, A. A., Heslenfeld, D. J., Johnsrude, I. S., Versfeld, N. J., and Kramer, S. E. (2014). The eye as a window to the listening brain: neural correlates of pupil size as a measure of cognitive listening load. *NeuroImage (Orlando, Fla.)* 101, 76–86. doi: 10.1016/j.neuroimage.2014.06.069

Conflict of Interest: 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 | Frontiers in Psychology

ORIGINAL RESEARCH
published: 01 April 2022
doi: 10.3389/fpsyg.2022.837644

Lexical Effects on the Perceived Clarity of Noise-Vocoded Speech in Younger and Older Listeners

Terrin N. Tamati[1,2], Victoria A. Sevich[3], Emily M. Clausing[1] and Aaron C. Moberly[1]*

[1] Department of Otolaryngology – Head and Neck Surgery, The Ohio State University Wexner Medical Center, Columbus, OH, United States, [2] Department of Otorhinolaryngology/Head and Neck Surgery, University Medical Center Groningen, University of Groningen, Groningen, Netherlands, [3] Department of Speech and Hearing Science, The Ohio State University, Columbus, OH, United States

When listening to degraded speech, such as speech delivered by a cochlear implant (CI), listeners make use of top-down linguistic knowledge to facilitate speech recognition. Lexical knowledge supports speech recognition and enhances the perceived clarity of speech. Yet, the extent to which lexical knowledge can be used to effectively compensate for degraded input may depend on the degree of degradation and the listener's age. The current study investigated lexical effects in the compensation for speech that was degraded via noise-vocoding in younger and older listeners. In an online experiment, younger and older normal-hearing (NH) listeners rated the clarity of noise-vocoded sentences on a scale from 1 ("very unclear") to 7 ("completely clear"). Lexical information was provided by matching text primes and the lexical content of the target utterance. Half of the sentences were preceded by a matching text prime, while half were preceded by a non-matching prime. Each sentence also consisted of three key words of high or low lexical frequency and neighborhood density. Sentences were processed to simulate CI hearing, using an eight-channel noise vocoder with varying filter slopes. Results showed that lexical information impacted the perceived clarity of noise-vocoded speech. Noise-vocoded speech was perceived as clearer when preceded by a matching prime, and when sentences included key words with high lexical frequency and low neighborhood density. However, the strength of the lexical effects depended on the level of degradation. Matching text primes had a greater impact for speech with poorer spectral resolution, but lexical content had a smaller impact for speech with poorer spectral resolution. Finally, lexical information appeared to benefit both younger and older listeners. Findings demonstrate that lexical knowledge can be employed by younger and older listeners in cognitive compensation during the processing of noise-vocoded speech. However, lexical content may not be as reliable when the signal is highly degraded. Clinical implications are that for adult CI users, lexical knowledge might be used to compensate for the degraded speech signal, regardless of age, but some CI users may be hindered by a relatively poor signal.

Keywords: speech clarity, noise-vocoded speech, priming, lexical properties, aging

Edited by:
Leah Fostick,
Ariel University, Israel

Reviewed by:
Karen Banai,
University of Haifa, Israel
Tobias Weissgerber,
University Hospital Frankfurt, Germany

****Correspondence:***
Terrin N. Tamati
terrin.tamati@osumc.edu

Specialty section:
This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Psychology

Received: *16 December 2021*
Accepted: *16 February 2022*
Published: *01 April 2022*

Citation:
Tamati TN, Sevich VA, Clausing EM and Moberly AC (2022) Lexical Effects on the Perceived Clarity of Noise-Vocoded Speech in Younger and Older Listeners. Front. Psychol. 13:837644. doi: 10.3389/fpsyg.2022.837644

INTRODUCTION

An important and distinctive property of speech perception is its robustness in the face of a wide range of adverse and challenging conditions. Successful recognition of a spoken word involves rapid mapping of the acoustic signal onto lexical representations stored in long-term memory (e.g., McClelland and Elman, 1986; Norris, 1994; Luce and Pisoni, 1998). In favorable listening conditions, lexical access occurs rapidly and automatically, with minimal recruitment of cognitive processing to disambiguate the message. In everyday, real-world environments, however, the speech signal is often distorted by environmental degradations, such as background noise or competing speech, as well as source degradations from variability arising from talkers with different developmental, social, and language histories (e.g., Mattys et al., 2012; Gilbert et al., 2013). Further, hearing-impaired listeners must also cope with additional degradations due to reduced audibility and/or distortions specific to their type, degree, and configuration of hearing loss. Even rehabilitative devices, such as hearing aids or cochlear implants (CIs), can preserve or introduce spectral degradations, despite partially restoring audibility. As a result of these combined sources of adversity, speech recognition in real-world conditions is challenging (e.g., Johnson et al., 2016; Meister et al., 2016; Hughes et al., 2018; Janse and Andringa, 2021), and resolving the increased ambiguity arising from these adverse factors requires the recruitment of cognitive mechanisms, such as attention, semantic and syntactic constraints, and lexical knowledge (e.g., Pichora-Fuller, 2008; Başkent et al., 2016a; Koeritzer et al., 2018). The effective use of cognitive processes and linguistic knowledge to recognize degraded speech likely depends on both bottom-up signal quality and the top-down cognitive-linguistic skills of the individual listener (e.g., Rönnberg et al., 2013; Başkent et al., 2016a; Moberly et al., 2021). Still, it is relatively unclear how these bottom-up and top-down processes interact to impact speech recognition, and further how the contribution of these factors may depend on the age of the listener. The current study explores the contribution of bottom-up and top-down factors – and their interaction – to the perceived clarity of noise-vocoded speech in younger and older adults with normal hearing (NH).

Recognition of Degraded Speech

Top-down mechanisms are especially relevant for hearing impaired adults with CIs. Adult CI users must achieve successful daily communication relying on speech signals that are heavily reduced in acoustic-phonetic detail compared to what is typically available to NH listeners, due to the limitations of the electrode-nerve interface and relatively broad electrical stimulation of the auditory nerve (for a review, see Başkent et al. (2016b)). This reduced spectral resolution limits the accurate recognition of speech in CI users (Henry et al., 2005). CI users may achieve accurate recognition of the degraded speech delivered by the device, but do so by relying on predictive coding and downstream cognitive resources (e.g., Pals et al., 2013; Bhargava et al., 2014; Winn et al., 2015; Başkent et al., 2016a). However, individual CI users display variability in spectral resolution across the electrode array (Won et al., 2007), which may be related to auditory nerve health, electrode placement, or other device or surgical factors (e.g., Blamey et al., 2013; Başkent et al., 2016b). Poorer spectral resolution in CI users may contribute to increased difficulty in recognizing speech (Henry et al., 2005; Won et al., 2007; Moberly et al., 2018b) and impact the ability to effectively use top-down resources (Bhargava et al., 2014; Pals et al., 2020).

Increased signal degradation may result in greater relative reliance on top-down cognitive-linguistic resources. For example, Pals et al. (2013) examined listening effort in the recognition of noise-vocoded speech. Noise-vocoding is commonly used to simulate – albeit imperfectly – the signal delivered by a CI and to introduce varying degrees of spectral degradation experimentally. In their study, increasing spectral resolution in the noise-vocoder simulations of CI hearing resulted in reduced response times in a dual-task paradigm, suggesting that listening effort decreases with increased signal quality. In a later study, Pals et al. (2020) examined the effect of the number of spectral channels (i.e., spectral resolution) on speech comprehension and listening effort in CI users. They found that increasing the number of spectral channels leads to an improvement in speech comprehension and response times in the sentence verification task, suggesting increased signal quality improves speech comprehension and listening effort. Interestingly, this effect was not observed in the dual-task paradigm, which the authors interpreted as evidence that changes in listening effort as a function of signal degradation may not be well reflected in tasks assessing speech recognition accuracy. Similarly, conventional measures of speech recognition accuracy may not be as sensitive to subtle differences in signal degradation and listening effort compared to measures that capture the time course and processes underlying speech perception and spoken word recognition (e.g., Başkent et al., 2016a; Pisoni et al., 2017; Moberly et al., 2018a; Winn and Teece, 2021). Measures involving subjective assessment of speech clarity may also be more sensitive to differences in signal quality since they would allow the listener to make more subtle distinctions between degraded signals (e.g., Sohoglu et al., 2014), even when using a wide range of degrees of degradation that may produce ceiling and/or floor effects in a word or sentence recognition task.

Lexical Knowledge in Degraded Speech Perception

To cope with degraded speech, listeners utilize several linguistic resources, including semantic context, syntactic structure, and lexical information (e.g., Pichora-Fuller, 2008; Başkent et al., 2016a; Wagner et al., 2016; Koeritzer et al., 2018). Regarding lexical information, listeners make use of linguistic context providing the lexical and phonological form of an utterance to make predictions about its content. The perceptual processing of speech is facilitated when a listener is provided with text that partially or completely matches the target utterance prior to its auditory presentation (e.g., Goldinger et al., 1992; Buchwald et al., 2009; Chng et al., 2019). Form-based prediction from exact matching text provides specific information about the lexical and phonological content of an upcoming utterance and allows for the activation of the lexical items in that utterance. In this manner, top-down lexical and phonological information provided visually by matching text primes enhances the perception of noise-vocoded speech (Davis et al., 2005; Hervais-Adelman et al., 2011; Wild et al., 2012; Signoret et al., 2018; Signoret and Rudner, 2019). Recently, Signoret et al. (2018) used a speech clarity rating task to assess the effects of bottom-up spectral resolution from acoustic noise-vocoding and top-down form-based prediction from matching text primes as well as meaning-based prediction from supportive semantic context on the perceived clarity of degraded speech in NH young to middle aged adults. The authors found that speech clarity ratings were sensitive to differences in the spectral resolution of the noise-vocoded speech (manipulated in that study by the number of vocoder channels). Moreover, they found evidence for independent and additive effects of form- and meaning-based prediction on clarity ratings. Together, these previous studies also demonstrate that a speech clarity rating task may be a sensitive and useful tool for assessing top-down effects on the perception of degraded speech.

The lexical properties of the words within an utterance also influence the speed and accuracy of spoken word recognition (e.g., Luce and Pisoni, 1998). According to most accounts, spoken word recognition involves the activation of a set of candidate words including the target and words that are phonologically-similar to the target. Words that differ from the target word by a single phoneme that is substituted, deleted, or added are considered to share phonological similarity and form part of the target word's phonological neighborhood (Luce and Pisoni, 1998). As more information becomes available, the target word is selected from the candidate words, while competitors must be inhibited (e.g., Marslen-Wilson, 1987; Luce and Pisoni, 1998). Two lexical properties – lexical frequency (i.e., frequency of occurrence in a spoken language) and neighborhood density (i.e., number of phonologically-similar lexical neighbors) – play key roles in the discrimination and selection of the target item. Words with higher lexical frequency and fewer neighbors ("easy" words) are easier to recognize than words with lower lexical frequency and more neighbors ("hard" words) since there is greater activation of the target word and less competition from neighbors. Accordingly, easy words have consistently been found to be more quickly and accurately recognized than hard words for NH listeners, particularly in the presence of noise or other sources of adversity (e.g., Howes, 1957; Savin, 1963; Sommers et al., 1997; Bradlow and Pisoni, 1999; Taler et al., 2010). Effects of lexical frequency and neighborhood density have also been observed for NH listeners with noise-vocoded speech (Tamati et al., 2020b) and CI users (Tamati and Moberly, 2021). Thus, the lexical content of an utterance may be a source of top-down compensatory information that has a relatively strong impact on the recognition of degraded speech.

Interactions of Bottom-Up and Top-Down Processing

Relative reliance on top-down compensatory mechanisms in speech understanding may depend on the degree of degradation of the speech signal. Listeners rely more on top-down mechanisms to a certain degree when speech is degraded by noise or other sources of adversity (e.g., Kalikow et al., 1977; Luce and Pisoni, 1998; Vitevitch and Luce, 1999; Mattys et al., 2012). However, reliance on top-down processing may decrease when the degree of degradation of the speech signal is more extreme (Boothroyd and Nittrouer, 1988; Mattys et al., 2009; Clopper, 2012; Bhargava et al., 2014; Gelfand et al., 2014). Linguistic information conveyed by a severely degraded signal may be undetectable or misleading (Samuel, 1981; Król and El-Deredy, 2011; Bhargava et al., 2014; Sohoglu et al., 2014), resulting in reduced reliance on higher-level linguistic knowledge and greater reliance on lower-level segmental cues. As such, the speech signal must provide sufficient acoustic-phonetic detail to support higher-level processing (Aydelott and Bates, 2004; Mattys et al., 2005, 2009; Clopper, 2012). Interestingly, in the study by Signoret et al. (2018) described above, the authors observed that form- (matching text primes) and meaning-based prediction (semantic context) had greater effects for more degraded signals compared to more favorable signals. In a follow-up study, Signoret and Rudner (2019) also found evidence for the interaction between top-down and bottom-up processes in speech clarity ratings in a group of older, hearing-impaired adults. With less degraded speech, older, hearing-impaired listeners benefited from semantic context. However, with more degraded speech, the benefit from semantic context was observed only when matching text primes preceded the sentence. Further, unlike findings in the original Signoret et al. (2018) study of younger listeners, benefits from form- and meaning-based prediction were not related to working memory capacity in the older, hearing-impaired listeners, suggesting that they may have exceeded their available resources to effectively process the degraded speech.

Similarly, findings from previous studies examining variability in speech recognition outcomes in adult CI users demonstrate that some CI users may be able to more effectively use top-down compensation (Bhargava et al., 2014; Moberly et al., 2014, 2016; Başkent et al., 2016a). Relatively poorer performing CI users have demonstrated a reduced ability to take advantage of top-down compensatory mechanisms (e.g., Liu et al., 2004; Bhargava et al., 2014; Başkent et al., 2016a), suggesting a reduced role of cognitive-linguistic abilities for poorer performers. Additionally, Moberly et al. (2021) found that the contribution of cognitive-linguistic abilities to speech recognition outcomes in adult CI users depended on individual bottom-up auditory sensitivity. Cognitive-linguistic abilities contributed less to speech recognition outcomes for adult CI users with poor auditory spectro-temporal resolution compared to CI users with better auditory resolution. Similarly, specifically comparing performance between groups of CI users with the poorest and best outcomes, Tamati et al. (2020a) suggested that top-down processes may play a limited role in speech recognition in CI users with the poorest bottom-up auditory sensitivity. However, although many adult CI users are typically of advanced age, these studies did not consider how aging may have contributed to individual differences in top-down compensation. Thus, for individual adult CI users, the ability to use top-down compensatory mechanisms to recognize the degraded signal delivered by a CI depends on cognitive-linguistic ability and, crucially, on the quality of the signal processed by the implant and delivered to the auditory cortex. Yet, it is still unknown how aging may alter the use of top-down compensatory strategies for degraded speech understanding.

The Effects of Aging on Top-Down Compensation

Top-down compensation for degraded speech among older adults may be impacted by age-related declines in neurocognitive functioning and auditory sensitivity. Older adults with "age-normal" hearing (i.e., normal or near-normal thresholds to tones on audiometric testing) demonstrate poorer spectro-temporal processing of auditory input (Fitzgibbons and Gordon-Salant, 1994; Schmiedt, 2010; Tun et al., 2012), as well as aging-related declines in neurocognitive functions of working memory capacity, inhibition-concentration, information-processing speed, and non-verbal reasoning (i.e., fluid intelligence). These age-related declines in top-down cognitive functioning and bottom-up auditory processes may contribute to overall poorer speech recognition abilities compared to younger adults (Pichora-Fuller and Singh, 2006; Arehart et al., 2013). Further, older listeners may be even more greatly impacted by adverse conditions, such as speech degraded by vocoding (Rosemann et al., 2017; Moberly et al., 2018b).

Some processes that may help support the perception of degraded speech are fortunately maintained during aging. Specifically, older listeners may rely upon prior knowledge (i.e., crystallized intelligence – knowledge previously acquired through prior learning and experiences, such as vocabulary knowledge) to enhance the processing of degraded speech. In contrast with fluid intelligence, crystallized intelligence is typically maintained in older age (Salthouse, 1993; Wingfield et al., 1994; Ryan et al., 2000; Park et al., 2002). Previous findings suggest that older adults may take advantage of crystallized intelligence in adverse listening conditions to the same extent – or possibly even more so – than younger listeners (e.g., Balota and Duchek, 1991; Wingfield et al., 1994; Pichora-Fuller et al., 1995; Valencia-Laver and Light, 2000; Daneman et al., 2006; Sheldon et al., 2008). Top-down compensation in older adults may therefore specifically involve reliance on linguistic knowledge, such as through use of supportive semantic or syntactic context (e.g., Pichora-Fuller, 2008) and lexical information (e.g., Schneider et al., 2016), during the recognition of degraded speech.

For older listeners, lexical knowledge may play an important role in the processing of degraded speech. Some previous studies suggest that older adults may benefit at least as much, if not more, as younger adults from exact or partially matching auditory or text primes (e.g., Wu et al., 2012; Getzmann et al., 2014; Freyman et al., 2017; Ouyang et al., 2020). Differences among older and younger listeners may arise from changes in lexical processing due to age-related declines in the top-down processing of speech (e.g., Federmeier et al., 2003) as well as increases in or maintenance of vocabulary knowledge (Verhaeghen, 2003) across the lifespan. Previous studies examining lexical competition in speech recognition suggest that older listeners display difficulty in resolving lexical competition during speech recognition (Sommers, 1996; Sommers and Danielson, 1999; Helfner and Jesse, 2015), potentially due to age-related declines in inhibitory control as well as increases in vocabulary size. Older adults show less accurate recognition of words with high neighborhood density in noise compared to younger listeners (Sommers, 1996; Sommers and Danielson, 1999). Examining the effects of lexical competition on word-in-sentence recognition, Taler et al. (2010) found that difference scores between accuracy for words with high and low neighborhood density in challenging conditions (lower SNR of −3 dB) were negatively related to inhibitory control across younger and older listeners, demonstrating that those with stronger inhibitory control were less affected by density effects. Further, the recognition of words with high neighborhood density, but not words with low neighborhood density, relates to stronger inhibitory control (Green and Barber, 1981, 1983; Jerger et al., 1993). Finally, increases in vocabulary size in aging may result in increased lexical competition in older adults (e.g., Salthouse, 2004; McAuliffe et al., 2013; Ramscar et al., 2014; Carroll et al., 2016). Thus, age-related changes in the top-down processing of speech may result in decreased lexical discriminability for words with many phonologically-similar neighbors.

Age-related changes in the use of lexical frequency information may also contribute to difficulties in resolving lexical competition during speech recognition. Results from Taler et al. (2010) suggest that lexical frequency effects on word-in-sentence recognition are similar across the lifespan. In that study, both older and younger listeners responded more accurately and quickly to sentences containing high-frequency words than low-frequency words. However, studies using other approaches suggest that older adults rely more heavily on lexical frequency than younger adults. Older adults appear to show increased activation of high frequency target words (and competitors) and less competition from low frequency competitors (Revill and Spieler, 2012). In an eye-tracking study, Revill and Spieler (2012) found that older adults were more likely to fixate high-frequency phonological competitors compared to younger listeners when listening to speech degraded with white noise; in contrast, younger adults were not more likely to fixate high-frequency competitors. Similarly, results in visual word processing demonstrate that older readers show stronger effects of word frequency than younger readers (Spieler and Balota, 2000; Balota et al., 2004). Together, these findings suggest that lexical effects (originating from matching text primes and/or the lexical content of the target stimulus) may have a greater impact on speech processing in older listeners.

The Current Study

The current study investigated the top-down cognitive-linguistic and bottom-up sensory factors that affect the perceived clarity of speech in NH younger and older adults using an online speech clarity rating task. Speech was degraded using acoustic noise-vocoder simulations of CI hearing. The use of simulations allows for the signal parameters to be well controlled in order ensure that NH listeners experience similar degrees of signal degradation. Additionally, the linguistic and hearing histories of NH listeners can be better controlled, in contrast with typical adult CI users who vary in age, durations of deafness, length of CI use, and etiology of hearing loss, which may influence overall speech recognition abilities (e.g., Blamey et al., 2013). Greater control over these factors facilitates the evaluation of how bottom-up and top-down processing impacts speech recognition. Finally, findings using noise-vocoded speech have potential clinical relevance for understanding speech recognition outcomes in CI users, providing valuable insight into how spectral degradation affects speech recognition outcomes (e.g., Friesen et al., 2001).

The main goal of the current study was to investigate the top-down cognitive-linguistic factors that affect the perceived clarity of noise-vocoded speech, and how these factors may interact with bottom-up sensory factors. Given that the current study was administered online, our first goal was to determine if speech clarity ratings provided within an online experimental procedure would be consistent with previous findings obtained with in-person experimental procedures. In line with previous studies (Signoret et al., 2018; Signoret and Rudner, 2019), we sought to evaluate if online speech clarity ratings for 8-channel acoustic noise-vocoder simulations of CI hearing are sensitive to signal quality differences. To investigate the effect of spectral resolution on speech clarity, the current study manipulated the sharpness of the slope of the bandpass filters to simulate current spread in the cochlea. The amount of spread of excitation in the cochlea determines the extent to which individual stimulation channels of the implant interact (e.g., Black and Clark, 1980; Bingabr et al., 2008; Gaudrain and Başkent, 2015; Koelewijn et al., 2021). Three vocoder conditions were included to simulate low spread (LS), medium spread (MS), and high spread (HS) of excitation (and decreasing spectral resolution, respectively), in order to obtain varying degrees of degradation. In prior studies, simulating electrical current spread in the cochlea by systematically varying synthesis filter slopes has yielded a wide range of performance on speech recognition accuracy in NH listeners (Bingabr et al., 2008; Oxenham and Kreft, 2014; Winn et al., 2016; Mehta et al., 2020). For example, Bingabr et al. (2008) found that listeners achieved more accurate recognition of 8-channel vocoded words with steeper filter slopes (lower spread, higher spectral resolution): accuracy increased from about 40–80% as filter slopes increased incrementally from 14 dB/octave (lowest spectral resolution) to 110 dB/octave (highest spectral resolution). Since more intelligible speech is correlated with higher ratings of speech clarity (Eisenberg et al., 1998), we similarly expected to find increasing ratings of speech clarity as we increased synthesis filter slopes (i.e., provided more favorable spectral resolution). If the online speech clarity ratings are consistent with previous in-person results, increasing spectral resolution would be expected to result in higher perceived clarity (i.e., LS < MS < HS).

Second, we examined the effects of form-based text priming and the effects of lexical content (lexical frequency and neighborhood density) on the perceived clarity of noise-vocoded speech. To do so, we first attempted to replicate the effect of matching text primes observed in previous studies by Signoret et al. (2018) and Signoret and Rudner (2019) within the online experimental procedure. Text primes that were either matching (i.e., the text prime and the target sentence were the same) or non-matching (i.e., the prime and the target sentence were different) were presented prior to a target sentence. Consistent with the findings from these previous studies, we expected that matching primes would enhance the perceived clarity of vocoded target sentences, compared to non-matching primes.

Expanding on the earlier findings, we also examined the effects of lexical frequency and neighborhood density on the perceived clarity of noise-vocoded sentences. Sentences used in the current study were from the Veteran's Affairs Sentence Test (VAST; Bell and Wilson, 2001), which was developed to control for frequency of word use and lexical confusability, based on Luce and Pisoni (1998). Each VAST sentence contained key words that had relatively (1) high or low lexical frequency and (2) high or low neighborhood density, resulting in four sentence types: high lexical frequency, low neighborhood density (HL); high lexical frequency, high neighborhood density (HH); low lexical frequency, low neighborhood density (LL); and low lexical frequency, high neighborhood density (LH). Based on previous findings regarding the effects of lexical frequency and neighborhood density in the recognition of noise-vocoded speech or in CI users (e.g., Tamati et al., 2020b; Tamati and Moberly, 2021), we expected that the lexical characteristics of the key words of a sentence would determine its perceived clarity. That is, we expected that sentences with high frequency key words would be perceived as clearer than sentences with low frequency key words, and that sentences with key words with low neighborhood density would be perceived as clearer than sentences with key words with high neighborhood density.

We further predicted an interaction between bottom-up signal quality (i.e., vocoder condition) and top-down cognitive-linguistic factors (i.e., priming and lexical content). Previous findings suggest decreased reliance on top-down processing under conditions of severe spectro-temporal degradation (e.g., Bhargava et al., 2014; Başkent et al., 2016a; Moberly et al., 2021). In the current study, vocoder conditions were designed to simulate decreasing degrees of spectral resolution. If top-down processing contributes less when the signal is more severely degraded, then lexical knowledge would be expected to contribute less to the perceived clarity of sentences with relatively poor signal quality (HS), and contribute relatively more for sentences with relatively more favorable signal quality (MS and LS). That is, lexical information should demonstrate a relatively stronger effect on perceived speech clarity in the MS and LS conditions compared to the HS condition (i.e., larger differences between priming conditions and sentence types). However, the two sources of lexical information in the current study (i.e., matching text primes presented visually and lexical content presented auditorily) differ by their susceptibility to signal degradation; as such, they may differ in their contributions under severely degraded conditions.

Finally, the current study sought to assess the impact of aging on how top-down cognitive-linguistic factors contribute to the perceived clarity of noise-vocoded speech. Previous research from Signoret et al. (2018) and Signoret and Rudner (2019) found potential differences in the interaction of top-down and bottom-up processes in younger versus older, hearing-impaired adults, the latter of whom seemed to exhibit less top-down compensation with degrees of degradation at which the younger adults had shown top-down effects. The findings of the two studies demonstrating reduced top-down compensation in older, hearing-impaired adults under conditions of more severe degradation suggest that there may be group differences in top-down compensation attributable to hearing impairment and/or age. Yet, because these two factors were conflated in the second study, it is unclear if aging alone impacts top-down processing and its interactions with signal quality. Based on previous findings on lexical effects on speech recognition across the lifespan, it was expected that older adults would be able to utilize top-down lexical knowledge to the same extent as, if not more than, younger adults, at least in conditions of more favorable signal quality. However, if the use of top-down lexical knowledge is restricted by poorer auditory and/or cognitive functioning, then older adults would not demonstrate the benefits from matching text primes and lexical content in conditions of poorer signal quality. Moreover, if poorer neurocognitive functioning contributes to this deficiency, then older adults may be limited in their use of lexical information, regardless of whether that information is delivered visually (i.e., matching text primes) or auditorily (i.e., lexical content). In contrast, if poorer auditory sensitivity contributes to this deficiency, then older adults would be expected to demonstrate less effective use of lexical knowledge in the perceived clarity of noise-vocoded speech specifically when relying exclusively on auditory information, thereby resulting in a relatively smaller effect of lexical content.

To summarize, this study tested the following hypotheses. First, decreasing spectral resolution would lead to lower perceived clarity for noise-vocoded speech (HS < MS < LS). Second, if adults capitalize upon top-down lexical knowledge, both matching text primes and the lexical content of the target utterance would enhance the perceived speech clarity for noise-vocoded speech. Third, if spectral resolution and top-down processes interact as predicted above, we would see the greatest impact of top-down processing in conditions of higher spectral resolution. Our final hypothesis related to aging was that older adults would be able to utilize lexical information in conditions of more favorable signal quality, but its use would be relatively more restricted in conditions of poorer signal quality. We further explored if lexical information delivered visually (i.e., matching text prime) or auditorily (i.e., lexical content) would differentially impact speech clarity in older adults, depending on age-related declines in auditory sensitivity or neurocognitive functioning.

MATERIALS AND METHODS

Participants

A total of 36 younger adults (17 female) and 38 older adults (26 female) were recruited for the current study. Younger participants were between the ages of 20–39 years and older participants were between the ages of 50–77 years, all with self-reported NH. All participants were recruited from the Prolific recruitment service (Prolific, 2021), an international research recruitment service. During testing, participants completed a short headphone screener to ensure the use of good-quality headphones during testing. Prior to analysis, six younger participants and eight older participants were excluded for failing a headphone screener. Thirty younger and older participants passed the headphone screener, with a score of ≥ 5 correct answers out of 6. The younger listener group (YNH) consisted of the remaining 30 younger participants (12 female), who were between the ages of 19 and 39 years (*M* = 29.8, *SD* = 5.8). The older listener group (ONH) consisted of the remaining 30 older participants (21 female), who were between the ages of 50 and 71 years (*M* = 57.3, *SD* = 5.8). All participants were native speakers of American English with no history of speech or language disorders. All participants provided electronic informed written consent prior to participation and received $7.50 for approximately 45 min of their time. Institutional Review Board (IRB) approval was obtained.

Materials

Stimulus materials consisted of 144 sentences originating from the VAST sentence materials (Bell and Wilson, 2001), later recorded as part of the Multi-talker Corpus of Foreign-Accented English (MCFAE; Tamati et al., 2011). Sentences were produced by a female native speaker of American English from the Midland dialect region. At the time of the recording collection, the talker was 22 years old and reported no prior history of speech or hearing disorders.

Each VAST sentence contained three key words; all key words were either high or low lexical frequency and either high or low phonological neighborhood density. The total set of sentences contained 36 sentences with high lexical frequency and high neighborhood density key words (HH), 36 sentences with high lexical frequency and low neighborhood density key words (HL), 36 sentences with low lexical frequency and high neighborhood density key words (LH), and 36 sentences with low lexical frequency and low neighborhood density key words (LL). The sentence materials and key lexical properties are provided in the **Supplementary Materials**.

The three vocoder conditions were created by processing sentences through an 8-channel noise-band vocoder in Matlab with code maintained by the dB SPL lab at the University Medical Center Groningen (e.g., Gaudrain and Başkent, 2015). For all vocoding conditions, the original signal was filtered into 8 analysis bands between 150 and 7,000 Hz, using 12th order (72 dB/oct.), zero-phase Butterworth filters. The bands corresponded to evenly spaced regions of the cochlea using Greenwood's frequency-to-place mapping function (Greenwood, 1990). The frequency cutoffs of individual bands were 150–301, 301–523, 523–852, 852–1,338, 1,338–2,056, 2,056–3,117, 3,117–4,684, and 4,684–7,000 Hz. The synthesis filters were 12th order filters (72 dB/octave) for the LS condition, 8th order filters (48 dB/octave) for the MS condition, and 4th order filters (24 dB/octave) for the HS vocoding condition, in order of decreasing spectral resolution. The synthesis filters had the same cutoff frequencies as the analysis filters. From each analysis band, the temporal envelope was extracted by half-wave rectification and low-pass filtering with a cutoff frequency of 300 Hz, using a zero-phase 4th order Butterworth filter (Gaudrain and Başkent, 2015). Noise carriers in each channel were modulated with the corresponding extracted envelope, which were then filtered by the synthesis filters. The modulated noise bands from all vocoder channels were added together to construct the stimuli. After processing, all stimuli were normalized to the same root mean square power.

Procedure

Participants completed the experiment on the gorilla.sc platform using their own desktop or laptop devices and headphones. They were asked to sit in a quiet room and use good-quality headphones during the experiment. Prior to starting the speech clarity task, participants completed a language background questionnaire, the headphone screener, and a familiarity block. The online headphone screener consisted of a three-alternative forced-choice task in which participants listened to three white noise sounds (all 1,000 ms) – one of which contains a faint tone – at comfortable level and responded as to which sound contained the tone (Milne et al., 2020). The headphone test was designed based on Huggins Pitch, a dichotic pitch percept that should be detectable only when using headphones. For each trial, two intervals contained diotically-presented white noise. The third interval contained the target Huggins Pitch stimulus. The percept of pitch was generated by presenting a white noise stimulus to one ear and the same white noise with a phase shift of 180 degrees over a narrow frequency band around the center frequency of 600 Hz to the other ear. The result of this manipulation is the perception of a tone with the pitch of the center frequency of the phase-shifted band (i.e., 600 Hz) in noise. For the purpose of the current study, participants who scored <5 correct answers out of 6 were considered to have failed the headphone screener and were excluded from the analysis. Although not designed to screen for other aspects of the listener's environment, participants completing the experiment in a noisy environment and/or participants with hearing impairment may also fail this screener (e.g., Santurette and Dau, 2007). During the familiarity block, participants were able to gain familiarity with noise-vocoded speech and the ratings scale. Three vocoded sentences were presented to provide references for the ratings scale of 1 ("very unclear") to 7 ("completely clear"). A LH sentence in the HS vocoder condition was used as a reference for low clarity, a HL sentence in the LS condition was used as a reference for high clarity, and a HH sentence in the MS condition was used as a reference for the middle range.

On each trial of the main speech clarity task, listeners were presented with a single sentence and were asked to rate the clarity of each sentence on a scale from 1 ("very unclear") to 7 ("completely clear"). The sentence was always preceded by a 500 ms fixation cross on the computer screen and a matching or non-matching text prime. The text prime appeared on the screen for 2.5 s in order to allow the participant enough time to read the text. The matching prime consisted of the word-by-word orthographic transcription of the target sentence. The non-matching prime was created by randomly reorganizing the letters of the original prime into nonsense words; non-matching primes were controlled to ensure that no real words resulted from the randomization. After reading the text prime, listeners were presented with the target sentence and responded by clicking one of seven numerical response options. Forty-eight sentences (12 of each HH, HL, LH, and LL) were presented in each of the LS, MS, and HS vocoder conditions. Half of the sentences (6 of each sentence type in each vocoder condition) were preceded by a matching text prime, while the other half of the sentences were preceded by a non-matching text prime. Rating responses were recorded and coded by sentence type, vocoder, and text prime condition.

RESULTS

A mixed ANOVA on speech clarity ratings was carried out with vocoder condition (LS, MS, and HS), priming (matching and non-matching), and sentence type (HL, HH, LL, and LH) as within-subject factors and listener group (YNH and ONH) as the between-subject factor. An alpha of 0.05 was used. *Post hoc* Tukey tests were used to explore the significant main effects and interactions. Significant main effects of vocoder [$F(2,58) = 414.09$, $p < 0.001$, $\eta_p^2 = 0.88$], priming [$F(1,58) = 150.61$, $p < 0.001$, $\eta_p^2 = 0.72$], and sentence type [$F(3,174) = 84.61$, $p < 0.001$, $\eta_p^2 = 0.59$] emerged, indicating that both the bottom-up factor of vocoder (i.e., spectral resolution) and the top-down factors of priming and sentence type impacted the perceived clarity of noise-vocoded speech. For vocoder condition, clarity ratings were significantly higher (i.e., perceived as clearer) in the LS ($M = 4.91$, $SD = 0.76$) and MS conditions ($M = 5.04$, $SD = 0.77$) compared to the HS condition ($M = 3.43$, $SD = 0.83$) (all $p < 0.001$). However, the LS and MS conditions were not significantly different from one another. For priming condition, clarity ratings were also significantly higher with matching ($M = 4.96$, $SD = 0.83$) than non-matching text primes ($M = 3.96$, $SD = 0.77$) ($p < 0.001$). For sentence type condition, clarity ratings were significantly higher in the HL condition ($M = 4.68$, $SD = 0.75$) compared to the LH conditions ($M = 4.09$, $SD = 0.79$) ($p < 0.001$). Clarity ratings were also significantly higher in the HH ($M = 4.63$, $SD = 0.77$) than the LH condition ($p < 0.001$), and in the LL ($M = 4.45$, $SD = 0.74$) compared to the LH condition ($p < 0.001$). No other comparisons reached significance. These results confirm the effects of the spectral resolution, matching text primes, and the lexical properties of frequency and neighborhood density on clarity ratings for noise-vocoded speech. The main effect of group was not significant (YNH: $M = 4.43$, $SD = 0.68$; ONH: $M = 4.49$, $SD = 0.81$), suggesting a lack of overall differences in clarity ratings between younger and older adults.

To facilitate interpretation of the interactions among the factors, the mean clarity ratings across vocoder conditions (LS, MS, and HS) in each priming condition (matching, non-matching) and sentence type (HL, HH, LL, and LH) are shown for YNH listeners in **Figure 1** and ONH listeners in **Figure 2**. Significant two-way interactions of vocoder × priming [$F(2,116) = 9.17$, $p < 0.001$, $\eta_p^2 = 0.39$] and vocoder × sentence type [$F(6,348) = 7.84$, $p < 0.001$, $\eta_p^2 = 0.12$] were uncovered, suggesting that the strength of effects of the top-down factors of priming and sentence type depended on the bottom-up factor of vocoder (i.e., spectral resolution). For priming condition, clarity ratings were higher with matching than non-matching

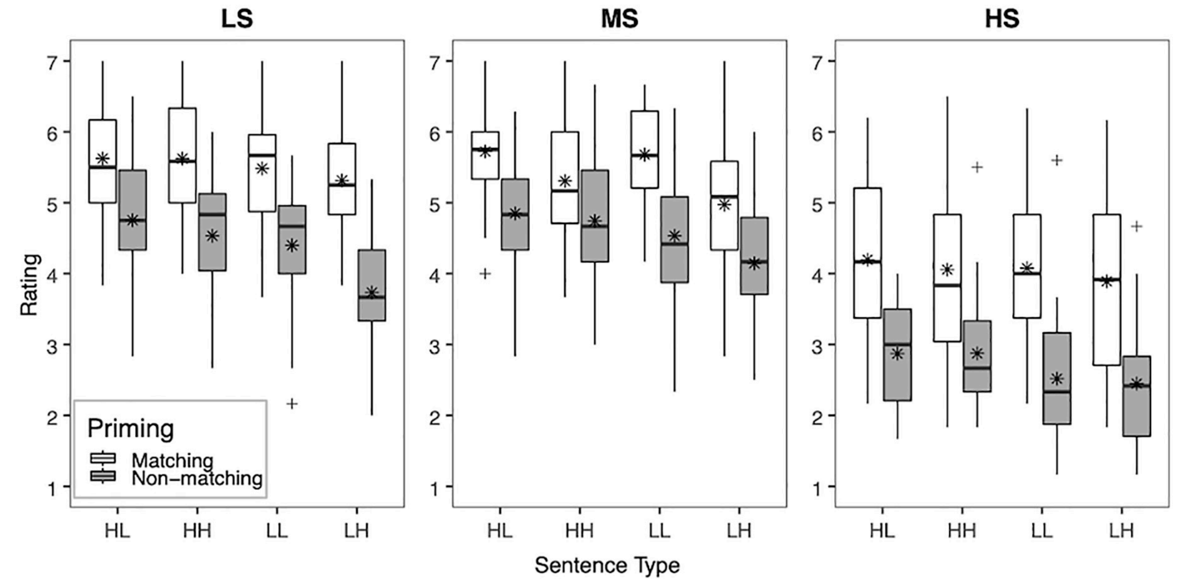


FIGURE 1 | Box plot showing the mean clarity ratings for NH younger adults (YNH). Clarity ratings (1 "very unclear" to 7 "completely clear") are plotted across all three vocoder conditions (LS, MS, and HS), priming condition (matching and non-matching), and sentence type (HL, HH, LL, and LH). The boxes extend from the lower to the upper quartile (the interquartile range, IQ), the solid midline indicates the median, and the star indicates the mean. The whiskers indicate the highest and lowest values no greater than 1.5 times the IQ, and the plus signs indicate outliers, which are defined as data points larger than 1.5 times the IQ.

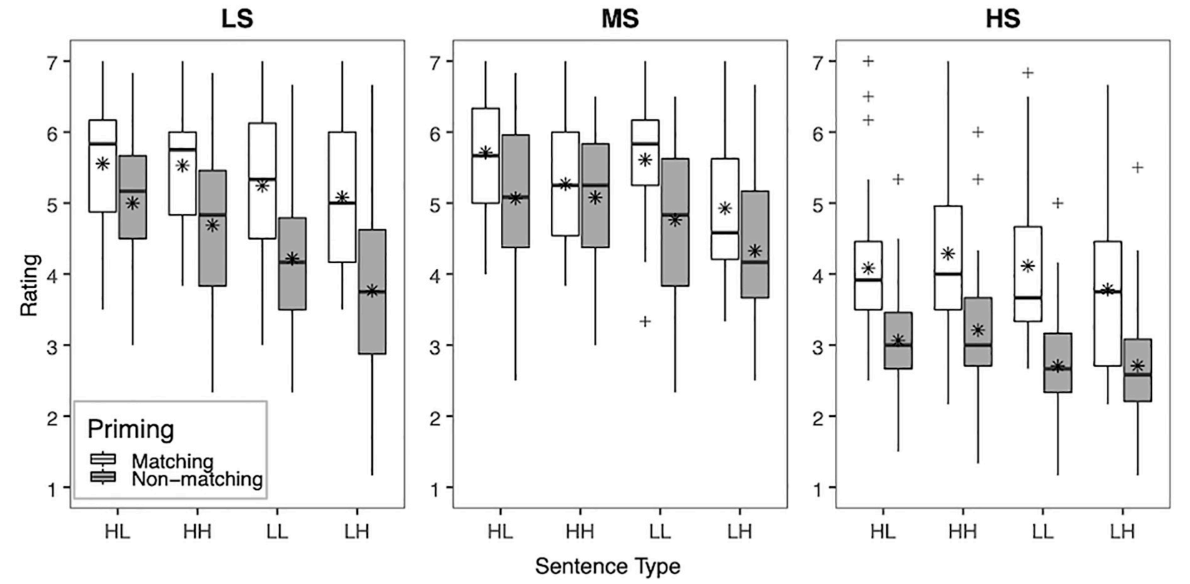


FIGURE 2 | Box plot showing the mean clarity ratings for NH older adults (ONH). Clarity ratings (1 "very unclear" to 7 "completely clear") are plotted across all three vocoder conditions (LS, MS, and HS), priming condition (matching and non-matching), and sentence type (HL, HH, LL, and LH). The boxes extend from the lower to the upper quartile (the interquartile range, IQ), the solid midline indicates the median, and the star indicates the mean. The whiskers indicate the highest and lowest values no greater than 1.5 times the IQ, and the plus signs indicate outliers, which are defined as data points larger than 1.5 times the IQ.

primes for all vocoder conditions (all $p < 0.001$). The effect of sentence type differed by vocoder condition. For the HS condition (i.e., low spectral resolution), no significant differences among sentence types emerged. For the MS condition (i.e., medium spectral resolution), ratings were higher for the HL, HH, and LL conditions than the LH condition (all $p \leq 0.002$), but no other comparison reached significance. For the LS condition (i.e., highest spectral resolution), ratings were higher for the HL and HH conditions than the LH condition (all $p < 0.001$), but no other comparisons among sentence types reached significance.

A significant two-way priming × sentence type interaction [$F(3,174) = 11.03$, $p < 0.001$, $\eta_p^2 = 0.16$] also emerged, suggesting that the effect of sentence type depended on the availability of visual text priming. Overall, for the matching priming condition, no comparison reached significance. For the non-matching priming condition, clarity ratings were significantly higher for the HL and HH conditions than the LH condition (all $p < 0.001$), and for the HL condition than the LL condition ($p = 0.036$). However, the three-way interaction of vocoder × priming × sentence type was also significant [$F(6,348) = 7.12$, $p < 0.001$, $\eta_p^2 = 0.11$], suggesting that the effect of lexical content depended both on the priming and vocoder condition. For the HS condition (i.e., low spectral resolution), no significant differences among sentence types emerged for either the matching or non-matching priming condition. For the MS condition (i.e., medium spectral resolution), clarity ratings for the LL condition were higher than for the LH condition for the matching priming condition ($p = 0.035$). For the non-matching priming condition, clarity ratings for the HL condition were higher than for the LH condition ($p = 0.023$). For the LS condition (i.e., highest spectral resolution), no significant differences among sentence types emerged for the matching priming condition. For the non-matching priming condition, clarity ratings for the HL and HH conditions were higher than for the LH condition (all $p < 0.001$). To summarize, these comparisons suggest that sentence type had little to no effect on clarity ratings for the HS vocoder condition. However, sentence type had an effect on clarity ratings for both matching and non-matching priming conditions for the MS condition, and for non-matching primes for the LS condition. Additionally, the most consistent difference among sentence types was observed between HL and LH conditions.

DISCUSSION

The current study investigated top-down lexical effects on the perceived clarity of noise-vocoded speech, and interactions with bottom-up signal quality, in NH younger and older adults. More specifically, the current study examined form-based priming, by introducing matching or non-matching text primes presented prior to the target utterance, as well as the lexical content of the target utterance, by varying the lexical frequency and neighborhood density of key words. Given that the current study was conducted online, we also sought to determine if speech clarity ratings from the online task were consistent with previous studies using an in-person experimental procedure (e.g., Signoret et al., 2018; Signoret and Rudner, 2019).

Examining the effects of bottom-up signal quality, we hypothesized that decreasing spectral resolution would result in lower perceived clarity of noise-vocoded speech. Consistent with our hypothesis, a main effect of vocoder condition demonstrated that increased spectral resolution enhanced the perceived clarity of noise-vocoded speech in both NH younger and older adults. These findings are broadly consistent with results from previous in-person studies (Signoret et al., 2018; Signoret and Rudner, 2019), in which spectral resolution was varied by manipulation of the number of vocoder channels. In the current study, an 8-channel vocoder was used to simulate the same number of electrode contact points in each vocoder condition, but spectral resolution was varied by manipulating the sharpness of the bandpass filter slopes to simulate low, medium, and high spread of excitation in the cochlea. Decreased spectral resolution via simulation of increased channel interaction has been found to result in less accurate speech recognition (Fu and Nogaki, 2005; Bingabr et al., 2008; Wang et al., 2012; Oxenham and Kreft, 2014; Winn et al., 2016; Mehta et al., 2020), less accurate pitch perception (Crew et al., 2012; Mehta and Oxenham, 2017), increased listening effort (Winn et al., 2016), and limitations in the perception of non-linguistic aspects of speech, such as voice cue perception (Gaudrain and Başkent, 2015; Koelewijn et al., 2021). In the current study, while there was a main effect of vocoder condition, only differences between the condition with the worst spectral resolution (HS; 4th order, 24 dB/octave) and the conditions with increasingly more favorable spectral resolutions (MS and LS; 8th order, 48 dB/octave and 12th order, 72 dB/octave, respectively) emerged. Similarly, Gaudrain and Başkent (2015) found that improving spectral resolution by increasing the filter order from 4 (24 dB/octave) to 8 (48 dB/octave) in a 12-channel noise-vocoder improved perception of vocal tract length cues, but further increasing to 12 (72 dB/octave) did not improve perception. In the current study, sharpening the filter slopes beyond 8 (48 dB/octave) also did not drastically enhance the perceived clarity of noise-vocoded speech when using eight channels. Although consistent with previous in-person studies, the extent to which clarity ratings were impacted by the online administration of the task is unclear. With the online study, the testing environment and equipment were not controlled. Additionally, participants' hearing thresholds were not evaluated. Although participants completed a headphone screener, the screener was not designed to specifically evaluate these factors. It is possible that better controlling for these factors in an in-person setting would result in clarity ratings that are more sensitive to subtle differences in spectral resolution. Here, findings suggest that younger and older adults perceived 8-channel noise-vocoded speech as clearer with improved spectral resolution introduced by increasing the sharpness of the filter slopes from the HS to MS and LS vocoder conditions. More controlled studies should be carried out in the future to further investigate the effects of bottom-up signal quality on the perceived clarity of noise-vocoded speech.

To investigate how top-down lexical knowledge affects the perceived clarity of noise-vocoded speech, we explored the effects of form-based prediction and varying lexical content (i.e., lexical frequency and neighborhood density) on the perceived clarity of noise-vocoded speech. Consistent with previous in-person studies (Wild et al., 2012; Signoret et al., 2018; Signoret and Rudner, 2019), results demonstrate a clarity-enhancing effect of form-based prediction. In their studies, Signoret et al. (2018) and Signoret and Rudner (2019) similarly demonstrated that matching text presented visually prior to the target utterance enhances the clarity of noise-vocoded speech in NH younger and hearing-impaired older listeners. Here, degraded speech was perceived as being clearer when a matching text prime had been presented prior to its auditory presentation, compared to when degraded speech was preceded by a random, meaningless assortment of letters. Moreover, the benefit from matching text primes was observed in all vocoder conditions, for both younger and older adults. Matching text provides identical lexical and phonological content of the utterance, and reliably facilitates the activation of the exact lexical items in the target utterance. As such, presenting text that partially or exactly matches the lexical and phonological content of the target utterances enhances the recognition and perceived clarity of degraded speech by allowing the listener to generate expectations about the upcoming target utterance (Wild et al., 2012; Sohoglu et al., 2014; Signoret et al., 2018; Signoret and Rudner, 2019).

To expand upon previous findings and further investigate top-down lexical effects, we also examined how the lexical content of the target utterance impacts the perceived clarity of noise-vocoded speech. We tested the hypothesis that sentences containing high frequency words with few phonological neighbors (low neighborhood density) would be perceived as clearer than sentences containing low frequency words with many phonological neighbors (high neighborhood density). Consistent with our hypothesis, we found a significant main effect of sentence type in the overall analyses across vocoder conditions, as well as significant effects of sentence type in each vocoder condition when exploring the interactions. Although sentence types that emerged as significantly different from one another varied by vocoder and priming conditions, sentences containing lexically easy words (high lexical frequency and low neighborhood density) were consistently perceived as clearer than sentences containing lexically hard words (low lexical frequency and high neighborhood density) by both younger and older listener groups. For example, HL, HH, and LL sentences were overall rated as clearer than LH sentences. Given that the most consistent difference emerged between the easy (i.e., HL) and hard (i.e., LH) sentences, these findings suggest a role for both lexical frequency and neighborhood density in the perceived clarity of noise-vocoded speech for younger and older adults.

Additionally, the lexical content of sentence key words (i.e., lexical frequency and neighborhood density) contributed to clarity ratings both when supportive visual information was available (matching text primes) and when listeners had to rely on auditory information alone (non-matching text primes). Thus, lexical content was utilized with the support of both combined visual and auditory information as well as auditory information alone in younger and older adults. The overall effects of lexical frequency and neighborhood density are consistent with existing accounts of spoken word recognition that emphasize the integration of top-down lexical knowledge with bottom-up acoustic-phonetic details during spoken word recognition, such as the Neighborhood Activation Model (NAM; Luce et al., 1990; Luce and Pisoni, 1998). Previous findings consistent with these accounts have demonstrated that lexically easy words are recognized or discriminated more accurately and faster than hard words under noise-vocoding (e.g., Tamati et al., 2020b) as well as in hearing-impaired listeners with or without cochlear implants (e.g., Dirks et al., 2001; Takayanagi et al., 2002; Tamati et al., 2021). Further, some evidence suggests that lexical content differentially facilitates reaction time in shadowing tasks when participants are presented with partially or exactly matching auditory primes (Dufour and Peereman, 2004), since easy target words are quickly activated from the prime prior to hearing the target and remain activated due to phonological overlap between the prime and target. Our findings extend upon these previous studies by demonstrating that these lexical properties impact the perceived clarity of noise-vocoded speech in younger and older adults.

Another goal of the study was to investigate how top-down lexical knowledge interacts with bottom-up signal quality. We hypothesized that an interaction between bottom-up and top-down processing would result in a decreased contribution of lexical knowledge on perceived clarity of sentences with relatively poor signal quality (HS), relative to conditions with relatively better quality (MS and LS). Indeed, the contribution of lexical knowledge described above appears to vary based on the degree of degradation of the noise-vocoded speech. However, in contrast with our initial hypothesis, the benefit from matching text primes was observed in all vocoder conditions, for both younger and older adults. However, a greater relative effect of priming appeared to emerge for the HS condition, which provided the most degraded spectral resolution, as can be seen in **Figures 1**, **2**. Similarly, Signoret et al. (2018) found that form-based prediction had a stronger effect at lower degrees of signal quality (3-channel noise vocoder). Although not designed to test these accounts, our results are consistent with accounts of degraded speech recognition, such as the Ease of Language Understanding Model (ELU; Rönnberg et al., 2013), which emphasizes the role of top-down processing when bottom-up processing is insufficient. It is worth pointing out that form-based predictions about the upcoming target utterance were generated based on visual text information, which was not degraded either visually or auditorily. As such, the matching text prime provided a reliable source of lexical information that enhanced the clarity of the noise-vocoded speech, regardless of the degree of degradation of the target utterance. Similarly, other sources of linguistic information that remain unaltered despite degradation in signal quality, such as visual contextual cues relating to the setting of a conversation (e.g., formal or informal; Brouwer et al., 2012) or text information from subtitles (Mitterer and McQueen, 2009), may be relied upon to enhance speech clarity and facilitate spoken word recognition in real-world, adverse conditions.

Broadly consistent with our initial hypothesis, the lexical content of the utterance appeared to contribute less to the perceived clarity of noise-vocoded speech in the HS condition, where speech was more degraded, and relatively more in the MS and LS conditions, where speech was less degraded. While sentence type (i.e., HL, HH, LL, and LH) was significant overall in the HS condition, differences did not emerge among individual sentence types. Insights into how bottom-up signal quality may have influenced the relative reliance on lexical content can be obtained by specifically examining the contribution of sentence type (i.e., lexical content) with and without matching text primes. In the LS condition (higher spectral resolution), sentence type only contributed to perceived sentence clarity without matching text primes; easy words were perceived as clearer than hard words only when participants had to rely upon the auditory signal alone for both younger and older adults. In the MS condition (middle spectral resolution), sentence type influenced perceived speech clarity both with and without matching primes. Overall, these findings demonstrate that sentence type had less influence on perceived clarity when spectral resolution was poor and potentially when combined conditions were among the most favorable (highest spectral resolution combined with matching text primes). Thus, top-down use of lexical content may be most relevant in conditions of moderate degradation.

These findings are largely consistent with previous research showing that top-down compensation may become less effective when the degree of degradation of the speech signal is more extreme (Samuel, 1981; Król and El-Deredy, 2011; Bhargava et al., 2014; Sohoglu et al., 2014), and reliance on top-down processing may decrease (Mattys et al., 2009; Clopper, 2012). Similarly, in the HS condition in the current study, the degraded speech signal likely did not provide sufficient acoustic-phonetic detail to support the robust use of top-down lexical knowledge (Aydelott and Bates, 2004; Mattys et al., 2005, 2009; Clopper, 2012). In other words, in the current study, listeners did not rely on the lexical content to the same extent in conditions of poor signal quality compared to conditions of more favorable signal quality, when lexical information was delivered solely by the degraded target utterance. These findings are consistent with previous studies showing that individual CI users with poorer bottom-up signal quality may less effectively employ top-down compensatory mechanisms to process the degraded speech delivered by the CI (Bhargava et al., 2014; Tamati et al., 2020a; Moberly et al., 2021). In contrast with matching text primes, lexical content delivered by a degraded speech signal, as well as other forms of top-down linguistic information such as semantic context, may be susceptible to bottom-up signal quality and may not be engaged to facilitate speech understanding as effectively in real-world, adverse conditions. Taken together, our findings suggest that the HS condition provided such a poor signal that only matching text primes could largely be relied upon to enhance perceived speech clarity; in contrast, in the LS condition, the lexical content of auditorily presented target utterance could be relied upon to a greater extent. Thus, top-down lexical knowledge was employed with the support of both combined visual and auditory information as well as auditory information alone in younger and older adults, and further interacts with bottom-up signal quality to impact the perceived clarity of noise-vocoded speech. However, these findings should be interpreted with caution. As mentioned above, the current study did not control for several factors, including the testing environment (e.g., noise or distractors) or the audiometric thresholds of the listeners, that could have impacted the quality of the signal conveyed to individual listeners and the relative reliance on top-down mechanisms. Future studies that better control for these factors are needed to shed more light on the interaction of bottom-up and top-down processing.

The final hypothesis tested in the current study related to the effects of aging on top-down processing. Overall, we predicted that older adults would effectively utilize top-down lexical knowledge to the same extent, if not more, than younger adults, at least in conditions of more favorable signal quality. However, we further predicted that if the use of top-down lexical knowledge is restricted by poorer auditory and/or cognitive functioning, then older adults would not demonstrate strong clarity-enhancing effects of matching text primes and lexical content in conditions of poorer signal quality. The general finding here did not support that hypothesis. Instead, both younger and older listener groups showed similar effects of matching text primes and lexical content on speech clarity ratings consistently across degrees of signal degradation. Regarding form-based prediction, the clarity-enhancing benefit observed from matching text primes was similar in the younger and older listener groups. Our findings provide additional evidence that degraded speech is perceived as clearer when the listener is provided with text that matches the target utterance prior to its auditory presentation (e.g., Sohoglu et al., 2012, 2014; Wild et al., 2012; Wu et al., 2012; Getzmann et al., 2014; Signoret et al., 2018; Signoret and Rudner, 2019). Additionally, Signoret et al. (2018) found the ability to use form-based prediction as well as semantic context was related to working memory capacity, suggesting a role for cognitive abilities in top-down compensation and a potential means by which aging could affect the perceived clarity of noise-vocoded speech. Although outside the scope of the current study, one potential explanation for the similarity of form-based prediction between the younger and older listeners could be that working memory capacity did not differ between our groups, or working memory capacity was not implicated within the speech clarity rating paradigm used currently. Moreover, while conducting the study using an online experimental protocol may have enabled recruitment from a larger participant pool, both the younger and older adults would have been comfortable with technology and online research. Thus, findings may not generalize broadly to other populations Yet, importantly, the current study expands on that literature to show that these priming effects do not appear to deteriorate significantly with aging.

Regarding lexical content, the results of the current study also showed that the younger and older adults appeared to demonstrate similar combined effects of lexical frequency and neighborhood density. These findings are in line with work by Taler et al. (2010), who examined the effects of lexical competition on word-in-sentence recognition. There, both groups of older and younger adults recognized words in sentences more accurately and quickly for sentences containing high frequency words (vs. low frequency words) as well as for sentences containing words with low neighborhood density (vs. high neighborhood density). In contrast, our findings differ somewhat from studies suggesting that older listeners display more difficulty in resolving lexical competition during speech recognition (Sommers, 1996; Sommers and Danielson, 1999; Helfner and Jesse, 2015). In addition, some previous studies have suggested that older adults rely more heavily on lexical frequency than younger adults in both auditory speech perception (Revill and Spieler, 2012) and visual word processing (Spieler and Balota, 2000; Balota et al., 2004). Notably, several of the studies examining neighborhood density effects have at least partially attributed these age-related differences to poorer inhibitory control in older listeners (e.g., Sommers and Danielson, 1999). For example, in the Taler et al. (2010) study, difference scores between accuracy for words with high and low neighborhood density at a lower SNR (−3 dB SNR) were negatively related to inhibitory control across all listeners. Additionally, changes in lexical processing across the lifespan may also be attributable to increases in vocabulary size with aging (e.g., Salthouse, 2004; McAuliffe et al., 2013; Ramscar et al., 2014; Carroll et al., 2016). Although we did not assess inhibitory control or vocabulary size in this study, a potential explanation for the similarity of neighborhood density effects between the younger and older listeners could be that inhibitory control or vocabulary size did not differ substantially between groups. Relatedly, the older participants in the current study were slightly of a younger age, with a mean age of 57.3 years and a range of 50–71 years, compared to the studies that have observed differences in lexical processing between younger and older listeners. Previous studies have included groups of older adults with mean ages of around 65–75 years. Finally, another possibility is that the current outcome measure – online perceptual ratings of speech clarity – was not sensitive to differences in inhibitory control or vocabulary size between the younger and older listening adults. In Taler et al. (2010), for example, the dependent measures were response times and accuracy in sentence recognition tasks, both of which may involve different levels of lexical-phonological processing than would be expected in the speech clarity rating paradigm used presently. Future studies examining the effects of aging on top-down compensation should consider using alternative measures, and including a wider age range of older participants.

In addition to examining the overall impact of aging on top-down compensation, we also considered two additional alternative hypotheses: that if poorer neurocognitive functioning contributes to an aging-related deficiency in top-down processing, then older adults would be limited in their use of lexical information, regardless of the modality of the source (i.e., matching text primes presented visually and lexical content presented auditorily). The alternative hypothesis was that if poorer auditory sensitivity contributes to an aging-related deficiency, then older adults should not show strong effects of lexical content on the clarity of noise-vocoded speech specifically when relying exclusively on auditory information (i.e., with a non-matching prime). In other words, there may be differences between age groups in which top-down mechanisms would enhance speech clarity. Previous studies from Signoret et al. (2018) and Signoret and Rudner (2019) identified potential differences in the interaction of top-down and bottom-up processes in younger and older, hearing-impaired adults, who seemed to exhibit less top-down compensation with more severe degrees of degradation, at which the younger adults had benefited. In contrast, our results suggested overall very similar effects of matching text primes and lexical content for the younger and older groups.

More generally, findings from this study provide additional evidence that older listeners can effectively enhance the processing of a novel form of degraded speech by making use of their crystallized intelligence (here and lexical knowledge), which has been found to be maintained into older age (Salthouse, 1993; Wingfield et al., 1994; Ryan et al., 2000; Park et al., 2002). Our findings are therefore consistent with previous studies showing that older adults can capitalize on crystallized intelligence in adverse listening conditions to the same extent as younger listeners (e.g., Balota and Duchek, 1991; Wingfield et al., 1994; Pichora-Fuller et al., 1995; Valencia-Laver and Light, 2000; Daneman et al., 2006; Sheldon et al., 2008). However, more research should be carried out to better understand top-down mechanisms and how their effective use may depend on bottom-up signal quality and age, particularly for older CI users who must deal with a degraded speech signal as part of their normal, daily communication.

Clinical Implications for Adult Cochlear Implant Users

The findings from the current study may have implications for understanding and addressing the vast individual differences in speech recognition outcomes observed among adult CI users (Lazard et al., 2012; Lenarz et al., 2012; Blamey et al., 2013). First, the preservation of top-down processing with advancing age is highly significant because it suggests that older listeners compensate for degraded listening conditions using their long-term linguistic knowledge. Therefore, targeting the use of linguistic context in understanding speech in rehabilitative training may be effective in helping adult CI users across the lifespan achieve real-world communication success. Second, our findings suggest that the effective use of some top-down compensatory mechanisms may crucially depend on bottom-up signal quality. This finding is clinically relevant since it could suggest that some top-down compensatory strategies across individual CI users may crucially depend on the quality of the bottom-up input. More specifically, similar to findings from Tamati et al. (2020a) and Moberly et al. (2021), individual CI users with poor bottom-up auditory input may not be able to take advantage of some top-down resources to effectively compensate for the degraded speech delivered by their CIs.

However, the relevance of the findings to CI users should be interpreted with caution. The current study used acoustic noise-vocoder CI simulations to simulate degraded speech that captures functional performance of adult CI users. Acoustic simulations capture the basic signal processing steps of CIs (Loizou, 1998) and, for some spoken word recognition tasks, the performance ranges of actual CI users (e.g., Friesen et al., 2001). However, there are many factors that additionally affect speech perception in CI users (Başkent et al., 2016b), including the severity and duration of deafness prior to implantation, and duration of CI use (e.g., Blamey et al., 2013). Increased severity and longer durations of deafness prior to implantation have been linked to weak phonological processing and poorer speech recognition outcomes in adult CI users (Lyxell et al., 1998; Lazard et al., 2010; Lazard and Giraud, 2017; Tamati et al., 2021). Weakened phonological processing may impact the structure and organization of the mental lexicon, thereby altering how or the extent to which listeners utilize lexical knowledge, issues we would not expect to face when testing NH younger and older adults. Additionally, CI users appear to benefit from experience using their devices to more effectively use top-down compensatory mechanisms (e.g., Winn et al., 2012; Fuller et al., 2014; Bhargava et al., 2016). In contrast, the NH listeners in the current study had minimal experience with noise-vocoded speech prior to testing. NH adults typically adapt quickly to noise-vocoded speech and reach a stable level of recognition accuracy with a small number of sentences (Davis et al., 2005; Hervais-Adelman et al., 2008; Huyck et al., 2017), particularly when presented with matching text primes (Davis et al., 2005). Nevertheless, the novelty of this form of degradation may have altered the reliance on top-down lexical knowledge. Thus, additional studies, possibly involving more diverse younger and older listeners with or without CIs, along with measures of demographic and cognitive-linguistic abilities, are needed to better understand the roles of top-down and bottom-up processing on the perception of degraded speech.

CONCLUSION

The current study examined how top-down cognitive-linguistic and bottom-up sensory factors affect the perceived clarity of speech in younger and older adults using an online speech clarity task. Findings demonstrate that both younger and older adults were able to effectively use lexical knowledge to enhance the clarity of noise-vocoded speech. In particular, listeners perceived the speech as clearer when preceded by an exact matching text prime and when the target utterance contained lexically easy words (i.e., high lexical frequency and low neighborhood density) compared to hard words (i.e., low lexical frequency, high neighborhood density). However, the effective use of top-down lexical knowledge appeared to depend on the bottom-signal quality. While matching text primes provided a relatively greater enhancement of more degraded speech, lexical content had a greater impact with more moderately degraded speech. Importantly, these findings also show that older adults make use of lexical knowledge to a similar degree as the younger listeners. Taken together, these findings emphasize the interactive nature of bottom-up and top-down processes in the perception of degraded speech. Further, findings suggest that lexical knowledge could be effectively used to enhance speech understanding in adult CI users across the lifespan, but some CI users may be hindered by a relatively poor signal.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation, to any qualified researcher.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Institutional Review Board at Ohio State University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

TT designed and conducted the study. TT, VS, and AM contributed to data analysis and interpretation. All authors contributed to the writing of the manuscript and approved of the final version of the manuscript for submission.

FUNDING

Preparation of this manuscript was supported in part by VENI Grant No. 275-89-035 from the Netherlands Organization for Scientific Research (NWO) to TT, funding from the President's Postdoctoral Scholars Program (PPSP) at The Ohio State University awarded to TT, Career Development Award 5K23DC015539-02 from the National Institutes of Health, National Institute on Deafness and Other Communication Disorders to AM, and R01DC019088 from the National Institutes of Health, National Institute on Deafness and Other Communication Disorders to AM.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2022.837644/full#supplementary-material

REFERENCES

Arehart, K. H., Souza, P., Baca, R., and Kates, J. M. (2013). Working memory, age and hearing loss: susceptibility to hearing aid distortion. *Ear Hear.* 34, 251–260. doi: 10.1097/AUD.0b013e318271aa5e

Aydelott, J., and Bates, E. (2004). Effects of acoustical distortion and semantic context on lexical access. *Lang. Cogn. Process.* 19, 29–56.

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., and Yap, M. J. (2004). Visual word recognition of single-syllable words. *J. Exp. Psychol.* 133, 283–316. doi: 10.1037/0096-3445.133.2.283

Balota, D. A., and Duchek, J. M. (1991). Semantic priming effects, lexical repetition effects, and contextual disambiguation effects in healthy aged individuals and individuals with senile dementia of the Alzheimer type. *Brain Lang.* 40, 181–201. doi: 10.1016/0093-934x(91)90124-j

Başkent, D., Clarke, J., Pals, C., Benard, M. R., Bhargava, P., Saija, J., et al. (2016a). Cognitive compensation of speech perception with hearing impairment, cochlear implants, and aging: how and to what degree can it be achieved? *Trends Hear.* 20, 1–16.

Başkent, D., Gaudrain, E., Tamati, T. N., and Wagner, A. (2016b). "Perception and psychoacoustics of speech in cochlear implant users," in *Scientific Foundations of Audiology: Perspectives from Physics, Biology, Modeling, and Medicine*, eds A. T. Cacace, E. de Kleine, A. G. Holt, and P. van Dijk (San Diego, CA: Plural Publishing), 285–319.

Bell, T. S., and Wilson, R. H. (2001). Sentence materials based on frequency of word use and lexical confusability. *J. Am. Acad. Audiol.* 12:514.

Bhargava, P., Gaudrain, E., and Başkent, D. (2014). Top-down restoration of speech in cochlear implant users. *Hear. Res.* 309, 113–123. doi: 10.1016/j.heares.2013.12.003

Bhargava, P., Gaudrain, E., and Başkent, D. (2016). The intelligibility of interrupted speech: cochlear implant users and normal hearing listeners. *J. Assoc. Res. Otolaryngol.* 17, 475–491. doi: 10.1007/s10162-016-0565-9

Bingabr, M., Espinoza-Varas, B., and Loizou, P. C. (2008). Simulating the effect of spread of excitation in cochlear implants. *Hear. Res.* 241, 73–79. doi: 10.1016/j.heares.2008.04.012

Black, R. C., and Clark, G. M. (1980). Differential electrical excitation of the auditory nerve. *J. Acoust. Soc. Am.* 67, 868–874. doi: 10.1121/1.383966

Blamey, P., Artieres, F., Başkent, D., Bergeron, F., Beynon, A., Burke, E., et al. (2013). Factors affecting auditory performance of postlinguistically deaf adults using cochlear implants: an update with 2251 patients. *Audiol. Neurotol.* 18, 36–47. doi: 10.1159/000343189

Boothroyd, A., and Nittrouer, S. (1988). Mathematical treatment of context effects in phoneme and word recognition. *J. Acoust. Soc. Am.* 84, 101–114. doi: 10.1121/1.396976

Bradlow, A. R., and Pisoni, D. B. (1999). Recognition of spoken words by native and non-native listeners: talker-, listener-, and item-related factors. *J. Acoust. Soc. Am.* 106, 2074–2085. doi: 10.1121/1.427952

Brouwer, S., Mitterer, H., and Huettig, F. (2012). Speech reductions change the dynamics of competition during spoken word recognition. *Lang. Cog. Process.* 27, 539–571.

Buchwald, A. B., Winters, S. J., and Pisoni, D. B. (2009). Visual speech primes open-set recognition of spoken words. *Lang. Cog. Process.* 24, 580–610. doi: 10.1080/01690960802536357

Carroll, R., Warzybok, A., Kollmeier, B., and Ruigendijk, E. (2016). Age-related differences in lexical access related to speech recognition in noise. *Front. Psychol.* 7:990. doi: 10.3389/fpsyg.2016.00990

Chng, K. Y., Yap, M. J., and Goh, W. D. (2019). Cross-modal masked repetition and semantic priming in auditory lexical decision. *Psychon. Bull. Rev.* 26, 599–608. doi: 10.3758/s13423-018-1540-8

Clopper, C. G. (2012). Effects of dialect variation on the semantic predictability benefit. *Lang. Cogn. Process.* 27, 1002–1020.

Crew, J. D., Galvin, J. J., and Fu, Q. J. (2012). Channel interaction limits melodic pitch perception in simulated cochlear implants. *J. Acoust. Soc. Am.* 132, EL429–EL435. doi: 10.1121/1.4758770

Daneman, M., Hannon, B., and Burton, C. (2006). Are there age-related differences in shallow semantic processing of text? Evidence from eye movements. *Discourse Process.* 42, 177–203.

Davis, M. H., Johnsrude, I. S., Hervais-Adelman, A., Taylor, K., and McGettigan, C. (2005). Lexical information drives perceptual learning of distorted speech: evidence from the comprehension of noise-vocoded sentences. *J. Exp. Psychol. Gen.* 134, 222–241. doi: 10.1037/0096-3445.134.2.222

Dirks, D. D., Takayanagi, S., and Moshfegh, A. (2001). Effects of lexical factors on word recognition among normal-hearing and hearing-impaired listeners. *J. Am. Acad. Audiol.* 12, 233–244.

Dufour, S., and Peereman, R. (2004). Phonological priming in auditory word recognition: initial overlap facilitation effect varies as a function of target word frequency. *Curre. Psychol. Lett. Behav. Brain Cogn.* 3:14.

Eisenberg, L. S., Dirks, D. D., Takayanagi, S., and Martinez, A. S. (1998). Subjective judgments of clarity and intelligibility for filtered stimuli with equivalent speech intelligibility index predictions. *J. Speech Lang.* 41, 327–339. doi: 10.1044/jslhr.4102.327

Federmeier, K. D., Van Petten, C., Schwartz, T. J., and Kutas, M. (2003). Sounds, words, sentences: age-related changes across levels of language processing. *Psychol. Aging* 18, 858–872. doi: 10.1037/0882-7974.18.4.858

Fitzgibbons, P. J., and Gordon-Salant, S. (1994). Age effects on measures of auditory duration discrimination. *J. Speech Lang. Hear.* 37, 662–670. doi: 10.1044/jshr.3703.662

Freyman, R. L., Terpening, J., Costanzi, A. C., and Helfer, K. S. (2017). The effect of aging and priming on same/different judgements between text and partially masked speech. *Ear Hear.* 38, 672–680. doi: 10.1097/AUD.0000000000000450

Friesen, L. M., Shannon, R. V., Başkent, D., and Wang, X. (2001). Speech recognition in noise as a function of the number of spectral channels: comparison of acoustic hearing and cochlear implants. *J. Acoust. Soc. Am.* 110, 1150–1163. doi: 10.1121/1.1381538

Fu, Q. J., and Nogaki, G. (2005). Noise susceptibility of cochlear implant users: the role of spectral resolution and smearing. *J. Assoc. Res. Otolaryngol.* 6, 19–27. doi: 10.1007/s10162-004-5024-3

Fuller, C. D., Gaudrain, E., Clarke, J. N., Galvin, J. J., Fu, Q.-J., Free, R. H., et al. (2014). Gender categorization is abnormal in cochlear implant users. *J. Asooc. Res. Otolaryngol.* 15, 1037–1048. doi: 10.1007/s10162-014-0483-7

Gaudrain, E., and Başkent, D. (2015). Factors limiting vocal-tract length discrimination in cochlear implant situations. *J. Acoust. Soc. Am.* 137, 1298–1308. doi: 10.1121/1.4908235

Gelfand, J. T., Christie, R. E., and Gelfand, S. A. (2014). Large-corpus phoneme and word recognition and the generality of lexical context in CVC word perception. *J. Speech Lang. Hear. Res.* 57, 297–307.doi: 10.1044/1092-4388(2013/12-0183)

Getzmann, S., Lewald, J., and Falkenstein, M. (2014). Using auditory pre-information to solve the cocktail-party problem: electrophysiological evidence for age-specific differences. *Front. Neurosci.* 8:413. doi: 10.3389/fnins.2014.00413

Gilbert, J. L., Tamati, T. N., and Pisoni, D. B. (2013). Development, reliability and validity of presto: a new high-variability sentence recognition test. *J. Am. Acad. Audiol.* 24, 1–11. doi: 10.3766/jaaa.24.1.4

Goldinger, S. D., Luce, P. A., Pisoni, D. B., and Marcario, J. K. (1992). Form-based priming in spoken word recognition: the roles of competition and bias. *J. Exp. Psychol. Learn. Mem. Cogn.* 18, 1211–1238. doi: 10.1037//0278-7393.18.6.1211

Green, E. J., and Barber, P. J. (1981). An auditory stroop effect with judgements of speaker gender. *Atten. Percept. Psychophys.* 30, 459–466. doi: 10.3758/bf03204842

Green, E. J., and Barber, P. J. (1983). Interference effects in an auditory stroop task: congruence and correspondence. *Acta Psychol.* 53, 183–194. doi: 10.1016/0001-6918(83)90001-x

Greenwood, D. D. (1990). A cochlear frequency-position function for several species- 29 years later. *J. Acoust. Soc. Am.* 87, 2592–2605. doi: 10.1121/1.399052

Helfner, K. S., and Jesse, A. (2015). Lexical influences on competing speech perception in younger, middle-aged, and older adults. *J. Acoust. Soc. Am.* 138, 363–376. doi: 10.1121/1.4923155

Henry, B. A., Turner, C. W., and Behrens, A. (2005). Spectral peak resolution and speech recognition in quiet: normal hearing, hearing impaired, and cochlear implant listeners. *J. Acoust. Soc. Am.* 118, 1111–1121. doi: 10.1121/1.1944567

Hervais-Adelman, A. G., Davis, M. H., Johnsrude, I. S., and Carlyon, R. P. (2008). Perceptual learning of noise vocoded words: effects of feedback and lexicality. *J. Exp. Psychol. Hum. Percept. Perform.* 34, 460–474. doi: 10.1037/0096-1523.34.2.460

Hervais-Adelman, A. G., Davis, M. H., Johnsrude, I. S., Taylor, K. J., and Carlyon, R. P. (2011). Generalization of perceptual learning of vocoded speech. *J. Exp. Psychol. Hum. Percept. Perform.* 37, 283–295. doi: 10.1037/a0020772

Howes, D. (1957). On the relation between the intelligibility and frequency of occurrence of English words. *J. Acoust. Soc. Am.* 29, 296–305.

Hughes, S. E., Hutchings, H. A., Rapport, F. L., McMahon, C. M., and Boisvert, I. (2018). Social connectedness and perceived listening effort in adult cochlear implant users: a grounded theory to establish content validity for a new patient reported outcome measure. *Ear Hear.* 39, 922–934. doi: 10.1097/AUD.0000000000000553

Huyck, J. J., Smith, R. H., Hawkins, S., and Johnsrude, I. S. (2017). Generalization of perceptual learning of degraded speech across talkers. *J. Speech Lang. Hear. Res.* 60, 3334–3341. doi: 10.1044/2017_JSLHR-H-16-0300

Janse, E., and Andringa, S. (2021). The roles of cognitive abilities and hearing acuity in older adults' recognition of words taken from fast and spectrally reduced speech. *Appl. Psycholinguist.* 42, 763–790.

Jerger, J., Silman, S., Lew, H. L., and Chmiel, R. (1993). Case studies in binaural interference: converging evidence from behavioral and electrophysiologic measures. *J. Am. Acad. Audiol.* 4, 122–131.

Johnson, J. A., Jingjing, X., and Cox, R. M. (2016). Impact of hearing aid technology on outcomes in daily life II: speech understanding and listening effort. *Ear Hear.* 37, 529–540. doi: 10.1097/AUD.0000000000000327

Kalikow, D. N., Stevens, K. N., and Elliot, L. L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *J. Acoust. Soc. Am.* 61, 1337–1351. doi: 10.1121/1.381436

Koelewijn, T., Gaudrain, E., Tamati, T., and Başkent, D. (2021). The effects of lexical content, acoustic and linguistic variability, and vocoding on voice cue perception. *J. Acoust. Soc. Am.* 150:1620. doi: 10.1121/10.0005938

Koeritzer, M. A., Rogers, C. S., Van Engen, K. J., and Peelle, J. E. (2018). The impact of age, background noise, semantic ambiguity, and hearing loss on recognition memory for spoken sentences. *J. Speech Lang.* 61, 740–751. doi: 10.1044/2017_JSLHR-H-17-0077

Król, M. E., and El-Deredy, W. (2011). When believing is seeing: the role of predictions in shaping visual perception. *Q. J. Exp. Psychol.* 64, 1743–1771. doi: 10.1080/17470218.2011.559587

Lazard, D. S., and Giraud, A. (2017). Faster phonological processing and right occipito-temporal coupling in deaf adults signal poor cochlear implant outcome. *Nat. Commun.* 8:14872. doi: 10.1038/ncomms14872

Lazard, D. S., Lee, H. J., Gaebler, M., Kell, C. A., Truy, E., and Giraud, A. L. (2010). Phonological processing in post-lingual deafness and cochlear implant outcome. *Neuroimage* 49, 3443–3451. doi: 10.1016/j.neuroimage.2009.11.013

Lazard, D. S., Vincent, C., Venail, F., Van de Heyning, P., Truy, E., Sterkers, O., et al. (2012). Pre-, per- and postoperative factors affecting performance of postlingually deaf adults using cochlear implants: a new conceptual model over time. *PLoS One* 7:e48739. doi: 10.1371/journal.pone.0048739

Lenarz, M., Sönmez, H., Joseph, G., Büchner, A., and Lenarz, T. (2012). Long-term performance of cochlear implants in postlingually deafened adults. *Otolaryngol. Head Neck Surg.* 147, 112–118. doi: 10.1177/0194599812438041

Liu, S., Del Rio, E., Bradlow, A. R., and Zeng, F.-G. (2004). Clear speech perception in acoustic and electric hearing. *J. Acoust. Soc. Am.* 116, 2374–2383. doi: 10.1121/1.1787528

Loizou, P. C. (1998). Mimicking the human ear. *IEEE Signal Process. Mag.* 15, 101–130. doi: 10.1109/TBCAS.2012.2219530

Luce, P. A., and Pisoni, D. B. (1998). Recognizing spoken words: the neighborhood activation model. *Ear Hear.* 19, 1–36.

Luce, P. A., Pisoni, D. B., and Goldinger, S. D. (1990). "Similarity neighborhoods of spoken words," in *Cognitive Models of Speech Processing: Psycholinguistic and Computational Perspectives*, ed. G. T. M. Altmann (Cambridge, MA: MIT Press), 122–147.

Lyxell, B., Andersson, J., Andersson, U., Arlinger, S., Bredberg, F., and Harder, H. (1998). Phonological representation and speech understanding with cochlear implants in deafened adults. *Scand. J. Psychol.* 39, 175–179. doi: 10.1111/1467-9450.393075

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition* 25, 71–102.

Mattys, S. L., Brooks, J., and Cooke, M. (2009). Recognizing speech under a processing load: dissociating energetic from informational factors. *Cogn. Psychol.* 59, 203–243. doi: 10.1016/j.cogpsych.2009.04.001

Mattys, S. L., Davis, M. H., Bradlow, A. R., and Scott, S. K. (2012). Speech recognition in adverse conditions: a review. *Lang. Cogn. Process.* 27, 953–978.

Mattys, S. L., White, L., and Melhorn, J. F. (2005). Integration of multiple speech segmentation cues: a hierarchical framework. *J. Exp. Psychol. Gen.* 134, 477–500. doi: 10.1037/0096-3445.134.4.477

McAuliffe, M. J., Gibson, E. M., Kerr, S. E., Anderson, T., and LaShell, P. J. (2013). Vocabulary influences older and younger listeners' processing of dysarthric speech. *J. Acoust. Soc. Am.* 134, 1358–1368. doi: 10.1121/1.4812764

McClelland, J. L., and Elman, J. L. (1986). The TRACE model of speech perception. *Cogn. Psychol.* 18, 1–86.

Mehta, A. H., Lu, H., and Oxenham, A. J. (2020). The perception of multiple simultaneous pitches as a function of number of spectral channels and spectral spread in a noise-excited envelope vocoder. *J. Assoc. Res. Otolaryngol.* 21, 61–72. doi: 10.1007/s10162-019-00738-y

Mehta, A. H., and Oxenham, A. J. (2017). Vocoder simulations explain complex pitch perception limitations experienced by cochlear implant users. *J. Assoc. Res. Otolaryngol.* 18, 789–802. doi: 10.1007/s10162-017-0632-x

Meister, H., Schreitmüller, S., Ortmann, M., Rählmann, S., and Walger, M. (2016). Effects of hearing loss and cognitive load on speech recognition with competing talkers. *Front. Psychol.* 7:301. doi: 10.3389/fpsyg.2016.00301

Milne, A. E., Bianco, R., Poole, K. C., Zhao, S., Oxenham, A. J., Billig, A. J., et al. (2020). An online headphone screening test based on dichotic pitch. *Behav. Res. Methods* 53, 1551–1562. doi: 10.3758/s13428-020-01514-0

Mitterer, H., and McQueen, J. M. (2009). Processing reduced word-forms in speech perception using probabilistic knowledge about speech production. *J. Exp. Psychol. Hum. Percept. Perform.* 35, 244–263. doi: 10.1037/a0012730

Moberly, A. C., Castellanos, I., Vasil, K., Adunka, O. F., and Pisoni, D. B. (2018a). 'Product' versus 'process' measures in assessing speech recognition outcomes in adults with cochlear implants. *Otol. Neurotol.* 39, e195–e202. doi: 10.1097/MAO.0000000000001694

Moberly, A. C., Lewis, J. H., Vasil, K. J., Ray, C., and Tamati, T. N. (2021). Bottom-up signal quality impacts the role of top-down cognitive-linguistic processing during speech recognition by adults with cochlear implants. *Otol. Neurotol.* 42, S33–S41. doi: 10.1097/MAO.0000000000003377

Moberly, A. C., Lowenstein, J. H., and Nittrouer, S. (2016). Word recognition variability with cochlear implants: "perceptual attention" versus "auditory sensitivity". *Ear Hear.* 37, 14–26. doi: 10.1097/AUD.0000000000000204

Moberly, A. C., Lowenstein, J. H., Tarr, E., Caldwell-Tarr, A., Welling, D. B., and Nittrouer, S. (2014). Do adults with cochlear implants rely on different acoustic cues for phoneme perception than adults with normal hearing? *J. Speech Lang. Hear. Res.* 57, 566–582. doi: 10.1044/2014_JSLHR-H-12-0323

Moberly, A. C., Vasil, K. J., Wucinich, T. L., Safdar, N., Boyce, L., Roup, C., et al. (2018b). How does aging affect recognition of spectrally degraded speech? *Laryngoscope* 128, S1–S16. doi: 10.1002/lary.27457

Norris, D. (1994). Shortlist: a connectionist model of continuous speech recognition. *Cognition* 52, 189–234. doi: 10.1037/0033-295X.115.2.357

Ouyang, M., Cai, X., and Zhang, Q. (2020). Aging effects on phonological and semantic priming in the tip-of-the-tongue: evidence from a two-step approach. *Front. Psychol.* 11:338. doi: 10.3389/fpsyg.2020.00338

Oxenham, A. J., and Kreft, H. A. (2014). Speech perception in tones and noise via cochlear implants reveals influence of spectral resolution on temporal processing. *Trends Hear.* 18, 1–14. doi: 10.1177/2331216514553783

Pals, C., Sarampalis, A., and Başkent, D. (2013). Listening effort with cochlear implant simulations. *J. Speech Lang. Hear. Res.* 56, 1075–1084. doi: 10.1044/1092-4388(2012/12-0074)

Pals, C., Sarampalis, A., Beynon, A., Stainsby, T., and Başkent, D. (2020). Effect of spectral channels on speech recognition, comprehension, and listening effort in cochlear-implant users. *Trends Hear.* 24, 1–15. doi: 10.1177/2331216520904617

Park, D. C., Lautenschlager, G. J., and Hedden, T. (2002). Models of visuospatial and verbal memory across the lifespan. *Psychol. Aging* 17, 299–320.

Pichora-Fuller, M. K. (2008). Use of supportive context by younger and older adult listeners: balancing bottom-up and top-down information processing. *Int. J. Audiol.* 47, S72–S82. doi: 10.1080/14992020802307404

Pichora-Fuller, M. K., Schneider, B. A., and Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *J. Acoust. Soc. Am.* 97, 593–608. doi: 10.1121/1.412282

Pichora-Fuller, M. K., and Singh, G. (2006). Effects of age on auditory and cognitive processing: implications for hearing aid fitting and audiologic rehabilitation. *Trends Hear.* 10, 29–59. doi: 10.1177/108471380601000103

Pisoni, D. B., Kronenberger, W. G., Harris, M. S., and Moberly, A. C. (2017). Three challenges for future research on cochlear implants. *World J. Otorhinolaryngol. Head Neck Surg.* 3, 240–254. doi: 10.1016/j.wjorl.2017.12.010

Prolific (2021). *Online Participant Recruitment*. Available online at: www.prolific.co (accessed on December 13, 2021)

Ramscar, M., Hendrix, P., Shaoul, C., Milin, P., and Baayen, H. (2014). The myth of cognitive decline: non-linear dynamics of lifelong learning. *Top. Cogn. Sci.* 6, 5–42. doi: 10.1111/tops.12078

Revill, K. P., and Spieler, D. H. (2012). The effect of lexical frequency on spoken word recognition in young and older listeners. *Psychol. Aging* 27, 80–87. doi: 10.1037/a0024113

Rönnberg, J., Lunner, T., Zekveld, A., Sörqvist, P., Danielsson, H., Lyxell, B., et al. (2013). The ease of language understanding (ELU) model: theoretical, empirical, and clinical advances. *Front. Syst. Neurosci.* 7:31. doi: 10.3389/fnsys.2013.00031

Rosemann, S., Giessing, C., Özyurt, J., Carroll, R., Puschmann, S., and Thiel, C. M. (2017). The contribution of cognitive factors to individual differences in understanding noise-vocoded speech in young and older adults. *Front. Hum. Neurosci.* 11:294. doi: 10.3389/fnhum.2017.00294

Ryan, J. J., Sattler, J. M., and Lopez, S. J. (2000). Age effects on wechsler adult intelligence scale-III subtests. *Arch. Clin. Neuropsychol.* 15, 311–317.

Salthouse, T. A. (1993). "Effects of aging on verbal abilities: examination of the psychometric literature," in *Language, Memory, and Aging*, eds L. L. Light and D. M. Burke (Cambridge, MA: Cambridge University Press), 17–35.

Salthouse, T. A. (2004). What and when of cognitive aging. *Curr. Dir. Psychol. Sci.* 13, 140–144.

Samuel, A. G. (1981). Phonemic restoration: insights from a new methodology. *J. Exp. Psychol. Gen.* 110, 474–494. doi: 10.1037//0096-3445.110.4.474

Santurette, S., and Dau, T. (2007). Binaural pitch perception in normal-hearing listeners and hearing impaired listeners. *Hear. Res.* 223, 29–47.

Savin, H. B. (1963). Word-frequency effect and errors in the perception of speech. *J. Acoust. Soc. Am.* 35, 200–206.

Schmiedt, R. A. (2010). The physiology of cochlear presbycusis. *Aging Audit. System* 34, 9–38.

Schneider, B. A., Avivi-Reich, M., and Daneman, M. (2016). How spoken language comprehension is achieved by older listeners in difficult listening situations. *Exp. Aging Res.* 42, 31–49. doi: 10.1080/0361073X.2016.1108749

Sheldon, S., Pichora-Fuller, M. K., and Schneider, B. A. (2008). Priming and sentence context support listening to noise-vocoded speech by younger and older adults. *J. Acoust. Soc. Am.* 123, 489–499. doi: 10.1121/1.2783762

Signoret, C., Johnsrude, I., Classon, E., and Rudner, M. (2018). Combined effects of form- and meaning-based predicability on perceived clarity of speech. *J. Exp. Psychol. Hum. Percept. Perform.* 44, 277–285. doi: 10.1037/xhp0000442

Signoret, C., and Rudner, M. (2019). Hearing impairment and perceived clarity of predictable speech. *Ear Hear.* 40, 1140–1148. doi: 10.1097/AUD.0000000000000689

Sohoglu, E., Peelle, J. E., Carlyon, R. P., and Davis, M. H. (2012). Predictive top-down integration of prior knowledge during speech perception. *J. Neurosci.* 32, 8443–8453. doi: 10.1523/JNEUROSCI.5069-11.2012

Sohoglu, E., Peelle, J. E., Carlyon, R. P., and Davis, M. H. (2014). Top-down influences of written text on perceived clarity of degraded speech. *J. Exp. Psychol. Hum. Percept. Perform.* 40, 186–199. doi: 10.1037/a0033206

Sommers, M. S. (1996). The structural organization of the mental lexicon and its contribution to age-related declines in spoken-word recognition. *Psychol. Aging* 11, 333–341. doi: 10.1037//0882-7974.11.2.333

Sommers, M. S., and Danielson, S. M. (1999). Inhibitory processes and spoken word recognition in young and older adults: the interaction of lexical competition and semantic context. *Psychol. Aging* 14, 458–472. doi: 10.1037//0882-7974.14.3.458

Sommers, M. S., Kirk, K. I., and Pisoni, D. B. (1997). Some considerations in evaluating spoken word recognition by normal-hearing, noise-masked normal hearing, and cochlear implant listeners. I: the effects of response format. *Ear Hear.* 18, 89–99. doi: 10.1097/00003446-199704000-00001

Spieler, D. H., and Balota, D. A. (2000). Factors influencing word naming in younger and older adults. *Psychol. Aging* 15, 225–231. doi: 10.1037//0882-7974.15.2.225

Takayanagi, S., Dirks, D. D., and Moshfegh, A. (2002). Lexical and talker effects on word recognition among native and non-native listeners with normal and impaired hearing. *J. Speech Lang. Hear. Res.* 45, 585–597. doi: 10.1044/1092-4388(2002/047)

Taler, V., Aaron, G. P., Steinmetz, L. G., and Pisoni, D. B. (2010). Lexical neighborhood density effects on spoken word recognition and production in healthy aging. *J. Gerontol.* 65B, 551–560. doi: 10.1093/geronb/gbq039

Tamati, T. N., Gilbert, J. L., and Pisoni, D. B. (2011). Individual differences in spoken word recognition: regional dialect variation. *J. Acoust. Soc. Am.* 129:2682. doi: 10.3766/jaaa.25.9.9

Tamati, T. N., and Moberly, A. C. (2021). Talker adaptation and lexical difficulty impact word recognition in adults with cochlear implants. *Audiol. Neurotol.* Online ahead of print., doi: 10.1159/000518643

Tamati, T. N., Sijp, L., and Başkent, D. (2020b). Talker variability in word recognition under cochlear implant simulation: does talker gender matter? *J. Acoust. Soc. Am.* 147:EL370. doi: 10.1121/10.0001097

Tamati, T. N., Ray, C., Vasil, K. J., Pisoni, D. B., and Moberly, A. C. (2020a). High- and low-performing adult cochlear implant users on high-variability sentence recognition: differences in auditory spectral resolution and neurocognitive functioning. *J. Am. Acad. Audiol.* 31, 324–335. doi: 10.3766/jaaa.18106

Tamati, T. N., Vasil, K. J., Kronenberger, W. G., Pisoni, D. B., Moberly, A. C., and Ray, C. (2021). Word and nonword reading efficiency in postlingually deafened adult cochlear implant users. *Otol. Neurotol.* 42, E272–E278. doi: 10.1097/MAO.0000000000002925

Tun, P. A., Williams, V. A., Small, B. J., and Hafter, E. R. (2012). The effects of aging on auditory processing and cognition. *Am. J. Audiol.* 21, 344–350.

Valencia-Laver, D. L., and Light, L. L. (2000). The occurrence of causal bridging and predictive inferences in young and older adults. *Discourse Process.* 30, 27–56.

Verhaeghen, P. (2003). Aging and vocabulary score: a meta-analysis. *Psychol. Aging* 18, 332–339. doi: 10.1037/0882-7974.18.2.332

Vitevitch, M. S., and Luce, P. A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *J. Mem. Lang.* 40, 374–408. doi: 10.1006/brln.1999.2116

Wagner, A., Pals, C., de Blecourt, C. M., Sarampalis, A., and Başkent, D. (2016). Does signal degradation affect top-down processing of speech? *Adv. Exp. Med. Biol.* 894, 297–306. doi: 10.1007/978-3-319-25474-6_31

Wang, N., Kreft, H., and Oxenham, A. J. (2012). Vowel enhancement effects in cochlear-implant users. *J. Acoust. Soc. Am.* 131, EL421–EL426. doi: 10.1121/1.4710838

Wild, C. J., Yusuf, A., Wilson, D. E., Peelle, J. E., Davis, M. H., and Johnsrude, I. S. (2012). Effortful listening: the processing of degraded speech depends critically on attention. *J. Neurosci.* 32, 14010–14021. doi: 10.1523/JNEUROSCI.1528-12.2012

Wingfield, A., Alexander, A. H., and Cavigelli, S. (1994). Does memory constrain utilization of top-down information in spoken word recognition? Evidence from normal aging. *Lang. Speech.* 37, 221–235. doi: 10.1177/002383099403700301

Winn, M. B., Chatterjee, M., and Idsardi, W. J. (2012). The use of acoustic cues for phonetic identification: effects of spectral degradation and electric hearing. *J. Acoust. Soc. Am.* 131, 1465–1479. doi: 10.1121/1.3672705

Winn, M. B., Edwards, J. R., and Litovsky, R. Y. (2015). The impact of auditory spectral resolution on listening effort revealed by pupil dilation. *Ear Hear.* 36, 153–165. doi: 10.1097/AUD.0000000000000145

Winn, M. B., Edwards, J. R., and Litovsky, R. Y. (2016). The impact of auditory spectral resolution on listening effort revealed by pupil dilation. *Ear Hear.* 36, 153–165.

Winn, M. B., and Teece, K. H. (2021). Listening effort is not the same as speech intelligibility score. *Trends Hear.* 25, 1–26. doi: 10.1177/23312165211027688

Won, J. H., Drennan, W. R., and Rubinstein, J. T. (2007). Spectral-ripple resolution correlates with speech perception in noise in cochlear implant users. *J. Assoc. Res. Otolaryngol.* 8, 384–392. doi: 10.1007/s10162-007-0085-8

Wu, M., Li, H., Hong, Z., Xian, X., Li, J., Wu, X., et al. (2012). Effects of aging on the ability to benefit from prior knowledge of message content in masked speech recognition. *Speech Commun.* 54, 529–542. doi: 10.1016/j.specom.2011.11.003

Conflict of Interest: AM and TT have received grant funding support from Cochlear Americas for unrelated investigator-initiated research studies. AM has served as a paid consultant for Cochlear Americas and Advanced Bionics and is CMO and on Board of Directors for Otologic Technologies.

The remaining 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 | Frontiers in Psychology

ORIGINAL RESEARCH
published: 11 April 2022
doi: 10.3389/fpsyg.2022.841466

One Size Does Not Fit All: Examining the Effects of Working Memory Capacity on Spoken Word Recognition in Older Adults Using Eye Tracking

Gal Nitsan[1,2†], **Karen Banai**[2†] and **Boaz M. Ben-David**[1,3,4*†]

[1] Baruch Ivcher School of Psychology, Reichman University (IDC), Herzliya, Israel, [2] Department of Communication Sciences and Disorders, University of Haifa, Haifa, Israel, [3] Department of Speech-Language Pathology, University of Toronto, Toronto, ON, Canada, [4] Toronto Rehabilitation Institute, University Health Networks, Toronto, ON, Canada

Edited by:
Leah Fostick,
Ariel University, Israel

Reviewed by:
Subong Kim,
Purdue University, United States
Likan Zhan,
Beijing Language and Culture University, China

***Correspondence:**
Boaz M. Ben-David
boaz.ben.david@idc.ac.il

†ORCID:
Gal Nitsan
orcid.org/0000-0001-8567-2823
Karen Banai
orcid.org/0000-0002-2990-0470
Boaz M. Ben-David
orcid.org/0000-0002-0392-962X

Specialty section:
This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Psychology

Received: 22 December 2021
Accepted: 14 March 2022
Published: 11 April 2022

Citation:
Nitsan G, Banai K and Ben-David BM (2022) One Size Does Not Fit All: Examining the Effects of Working Memory Capacity on Spoken Word Recognition in Older Adults Using Eye Tracking. Front. Psychol. 13:841466. doi: 10.3389/fpsyg.2022.841466

Difficulties understanding speech form one of the most prevalent complaints among older adults. Successful speech perception depends on top-down linguistic and cognitive processes that interact with the bottom-up sensory processing of the incoming acoustic information. The relative roles of these processes in age-related difficulties in speech perception, especially when listening conditions are not ideal, are still unclear. In the current study, we asked whether older adults with a larger working memory capacity process speech more efficiently than peers with lower capacity when speech is presented in noise, with another task performed in tandem. Using the Eye-tracking of Word Identification in Noise Under Memory Increased Load (E-WINDMIL) an adapted version of the "visual world" paradigm, 36 older listeners were asked to follow spoken instructions presented in background noise, while retaining digits for later recall under low (single-digit) or high (four-digits) memory load. In critical trials, instructions (e.g., "point at the candle") directed listeners' gaze to pictures of objects whose names shared onset or offset sounds with the name of a competitor that was displayed on the screen at the same time (e.g., candy or sandal). We compared listeners with different memory capacities on the time course for spoken word recognition under the two memory loads by testing eye-fixations on a named object, relative to fixations on an object whose name shared phonology with the named object. Results indicated two trends. (1) For older adults with lower working memory capacity, increased memory load did not affect online speech processing, however, it impaired offline word recognition accuracy. (2) The reverse pattern was observed for older adults with higher working memory capacity: increased task difficulty significantly decreases online speech processing efficiency but had no effect on offline word recognition accuracy. Results suggest that in older adults, adaptation to adverse listening conditions is at least partially supported by cognitive reserve. Therefore, additional cognitive capacity may lead to greater resilience of older listeners to adverse listening conditions. The differential effects documented by eye movements and accuracy highlight the importance of using both online and offline measures of speech processing to explore age-related changes in speech perception.

Keywords: speech perception, working memory, aging, word recognition, eye-tracking, visual world paradigm, cognitive hearing science

INTRODUCTION

A recent report by the World Health Organization (2021) emphasizes the importance of functional ability as a key to healthy aging. It suggests that preserving the abilities to build and maintain relationships and to grow learn and make decisions all promote well-being and healthy aging. These functional abilities depend heavily on successful speech perception. Indeed, difficulties understanding speech are one of the most prevalent complaints among older adults, especially in daily listening situations when listening conditions are not ideal (e.g., Abrams and Farrell, 2011). Although hearing deficits are a main source of difficulty in speech perception (Humes et al., 1994; Humes, 2021), successful speech perception also depends on the interaction of bottom-up hearing related factors and top-down linguistic and cognitive processes (Sommers, 2005; Zekveld et al., 2006; Pichora-Fuller, 2008; Rogers and Peelle, 2021). Furthermore, difficulties in speech perception are also observed among older adults with relatively preserved hearing (Sommers and Danielson, 1999; Fostick et al., 2013; Lash et al., 2013). Our goal is to test whether older listeners with a higher working memory capacity process speech in adverse conditions more efficiently than peers with lower capacity.

Previous studies in cognitive hearing science reported an association between individual differences in cognitive factors and differences in speech perception, even in young and healthy hearing populations. One consistent finding is that these differences are pronounced mainly when using complex testing materials (i.e., sentences, connected discourse comprehension, conversational situations; e.g., Heinrich et al., 2015; Dryden et al., 2017; Meister, 2017). For example, by comparing performance of older listeners across a wide range of speech perception tests differing in complexity, Heinrich et al. (2015) showed that the contribution of cognition increases as the complexity of the speech perception task increases. That is, for older adults, cognitive factors predict sentence perception to a larger extent than single spoken word perception. Of the many cognitive constructs tested, working memory has been widely recognized as related to differences in speech perception abilities, especially in adverse listening condition for older adults (see Akeroyd, 2008; Besser et al., 2013; Dryden et al., 2017 for relevant reviews). In particular, the storage and processing components of working memory play an important role in sentence processing as the listener is required to correctly encode the speech sounds, identify them as words, and then retain the string of words in memory until the sentence is fully heard (Daneman and Carpenter, 1980; Pichora-Fuller et al., 1995; Daneman and Merikle, 1996; Rönnberg et al., 2008). Working memory has also been linked with inhibition of irrelevant information (Awh and Vogel, 2008). The latter is directly related to successful speech perception, where the listener needs to continuously inhibit irrelevant lexical items from his/her mental lexicon to allow correct word recognition. For example, Janse (2012) showed that when speech is presented in background noise, poor inhibitory abilities lead to greater interference by the competing noise which impairs speech perception of older adults.

Contrary to the agreement regarding the association between working memory, aging and spoken sentence processing, only little and mixed evidence is available on this association at the single word level. This is of special importance because lexical ambiguities frequently occur in daily life. For example, cell phones may distort a critical portion of the incoming signal. Consider the sentence "Grandpa! Have you seen the *dog*?" The word *dog* may be mistaken for *doll* (as the two share onset sounds, e.g., see Allopenna et al., 1998; Onset Cohort model, Marslen-Wilson, 1990; Shortlist, Norris, 1994) which can lead to miscommunication with severe consequences on future social participation. Despite these challenges, listeners appear to recognize words with little effort. Moreover, studying the effects of working memory at the single word level has theoretical implications. As spoken sentence processing involves many intervening factors, they may inflate the effects of working memory. Among the abilities necessary to understand sentences are sustained attention for the duration of the sentence and maintaining a running memory of the input to relate what is being heard to what has just been heard and to integrate it with what is about to be heard (Ayasse et al., 2017; Harel-Arbeli et al., 2021). Further, spoken context processing may be more influenced by linguistic experience and vocabulary than the processing of a single spoken word (Stine-Morrow et al., 2006; Borovsky et al., 2012; Ben-David et al., 2015; Kavé and Halamish, 2015). Thus, the aforementioned effects of working memory on the sentence level may reflect other processes. However, if effects are found at a single word level, that would indicate that working memory is involved at very early and basic levels of lexical access.

There is mixed evidence in the literature with regards to the effects of working memory and single spoken word recognition in aging. For example, Heinrich and Knight (2016) found that older adults' performance in a visual working memory task significantly correlated with their performance on a word in noise (WIN) recognition task, irrespective of the noise level [both in low and high signal to noise ratios (SNRs)]. Gordon-Salant and Cole (2016) found similar results with both young and older adults, correlating auditory working memory capacity, with single-word recognition in noise. Conversely, other studies failed to find this correlation on the single word level.

For example, Parbery-Clark et al. (2011) did not find auditory working memory performance to correlate with performance on the WIN test for older adults with and without hearing loss. Similar findings were reported by Smith and Pichora-Fuller (2015) who failed to find a correlation between auditory and visual working memory performance and scores on the WIN test.

A possible explanation for these contradictory findings may stem from the use of offline measures to gauge word recognition (such as accuracy or SNR to achieve 50% recognition). Offline measures test the result of successful (or unsuccessful) word recognition, after the entire word has been heard, processed and a response has been made. It gages the final outcome of the process, and it cannot reveal the early processes underlying online speech processing. Additionally, previous works showed that this association between working memory and word recognition might differ depending on whether verbal or non-verbal measures of working memory are used and the modality of working memory tasks: auditory or visual. There is some evidence to suggest that auditory working memory plays a greater role in speech perception than visual working memory (Baldwin and Ash, 2011; Smith and Pichora-Fuller, 2015; Smith et al., 2016; Kim et al., 2020). Finally, none of the studies listed above tapped cognitive resources while performing speech recognition task, they only measured the correlation between performance on these separate measures. Direct manipulations of the memory load can allow us to better assess the causal relationship between reduced cognitive capacity and spoken word processing in aging.

The Current Study

In the current study, we examined the role of working memory capacity in spoken word recognition in adverse conditions for older adults. We hypothesized that older listeners with a larger working memory capacity would process speech more efficiently than their peers with a lower capacity; this is tested when speech is presented in noise, with another working memory demanding task performed in tandem. As listeners with lower working memory capacity already have fewer cognitive resources, we expect that the effects of increased load would be especially detrimental for their spoken word processing. This was tested using an adapted version of the eye-tracking "visual world" paradigm, coined the Eye-tracking of Word Identification in Noise Under Memory Increased Load (E-WINDMIL; Hadar et al., 2016; Nitsan et al., 2019). This paradigm was found to have significant test retest reliability for older adults (Baharav et al., 2021). In the E-WINDMIL listeners are instructed to press on one of four objects displayed on the monitor in response to spoken instructions presented in noise. They performed the speech recognition task while retaining for later recall either low (a single spoken digit) or high (four-digits) memory-load. In experimental trials, the named object shares phonology with the name of one of other presented objects. We compared eye-fixations on the named spoken target word, relative to fixations on its phonological competitor, as the word unfolded in time (online). Studies demonstrated that under adverse conditions, spoken word recognition dynamics differ significantly between situations in which the names of the target objects and competitors share an onset and those in which they share an offset in young adults (McQueen and Huettig, 2012; Brouwer and Bradlow, 2016; Hadar et al., 2016), young adults with higher and lower working memory capacity (Nitsan et al., 2019), older adults (Ben-David et al., 2011), and hearing impaired listeners (McMurray et al., 2017). Therefore, the two types of phonological competition will be analyzed separately in the present study, and our analysis will focus on the onset overlap trials.

MATERIALS AND METHODS

Participants

Thirty-eight older adults were recruited from Reichman University's (IDC) older adult volunteer pool. Of this group, two were excluded due to loss of eye-tracking signal. Thus, the final group for analysis included 36 participants (M_{age} = 67.9 years, SD = 3.2, 20 females). All participants met the research inclusion criteria (see **Table 1** for details). Participants were paid 35 NIS (approximately $10) for their participation. The number of participants was based on previous studies using a highly similar paradigm (Nitsan et al., 2019; Baharav et al., 2021).

Working Memory

Working memory span was assessed using the forward digit span subtest (Hebrew version of WAIS-III (*Goodman, 2001). To measure the participants' memory spans, sets of random digits were read aloud at a rate of one per second and they were instructed to repeat them, in the order in which they had been heard. The first list contained two digits, and the number of digits presented for recall increased gradually until the individual was no longer able to recall correctly. Two lists of each length were presented (e.g., two lists of three digits and then two lists of four digits, etc.). A single point was assigned to each list the participant correctly remembered (range of 0–16). Participants were divided into two subgroups based on their digit span scores (range 5–13). The lower-capacity subgroup consisted of 18 participants with a span score of five to nine (M = 7.9, SD = 1.1). The higher-capacity subgroup consisted of 18 participants with a span score of 10–13 (M = 10.8, SD = 0.89). The two groups did not differ in most individual characteristics, but differed on hearings status, with slightly better audiometric thresholds for the lower-capacity group (see **Table 2**).

Procedure

The experiment was administered individually in a dedicated sound attenuated booth (Iac Acoustics). Participants were seated 60 cm from a computer screen with their head placed in a customized chin rest to stabilize head movement. Each participant's dominant eye was calibrated to ensure that throughout the course of the trial participants' online eye-gaze position was recorded. A table mounted SR EyeLink 1000 eye-tracker in the "tower mount" configuration was used (SR Research Ltd., Kanata, ON, Canada). Eye-gaze position was recorded *via* the EyeLink software at a rate of 500 Hz.

During the experiment, two tasks were presented: spoken word recognition and digit recall (working memory load), conducted in a dual task situation. Trials began with a visual cue

TABLE 1 | Inclusion criteria for participant recruitment.

	Inclusion criteria
Language background	Proficient Hebrew speakers (no early bilinguals were included) assessed by a self-report and a score within the normal range in the WAIS-III Hebrew vocabulary subtest.
Hearing	Symmetrical air-conduction hearing thresholds expressed as pure tone averages (0.5, 1, and 2 kHz) of ≤25 dB HL in each ear, no reported history of auditory pathology. Audiometric assessment was conducted using a MAICO MA-51 audiometer using standard audiometric procedures in a sound attenuating testing booth.
Vision	Normal or corrected to normal visual acuity and color vision assessed by the Landolt-C charts and the Ishihara charts.
Cognition	Clinically normal scores for their age range on the MoCA cognitive screening test (≥22), and on the forward (≥5) and backward (≥4) digit span subtests (Hebrew version of WAIS-III; *Goodman, 2001).

TABLE 2 | Background information by working memory capacity group.

	Lower capacity	Higher capacity	Group comparison
N	18	18	
Age: mean (SD), years	68.5 (2.7)	67.5 (3.6)	t = 0.95, p = 0.35
Gender: count, females	9	11	χ^2 = 0.25, p = 0.62
Hearing: mean (SD), 0.5, 1, and 2 kHz	15.1 (4.4)	18.7 (4.1)	t = 2.5, p = 0.02
Years of education: mean (SD)	16.5 (3.2)	16.2 (3.4)	t = 0.25, p = 0.8
MoCA: mean (SD)	25.5 (1.7)	26.3 (2.5)	t = 1.2, p = 0.25
Digit span: mean (SD)	7.9 (1.1)	10.8 (0.9)	t = 8.5, p < 0.001

of a black "play" triangle centered on the screen, immediately followed by the auditory presentation of the digit(s) preload through headphones, either one digit: low-load condition, or four digits: high-load condition. Participants were told to memorize these digits (in the order presented) for later recall. Then, a 3 × 3 grid with the four images would appear (**Figure 1A**). Participants were given 2 s to familiarize themselves with the four objects and their position on the computer screen. At the end of these 2 s a flickering fixation cross would appear in the center of the screen, once participants pressed the fixation cross to initiate the trial, the instruction sentence "point at the ___ [target word]," would be presented binaurally *via* the headphones. Selection of a named object was indicated by touching the object picture on the touch screen. Following the participant selection of a stimulus, a visual feedback signal: red highlight for an incorrect answer or green highlight for a correct answer, would appear in the square of the selected image. The visual display would then clear and a visual cue of a black circle would appear in the screen signaling participants to recall aloud the digit(s) preload from the beginning of the trial (**Figure 1B** illustrates the sequence of displays presented in each trial). The experimenter would then code the response (either correct or incorrect) online. Participants were instructed that speed and accuracy of both the object selection and digit recall were equally important. Participants completed 68 trials split into two trial blocks of the two memory load conditions (Low-load: one digit and High-load: four digits).

Each condition contained 34 trials of which two were practice trials, and 32 were experimental trials. The 32 trials in each condition were split such that 16 were "filler": target object name did not share any phonology with the surrounding objects, and 16 were "critical" trials in which 8 were phonological onset competitors (e.g., /aʁ.nav/–/aʁ.gaz/ rabbit and box, respectively), and 8 were phonological offset competitors (e.g., /xa.lon/–/ba.lon/ window and balloon, respectively).

Stimuli

Auditory Stimuli

Stimuli were taken from Nitsan et al. (2019), and contained both the object names of the visual stimuli, and the sentence "point at the ___ [target word]" in Hebrew using a plural generic form. All object names were disyllabic. Average target word duration, including the Hebrew article *ha-* (the), was 1078 ms, SD = 91 ms (Nitsan et al., 2019). Considering that the definite article in Hebrew is not a separate word but a prefix, the target word onset was adjusted for each word separately (see Hadar et al., 2016). The root mean square (RMS) intensity was equated across all recorded sentences. Files were mixed with a continuous steady-state speech spectrum noise (for full details, see Ezzatian et al., 2010) at a fixed 0 dB SNR based off of values for discrimination timeline in Ben-David et al. (2012). Stimuli were presented binaurally at 50 dB above individual pure tone average (PTA) *via* a MAICO MA-51 audiometer using TDH 39 supra-aural headphones.

Visual Display

On each trial participants were presented with a 3 × 3 grid with four images of objects positioned at the grid corners. The stimuli (images) were previously used by Hadar et al. (2016), Nitsan et al. (2019), and Baharav et al. (2021) studies and were confirmed as clearly identifiable and highly familiar. In all trials one of the four image names represented the spoken target word and a second image's name was a phonological competitor: sharing the initial syllable (onset sound overlap) or the final syllable (offset sound overlap) with the spoken target word. The remaining two objects presented on screen represented words that were phonologically and semantically unrelated to both the

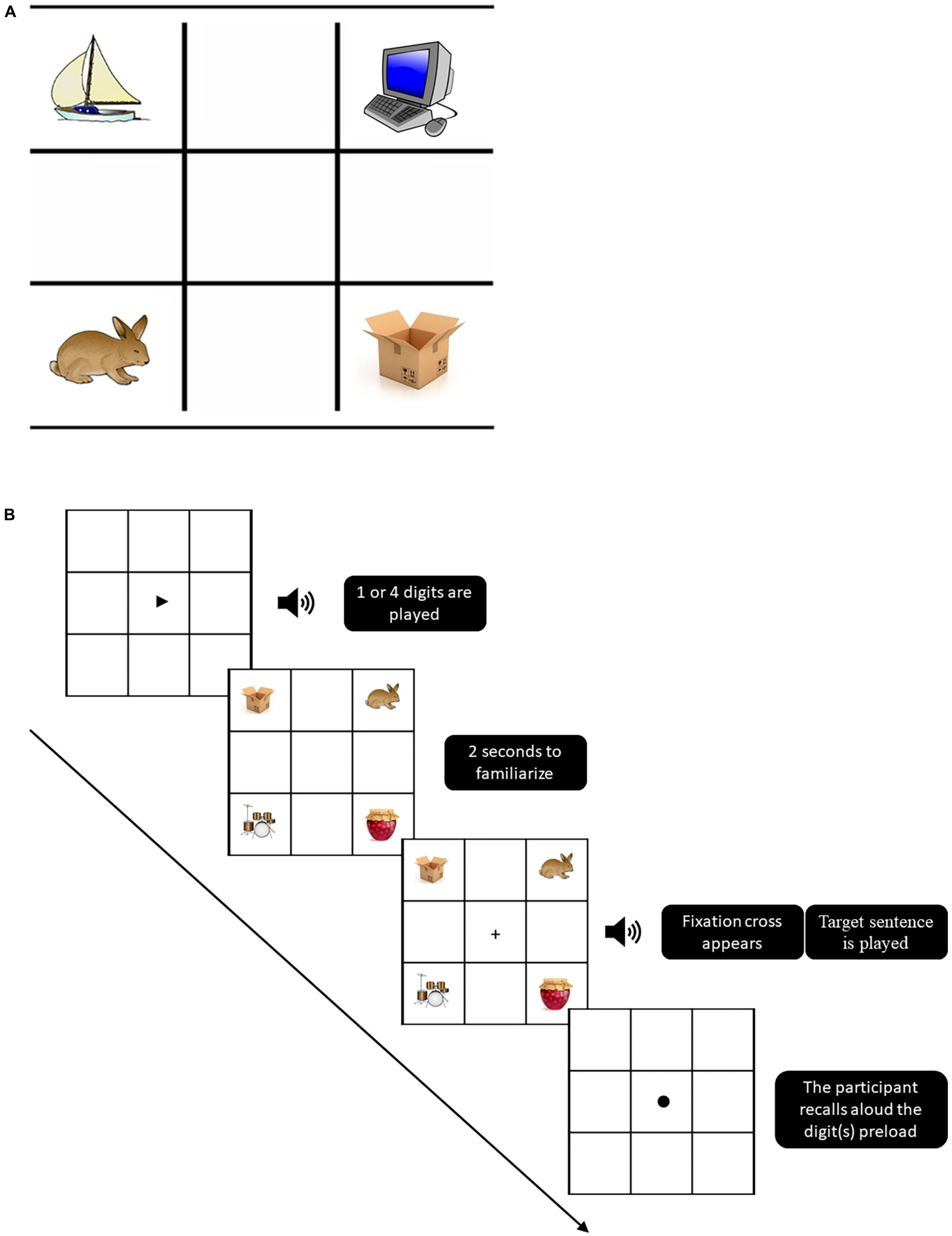


FIGURE 1 | (A) Example of an experimental display in Hebrew: the target word, /aʁ.nav/ (rabbit), is represented in the bottom left corner. The onset phonological competitor /aʁ.gaz/ (box), is represented in the bottom right corner. /si.ʁa/ and /max.ʃev/ (boat and computer, respectively) are unrelated distractors.
(B) Experimental task design: the sequence of displays presented in each trial.

target spoken word and phonological competitor. In critical trials the target word to be recognized was one of the two sound-sharing images. In addition to critical trials, filler trials were used to diminish participant expectation of phonetic resemblance between the words. Objects were presented twice during the experiment, once as a critical trial, and once as a filler trial in which one of the two phonologically "unrelated" items was used as the target word. To prevent implicit spatial learning, object positions on the screen were randomly rotated at each presentation (Farris-Trimble and McMurray, 2013).

Statistical Analysis

Growth curve analysis (GCA) (Mirman et al., 2008) was used to analyze the time course of fixation from word onset to 1200 ms after word onset (i.e., when target fixations had plateaued). To express listeners' ability to discriminate the target word from its phonological competitor, we calculated *target discrimination scores* (following: Arnold et al., 2003; Kaiser and Trueswell, 2008; Brown-Schmidt, 2009; Ben-David et al., 2011). To generate the *target discrimination scores*, the proportion of fixations on the competitor was subtracted from the proportion of fixations on the target within 20 ms time bins, starting from the word onset to 1200 ms post word onset. In this measure, the higher the value the better listeners can discriminate the target from its phonological competitor; values approaching zero reflect an inability to discriminate between the target and competitor words. The overall time course of *target discrimination score* was captured with a second-order (quadratic) orthogonal polynomial with fixed effects of capacity group (low vs. high capacity) and working memory load (low vs. high load) on all time terms, and participant random effects on all time terms. The low working memory load condition and the high-capacity group was treated as the reference (baseline) and relative parameters estimated for the high working memory load condition and low-capacity group. These baseline conditions were selected to reflect preserved cognition and the easiest listening condition in this study. The two phonological competition conditions (onset and offset overlap) were modeled separately. Statistical significance (p-values) for individual parameter estimates was assessed using the normal approximation.

Offline response accuracy was analyzed using multilevel modeling (Heck et al., 2013) with fixed effects of capacity group (low vs. high capacity) and working memory load (low vs. high load) on response accuracy, participants were included as random effects. All analyses were carried out in SPSS version 25.

RESULTS

Onset Overlap – Accuracy of Behavioral Responses

Eye-gaze analysis included only trials in which participants both correctly selected the corresponding object on the visual display (indicating correct spoken word recognition) and correctly recalled the working memory load digits (indicating correct digit recall). **Table 3** shows mean accuracy performance across conditions and reflects differential effect of increased load for each working memory capacity group. In the low-capacity group, increasing memory load from one (low load) to four (high load) digits significantly reduced their response accuracy. However, the same increase in task demands did not change response accuracy for the high-capacity group. These differences were confirmed using a multilevel model as detailed in the statistical analysis section. The analysis revealed a main effect of load $F(1,34) = 13.21$, $p = 0.001$ on response accuracy and a significant interaction of load and span $F(1,34) = 6.60$, $p = 0.015$. LSD-corrected pairwise comparisons were conducted to clarify the interaction. It confirmed that the interaction of working memory load and capacity group was due to participants from the low-capacity group being significantly less accurate when a high load was present compared to when a low load present $F(1,34) = 19.25$, $p < 0.001$. In the high-capacity group accuracy did not differ significantly between the two load conditions $F(1,34) = 0.57$, $p = 0.456$.

TABLE 3 | Mean percentage (and SEs) of trials in which target word was correctly selected and digits were correctly recalled.

	Low WM capacity	High WM capacity
Low WM load	100% (0.0)	97.9% (1.13)
High WM load	83.8% (4.85)	95.1% (2.05)

Low and high working memory (WM) load, indicate the two preload conditions, one and four digit/s, respectively.

Onset Overlap – Eye Gaze

The data and model fits are shown in **Figure 2**. Visual inspection of the left panel of **Figure 2A** shows that for listeners with lower working memory capacity, increasing task demands from low to high working memory load did not change the pattern and rate of target discrimination scores. In contrast, the right panel of **Figure 2B** indicates that for listeners with higher working memory capacity, increasing the working memory load delayed processing, suggesting less efficient spoken word processing. The results of the analysis as shown in **Table 4** confirm these observations. The analysis shows a significant effect of capacity group on the intercept and all polynomial time terms (linear and quadratic), suggesting that the rate of accumulating evidence from the unfolding spoken word differs between the two capacity groups. Working memory load was also found to have a significant effect on the linear and quadratic time terms, again suggesting a difference in evidence accumulation. Most importantly, the interaction between working memory load and capacity group on the linear and quadratic time terms was found to be significant.

A follow up model conducted separately for each capacity group revealed the source of this interaction (**Table 5**). In the low-capacity group, no significant effect of working memory load was evident; whereas in the high-capacity group the effect of working memory load on the linear and quadratic time terms was significant. The significant effect of working memory load on the linear term indicates a steeper slope, faster accumulation of evidence, under low working memory load. The effect of working memory load on the quadratic term further showcases

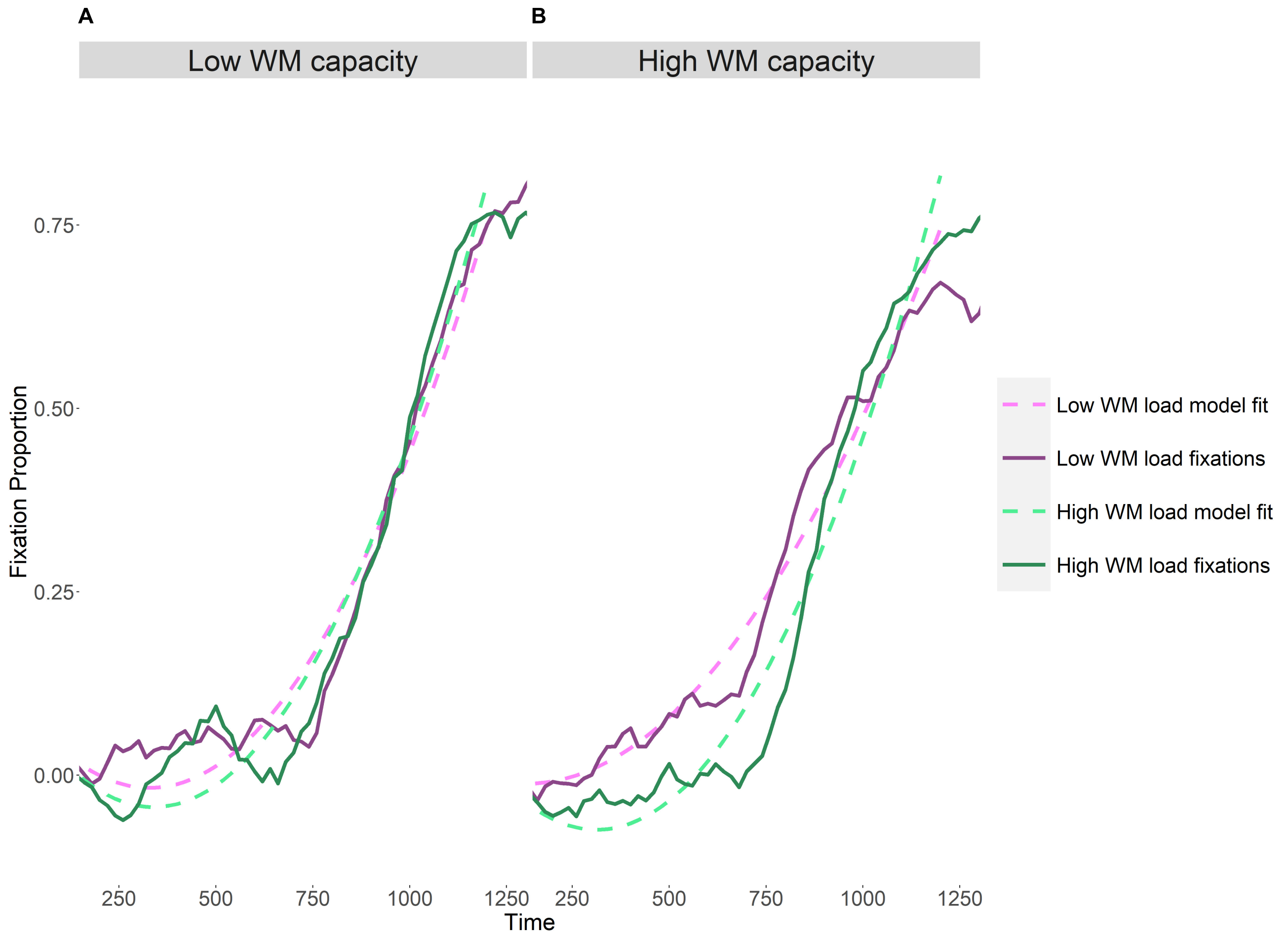


FIGURE 2 | Time-course of *target discrimination scores*. Fixations are shown as a subtraction, with phonological competitor fixations subtracted from the target fixations. The model fits (dashed lines) are plotted along with the observed fixation data (solid lines). Left panel **(A)** show the proportion of fixations for each load condition, one and four digits, respectively, for the low WM capacity group and panel **(B)** show the high WM capacity group.

TABLE 4 | Results of growth curve analysis (GCA) – onset overlap.

	Term	Estimate	SE	*t*-Value	*p*
Participant group (WM capacity)	Intercept	0.094	0.041	2.28	0.025
	Linear	−0.502	0.105	−4.79	< 0.001
	Quadratic	0.004	<0.001	4.26	< 0.001
Working memory load	Intercept	0.037	0.027	1.37	0.170
	Linear	−0.541	0.105	−5.16	< 0.001
	Quadratic	0.001	<0.001	5.62	< 0.001
Participant group (WM capacity) × working memory load	Intercept	−0.041	0.038	−1.07	0.287
	Linear	0.429	0.148	2.89	0.004
	Quadratic	−0.001	<0.001	−2.86	0.004

a difference in the change in the rate of evidence accumulation between the two load conditions.

In sum, eye-movement analyses of onset overlap trials indicate that for the higher working memory capacity group, an increase in working memory load slowed spoken word processing. This slowdown was not evident for the lower working memory capacity group.

The same analyses conducted for the onset overlap trials were replicated for the offset overlap condition. The effects noted in the eye-gaze for the onset overlap condition were not found in the offset overlap, but for the effect of working memory capacity group. Analysis of accuracy of behavioral responses in the offset overlap revealed that increasing memory load from one to four digits significantly reduced listeners' response accuracy regardless

TABLE 5 | Results of growth curve analysis (GCA) conducted separately for each WM capacity group.

	Term	Estimate	SE	*t*-Value	*p*
Low-capacity group – working memory load	Intercept	−0.004	0.029	−0.12	0.902
	Linear	−0.112	0.112	−0.10	0.319
	Quadratic	0.000	0.000	1.47	0.141
High-capacity group – working memory load	Intercept	0.037	0.025	1.49	0.137
	Linear	−0.541	0.967	−5.59	<0.001
	Quadratic	0.000	<0.001	6.09	<0.001

of span group membership. Additionally, is shows that overall listeners from the high-capacity group had higher response accuracy compared to listeners from the low-capacity group. The low-capacity group had a greater reduction in response accuracy compared to the high-capacity group. The full analysis is provided in **Appendix A**.

DISCUSSION

We investigated the efficacy with which older adults with different working memory capacities process a spoken word in adverse conditions. Both online (eye-tracking) and offline (behavioral response accuracy) measures for spoken word recognition were used. Consistent with our hypothesis, we report that increasing task demands had different effects on listeners with higher vs. lower working memory capacity when the target and competitor shared onset sounds. Overall, listeners with higher working memory capacity were able to maintain their offline response accuracy at maximal performance even when they were asked to retain four digits for later recall instead of only one digit (high and low working memory load, respectively). However, this increase in working memory load had slowed down their online spoken word processing, suggesting less efficient processing at the single word level. For listeners with lower working memory capacity, increasing task demands significantly reduced offline recognition accuracy (from ∼100 to ∼80%), with no effect on online word processing. In the offset sound sharing condition, increasing memory load from one to four digits significantly reduced listeners' offline response accuracy regardless of their working memory capacity without affecting their online processing.

Our results present a clear support for the involvement of cognition, and more specifically working memory, in speech perception for older adults, even in the processing of a single spoken word. The literature to-date is inconsistent with regards to this question. Some studies on older adults observed correlations between working memory scores and recognition of single words in noise (Gordon-Salant and Cole, 2016; Heinrich and Knight, 2016) while others did not (Parbery-Clark et al., 2011; Smith and Pichora-Fuller, 2015). The present study has the distinct advantage of directly manipulating memory load, testing the effect of reduced cognitive resources on spoken word processing in aging. By varying the number of digits to be remembered (one vs. four digits) we were able to temporarily deplete spare cognitive capacity while listeners performed a speech recognition task in noise. This momentary depletion led to changes in offline word recognition (for the lower-capacity group) and in online word processing (for the higher-capacity group). Note, if we were to test offline word recognition only, results would suggest that cognitive depletion mainly affects individuals with already low cognitive reserve. Indeed, previous works showed that increasing working memory load impairs language processing for clinical populations with reduced working memory capacity, such as people with aphasia, to a larger extent than for neurologically intact adults (Martin et al., 2012; Obermeyer et al., 2021). By using online measures, the current study shows the intricate effect of working memory depletion already at the single word level, even for individuals with larger cognitive reserves. Therefore, accessing and retrieving words from the mental lexicon when the input is degraded may require some available working memory resources even in healthy older adults with no signs of cognitive impairment. This link between cognition and speech processing in adverse listening conditions may stem from correlated activity across different brain regions. Indeed, spoken language processing rely on the joint activation of multiple cortical subsystems and several attempts were done to estimate its effectiveness by measuring cortical evoked responses (Gow, 2012). For example, Kim et al. (2021) suggested that changes in left supramarginal gyrus activity may be used as an independent predictor for speech processing efficiency.

In our analysis we found a differential effect of increasing working memory load for individuals with higher and lower working memory capacities. While increased load impaired offline accuracy for individuals with lower capacity, it affected online processing efficiency for individuals with higher capacity. According to the Framework for Understanding Effortful Listening (FUEL; Pichora-Fuller et al., 2016) speech processing depends on deployment of cognitive resources and therefore might be affected by differences in maximal capacity, especially under increased perceptual effort conditions such as in the presence of background noise and working memory load. It is possible that the listeners with lower working memory capacity were already using all their available resources in the low load condition in order to achieve maximal performance (100% accuracy). In other words, their online spoken word processing efficiency reflects their maximal ability. When facing increased task demands, they had no more available resources to allocate. Thus, with the same (maximal) word processing efficiency, as indicated by the online measures, their offline accuracy was significantly reduced. It is important to note that our analysis included only trials in which participants both correctly recognized the spoken word and correctly recalled the working memory load digits. Removing incorrect trials arguably removes the most challenging trials from the analysis which might lead to an under-estimation of the effects of increased load on individuals with lower working memory capacity. In contrast, listeners with higher working memory capacity were not using all their available resources in the low load condition. Consequently, when working memory load increased they still had some spare available resources to allocate to maintain their performance. But this came with a cost of slower online word processing.

Our results might be interpreted in light of the Ease of Language Understanding (ELU) model (Rönnberg et al., 2013). According to the model, understanding speech in adverse conditions is possible by drawing on central cognitive resources, mainly identified with working memory resources to compensate for the loss of automatic matching between the input and lexical representations when the input is degraded. Consistent with our findings, this model predicts that individuals with higher working memory capacity will be able to allocate these resources to maintain their offline performance. Changes in online processing could reflect either input degradation or the increased effort associated with the loss of automated word recognition.

In contrast to previous studies that relied on offline measures alone, the present study employed also online measures to track word processing as the acoustic signal unfolded over time. Standard measures of offline spoken word recognition accuracy do not capture the cost associated with maintaining a good level of performance. Our results highlight the importance of using both online and offline measures of speech processing to explore age-related changes in speech perception. The current study joins other studies that effectively used the visual world paradigm as a gauge of speech processing in adverse listening conditions (McQueen and Huettig, 2012; Helfer and Staub, 2014; Brouwer and Bradlow, 2016; McMurray et al., 2017). For example, McMurray et al. (2017) demonstrated that listeners with normal hearing process speech in a similar manner to that of cochlear implant users, when listening to severely degraded speech. In exploring the temporal dynamics of word recognition, authors could not only gauge the timing of target word recognition, but also determine the level and type of lexical competition that listeners were experiencing. Recent work from our lab also demonstrated that group-differences related to working memory load that were obscured in offline measures (e.g., accuracy) were uncovered when gaging online eye-tracking measures (Hadar et al., 2016; Nitsan et al., 2019; Harel-Arbeli et al., 2021).

Conclusions and Future Studies

The present data illustrate the differential effect of increasing task demands on spoken word recognition by listeners with higher vs. lower working memory capacity. Our findings suggest that additional cognitive capacity may lead to greater resilience of older listeners to adverse listening conditions. Future studies may wish to examine this paradigm using different types of adverse listening condition such as fast speech. Understanding accelerated speech is another predominant complaint among elderly listeners but little is known about its time course (Humes and Dubno, 2010; Banai and Lavie, 2020; Rotman et al., 2020). Studies should also consider carefully controlling for the possible effects of stress and stereotype threat on hearing assessments (Ben-David et al., 2018; Nagar et al., 2022). Another path for investigation is testing these findings in clinical populations with cognitive decline (noting the difficulties in adaptation, Tziraki et al., 2017) or hearing aids and cochlear implant users to better tailor hearing rehabilitation expectations (e.g., Taitelbaum-Swead et al., 2022). Future studies may also choose to further examine the effects of working memory load and span on brain activity involved in speech processing in aging.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the School of Psychology Review Board, Reichman University, and by the Ethics Committee of the Faculty of Social Welfare and Health Sciences, University of Haifa. The participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

GN, KB, and BB-D wrote the manuscript. GN was responsible of the analysis and interpretation of the data. KB contributed to the conceptualizing of the research question and interpreting the results. BB-D was responsible of the design of the paradigm, the analysis, and the interpretation of the results. BB-D was the corresponding author and the study was conducted in his lab. All authors had a prominent intellectual contribution to the study, are accountable for the data and approved the final version of the manuscript.

ACKNOWLEDGMENTS

We wish to thank Dalith Tal-Shir, Tali Livne, and Shai Baharv for their help in recruiting participants and collecting data.

REFERENCES

Abrams, L., and Farrell, M. T. (2011). "Language processing in normal aging," in *The Handbook of Psycholinguistic and Cognitive Processes: Perspectives in Communication Disorders*, eds J. Guendouzi, F. Loncke, and M. J. Williams (Hove: Psychology press), 49–73.

Akeroyd, M. (2008). Are individual differences in speech reception related to individual differences in cognitive ability? A survey of twenty experimental studies with normal and hearing-impaired adults. *Int. J. Audiol.* 47, 53–71. doi: 10.1080/14992020802301142

Allopenna, P. D., Magnuson, J. S., and Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: evidence for continuous mapping models. *J. Mem. Lang.* 38, 419–439. doi: 10.1006/jmla.1997.2558

Arnold, J. E., Fagnano, M., and Tanenhaus, M. K. (2003). Disfluencies Signal Theee, Um, New Information. *J. Psycholinguist. Res.* 32, 25–36. doi: 10.1023/A:1021980931292

Awh, E., and Vogel, E. K. (2008). The bouncer in the brain. *Nat. Neurosci.* 11, 5–6. doi: 10.1038/nn0108-5

Ayasse, N. D., Lash, A., and Wingfield, A. (2017). Effort not speed characterizes comprehension of spoken sentences by older adults with mild hearing impairment. *Front. Aging Neurosci.* 8:329. doi: 10.3389/fnagi.2016.00329

Baharav, S., Nitsan, G., and Ben-David, B. M. (2021). Commentary: working memory load affects processing time in spoken word recognition: test retest reliability of the E-WINDMIL eyetracking paradigm. *Front. Neurosci.* 15:663930 doi: 10.3389/fnins.2021.663930

Baldwin, C. L., and Ash, I. K. (2011). Impact of sensory acuity on auditory working memory span in young and old adults. *Psychol. Aging* 26, 85–91. doi: 10.1037/a0020360

Banai, K., and Lavie, L. (2020). Rapid perceptual learning and individual differences in speech perception: the good, the bad, and the sad. *Audit. Percept. Cogn.* 3, 201–211. doi: 10.1080/25742442.2021.1909400

Ben-David, B. M., Chambers, C. G., Daneman, M., Pichora-Fuller, M. K., Reingold, E. M., and Schneider, B. A. (2011). Effects of aging and noise on real-time spoken word recognition: evidence from eye movements. *J. Speech Lang. Hear. Res.* 54, 243–262. doi: 10.1044/1092-4388(2010/09-0233)

Ben-David, B. M., Erel, H., Goy, H., and Schneider, B. A. (2015). "Older is always better": age-related differences in vocabulary scores across 16 years. *Psychol. Aging* 30, 856–862. doi: 10.1037/pag0000051

Ben-David, B. M., Malkin, G., and Erel, H. (2018). "Ageism and neuropsychological tests," in *Contemporary Perspectives on Ageism*, eds L. Ayalon and C. Tesch-Römer (Cham: Springer), 277–297. doi: 10.1007/978-3-319-73820-8_17

Ben-David, B. M., Vania, Y. Y., and Schneider, B. A. (2012). Does it take older adults longer than younger adults to perceptually segregate a speech target from a background masker? *Hear. Res.* 290, 55–63.

Besser, J., Koelewijn, T., Zekveld, A. A., Kramer, S. E., and Festen, J. M. (2013). How linguistic closure and verbal working memory relate to speech recognition in noise—A review. *Trends Amplif.* 17, 75–93. doi: 10.1177/1084713813495459

Borovsky, A., Elman, J. L., and Fernald, A. (2012). Knowing a lot for one's age: vocabulary skill and not age is associated with anticipatory incremental sentence interpretation in children and adults. *J. Exp. Child Psychol.* 112, 417–436. doi: 10.1016/j.jecp.2012.01.005

Brouwer, S., and Bradlow, A. R. (2016). The temporal dynamics of spoken word recognition in adverse listening conditions. *J. Psycholinguist. Res.* 45, 1151–1160. doi: 10.1007/s10936-015-9396-9

Brown-Schmidt, S. (2009). The role of executive function in perspective taking during online language comprehension. *Psychon. Bull. Rev.* 16, 893–900. doi: 10.3758/PBR.16.5.893

Daneman, M., and Carpenter, P. A. (1980). Individual differences in working memory and reading. *J. Verb. Learn. Verb. Behav.* 19, 450–466.

Daneman, M., and Merikle, P. M. (1996). Working memory and language comprehension: a meta-analysis. *Psychon. Bull. Rev.* 3, 422–433. doi: 10.3758/bf03214546

Dryden, A., Allen, H. A., Henshaw, H., and Heinrich, A. (2017). The association between cognitive performance and speech-in-noise perception for adult listeners: a systematic literature review and meta-analysis. *Trends Hear.* 21:2331216517744675. doi: 10.1177/2331216517744675

Ezzatian, P., Avivi, M., and Schneider, B. A. (2010). Do nonnative listeners benefit as much as native listeners from spatial cues that release speech from masking? *Speech Commun.* 52, 919–929. doi: 10.1016/j.specom.2010.04.001

Farris-Trimble, A., and McMurray, B. (2013). Test–retest reliability of eye tracking in the visual world paradigm for the study of real-time spoken word recognition. *J. Speech Lang. Hear. Res.* 56, 1328–1345. doi: 10.1044/1092-4388(2012/12-0145)

Fostick, L., Ben-Artzi, E., and Babkoff, H. (2013). Aging and speech perception: beyond hearing threshold and cognitive ability. *J. Basic Clin. Physiol. Pharmacol.* 24, 175–183. doi: 10.1515/jbcpp-2013-0048

*Goodman, L. (2001). Translation of WAIS-III - wechsler adult intelligence scale. *Psych. Tech.* 1, 133–136.

Gordon-Salant, S., and Cole, S. S. (2016). Effects of age and working memory capacity on speech recognition performance in noise among listeners with normal hearing. *Ear Hear.* 37, 593–602. doi: 10.1097/AUD.0000000000000316

Gow, D. W. (2012). The cortical organization of lexical knowledge: a dual lexicon model of spoken language processing. *Brain Lang.* 121, 273–288. doi: 10.1016/j.bandl.2012.03.005

Hadar, B., Skrzypek, J. E., Wingfield, A., and Ben-David, B. M. (2016). Working memory load affects processing time in spoken word recognition: evidence from eye-movements. *Front. Neurosci.* 10:221 doi: 10.3389/fnins.2016.00221

Harel-Arbeli, T., Wingfield, A., Palgi, Y., and Ben-David, B. M. (2021). Age-related differences in the online processing of spoken semantic context and the effect of semantic competition: evidence from eye gaze. *J. Speech Lang. Hear. Res.* 64, 315–327. doi: 10.1044/2020_JSLHR-20-00142

Heck, R. H., Thomas, S. L., and Tabata, L. N. (2013). *Multilevel and Longitudinal Modeling with IBM SPSS*, 2nd Edn. Oxfordshire: Routledge, doi: 10.4324/9780203701249

Heinrich, A., Henshaw, H., and Ferguson, M. A. (2015). The relationship of speech intelligibility with hearing sensitivity, cognition, and perceived hearing difficulties varies for different speech perception tests. *Front. Psychol.* 6:782. doi: 10.3389/fpsyg.2015.00782

Heinrich, A., and Knight, S. (2016). "The contribution of auditory and cognitive factors to intelligibility of words and sentences in noise," in *Physiology, Psychoacoustics and Cognition in Normal and Impaired Hearing*, eds P. van Dijk, D. Başkent, E. Gaudrain, E. de Kleine, A. Wagner and C. Lanting (Cham: Springer), 37–45. doi: 10.1007/978-3-319-25474-6_5

Helfer, K., and Staub, A. (2014). Competing speech perception in older and younger adults: behavioral and eye-movement evidence. *Ear Hear.* 35, 161–170. doi: 10.1097/AUD.0b013e3182a830cf

Humes, L. E. (2021). Factors underlying individual differences in speech-recognition threshold (SRT) in noise among older adults. *Front. Aging Neurosci.* 13:383. doi: 10.3389/fnagi.2021.702739

Humes, L. E., and Dubno, J. R. (2010). "Factors affecting speech understanding in older adults," in *The Aging Auditory System*, eds S. Gordon-Salant, R. D. Frisina, A. N. Popper, and R. R. Fay (Cham: Springer), 211–257. doi: 10.1007/978-1-4419-0993-0_8

Humes, L. E., Watson, B. U., Christensen, L. A., Cokely, C. G., Halling, D. C., and Lee, L. (1994). Factors associated with individual differences in clinical measures of speech recognition among the elderly. *J. Speech Lang. Hear. Res.* 37, 465–474. doi: 10.1044/jshr.3702.465

Janse, E. (2012). A non-auditory measure of interference predicts distraction by competing speech in older adults. *Aging Neuropsychol. Cogn.* 19, 741–758. doi: 10.1080/13825585.2011.652590

Kaiser, E., and Trueswell, J. C. (2008). Interpreting pronouns and demonstratives in Finnish: evidence for a form-specific approach to reference resolution. *Lang. Cogn. Process.* 23, 709–748. doi: 10.1016/j.cognition.2009.03.010

Kavé, G., and Halamish, V. (2015). Doubly blessed: older adults know more vocabulary and know better what they know. *Psychol. Aging* 30, 68–73. doi: 10.1037/a0038669

Kim, S., Choi, I., Schwalje, A. T., Kim, K., and Lee, J. H. (2020). Auditory working memory explains variance in speech recognition in older listeners under adverse listening conditions. *Clin. Interv. Aging* 15, 395–406. doi: 10.2147/CIA.S241976

Kim, S., Schwalje, A. T., Liu, A. S., Gander, P. E., McMurray, B., Griffiths, T. D., et al. (2021). Pre- and post-target cortical processes predict speech-in-noise performance. *Neuroimage* 228:117699. doi: 10.1016/j.neuroimage.2020.117699

Lash, A., Rogers, C. S., Zoller, A., and Wingfield, A. (2013). Expectation and entropy in spoken word recognition: effects of age and hearing acuity. *Exp. Aging Res.* 39, 235–253. doi: 10.1080/0361073X.2013.779175

Marslen-Wilson, W. (1990). "Activation, competition, and frequency in lexical access," in *Cognitive Models of Speech Processing*, ed. G. T. M. Altmann (Cambridge, MA: The MIT Press), 148–172.

Martin, N., Kohen, F., Kalinyak-Fliszar, M., Soveri, A., and Laine, M. (2012). Effects of working memory load on processing of sounds and meanings of words in aphasia. *Aphasiology* 26, 462–493. doi: 10.1080/02687038.2011.619516

McMurray, B., Farris-Trimble, A., and Rigler, H. (2017). Waiting for lexical access: cochlear implants or severely degraded input lead listeners to process speech less incrementally. *Cognition* 169, 147–164. doi: 10.1016/j.cognition.2017.08.013

McQueen, J. M., and Huettig, F. (2012). Changing only the probability that spoken words will be distorted changes how they are recognized a. *J. Acoust. Soc. Am.* 131, 509–517. doi: 10.1121/1.3664087

Meister, H. (2017). Speech audiometry, speech perception, and cognitive functions. *HNO* 65, 1–4. doi: 10.1007/s00106-016-0250-7

Mirman, D., Dixon, J. A., and Magnuson, J. S. (2008). Statistical and computational models of the visual world paradigm: growth curves and individual differences. *J. Mem. Lang.* 59, 475–494. doi: 10.1016/j.jml.2007.11.006

Nagar, S., Mikulincer, M., Nitsan, G., and Ben-David, B. M. (2022). Safe and sound: the effects of experimentally priming the sense of attachment security on pure-tone audiometric thresholds among young and older adults. *Psychol. Sci.* [Online ahead of print]. doi: 10.1177/09567976211042008

Nitsan, G., Wingfield, A., Lavie, L., and Ben-David, B. M. (2019). Differences in working memory capacity affect online spoken word recognition: evidence from eye movements. *Trends Hear.* 23:233121651983962. doi: 10.1177/2331216519839624

Norris, D. (1994). Shortlist: a connectionist model of continuous speech recognition. *Cognition* 52, 189–234. doi: 10.1037/0033-295X.115.2.357

Obermeyer, J., Reinert, L., Kamen, R., Pritchard, D., Park, H., and Martin, N. (2021). Effect of working memory load and typicality on semantic processing in aphasia. *Am. J. Speech Lang. Pathol.* 31, 12–29. doi: 10.1044/2021_AJSLP-20-00283

Parbery-Clark, A., Strait, D. L., Anderson, S., Hittner, E., and Kraus, N. (2011). Musical experience and the aging auditory system: implications for cognitive abilities and hearing speech in noise. *PLoS One* 6:e18082. doi: 10.1371/journal.pone.0018082

Pichora-Fuller, M. K. (2008). Use of supportive context by younger and older adult listeners: balancing bottom-up and top-down information processing. *Int. J. Audiol.* 47, S72–S82. doi: 10.1080/14992020802307404

Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W., Humes, L. E., et al. (2016). Hearing impairment and cognitive energy: the Framework for Understanding Effortful Listening (FUEL). *Ear Hear.* 37, 5S–27S. doi: 10.1097/AUD.0000000000000312

Pichora-Fuller, M. K., Schneider, B. A., and Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *J. Acoust. Soc. Am.* 97, 593–608. doi: 10.1121/1.412282

Rogers, C. S., and Peelle, J. E. (2021). "Interactions between audition and cognition in hearing loss and aging," in *Speech Perception*, 1st Edn, eds L. Holt, J. Peelle, A. B. Coffin, A. N. Popper, and R. R. Fay (Cham: Springer International Publishing), doi: 10.31234/osf.io/d2bxw

Rönnberg, J., Lunner, T., Zekveld, A., Sörqvist, P., Danielsson, H., Lyxell, B., et al. (2013). The Ease of Language Understanding (ELU) model: theoretical, empirical, and clinical advances. *Front. Syst. Neurosci.* 7:31. doi: 10.3389/fnsys.2013.00031

Rönnberg, J., Rudner, M., Foo, C., and Lunner, T. (2008). Cognition counts: a working memory system for ease of language understanding (ELU). *Int. J. Audiol.* 47, S99–S105. doi: 10.1080/14992020802301167

Rotman, T., Lavie, L., and Banai, K. (2020). Rapid perceptual learning: a potential source of individual differences in speech perception under adverse conditions? *Trends Hear.* 24:2331216520930541. doi: 10.1177/2331216520930541

Smith, S. L., and Pichora-Fuller, M. K. (2015). Associations between speech understanding and auditory and visual tests of verbal working memory: effects of linguistic complexity, task, age, and hearing loss. *Front. Psychol.* 6:1394. doi: 10.3389/fpsyg.2015.01394

Smith, S. L., Pichora-Fuller, M. K., and Alexander, G. (2016). Development of the word auditory recognition and recall measure: a working memory test for use in rehabilitative audiology. *Ear Hear.* 37, e360–e376. doi: 10.1097/AUD.0000000000000329

Sommers, M. S. (2005). "Age-related changes in spoken word recognition," in *The Handbook of Speech Perception*, eds D. B. Pisoni and R. Remez (Hoboken, NJ: Blackwell Publishing Ltd), 469–493. doi: 10.1002/9780470757024.ch19

Sommers, M. S., and Danielson, S. M. (1999). Inhibitory processes and spoken word recognition in young and older adults: the interaction of lexical competition and semantic context. *Psychol. Aging* 14, 458–472. doi: 10.1037//0882-7974.14.3.458

Stine-Morrow, E. A. L., Shake, M. C., Miles, J. R., and Noh, S. R. (2006). Adult age differences in the effects of goals on self-regulated sentence processing. *Psychol. Aging* 21, 790–803. doi: 10.1037/0882-7974.21.4.790

Taitelbaum-Swead, R., Icht, M., and Ben-David, B. M. (2022). More than words: the relative roles of prosody and semantics in the perception of emotions in spoken language by postlingual cochlear implant users. *Ear Hear.* [Online ahead of print]. doi: 10.1097/AUD.0000000000001199

Tziraki, C., Berenbaum, R., Gross, D., Abikhzer, J., and Ben-David, B. M. (2017). Designing serious computer games for people with moderate and advanced dementia: interdisciplinary theory-driven pilot study. *JMIR Serious Games* 5:e6514. doi: 10.2196/games.6514

World Health Organization (2021). *Decade of Healthy Ageing: Baseline Report*. Available online at: https://www.who.int/publications-detail-redirect/9789240017900 [accessed on January 14, 2021]

Zekveld, A. A., Heslenfeld, D. J., Festen, J. M., and Schoonhoven, R. (2006). Top–down and bottom–up processes in speech comprehension. *Neuroimage* 32, 1826–1836. doi: 10.1016/j.neuroimage.2006.04.199

Conflict of Interest: 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.

APPENDIX A. FULL ANALYSIS OF THE OFFSET OVERLAP CONDITION

Eye Gaze

Unlike the onset overlap condition, the analysis of offset overlap trials showed only an effect of span group on the linear and quadratic time terms, suggesting differential online word processing between span groups. **Table A1** summarizes the results of the analysis.

TABLE A1 | Results of growth curve analysis (GCA) – offset overlap.

	Term	Estimate	SE	*t*-Value	*p*
Participant group (WM capacity)	Intercept	−0.053	0.039	−1.37	0.174
	Linear	0.333	0.099	3.36	< 0.001
	Quadratic	<0.001	<0.001	−3.13	0.002
Working memory load	Intercept	−0.013	0.026	−0.51	0.607
	Linear	0.064	0.099	0.644	0.519
	Quadratic	<0.001	<0.001	−0.78	0.438
Participant group (WM capacity) × working memory load	Intercept	0.004	0.036	0.11	0.915
	Linear	−0.132	0.140	−0.94	0.347
	Quadratic	<0.001	<0.001	1.09	0.275

Accuracy of Behavioral Responses

Eye-gaze analysis included only trials in which participants both correctly selected the corresponding object on the visual display (indicating correct spoken word recognition) and correctly recalled the working memory load digits (indicating correct digit recall). The analysis indicated a main effect of load $F(1,34) = 34.23$, $p < 0.001$ and span group $F(1,34) = 6.83$, $p = 0.013$ on response accuracy. These two effects suggest that increasing memory load from one to four digits significantly reduced listeners' response accuracy regardless of span group membership. Additionally, is shows that overall listeners from the high span group had higher response accuracy ($M = 95.83$ vs. $M = 89.48$) compared to listeners from the low span group. The two effects interacted significantly $F(1,34) = 5.59$, $p = 0.024$. LSD-corrected pairwise comparisons revealed that increasing memory load yielded greater reduction in response accuracy in the low span group $F(1,34) = 33.75$, $p < 0.001$ compared to the high span group $F(1,34) = 6.08$, $p = 0.019$ as shown in **Table A2**.

TABLE A2 | Mean percentage (and SEs) of trials in which target word was correctly selected and digits were correctly recalled.

	Low WM capacity	High WM capacity
Low WM load	99.3% (0.69)	100% (0.0)
High WM load	79.7% (4.08)	91.7% (2.48)

Low and high WM, indicate the two preload conditions, one and four digit/s, respectively.

frontiers | Frontiers in Psychology

ORIGINAL RESEARCH
published: 15 April 2022
doi: 10.3389/fpsyg.2022.845285

Age Differences in Speech Perception in Noise and Sound Localization in Individuals With Subjective Normal Hearing

Tobias Weissgerber[1]*, Carmen Müller[1], Timo Stöver[2] and Uwe Baumann[1]

[1] *Audiological Acoustics, Department of Oto-Rhino-Laryngology, University Hospital Frankfurt, Goethe University Frankfurt, Frankfurt am Main, Germany,* [2] *Department of Oto-Rhino-Laryngology, University Hospital Frankfurt, Goethe University Frankfurt, Frankfurt am Main, Germany*

Hearing loss in old age, which often goes untreated, has far-reaching consequences. Furthermore, reduction of cognitive abilities and dementia can also occur, which also affects quality of life. The aim of this study was to investigate the hearing performance of seniors without hearing complaints with respect to speech perception in noise and the ability to localize sounds. Results were tested for correlations with age and cognitive performance. The study included 40 subjects aged between 60 and 90 years (mean age: 69.3 years) with not self-reported hearing problems. The subjects were screened for dementia. Audiological tests included pure-tone audiometry and speech perception in two types of background noise (continuous and amplitude-modulated noise) which was either co-located or spatially separated (multi-source noise field, MSNF) from the target speech. Sound localization ability was assessed and hearing performance was self-evaluated by a questionnaire. Speech in noise and sound localization was compared with young normal hearing adults. Although considering themselves as hearing normal, 17 subjects had at least a mild hearing loss. There was a significant negative correlation between hearing loss and dementia screening (DemTect) score. Speech perception in noise decreased significantly with age. There were significant negative correlations between speech perception in noise and DemTect score for both spatial configurations. Mean SRTs obtained in the co-located noise condition with amplitude-modulated noise were on average 3.1 dB better than with continuous noise. This gap-listening effect was severely diminished compared to a younger normal hearing subject group. In continuous noise, spatial separation of speech and noise led to better SRTs compared to the co-located masker condition. SRTs in MSNF deteriorated in modulated noise compared to continuous noise by 2.6 dB. Highest impact of age was found for speech perception scores using noise stimuli with temporal modulation in binaural test conditions. Mean localization error was in the range of young adults. Mean amount of front/back confusions was 11.5% higher than for young adults. Speech perception tests in the presence of temporally modulated noise can serve as a screening method for early detection of hearing disorders in older adults. This allows for early prescription of hearing aids.

Keywords: speech perception, noise, sound localization, age, cognition

Edited by:
Leah Fostick,
Ariel University, Israel

Reviewed by:
Yuan Cheng,
East China Normal University, China
Christian Füllgrabe,
University College London,
United Kingdom

***Correspondence:**
Tobias Weissgerber
tobias.weissgerber@kgu.de

Specialty section:
This article was submitted to
Perception Science,
a section of the journal
Frontiers in Psychology

Received: *29 December 2021*
Accepted: *21 March 2022*
Published: *15 April 2022*

Citation:
Weissgerber T, Müller C, Stöver T and Baumann U (2022) Age Differences in Speech Perception in Noise and Sound Localization in Individuals With Subjective Normal Hearing. Front. Psychol. 13:845285. doi: 10.3389/fpsyg.2022.845285

INTRODUCTION

The World Health Organization (WHO) estimates that approximately one third of all people over the age of 65 suffer from a hearing loss, although the number of unreported cases is likely to be even higher, as it usually takes time before a progressive hearing loss is diagnosed (World Health Organization, 2017). From 2008 to 2030 the age group of people aged 65 and older will increase in Germany by around one third due to demographic change (Statistische Ämter des Bundes und der Länder, 2011). For the whole European Union, the estimated increase in that age group for the same time span is about 45% (European Commission, 2020). This means that hearing loss in particular will become more important in the elderly. In addition, lifestyles in later life have changed considerably. Today, there are numerous options for living arrangements such as nursing homes or assisted living. In order to continue participating in social life, more and more elderly rely on telecommunication or events such as senior citizens' meetings. The range of possible consequences of hearing impairment is wide, including social isolation and inability to work or psychosomatic disorders such as anxiety and depression (Garnefski and Kraaij, 2012).

Although hearing impairment has a major impact on the quality of life of the elderly. hearing loss in elderly subjects is frequently undetected and untreated (Völter et al., 2020). In a study population aged between 40 and 79 years with at least a mild hearing loss, only 6% were aware of any symptoms (Ramage-Morin et al., 2019).

It is known that speech perception in noise is impaired even in the presence of a mild hearing loss compared to age-matched normal-hearing individuals (Dubno et al., 1984). In the study of Dubno et al. (1984) it was also shown that older subjects with normal hearing (which was comparable to a younger subject group) showed decreased speech perception in noise for suprathreshold signal presentation.

Likewise, Meister et al. (2011) reported slightly degraded speech reception thresholds in quiet and continuous noise but substantial differences were found for modulated noise. It was reported by Füllgrabe (2013) that suprathreshold processing of temporal fine structure (TFS) declines with increasing age even when hearing sensitivity is (nearly) normal. In a subsequent study, Füllgrabe et al. (2015) showed decreased speech perception scores for consonants and sentences in noise in a study population aged older than 60 years with pure-tone thresholds matched to audiometrically normal-hearing younger (<30 years) adults. Furthermore, a correlation between consonant perception as well as speech perception in noise and sensitivity to TFS was shown. In concordance with Meister et al. (2011), a correlation between speech perception scores (consonants and sentences) and several cognitive measures was also reported by Füllgrabe et al. (2015).

In Summary, increasing age adversely affects the processing of both TFS and slowly varying envelope (ENV), whereby a stronger correlation with speech perception in noise was found for TFS. Therefore, spatial release from masking (SRM) is reduced for speech stimuli in older subjects with or without hearing loss. This implies that aged subjects will have difficulties relative to young normal-hearing subjects when trying to understand speech in the presence of interfering sounds coming from different directions in space, as is common in everyday life (Moore, 2021).

Therefore, the aim of the present study was to investigate speech perception under complex noise conditions with multiple noise sources in a cohort of older persons. Speech reception thresholds (SRTs) were assessed in two types of background noise (continuous and amplitude-modulated noise) which was either co-located or spatially separated (four sound sources, multi-source noise field, MSNF) from the target speech (frontal presentation).

It was expected that impaired TFS and ENV processing as well as a potentially reduced SRM effect would show an impact on SRTs as a function of age. In addition, spatial hearing ability was evaluated based on the accuracy of sound localization obtained for broadband noise stimuli. Screening of cognitive performance was assessed to identify potentially deteriorated results that might correlate with the hearing test battery outcomes. Finally, results from a cohort of seniors with no self-reported hearing loss (i.e., subjectively no known symptoms of hearing loss, normal to mild hearing loss) were compared with data obtained in previous studies using the same test setup in young adults with normal hearing (Weissgerber et al., 2017).

MATERIALS AND METHODS

Subjects

The study comprised a total of 40 subjects (28 female, 12 male). The subjects were aged between 60.1 and 89.7 years (mean age: 69.3 ± 7.1 years, mean age of the female subjects: 69.6 ± 7.1 years, mean age of the male subjects: 68.6 ± 7.5 years).

Subjects were recruited for participation via flyers and an advertisement at the grounds of the University Hospital Frankfurt. The three inclusion criteria mentioned in that advertisement were (1) aged 60 or higher, (2) no subjective awareness of any hearing problems, and (3) no use of hearing aids. Each of the subjects was a native German speaker, as the speech perception tests were conducted in German. The study was approved by the Ethics Committee of the Department of Medicine of the Goethe University in Frankfurt am Main, Germany (No. 164/13).

Before performing the tests, an ear inspection was performed and a tympanogram was obtained for each ear to exclude study candidates with eventual conductive hearing loss. The study tests required ~3 h per subject and were each conducted on 1 day. The order of study tests was randomized.

Screening for Dementia

The DemTect (Kessler et al., 2000) was used to check for a potential onset of dementia. The DemTect consists of five subtests, which are carried out in the form of a survey. A list of ten words is read out to the subject and then immediately queried. Afterward, the same list is again read out and queried again. At the end of the test the word list has to be repeated by the subject without being read out again in order to test verbal memory. Furthermore, there is a subtest on intellectual flexibility in which numbers have to be converted into text and vice versa. Finally, there is a subtest for word fluency, in which the subject has to list things that can be bought in the supermarket within 1 min. A further subtest on verbal memory and attention follows with the reproduction of a sequence of numbers read out in reverse order.

The results of the individual subtests are converted into age-corrected test scores, then summed up and expressed as a DemTect score (max. 18 points). The resulting test scores are independent of age and educational level and thus provide information on whether cognitive performance is age-appropriate (DemTect score: 13–18 points), slightly impaired (DemTect score: 9–12 points) or whether dementia is suspected (DemTect score: ≤8 points) (Kalbe et al., 2004). The duration of the DemTect is ~8–10 min.

Pure-Tone Audiometry

Pure-tone audiometry was performed in a sound-attenuated room to determine the subjects' individual hearing thresholds. Air conduction hearing thresholds were determined for pure-tones from 125 to 8,000 Hz for each ear of each subject using calibrated headphones. Bone conduction hearing thresholds were not determined because middle ear pathologies were excluded in advance both by patient history and by tympanometry. The pure-tone average (PTA) hearing loss was determined from the frequencies 500 Hz, 1, 2, and 4 kHz (PTA_4). Furthermore, a pure-tone average hearing loss for high frequencies was calculated as mean hearing loss of the frequencies 4, 6, and 8 kHz (PTA_high).

Speech Perception in Noise

Speech tests were conducted for different types of background noise in two spatial loudspeaker configurations to simulate everyday listening situations. The measurements were conducted in an anechoic chamber with dimensions 4.1 m × 2.6 m × 2.1 m (length × width × height). The system for sound playback consisted of 128 loudspeakers arranged in a rectangular array in the horizontal plane at a height of 1.20 m, which corresponded approximately to the ear height of the seated subjects. Further detailed information on the playback system is given in Weissgerber et al. (2015).

Speech reception thresholds (SRTs) in noise were determined with the German matrix test [Oldenburg Sentence Test, OLSA, Wagener et al. (1999)]. Each test list consisted of 20 sentences, which contained a noun, verb, numeral, adjective, and object. Noise level was kept constant at 65 dB SPL and speech level was set adaptively according to the number of words perceived correctly. Speech levels automatically increased when two or fewer words were perceived correctly and decreased when more than two words were correct. The step sizes for this adaptive procedure decreased with the number of inflection points as suggested by Brand and Kollmeier (2002). The result of the OLSA test is the SRT for 50% correct word understanding. Speech signal was always presented from the same direction of 0° frontally at a distance of 1.75 m from the subject. Four adjacent speakers of the playback system were used to obtain a sound pressure level with negligible distortion at the subject's position (Weissgerber et al., 2017).

The test was conducted in a closed set mode, i.e., the subject had to select the perceived words of the sentence on a matrix presented on a touchscreen. In order to become familiar with the task, a training trial with 30 test sentences was performed with each subject before the study test began. Subsequently, four test runs of the OLSA were performed with each subject in random order, differing in noise type and spatial noise configuration.

Two types of noise were used. The speech-shaped "OlNoise" is a temporally continuous noise whose long-term spectrum matches that of the word material of the matrix test (Wagener et al., 1999). This continuous noise is used to simulate background noise with low temporal modulation, such as the noise of a vacuum cleaner or a fan. The other test stimulus was amplitude-modulated, speech-shaped, fluctuating noise according to Fastl (1987) and Fastl and Zwicker (2007). The spectral distribution of the amplitude-modulation reaches maximum values at 4 Hz, which is consistent with many spoken syllables of Western speech.

Two spatial noise configurations were tested. In the first condition, noise was presented from the same direction as speech signal (0°, condition S0N0).

In the second condition, a diffuse noise (multi-source noise field, MSNF) was created by means of wave field synthesis (Berkhout, 1988). Four virtual noise sources were placed at the positions ±28.6° and ±151.4° with a distance of 1.25 m from the center of the subject's head (see **Figure 1**). The four virtual noise sources were temporally uncorrelated. The MSNF speaker configuration was proposed by Rader et al. (2013) to simulate everyday conversational situations in noisy environments, such as conversations in a restaurant, etc.

In the study of Weissgerber et al. (2017) data obtained in young adults with normal hearing (n = 14, mean age: 26.4 ± 5.4 years, range: 22–37.3 years) using the identical test setup was shown. The criterion for normal hearing was a pure-tone hearing loss lower than 25 dB HL between 0.25 and 8 kHz. The results obtained in the cohort of older subjects in the present work were compared with the data from Weissgerber et al. (2017).

Sound Localization

The test for sound localization took place in the same anechoic room with the same loudspeaker arrangement. LED chains with a total of 704 individual LEDs were mounted above the loudspeakers to indicate the direction of sound incidence.

The test stimulus was a white noise (high-pass filtered at 150 Hz) consisting of five pulses. Each pulse had a duration of 30 ms with a rise time of 3 ms followed by a pause of 70 ms [according to Seeber (2002)]. Before the test began, a blue LED lit up in front of the subject at the 0° position, at which the subject had to focus on. After hearing the test stimulus, the subject was first asked to indicate by means of a toggle switch that changed the LED color whether the sound was perceived from the front (red LED) or from behind (green LED). Subsequently, the subject should select the LED that corresponded to the perceived horizontal angle of incidence of the auditory event using a rotary encoder. The indication of a sound from behind

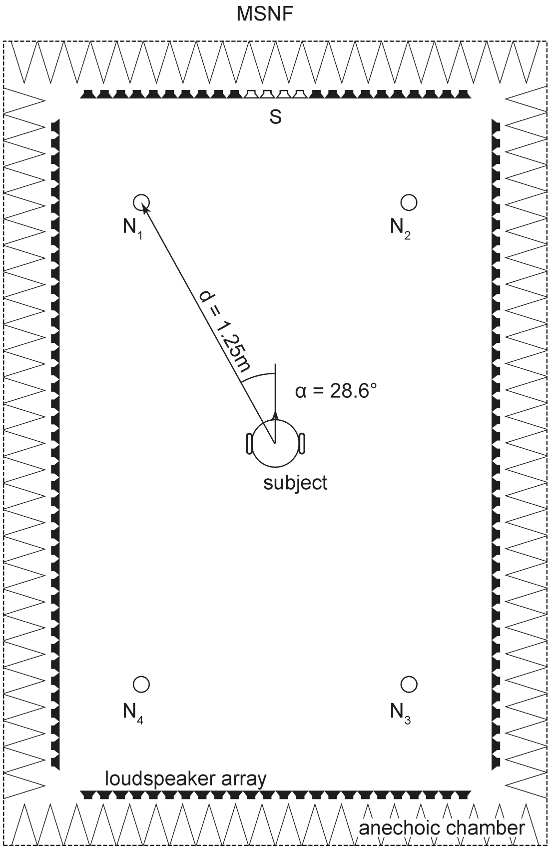


FIGURE 1 | Loudspeaker array with 128 independent loudspeaker channels in rectangular shape mounted in the horizontal plane at 1.20 m for measurement of speech perception in noise. The speech signal (S) was generated from front loudspeakers in the 0°-position. Four virtual noise sources (N1-4) for multi-source noise field (MSNF) were created by wave field synthesis.

was marked green *via* the toggle switch and mirrored to the front. Prior to the start of the test, a detailed introduction to the LED display system took place, as well as a training run in which each test loudspeaker was tested once. A total of seven test loudspeakers were tested between −60 and +60° (−59.2°, −42.1°, −21.2°, −2.5°, 16.8°, 42.1°, 59.2°) in front and back. Each of the 14 test loudspeakers was randomly selected five times (i.e., 70 trials) in order to measure the localization accuracy (mean error, i.e., deviation of presented angle and perceived angle) and uncertainty (dispersion of mean error) of localization. The test was performed in complete darkness. The duration of a whole test run was about 15 min.

The relative localization error was calculated for each subject by averaging the relative localization errors of each angle (both front and back). Furthermore, the percentage front-back confusions were calculated. Results were compared with data obtained in young adults with normal hearing ($n = 9$, mean age: 30.3 ± 6.1 years, pure-tone hearing loss lower than 25 dB HL for all test frequencies between 0.25 and 8 kHz) using the identical test setup (data unpublished so far).

Speech, Spatial, and Qualities of Hearing Scale Questionnaire

All 40 senior subjects completed the validated German-language version of the "Speech, Spatial, and Qualities of Hearing Scale" (SSQ) questionnaire (Kießling et al., 2011) to assess their subjective hearing performance. The SSQ questionnaire is a standardized and validated questionnaire consisting of 49 questions on various listening situations, which the subject answers by marking his subjectively assessed listening ability on a Likert scale of 0–10 points (Gatehouse and Noble, 2004). The questions are divided into three sections. The first section contains questions on speech perception in a wide variety of everyday listening situations. The second section includes questions on spatial hearing. The third section focuses on listening quality, e.g., naturalness, clarity and identifiability of a speaker.

Statistics

The collected data was processed and analyzed using the statistical program SPSS Statistics 27 (IBM, Armonk, NY, United States). The target variables of the various tests were checked for normal distribution. Since none of the target variables showed a normal distribution, further evaluation was carried out with non-parametric methods. The Spearman Rho test was used to examine the correlation between the test variables. For multiple comparisons p-values were adjusted using the Bonferroni-Holm method. Adjusted p-values < 0.05 were considered as statistically significant. Only adjusted p-values were given in the manuscript if not stated otherwise.

RESULTS

Dementia Screening

Individual results of the DemTect are shown in **Figure 2**. The median DemTect score of all subjects was 14.0 points (interquartile range IQR: 12.0 to 15.0 points). Sixty five percentage of the subjects had DemTect scores between 13 and 18 points, so that their cognitive performance can be classified as age-appropriate. The remaining 35% of the subjects had DemTect scores between 9 and 12 points and thus a mild cognitive impairment. None of the subjects had a DemTect score that indicated a suspicion of dementia.

Pure-Tone Audiometry

Pure-tone thresholds of the 40 subjects (for better illustration divided into two age groups: 60–74, 75–90 years) are shown in **Figure 3**. According to the classification of hearing loss published by the World Health Organization [WHO] (2001) 26 subjects had normal hearing in both ears and three subjects had normal hearing in one ear. Seven subjects had a mild hearing loss in both ears and five subjects in one ear. One subject had a moderate hearing loss in both ears and four subjects in one ear. None of the subjects had a severe/profound hearing loss or deafness in any ear.

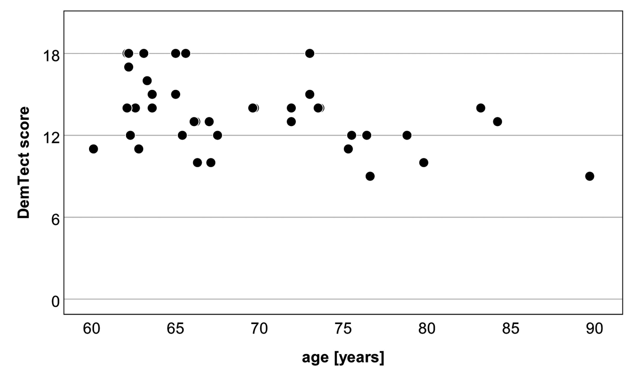


FIGURE 2 | Scatter plot of age and DemTect scores of the 40 test subjects. Thirteen–eighteen points: age-appropriate cognitive performance; 9–12 points: slight impairment; ≤8 points: suspicion for dementia.

Median PTA_4 was 18.75 dB HL (IQR: 12.5-27.5 dB HL), median PTA_high was 35 dB HL (IQR: 20-56.7 dB HL).

Speech Perception in Noise

The SRT results divided in two age groups 60–74 years and 75–90 years for the different test conditions are shown in **Figure 4**. The age group 60–74 years showed significantly lower mean SRTs than the older age group in all test conditions (S0N0 continuous noise: $Z = -3.715$, $p < 0.001$; S0N0 modulated noise: $Z = -3.515$, $p < 0.001$; MSNF continuous noise: $Z = -2.642$, $p = 0.016$; MSNF modulated noise: $Z = -2.236$, $p = 0.025$).

Boxplots of SRT results averaged over all older study participants were shown in **Figure 5** (gray boxes). The median SRT in S0N0 condition using continuous noise was −5.2 dB SNR (IQR: −4.5 to −5.8 dB SNR). Median SRT in S0N0 with modulated noise was significantly better than in continuous noise (difference: 3.1 dB, $Z = -5.216$, $p < 0.001$). The IQR of the SRTs obtained in modulated noise (−6.2 to −10.6 dB SNR) was 3.1 dB larger than the IQR for continuous noise.

The SRT in MSNF using continuous noise was −8.8 dB SNR (IQR: −8.1 to −9.7 dB SNR), 3.7 dB significantly lower than the SRT for the modulated noise ($Z = -5.276$, $p < 0.001$). The IQR of the SRTs obtained in MSNF with modulated noise (−3.0 to −7.4 dB SNR) was 2.8 dB larger than the IQR for continuous noise.

Speech reception thresholds with continuous noise were significantly higher in the S0N0 condition than in the MSNF condition (3.6 dB difference, $Z = -5.513$, $p < 0.001$). In the modulated noise condition, a significantly lower SRT was obtained in S0N0 condition compared with MSNF (3.2 dB difference, $Z = -4.678$, $p < 0.001$).

Comparison With Young Adults

Speech reception thresholds results obtained in young adults with normal hearing using the identical test setup (Weissgerber et al., 2017) are illustrated in **Figure 5** (white boxes). SRTs in the young normal hearing group were significantly lower for all test conditions ($p < 0.001$). Mean difference was lowest for test conditions using continuous noise (S0N0: 1.9 dB; MSNF: 1.3 dB). More important, in conditions with modulated noise large differences between younger and older subjects were found (S0N0: 9.8 dB; MSNF: 6.3 dB).

Sound Localization

The median relative localization error was 5.8° (interquartile range: 4.5–8.1°) and the median amount of front-back confusions was 12.9% (interquartile range: 5.0–31.1%).

Results of mean localization error and front-back confusions divided in two age groups 60–74 years and 75–90 years are provided as **Supplementary Material**. There was no significant difference between the age group 60–74 years and the age group 75–90 years for both measures localization error and front-back confusions. There was also no significant difference between

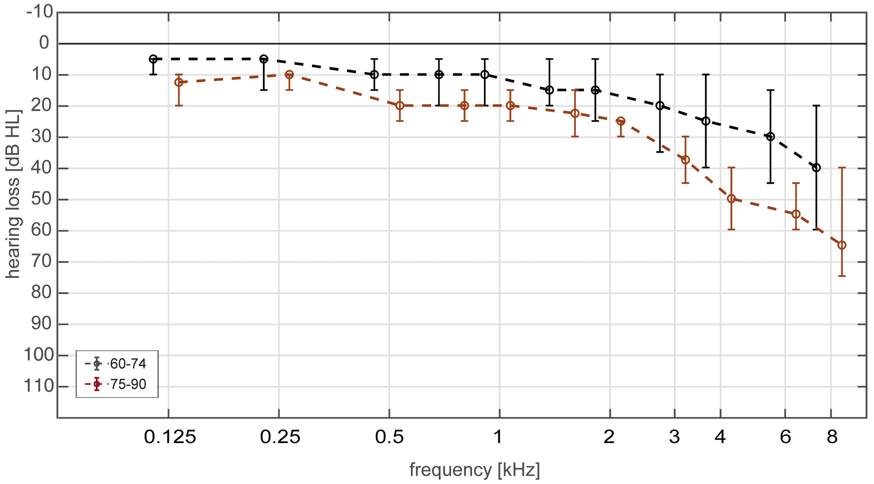


FIGURE 3 | Hearing thresholds (median and interquartile range) of 80 ears (40 subjects) divided into two age groups. Age groups were 60–74 (black) and 75–90 (red) years.

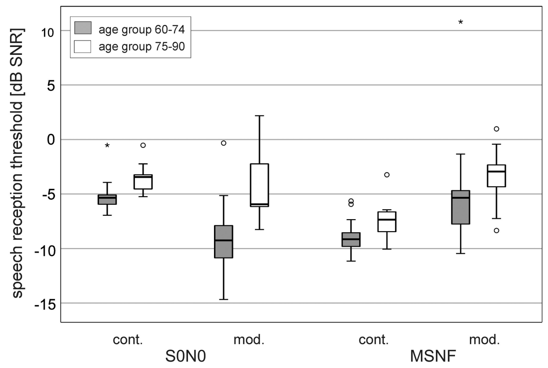


FIGURE 4 | Speech reception thresholds in spatial conditions S0N0 and multi-source noise field (MSNF) for the continuous (cont.) noise and modulated (mod.) noise divided into age groups 60–74 years (gray boxes) and 75–90 years (white boxes). * Outliers 3 times greater than interquartile range.

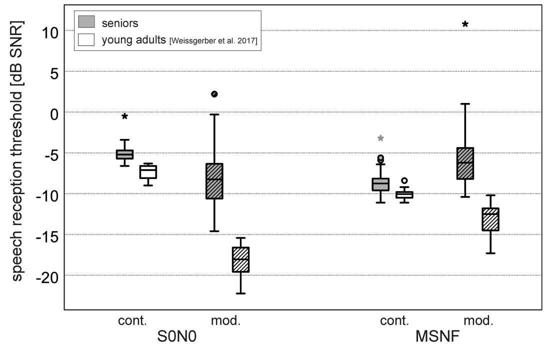


FIGURE 5 | Speech reception thresholds for the continuous noise (cont., filled boxes) and modulated noise (mod., hatched boxes) in spatial conditions S0N0 and multi-source noise field (MSNF). Gray boxes: seniors (present study), white boxes: data of young adults with normal hearing (n = 14, mean age: 26.4 ± 5.4 years, range: 22–37.3 years) using the identical test setup (Weissgerber et al., 2017). * Outliers 3 times greater than interquartile range.

the localization error of the seniors measured in this study and a young group with normal hearing (n = 9, mean age: 30.3 ± 6.1 years).

The results of front/back confusions are shown in **Figure 6**. Additionally, reference data obtained in young adults with normal hearing using the identical test setup is illustrated. The amount of front-back confusions was significantly worse in the test subjects of the present study (difference: 11.5%; Z = −3.213, $p < 0.001$) compared with young adults.

Correlations With Age

Correlations in the senior group between age and DemTect scores, pure-tone hearing thresholds, and speech in noise scores and were shown in **Table 1**. Additionally, correlations were calculated after partialling out mean high-frequency hearing loss (PTA_high).

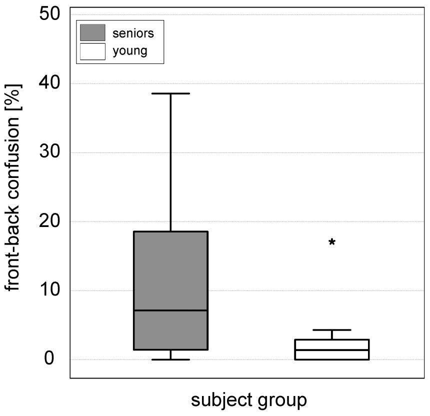


FIGURE 6 | Boxplots of front-back confusions [%] in the sound localization test for the senior study participants (gray box) compared to a reference group of nine young adults with normal hearing (white box). * Outliers 3 times greater than interquartile range.

TABLE 1 | Spearman correlation coefficients ρ for results on measures of DemTect, hearing loss, and speech perception in noise vs. age (first column).

	Age	Age PTA_high partialled out
DemTect	**–0.412**	–0.323
PTA_4	**0.367**	0.147
PTA_high	**0.353**	N/A
SRT S0N0 cont.	**0.539**	**0.44**
SRT S0N0 mod.	**0.426**	0.286
SRT MSNF cont.	**0.383**	0.289
SRT MSNF mod.	**0.398**	0.288

Correlation coefficients after partialling out high-frequency hearing loss PTA_high (second column). Gray values indicate non-significant correlations (p > 0.05). Values in black indicate significant results (p ≤ 0.05). Values in boldface indicate significant results after applying a Bonferroni-Holm correction. N/A: correlation non-applicable.

A significant negative correlation between age and DemTect score was found (ρ = −0.412, p = 0.04). There was significant correlation between age and PTA_4 (ρ = 0.367, p = 0.045) as well as between age and PTA_high (ρ = 0.353, p = 0.045). There was a significant correlation between age and S0N0 SRT for both types of noise (continuous: ρ = 0.539, $p < 0.001$, modulated: ρ = 0.426, p = 0.036). There was a significant correlation between age and MSNF SRT for continuous noise (ρ = 0.383, p = 0.045) and modulated noise (ρ = 0.398, p = 0.044).

After partialling out high-frequency hearing loss, only a significant correlation between age and S0N0 SRT in continuous noise (ρ = 0.44, p = 0.03) was found.

Correlations With Cognitive Performance

Correlations in the senior group between DemTect scores and both measures pure-tone hearing thresholds and speech in noise scores were shown in **Table 2**. Correlations were also calculated after partialling out mean high-frequency hearing loss (PTA_high).

There was a significant negative correlation between the DemTect and the S0N0 SRT in continuous noise (ρ = −0.461, p = 0.015) and between the DemTect score and the MSNF SRT in

TABLE 2 | Spearman correlation coefficients ρ for results on measures of hearing loss and speech perception in noise vs. DemTect score (first column).

	DemTect	DemTect PTA_high partialled out
PTA_4	−0.284	0.036
PTA_high	−0.375	N/A
SRT S0N0 cont.	**−0.461**	−0.325
SRT S0N0 mod.	−0.382	−0.212
SRT MSNF cont.	−0.203	−0.072
SRT MSNF mod.	**−0.502**	−0.404

Correlation coefficients after partialling out high-frequency hearing loss PTA_high (second column). Gray values indicate non-significant correlations (p > 0.05). Values in black indicate significant results (p ≤ 0.05). Values in boldface indicate significant results after applying a Bonferroni-Holm correction. N/A: correlation non-applicable.

modulated noise (ρ = −0.502, p = 0.006). A scatterplot showing individual DemTect scores and MSNF SRTs in modulated noise is provided as **Supplementary Material**. Furthermore, a negative correlation between PTA_high and DemTect score (ρ = −0.375, unadjusted p = 0.017) and between S0N0 SRT in modulated noise and DemTect score (ρ = −0.382, unadjusted p = 0.015) was found which failed to reach significance after Bonferroni-Holm correction (p = 0.06 for both correlation coefficients).

After partialling out high-frequency hearing loss, no significant correlations between DemTect score and all other measures were found.

Subjective Hearing Performance

In the evaluation of the SSQ questionnaire, both the obtained scores of the three sections of the questionnaire individually and the score of the whole questionnaire were analyzed. Boxplots of the SSQ results are shown in **Figure 7**. In the subjective evaluation of speech perception (SSQ section 1) the median score was 7.1 points (IQR: 5.9–8.4 points). The median of the subjective evaluation of spatial hearing (SSQ section 2) was 7.8 points (IQR: 6.5–9.0 points). The median score increased to 8.4 points (IQR: 7.4–9.1 points) for the subjective rating of hearing quality (SSQ section 3). The median score of the complete SSQ questionnaire was 7.8 points (IQR: 7.0–8.6 points).

Correlations between age and SSQ scores are shown in **Table 3**. There was no correlation between age and SSQ scores for speech and spatial perception in the senior study population. A correlation between on one hand age and on the other SSQ scores for sound quality (ρ = −0.364, unadjusted p = 0.021) and the total SSQ scores (ρ = −0.326, unadjusted p = 0.039) failed to reach significance after Bonferroni-Holm correction (p = 0.084 and p = 0.117). After partialling out high-frequency hearing loss, no significant correlations between age and SSQ scores were found.

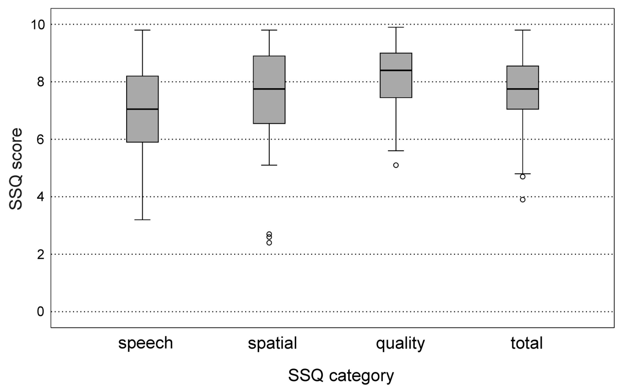


FIGURE 7 | Boxplots of speech, spatial, and qualities of hearing scale (SSQ) scores for the three subsets speech, spatial, quality, and for the mean total SSQ score.

TABLE 3 | Spearman correlation coefficients ρ for results on SSQ scores (speech, spatial, quality, and total SSQ score) vs. age (first column).

	Age	Age PTA_high partialled out
SSQ speech	−0.23	−0.113
SSQ spatial	−0.299	−0.246
SSQ quality	−0.364	−0.295
SSQ total	−0.326	−0.235

Correlation coefficients after partialling out high-frequency hearing loss PTA_high (second column). Gray values indicate non-significant correlations (p > 0.05). Values in black indicate significant results (p ≤ 0.05). After applying a Bonferroni-Holm correction no significant correlations were found.

TABLE 4 | Spearman correlation coefficients ρ for results on measures of hearing loss, speech perception in noise, and sound localization ability vs. SSQ scores of speech perception (first column), spatial hearing (second column) and total SSQ score (third column).

	SSQ speech	SSQ spatial	SSQ total
PTA_high	−0.373	−0.209	−0.343
SRT S0N0 cont.	**−0.433**	–	**−0.482**
SRT S0N0 mod.	−0.384	–	−0.403
SRT MSNF cont.	−0.285	−0.105	−0.242
SRT MSNF mod.	−0.388	−0.283	−0.357
localization error	–	−0.32	−0.316
front-back confusion	–	−0.145	−0.186

SSQ scores of speech perception were correlated with PTA_high and all SRTs in noise. SSQ scores of spatial perception were correlated with PTA_high, SRTs in MSNF, and localization measures. Gray values indicate non-significant correlations (p > 0.05). Values in black indicate significant results (p ≤ 0.05). Values in boldface indicate significant results after applying a Bonferroni-Holm correction. –: no correlations were calculated.

Correlation Between Subjective Hearing and Test Results

Spearman rank correlations were calculated between SSQ speech scores of the senior subject group and high-frequency hearing loss and SRTs in noise. SSQ spatial scores were analyzed for potential correlations with high-frequency hearing loss, SRTs in spatially separated noise and localization scores. Furthermore, correlations between mean total SSQ score and all hearing performance measures were calculated. The correlation coefficients are shown in **Table 4**.

A significant negative correlation (after applying a Bonferroni-Holm correction) was found between the SSQ score of speech perception and the SRT in configuration S0N0 for continuous noise (ρ = −0.433, $p = 0.025$). Negative correlations between SSQ score of speech and PTA_high (ρ = −0.373, unadjusted $p = 0.018$) and SRTs in modulated noise (S0N0: ρ = −0.384, unadjusted $p = 0.014$; MSNF: ρ = −0.388, unadjusted $p = 0.013$) failed to reach significance after applying a Bonferroni-Holm correction ($p = 0.052$ for all three correlations).

A negative correlation between the SSQ score for spatial hearing and localization error (ρ = −0.32, unadjusted $p = 0.044$) failed to reach significance after applying a Bonferroni-Holm correction ($p = 0.22$).

There was a significant negative correlation between the mean total SSQ score and the SRT in the configuration S0N0 in continuous noise (ρ = −0.482, $p = 0.014$). Negative correlations between total SSQ score and PTA_high (ρ = −0.343, unadjusted $p = 0.03$), SRTs in modulated noise (S0N0: ρ = −0.403, unadjusted $p = 0.01$; MSNF: ρ = −0.357, unadjusted $p = 0.024$) and localization error (ρ = −0.316, unadjusted $p = 0.047$) became non-significant (p = 0.12/0.06/0.12/0.141) after correction for multiple comparisons.

DISCUSSION

Speech Perception in Noise

In the senior subject group a significant impact of age on speech perception was found for both noise types and spatial noise configurations. Highest differences to young normal hearing adults were found in conditions with modulated noise. Furthermore, subjects with lower cognitive scores showed higher SRTs for both noise types in co-located target and masker condition and for modulated noise in the spatial masker condition MSNF. It should be noted that many of the senior subjects had age-related hearing loss. Therefore, especially high-frequency hearing loss was a confounding factor in this subjects which presumably accounted more for the decline in speech in noise as age itself. Accordingly, only one out of the four speech tests in noise correlated with age after partialling high-frequency hearing loss out. Still, even in the present cohort of seniors with mainly mild hearing loss, speech perception in certain situations of daily life (depending on signal-to-noise ratio) could be substantially degraded as reflected in the speech in noise tests using MSNF. The slope of the speech discrimination function for modulated noise in S0N0 and MSNF is ~6%/dB. Therefore, mean deterioration of speech perception in modulated noise in compared with young adults was 58.8% (S0N0) and more than 37 percent (MSNF). In conditions using continuous noise differences of about 18% (MSNF) and 31% (S0N0) were found. Since speech perception in clinical routine is mainly assessed in quiet or in continuous noise, it could be expected that deficits in speech perception in the elderly in everyday conditions are oftentimes underestimated.

In the co-located target and masker condition S0N0, mean SRTs in modulated noise were significantly lower than in continuous noise. The amplitude-modulated noise contains temporal gaps, which enables for release from masking (RM). Furthermore, Stone et al. (2011, 2012) showed that the inherent fluctuations in "continuous" noise also have a masking effect which leads to worse SRTs compared to temporally modulated noise with low modulation rates.

In our study RM was found to be 3.1 dB and, thus, 7.9 dB poorer than in a normal hearing adult group. A possible explanation for the reduced RM in seniors in the present study could be that these seniors had age-related hearing loss. As one result, audibility for sibilants is reduced. There are also deficits in frequency selection resulting in poorer separation of speech and noise and making temporal gaps more difficult or impossible to detect (Moore, 1985). Duquesnoy (1983) also found that older persons aged 75–88 years with presbycusis are less able to use the temporal gaps in fluctuating noise than normal-hearing persons. Peters et al. (1998) showed that both age and hearing impairment have a considerable influence on speech perception with highest impact in modulated noise. For young adults with normal hearing, RM was up to 4-7 dB whereas in older persons with hearing loss only 1.5 dB improvement was documented. Results from van Summers and Molis (2004) imply that signal audibility is not the major factor limiting RM in the presence of a mild to moderate hearing impairment. Hearing loss reduced the benefit from masker fluctuations for the majority of their study subjects even for an increase in presentation level of up to 30 dB. Rather, distortions in the processing of suprathreshold speech may account for reduced RM. Another aspect is a potential deterioration of temporal resolution. Results by Füllgrabe (2013) showed that temporal processing is reduced with increasing age even in the absence of a peripheral hearing loss. Sensitivity to temporal fine structure decreased in a monaural as well as binaural task with increasing age already beginning in early midlife.

The role of binaural processing is evident in the tested MSNF condition where speech and noise were spatially separated. Spatial release from masking (SRM) leads to improved SRTs compared to a co-located masker and target position due to the head shadow and binaural squelch effect. In the study by Duquesnoy (1983) it was found that binaural listening (noise signal from side, speech signal from front) could improve SRTs by 5–9 dB for young normal-hearing persons and by only 3–4 dB for elderly persons with age-related hearing loss. In the present study, SRTs for continuous noise were also significantly lower in MSNF than in S0N0 condition in elderly subjects. SRM in continuous noise was found to be in the same range than for young normal hearing adults. In the MSNF condition with four uncorrelated modulated noise sources the effect monaural unmasking is reduced compared to co-located speech and masker presentation. Even though temporal gaps are smaller than in single noise co-located masker condition, young normal hearing subjects still show 2.4 dB better SRTs than in MSNF with continuous noise (i.e., combined effects of monaural release from masking and SRM). On the other hand, MSNF SRTs in the senior group were 2.6 dB worse for modulated noise than for continuous noise in spite of the presence of temporal gaps. Thus, SRTs were even worse compared to continuous noise. This effect could be caused by distortions of binaural temporal processing in the senior subject group.

A review on the relationship between hearing loss and age and binaural processing is given by Moore (2021). Füllgrabe and Moore (2018) conducted a meta-analysis on the relations between binaural temporal fine structure sensitivity and hearing loss and age. Hearing loss and age were significantly negatively correlated to temporal fine structure sensitivity where age was a better predictor than audiometric threshold. Reduced temporal binaural processing could not solely explain the disruptive effect of modulated noise on SRM in MSNF. It is conceivable that in demanding binaural test conditions cognitive performance is more influential on auditory performance than in co-located masker conditions.

In the present study, DemTect scores correlated significantly with two out of four OLSA SRTs. During the OLSA task the subject has to remember five words before recalling them on a touch screen display. Working memory stores verbal information while processing that information (Baddeley and Hitch, 1974). If the verbal information matches information in the mental lexicon of long-term memory, the speech signal is recognized and processed. Hwang et al. (2017) showed that age and working memory capacity influences speech perception in noise in a group of hearing impaired subjects aged 24–80 years whereas no effect of age and only a low effect of working memory (subtest digit backward span) was found in a normal hearing group aged 27–73 years. Rudner et al. (2011) reported that hearing impaired subjects with higher working memory capacity performed better in speech perception in modulated noise. Deficits in cognitive performance could lead to an increased listening effort and may reduce the use of the temporal gaps for RM.

Sound Localization

The mean localization error in our study population was only slightly but not significantly increased compared with normal hearing young adults. This is in line with results reported by Otte et al. (2013) who measured sound localization in a group of normal hearing young adults (20–34 years) and older adults (63–80 years) with mild to moderate high-frequency hearing loss (mean hearing thresholds of subject group). However, the literature also contains studies showing that the ability to localize sound decreases with age. In such studies oftentimes localization ability was not only assessed for broad-band noise but also for narrow-band or highpass-filtered and lowpass-filtered noise stimuli (Abel et al., 2000; Dobreva et al., 2011; Freigang et al., 2015). Freigang et al. (2015) reported a decrease in sound localization accuracy for older adults which was most prominent for lateral sound source positions and high-frequency stimuli. Dobreva et al. (2011) reported a decrease in sound localization ability in the horizontal plane with age (for subjects groups aged 45–66 and 70–81 years in comparison to younger adults aged 19–41) for both broadband noise and narrowband noise. The discrepancy in the results might be due to differences in the methodology for measuring sound localization ability (e.g., angular span, pointer method, stimulus type and presentation level, amount of level roving, etc.) and in the distribution of age and hearing loss of the test subjects. The difference in mean localization error for broadband noise between younger and older subjects reported in the present study could be considered as clinically irrelevant. However, it cannot be ruled out that even in our subject group localization ability for higher frequency narrow-band sounds is deteriorated.

This is supported by the result of a significantly higher amount of front-back confusions compared with young adults which was confirmed in two other studies (Abel et al., 2000; Otte et al., 2013). The occurrence of front-back confusions seem to be directly related to the high-frequency hearing loss of elderly subjects. Interaural time differences and interaural level differences are the dominant cues for horizontal sound localization, whereas high-frequency monaural cues contribute significantly to vertical sound localization as well as to resolving front-back confusions. It was hypothesized that poor coding of interaural time differences in older subjects with presbycusis accounts for deficits in sound localization ability in the horizontal plane (Dobreva et al., 2011).

In the present study no significant correlation between the results of cognitive performance and the localization ability of sounds was found. Likewise, Neher et al. (2011) reported no cognitive measures as predictors for sound localization ability.

Cognition and Hearing Loss

Sixty-five percent of the subjects showed age-appropriate cognitive performance in the DemTect test, and 35% of the subjects showed a mild cognitive impairment. None of the subjects were suspected of having dementia. Since study subjects were recruited for participation via flyers and an advertisement at the grounds of the University Hospital Frankfurt, it is also conceivable that seniors with reduced cognitive performance who were already being cared for by nursing at home or in a retirement home may not have been reached at all. Therefore, our study group cannot be considered as representative.

Memory span of the working memory decreases with age (Cattell, 1971). This was also shown in the Berlin Aging Study (Lindenberger et al., 2010). In order to determine cognitive performance, 14 cognitive tests were performed, which could be assigned to five cognitive abilities. All five abilities were shown to decrease linearly with age, especially those abilities that belong to fluid intelligence. In our subject group without any severe cognitive impairment the cognitive-test performance also correlated significantly with age. This is surprising since the scores of the DemTect test are age-corrected.

On the other hand, the correlation between age and DemTect score vanished after partialling out high-frequency hearing loss and applying Bonferroni-Holm correction for multiple comparisons. It was also shown that subjects with higher age and more severe hearing loss (i.e., higher PTA_high) and/or with higher SRTs in noise (three out of four tests) tended to have lower DemTect scores. The Berlin Aging Study also reported that individuals with poorer hearing also had poorer cognitive performance (correlation $r = 0.5$). In a study by Lin et al. (2013) it was even observed that hearing loss can lead to an accelerated decline in cognitive performance by up to 30–40%. Study participants with hearing loss who did not wear hearing aids had slightly worse scores on cognitive tests than study participants with hearing aids.

However, it must also be considered that misunderstandings in verbal communication during the test procedure due to hearing loss could also impair test scores in the assessment of cognitive abilities. Füllgrabe (2020) showed that young participants with simulated hearing loss performed significantly worse in cognitive tasks using acoustically presented test items (forward digit span, backward digit span, listening span) than a control group with the same age without hearing loss. It was concluded that cognitive impairments could be overestimated in the presence of a hearing impairment.

Castiglione et al. (2019) introduced an audiological screening model of subjects at risk of cognitive decline with slight to moderate hearing loss. It could potentially be useful to screen elderly subjects with hearing loss for dementia on regular basis and to conduct hearing test in patients suffering from an onset of dementia.

Subjective Hearing Performance

The present study included subjects who described themselves as having normal hearing and who did not use hearing aids. However, only 26 of the 40 subjects had normal hearing in both ears and three subjects had normal hearing in one ear. In five of the 40 test persons, even the indication for a unilateral hearing aid provision was present, and five other test persons even had an indication for bilateral hearing aid provision.

Considering the correlations of SSQ scores with high-frequency pure-tone hearing loss and speech perception in noise (S0N0 in continuous noise) there was at least some awareness of the subjects for an own auditory deficit. Nevertheless, all of them described themselves as having no hearing problems. This suggests that not only the own perception of hearing loss is a problem for seniors, but also to accept hearing disabilities and deciding to seek help is a challenge. Carson (2016) reports that it takes an average of 7–10 years for a person to seek medical help after the first recognized signs of hearing loss. Since presbycusis is an age-related condition, its acceptance also means acceptance of aging, which for some seniors may mean a reduction in independence or a loss of control. According to Donahue et al. (2010) only one in five people suffering from age-related hearing loss sought professional help. Therefore, it would be advisable that seniors should undergo routine hearing screenings in order to detect hearing disorders as soon as possible, so that hearing aids can be prescribed at an early stage.

Potential Limitations of the Study

All subjects without self-reported hearing complaints were included. Since subjects suffering from so far unnoticed hearing loss were not excluded the study population is partly inhomogeneous. Another drawback is potentially that subjects with asymmetric hearing loss were included. A higher amount of subjects aged 80 and older would be desirable to extent the quality of correlation analysis. Furthermore, additional tests on temporal processing (e.g., on the perception of temporal fine structure) are in need to interpret deficits in speech perception in modulated noise or spatial noise conditions and its relation to cognitive performance.

CONCLUSION

Although no complaints about hearing ability were reported in the present study group of seniors, the results of the study support the hypotheses that hearing performance decreases with increasing age together with declining cognitive abilities even if not detected by the subject itself. This holds especially for speech perception in noise in complex conditions where intact binaural hearing is a mandatory requirement. Therefore, special attention should be given to hearing screening programs to improve the quality of life of older people. Speech perception tests using temporally modulated noise can serve as a screening method for early detection of hearing disorders in older adults. Hearing screening should also be mandatory for dementia patients, just as dementia screening is mandatory for seniors with known hearing loss. Further research is needed to investigate on the causality between dementia and hearing loss.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics Committee of the Department of Medicine of the Goethe-University Frankfurt am Main, Germany (No. 164/13). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

TW and UB contributed to the conception and design of the study. CM and TW organized and conducted the experiments, performed the statistical analysis, and wrote the first draft of the manuscript. All authors contributed to the interpretation of the data, manuscript revision, and read and approved the submitted version.

FUNDING

This work was funded by Cochlear Deutschland GmbH & Co. KG. This study was designed in collaboration with Cochlear Deutschland GmbH & Co. KG. The funders had no role in data collection and analysis and decision to publish or preparation of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2022.845285/full#supplementary-material

REFERENCES

Abel, S. M., Giguère, C., Consoli, A., and Papsin, B. C. (2000). The effect of aging on horizontal plane sound localization. *J. Acoust. Soc. Am.* 108, 743–752. doi: 10.1121/1.429607

Baddeley, A. D., and Hitch, G. (1974). "Working memory," in *The Psychology Of Learning And Motivation: Advances In Research And Theory*, ed. G. H. Bower (London: Academic Press), 47–89.

Berkhout, A. J. (1988). A holographic approach to acoustic control. *JAES* 36, 977–995.

Brand, T., and Kollmeier, B. (2002). Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests. *J. Acoust. Soc. Am.* 111, 2801–2810. doi: 10.1121/1.1479152

Carson, A. J. (2016). The decision-making spiral in seeking help for hearing problems. *Hear. J.* 69, 28–32.

Castiglione, A., Casa, M., Gallo, S., Sorrentino, F., Dhima, S., Cilia, D., et al. (2019). Correspondence between cognitive and audiological evaluations among the elderly: a preliminary report of an audiological screening model of subjects at risk of cognitive decline with slight to moderate hearing loss. *Front. Neurosci* 13:1279. doi: 10.3389/fnins.2019.01279

Cattell, R. B. (1971). *Abilities: Their Structure, Growth, And Action*. Boston: Houghton Mifflin.

Dobreva, M. S., O'Neill, W. E., and Paige, G. D. (2011). Influence of aging on human sound localization. *J. Neurophysiol.* 105, 2471–2486. doi: 10.1152/jn.00951.2010

Donahue, A., Dubno, J. R., and Beck, L. (2010). Guest editorial: accessible and affordable hearing health care for adults with mild to moderate hearing loss. *Ear Hear.* 31, 2–6. doi: 10.1097/AUD.0b013e3181cbc783

Dubno, J. R., Dirks, D. D., and Morgan, D. E. (1984). Effects of age and mild hearing loss on speech recognition in noise. *J. Acoust. Soc. Am.* 76, 87–96. doi: 10.1121/1.391011

Duquesnoy, A. J. (1983). Effect of a single interfering noise or speech source upon the binaural sentence intelligibility of aged persons. *J. Acoust. Soc. Am.* 74, 739–743. doi: 10.1121/1.389859

European Commission (2020). *European Commission Report on European Commission Report on the Impact of Demographic Change*. Available online at: https://ec.europa.eu/info/sites/default/files/demography_report_2020_n.pdf. (accessed February 14, 2022).

Fastl, H. (1987). Ein Störgeräusch für die Sprachaudiometrie [A background noise for speech audiometry]. *Audiol. Akustik* 26, 2–13.

Fastl, H., and Zwicker, E. (2007). *Psychoacoustics: Facts And Models*, 3rd Edn. Berlin: Springer.

Freigang, C., Richter, N., Rübsamen, R., and Ludwig, A. A. (2015). Age-related changes in sound localisation ability. *Cell Tissue Res.* 361, 371–386. doi: 10.1007/s00441-015-2230-8

Füllgrabe, C. (2013). Age-dependent changes in temporal-fine-structure processing in the absence of peripheral hearing loss. *Am. J. Audiol.* 22, 313–315. doi: 10.1044/1059-0889(2013/12-0070)

Füllgrabe, C. (2020). On the possible overestimation of cognitive decline: the impact of age-related hearing loss on cognitive-test performance. *Front. Neurosci.* 14:454. doi: 10.3389/fnins.2020.00454

Füllgrabe, C., and Moore, B. C. (2018). The association between the processing of binaural temporal-fine-structure information and audiometric threshold and age: a meta-analysis. *Trends Hear.* 22:2331216518797259. doi: 10.1177/2331216518797259

Füllgrabe, C., Moore, B. C., and Stone, M. A. (2015). Age-group differences in speech identification despite matched audiometrically normal hearing: contributions from auditory temporal processing and cognition. *Front. Aging Neurosci.* 6:347. doi: 10.3389/fnagi.2014.00347

Garnefski, N., and Kraaij, V. (2012). Cognitive coping and goal adjustment are associated with symptoms of depression and anxiety in people with acquired hearing loss. *Int. J. Audiol.* 51, 545–550. doi: 10.3109/14992027.2012.675628

Gatehouse, S., and Noble, W. (2004). The speech, spatial and qualities of hearing scale (SSQ). *Int. J. Audiol.* 43, 85–99.

Hwang, J. S., Kim, K. H., and Lee, J. H. (2017). Factors affecting sentence-in-noise recognition for normal hearing listeners and listeners with hearing loss. *J. Audiol. Otol.* 21, 81–87. doi: 10.7874/jao.2017.21.2.81

Kalbe, E., Kessler, J., Calabrese, P., Smith, R., Passmore, A. P., Brand, M., et al. (2004). DemTect: a new, sensitive cognitive screening test to support the diagnosis of mild cognitive impairment and early dementia. *Int. J. Geriatr. Psychiatry* 19, 136–143. doi: 10.1002/gps.1042

Kessler, J., Calabrese, P., Kalbe, E., and Berger, F. (2000). DemTect: ein neues screening-verfahren zur unterstützung der demenzdiagnostik. *Psycho* 19, 343–347.

Kießling, J., Grugel, L., Meister, H., and Meis, M. (2011). Übertragung der fragebögen SADL, ECHO und SSQ ins deutsche und deren evaluation. [German translations of questionnaires SADL, ECHO and SSQ and their evalution.]. *Z Audiol.* 50, 6–16.

Lin, F. R., Yaffe, K., Xia, J., Xue, Q.-L., Harris, T. B., Purchase-Helzner, E., et al. (2013). Hearing loss and cognitive decline in older adults. *JAMA Int. Med.* 173, 293–299.

Lindenberger, U., Smith, J., Mayer, K. U., and Baltes, P. B. (eds) (2010). *Die Berliner Altersstudie*. Berlin: Akad.-Verl.

Meister, H., Schreitmüller, S., Grugel, L., Landwehr, M., Wedel, H., von Walger, M., et al. (2011). Untersuchungen zum sprachverstehen und zu kognitiven fähigkeiten im alter [Examination of speech perception and cognitive functioning in the elderly]. *HNO* 59, 689–695.

Moore, B. C. (1985). Frequency selectivity and temporal resolution in normal and hearing-impaired listeners. *Br. J. Audiol.* 19, 189–201. doi: 10.3109/03005368509078973

Moore, B. C. (2021). Effects of hearing loss and age on the binaural processing of temporal envelope and temporal fine structure information. *Hear. Res.* 402:107991. doi: 10.1016/j.heares.2020.107991

Neher, T., Laugesen, S., Jensen, N. S., and Kragelund, L. (2011). Can basic auditory and cognitive measures predict hearing-impaired listeners' localization and spatial speech recognition abilities? *J. Acoust. Soc. Am.* 130, 1542–1558. doi: 10.1121/1.3608122

Otte, R. J., Agterberg, M. J. H., van Wanrooij, M. M., Snik, A. F. M., and van Opstal, A. J. (2013). Age-related hearing loss and ear morphology affect vertical but not horizontal sound-localization performance. *J. Assoc. Res. Otolaryngol.* 14, 261–273. doi: 10.1007/s10162-012-0367-7

Peters, R. W., Moore, B. C., and Baer, T. (1998). Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people. *J. Acoust. Soc. Am.* 103, 577–587. doi: 10.1121/1.421128

Rader, T., Fastl, H., and Baumann, U. (2013). Speech perception with combined electric-acoustic stimulation and bilateral cochlear implants in a multisource noise field. *Ear Hear.* 34, 324–332. doi: 10.1097/AUD.0b013e318272f189

Ramage-Morin, P. L., Banks, R., Pineault, D., and Atrach, M. (2019). Unperceived hearing loss among Canadians aged 40 to 79. *Health Rep.* 30, 11–20. doi: 10.25318/82-003-x201900800002-eng

Rudner, M., Rönnberg, J., and Lunner, T. (2011). Working memory supports listening in noise for persons with hearing impairment. *J. Am. Acad. Audiol.* 22, 156–167. doi: 10.3766/jaaa.22.3.4

Seeber, B. (2002). A new method for localization studies. *Acta Acust.* 88, 446–450.

Statistische Ämter des Bundes und der Länder (2011). *Demografischer Wandel in Deutschland: Heft 1 - Bevölkerungs- und Haushaltsentwicklung Im Bund Und In Den Ländern*. Available online at: https://www.destatis.de/GPStatistik/servlets/MCRFileNodeServlet/DEHeft_derivate_00012505/5871101119004.pdf. (accessed February 14, 2022).

Stone, M. A., Füllgrabe, C., and Moore, B. C. (2012). Notionally steady background noise acts primarily as a modulation masker of speech. *J. Acoust. Soc. Am.* 132, 317–326. doi: 10.1121/1.4725766

Stone, M. A., Füllgrabe, C., Mackinnon, R. C., and Moore, B. C. (2011). The importance for speech intelligibility of random fluctuations in "steady" background noise. *J. Acoust. Soc. Am.* 130, 2874–2881. doi: 10.1121/1.3641371

Summers, V., and Molis, M. R. (2004). Speech recognition in fluctuating and continuous maskers. *J. Speech Lang. Hear. Res.* 47, 245–256. doi: 10.1044/1092-4388(2004/020)

Völter, C., Götze, L., Dazert, S., Wirth, R., and Thomas, J. P. (2020). Impact of hearing loss on geriatric assessment. *Clin. Interv. Aging* 15, 2453–2467. doi: 10.2147/CIA.S281627

Wagener, K., Kühnel, V., and Kollmeier, B. (1999). Entwicklung und evaluation eines Satztests für die deutsche Sprache. Teil I: design des oldenburger satztests

[Development and evaluation of a german sentence test. part i: design of the oldenburg sentence test]. *Z Audiol.* 38, 4–15.

Weissgerber, T., Rader, T., and Baumann, U. (2015). Impact of a moving noise masker on speech perception in cochlear implant users. *PLoS One* 10:e0126133. doi: 10.1371/journal.pone.0126133

Weissgerber, T., Rader, T., and Baumann, U. (2017). Effectiveness of directional microphones in bilateral/bimodal cochlear implant users-impact of spatial and temporal noise characteristics. *Otol. Neurotol.* 38, e551–e557. doi: 10.1097/MAO.0000000000001524

World Health Organization (2017). *Deafness And Hearing Impairment: Fact Sheet No. 300.2004*. Available online at: http://www.who.int/mediacentre/factsheets/fs300/en/. (accessed December 27, 2021).

World Health Organization [WHO] (2001). *Grades Of Hearing Impairment*. Available online at: http://www.who.int/pbd/deafness/hearing_impairment_grades/en/. (accessed December 27, 2021).

Conflict of Interest: 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 | Frontiers in Neuroscience

BRIEF RESEARCH REPORT
published: 25 April 2022
doi: 10.3389/fnins.2022.846117

Age-Related Changes in the Perception of Emotions in Speech: Assessing Thresholds of Prosody and Semantics Recognition in Noise for Young and Older Adults

Yehuda I. Dor[1,2], Daniel Algom[1], Vered Shakuf[3] and Boaz M. Ben-David[2,4,5]*

[1] School of Psychological Sciences, Tel Aviv University, Tel Aviv, Israel, [2] Communication, Aging and Neuropsychology Lab (CANlab), Baruch Ivcher School of Psychology, Reichman University (IDC), Herzliya, Israel, [3] Department of Communications Disorders, Achva Academic College, Arugot, Israel, [4] Toronto Rehabilitation Institute, University Health Networks (UHN), Toronto, ON, Canada, [5] Department of Speech-Language Pathology, University of Toronto, Toronto, ON, Canada

Older adults process emotions in speech differently than do young adults. However, it is unclear whether these age-related changes impact all speech channels to the same extent, and whether they originate from a sensory or a cognitive source. The current study adopted a psychophysical approach to directly compare young and older adults' sensory thresholds for emotion recognition in two channels of spoken-emotions: prosody (tone) and semantics (words). A total of 29 young adults and 26 older adults listened to 50 spoken sentences presenting different combinations of emotions across prosody and semantics. They were asked to recognize the prosodic or semantic emotion, in separate tasks. Sentences were presented on the background of speech-spectrum noise ranging from SNR of −15 dB (difficult) to +5 dB (easy). Individual recognition thresholds were calculated (by fitting psychometric functions) separately for prosodic and semantic recognition. Results indicated that: (1). recognition thresholds were better for young over older adults, suggesting an age-related general decrease across channels; (2). recognition thresholds were better for prosody over semantics, suggesting a prosodic advantage; (3). importantly, the prosodic advantage in thresholds did not differ between age groups (thus a sensory source for age-related differences in spoken-emotions processing was not supported); and (4). larger failures of selective attention were found for older adults than for young adults, indicating that older adults experienced larger difficulties in inhibiting irrelevant information. Taken together, results do not support a sole sensory source, but rather an interplay of cognitive and sensory sources for age-related differences in spoken-emotions processing.

Keywords: auditory processing, speech perception, aging, semantics, emotions, noise, auditory sensory-cognitive interactions, prosody

OPEN ACCESS

Edited by:
Leah Fostick,
Ariel University, Israel

Reviewed by:
Uwe Baumann,
University Hospital Frankfurt, Germany
Patrizia Mancini,
Sapienza University of Rome, Italy

Correspondence:
Boaz M. Ben-David
boaz.ben.david@idc.ac.il

Specialty section:
This article was submitted to Perception Science, a section of the journal Frontiers in Neuroscience

Received: *30 December 2021*
Accepted: *14 March 2022*
Published: *25 April 2022*

Citation:
Dor YI, Algom D, Shakuf V and Ben-David BM (2022) Age-Related Changes in the Perception of Emotions in Speech: Assessing Thresholds of Prosody and Semantics Recognition in Noise for Young and Older Adults. Front. Neurosci. 16:846117. doi: 10.3389/fnins.2022.846117

INTRODUCTION

Communication in older age is essential to maintain quality of life, cognitive skills, and emotional wellbeing (Heinrich et al., 2016; Livingston et al., 2017). Abundant evidence suggests that speech processing is impaired in aging, with severe implications (Helfer et al., 2017). Specifically, the literature points to major age-related changes in the perception of emotions in spoken language

(Ben-David et al., 2019). However, it is not clear whether these changes are domain-specific or reflect a general age-related decline in emotion perception (Ruffman et al., 2008; Castro and Isaacowitz, 2019). In other words, do these changes stem from a specific deficit in processing of certain types of emotional channels (while processing of others is preserved), or from a general decrease in processing? In addition, there is debate on the mechanisms underlying these age-related changes; various sensory, cognitive, affective, and neural factors have been considered (Mather, 2016; Helfer et al., 2017; Ben-David et al., 2019).

In spoken language, emotions are presented via two main channels: (a) emotional semantics – the emotional meaning of spoken words or a complete sentence (segmental speech information); (b) emotional prosody – the tone of speech (suprasegmental speech information), composed of vocal cues such as stress, rhythm, and pitch. Processing of emotional speech is therefore a complex and dynamic integration of information, which may be congruent or incongruent, from these two channels. Significant age-related changes are indicated when incongruent prosody-semantics emotional combinations are presented. Specifically, when asked to integrate the two channels, young adults rely mainly on emotional prosody, while older adults weigh the two channels more equally (Dupuis and Pichora-Fuller, 2010; Ben-David et al., 2019). In addition, when listeners are asked to focus on only one speech channel, larger failures of selective attention are found for older adults than for young adults (Ben-David et al., 2019). In other words, the same spoken emotional sentences are interpreted differently by older and young listeners.

Mainly, cognitive and sensory sources have been suggested for these age-related differences (Ben-David et al., 2019). Following a cognitive source, age-related differences in executive functions, especially inhibition (Hasher and Zacks, 1988), are at the basis of changes in spoken emotion processing (Wingfield and Tun, 2001; Harel-Arbeli et al., 2021). Namely, both older and young adults may implicitly adopt the same weighting schematics – i.e., more weight to the prosodic than to the semantic channel. However, older adults might find it more difficult to inhibit the semantic information, processing it to a larger extent than intended.

An alternative sensory source lies in the relative imbalance between dimensions. The literature suggests that when one dimension becomes more perceptually salient than the other, the system is biased to rely on the first (Melara and Algom, 2003). Accordingly, young adults may be biased to process the prosody over the semantics, because emotional prosody is more sensory salient than emotional semantics. However, if this dimensional imbalance is reduced for older adults, the prosodic bias might be diminished as well (for a discussion on age-related sensory and dimensional-imbalance changes, see Ben-David and Schneider, 2009, 2010).

Some evidence in the literature may support this sensory source, with a specific age-related deficit in *prosodic processing* that might not be accompanied by a similar deficit in spoken-word processing. Indeed, age-related decrease in the recognition of prosodic information has been widely reported, both in quiet and in noise (Dmitrieva and Gelman, 2012; Lambrecht et al., 2012; Dupuis and Pichora-Fuller, 2014, 2015; Ben-David et al., 2019), suggesting a specific deficit in decoding emotional prosody in aging (Orbelo et al., 2005; Mitchell, 2007; Mitchell and Kingston, 2011). This prosodic deficit may relate to senescent changes in auditory brain areas and neural activity patterns (Orbelo et al., 2005; Giroud et al., 2019; Myers et al., 2019; Grandjean, 2021). However, there are mixed findings in the literature regarding the extent of age-related changes in *semantic processing*. While some studies have found a decline in older adults' ability to extract the emotional meaning from words (Grunwald et al., 1999; Isaacowitz et al., 2007), other studies have maintained that semantic processing is preserved, at least when speech is presented in ideal listening conditions (Phillips et al., 2002; Ben-David et al., 2019). In sum, an age-related decrease in sensory dimensional imbalance may be the source for the age-related decrease in prosodic bias.

In the current study, we adopted a psychophysical approach to test the sensory base of age-related differences in processing of spoken emotions. Following the results obtained by Ben-David et al. (2019), we directly asked older and young listeners to recognize the prosodic emotion and semantic emotion of 50 spoken sentences in separate trials. Sentences were presented in five different signal-to-noise-ratios (SNRs) to calculate emotional recognition thresholds. Take, for example, the semantically happy sentence "I won the lottery" spoken with sad prosody. In previous studies, young adults were found to judge this sentence to convey mostly sadness (prosody), whereas older adults judged the sentence to present a similar extent of happiness (semantics) and sadness (prosody; Ben-David et al., 2019). A sensory source would be supported if a larger prosodic advantage in thresholds were to be found for young over older adults. A cognitive source would be supported if larger failures of selective attention were to be found for older adults, as gauged by accuracy differences between congruent and incongruent sentences. Note, the two sources are not mutually exclusive.

The following hypotheses were made:

1. *Age-related advantage*: Recognition thresholds and accuracy would be lower (i.e., better) for young than for older adults.
2. *Prosodic advantage*: Across age groups, recognition thresholds for emotional prosody would be lower (i.e., better) than for emotional semantics.
3. *Age-related differences in prosodic advantage*: As the literature is not clear, we did not wish to make an a-priori hypothesis as to whether the advantage in prosodic over semantic recognition thresholds would be affected by age group or not.
4. *Failures of selective attention*: Selective attention failures would be larger for older adults.

MATERIALS AND METHODS

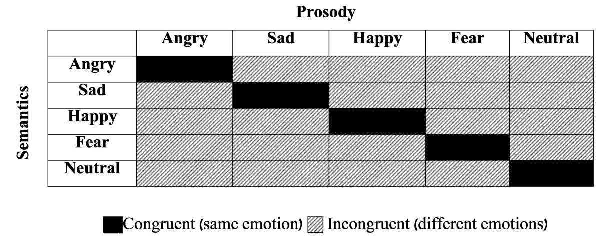
Participants

A total of 26 older adults from the community (16 women; 58-75 years old, *M* = 65.76 years, *SD* = 4.80) and 29 young adults, undergraduate students from Reichman University (24 women; 22-27 years old, *M* = 25.40 years, *SD* = 1.17) were recruited for this study and met the following inclusion criteria:

(a) native Hebrew speakers as assessed by self-reports (Ben-David and Icht, 2018), and verified by above-average standard scores for their age range on a vocabulary test (subscale of the WAIS-III, Goodman, 2001), as language proficiency is related to processing of emotional semantics (Phillips et al., 2002); (b) good ocular health; no auditory, cognitive or language problems, and without any medical or mental conditions related to emotional processing as assessed by self-reports (Nitsan et al., 2019); (c) no indication of clinical depression as assessed by self-reports (older: GDS, Zalsman et al., 1998; young: DASS-21, Henry and Crawford, 2005); and (d) pure-tone air-conduction thresholds within clinically normal limits for their age group, for 500, 1,000, and 2,000 Hz (average pure-tone thresholds ≤ 15 dB HL for young, and ≤ 25 dB HL for older adults, difference between ears < 20 dB HL). Note, groups were matched on years of education (M = 14.23 and 14.19 for young and older adults, respectively), taken as a reliable gauge for linguistic skills (Kaufman et al., 1989; Ben-David et al., 2015). Young adults participated in the study for partial course credit, and older adults were compensated by the equivalent of $10. From the final dataset, we excluded data of two young participants who did not follow the instructions, and of four older adults who exhibited very low recognition rates (< 50% correct recognition in the easiest SNR). A detailed description of the demographic and audiological characteristics of participants can be found in **Supplementary Appendix A**.

Stimuli

The stimulus set was made of 50 spoken sentences taken from the Test for Rating of Emotions in Speech (T-RES; Ben-David et al., 2016, 2019), which presents emotional semantic and prosodic content in different combinations from trial to trial. Five different emotions were used: Anger, Happiness, Sadness, Fear, and Neutrality. Each semantic category was represented in each of the tested prosodies, generating a 5 (semantics) *5 (prosody) matrix (see **Figure 1**). The experimental set consisted of two sentences in each of the 25 different combinations of emotional semantics and prosody. Ten sentences were congruent (e.g., semantically angry semantics such as "Get out of my room" spoken with congruent angry prosody; black cells in **Figure 1**) and 40 were incongruent (e.g., semantically happy semantics such as “I won the lottery” spoken with incongruent sad prosody; gray cells in **Figure 1**). All spoken sentences were recorded by a professional radio drama actress; digital audio files were equated with respect to their duration and root-mean-square amplitude (before they were mixed with noise).



Semantics \ Prosody	Angry	Sad	Happy	Fear	Neutral
Angry	■	▨	▨	▨	▨
Sad	▨	■	▨	▨	▨
Happy	▨	▨	■	▨	▨
Fear	▨	▨	▨	■	▨
Neutral	▨	▨	▨	▨	■

■ Congruent (same emotion) ▨ Incongruent (different emotions)

FIGURE 1 | General design of the Stimuli. All combinations of prosodic and semantic emotions are presented. Each cell represents two different sentences used in this study. Black cells: congruent sentences (same emotion in both speech channels). Gray cells: incongruent sentences (different emotions in semantics and prosody).

Reliability, Sensitivity, and Validity

We used the Hebrew version of the T-RES sentences. Content validity (Chan, 2014) was confirmed by verifying that all semantic stimuli were distinctive in their categories and exemplars of their respective semantic categories for both young and older listeners, and equated on main linguistic characteristics. For full details on the procedure for stimuli selection, see Ben-David et al. (2011b, 2013, 2019). A recent study from our lab has further shown that the discrete prosodic emotions are clearly distinct in acoustic characteristics (mean F0 and speech rate; Carl et al., 2022) in this set. The T-RES reliability was confirmed as data for young adult undergraduates were found to be equivalent across studies and platforms (Ben-David et al., 2021). The T-RES stimuli were also found to be valid and sensitive in detecting population-related differences in various studies. For example, expected differences in spoken emotional processing were found when comparing cochlear implant users and their peers (Taitelbaum-Swead et al., 2022).

Sentence Division and Combination With Noise

The final set was divided into five subsets of ten sentences each, with each subset consisting of two congruent and eight incongruent sentences. Each of the five emotional prosodies and each of the five emotional semantic categories was represented twice in each subset (see **Supplementary Appendix B**). Using PRAAT software (Boersma and Weenink, 2019), stimuli in each subset were combined with a different level of background speech-spectrum noise using a standard steady-state noise masker taken from the Revised Speech Perception in Noise test (Bilger et al., 1984; for spectral analysis of this noise, see Figure 6 in Ben-David et al., 2012). Five SNR levels were used: −15 dB, −10 dB, −5 dB, 0 dB, and +5 dB; creating a scale from the most difficult SNR (−15 dB) to the easiest SNR (+5 dB).

Procedure and Apparatus

Upon arrival, all participants received a short explanation regarding the experimental task and signed an informed consent form. Participants completed the self-reports and the vocabulary test. Next, they were seated in an IAC sound-attenuated booth and performed the pure-tone hearing thresholds test. All auditory stimuli were presented via MAC-51 audiometer headphones. Spoken sentences (experimental task) were presented 40 dB above individual audiometric thresholds (pure-tone average) in quiet, to partially mitigate age-related differences in auditory thresholds. Instructions were presented on a 17-in. flat color monitor.

Experimental Session

The experimental session consisted of two five-alternative-forced-choice (5-AFC) tasks. In both, participants were instructed to recognize the emotion presented, choosing one of five options (anger, happiness, sadness, fear, and neutrality) by pressing a designated key on the keyboard. Listeners were asked to recognize only the emotion presented by the semantics in the Semantics-recognition task, or only by prosodics in the Prosody-recognition task. Each task consisted of five blocks of ten spoken sentences each, with different levels of SNR in each block. The order of tasks (Semantics-recognition or Prosody-recognition) and the order of blocks in each task were counterbalanced across participants, using a Latin-square design. The order of sentences within each block was fully randomized. The whole session (two tasks with 100 sentences in total) lasted less than 30 min. Participants were given the option to take short breaks before the session, or between the tasks, if needed.

Data Fitting and Psychometric Functions

For each participant, five recognition-accuracy rates were calculated separately for prosody and semantics, based on average accuracy across the ten sentences in each of the five SNRs. Using a customized MATLAB script (McMurray, 2017), data were fitted to the logistic psychometric function of the form,

$$f(x) = A + \frac{L-A}{1+e^{-k(x-x_0)}}, \tag{1}$$

where *f(x)* represents recognition-accuracy rates, *x* is the SNR in dB, *L* and *A* are the upper and lower asymptotes of the function, respectively. Most importantly, the parameter x_0 represents the function's crossover point, or the *x* value that corresponds to middle performance between the boundaries of the function. The crossover point is taken to represent the point at which the rate of increase in recognition as a function of SNR begins to decrease. As such, the value of x_0 can serve as an index for individual recognition statistical threshold (Ben-David et al., 2012; Morgan, 2021). Finally, *k* represents the function's slope at x_0.

The lower asymptote of the function (*A*) for all conditions was pre-defined as 0.2 (chance level) using two techniques: (1) All performance levels averaging under 0.2 were corrected to 0.2 to avoid function estimations below chance level (1.4% of the data corrected). (2) We added an estimation level of 0.2 recognition rates (chance level) for an SNR of −20 dB, to correspond to the function's predicted lower bound. However, we chose not to pre-define the upper bound of the function (i.e., maximum recognition rates, see Morgan, 2021), as even without any background noise emotional recognition rates are not expected to reach 100%, especially for older adults (see Ruffman et al., 2008; Ben-David et al., 2019). Hence, the three other parameters (x_0, *L*, and *k*) were estimated based on our data. Correlations between actual data and the values predicted by the psychometric function were high (Mean correlation, 0.98–0.95), indicating a very good fit (McMurray, 2017) for both young and older adults. For full details regarding recognition rates, fitted psychometric functions' parameters, and quality of fits, see **Supplementary Appendices C** and **D**.

Statistical Analysis

All analyses of the thresholds, maximum asymptotes, and slopes (x_0, *L*, and *k*, taken from the psychometric function) included mixed linear modeling, MLM (SPSS Statistics 20; IBM Corp, 2011), with each serving as the dependent variable in different models. Group (young adults vs. older adults) was the between participant variable and Speech Channel (Prosody-rating vs. Semantics-rating) was the within participant variable. To test Selective Attention, the same MLM model was used, with recognition-accuracy rates (averaged across all SNRs) as the dependent variable, and the Selective Attention factor (congruent vs. incongruent sentences) added as another within participant factor.

RESULTS

Analysis of Thresholds and Recognition Rates

Table 1A presents the full MLM analyses of recognition thresholds, maximum asymptotes, and slopes. Results indicated a significant main effect for Age Group, *F*(1,47) = 14.57, $p < 0.001$, suggesting lower recognition thresholds for young, compared to older adults (average thresholds of −9.57 dB vs. −7.37 dB, respectively). A significant main effect was also found for Speech Channel, *F*(1,47) = 74.98, $p < 0.001$, suggesting lower recognition thresholds for emotions in prosody, compared to semantics (average thresholds of −10.09 dB vs. −6.85 dB, respectively). However, the interaction of the two factors was not significant, *F*(1,47) = 1.24, *p* = 0.27, indicating that the prosodic threshold advantage was similar for both age groups (left column of **Table 1A**). When using the same model to test differences in maximum asymptotes (i.e., maximal recognition rates under minimal noise) significant main effects were found for Age Group, *F*(1,47) = 27.62, $p < 0.001$, and for Speech Channel, *F*(1,47) = 5.34, *p* = 0.025, without a significant interaction between the two, *F*(1,47) = 0.304, *p* = 0.584 (middle column of **Table 1A**). When the same model was used to test differences in slopes, none of the tested effects were significant, indicating similar growth rates across all conditions (right column of **Table 1A**). When we excluded from analysis all psychometric functions whose fit quality was less than 0.9 (excluding seven functions, 7% of data), the result pattern remained the same (see **Supplementary Appendix E**).

To sum, our first and second hypotheses were confirmed: Young adults' recognition thresholds were lower (better) than those of older adults (a difference of about 2.2 dB), and prosodic emotions yielded lower recognition thresholds than did semantics emotions (a difference of about 3.3 dB). Critically, regarding our third hypothesis, the relative extent of the advantage of prosody over semantics was highly similar for older and young adults. Namely, prosodic thresholds were better than semantic thresholds by about a third, 32.16%, and 32.06% (−8.98 vs. −5.96 dB SNR; and, −11.40 vs. −7.74 dB SNR) for older and young adults, respectively. These results and the estimated psychometric functions in different Age Groups and Speech Channels are visually presented in **Figure 2**.

TABLE 1 | Model Summary and results of MLM analyses.

A: Psychometric Function's Parameters

	Threshold	**Max recognition**	**Slope**
Age Group	$F(1,47) = 14.57, p < 0.001$	$F(1,47) = 27.62, p < 0.001$	$F(1,47) = 1.17, p = 0.285$
Speech Channel	$F(1,47) = 74.98, p < 0.001$	$F(1,47) = 5.34, p = 0.025$	$F(1,47) = 1.04, p = 0.313$
Age Group X Speech Channel	$F(1,47) = 1.24, p = 0.272$	$F(1,47) = 0.304, p = 0.584$	$F(1,47) = 0.233, p = 0.632$
Model Summary	BIC = 448.87	BIC = −135.16	BIC = −53.79

B: Selective Attention

	Recognition rates
Age Group	$F(1,42.82) = 20.45, p < 0.001$
Speech Channel	$F(1,40.64) = 12.88, p = 0.001$
Selective Attention	$F(1,44.04) = 55.6, p < 0.001$
Age Group X Speech Channel	$F(1,40.64) = 0.809, p = 0.374$
Age Group X Selective Attention	$F(1,44.04) = 10.2, p = 0.003$
Speech Channel X Selective Attention	$F(1,32.13) = 2.76, p = 0.106$
Age Group X Speech Channel X Selective Attention	$F(1,32.13) = 3.1, p = 0.088$
Model Summary	BIC = −161.72

Top panel: analysis of individual psychometric functions' parameters (left column: Thresholds, x_0 parameter; middle column: Max recognition, L parameter, maximum asymptote; right column: Slope, k parameter) for all data. Bottom panel: analysis of Selective Attention effects (difference between recognition rates of emotions in congruent and incongruent sentences). Significant effects are shaded.

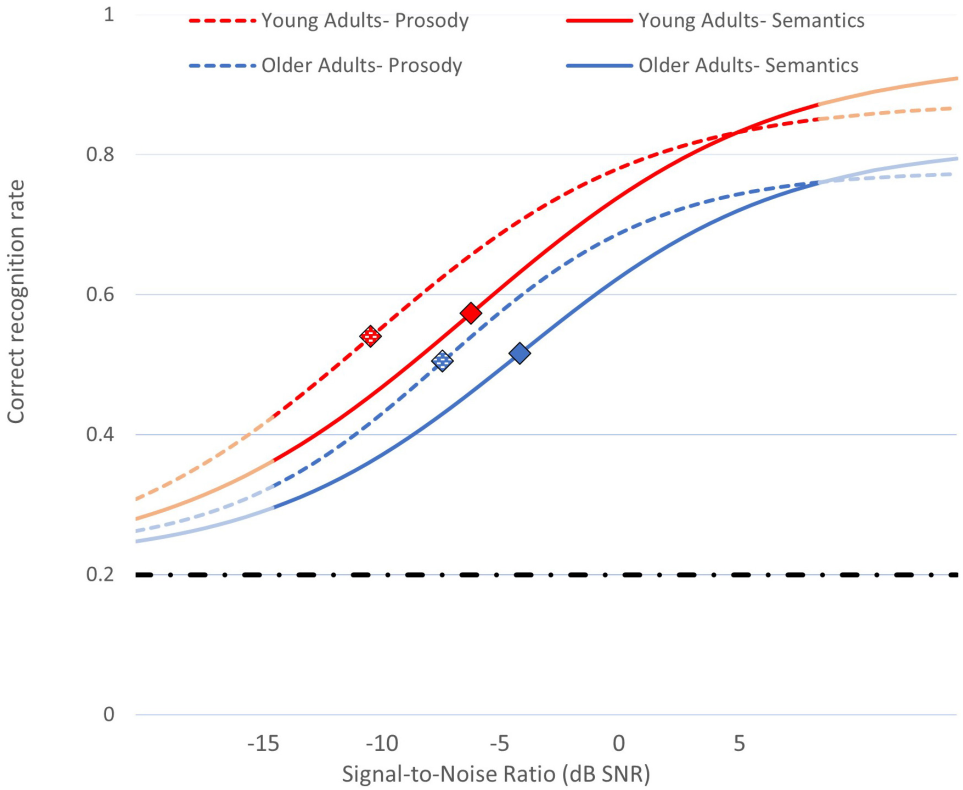


FIGURE 2 | Psychometric functions for recognition of emotions in speech in different SNRs, averaged across participants. Blue lines: older adults; red lines: young adults. Dashed lines: recognition of emotional prosody; full lines: recognition of emotional semantics. Diamond-shaped markers indicate statistical recognition thresholds for each condition. Light-blue and light-red lines represent extrapolations of the functions beyond the SNRs tested in the study. The dashed-and-dot horizontal line indicates the functions' minimal asymptote (0.2 - chance level).

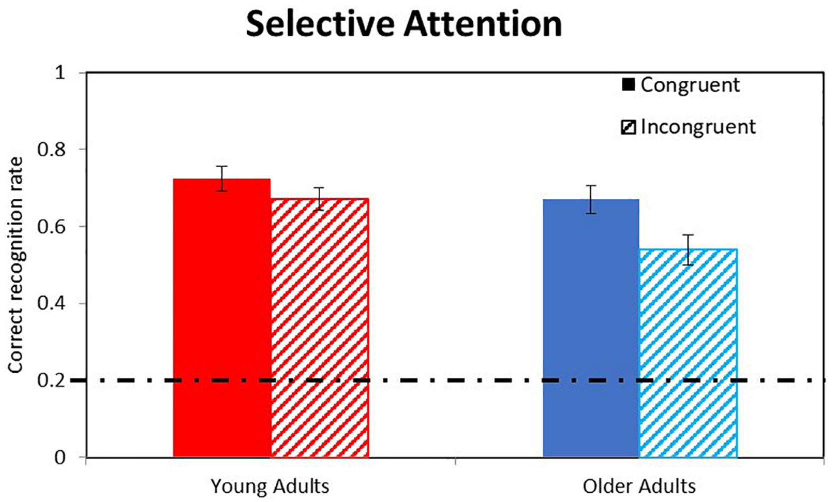


FIGURE 3 | Analysis of Selective Attention effects. Bars indicate correct recognition rates for congruent (full) vs. incongruent (dashed) sentences in young (red) and older (blue) adults averaged across different SNRs. Error bars indicate 95% CI of their respective means (MLM estimates). The dashed-and-dot horizontal line indicates the chance level (0.2).

Analysis of Selective Attention Failures

Table 1B presents the full MLM analyses of the Selective Attention factor. Results show significant main effect for Age Group, $F(1,42.82) = 20.45$, $p < 0.001$, and for Speech Channel, $F(1,40.64) = 12.88$, $p = 0.001$, with no significant interaction between the two, $F(1,40.64) = 0.809$, $p = 0.374$, conceptually replicating the results reported above. Most importantly, we found a significant main effect for Selective Attention, $F(1,44.04) = 55.6$, $p < 0.001$, that significantly interacted with Age Group, $F(1,44.04) = 10.2$, $p = 0.003$, reflecting larger failures of selective attention for older adults.

To sum, our fourth hypothesis was confirmed: Recognition rates were better for congruent than for incongruent sentences (correct recognition rates of 0.697 vs. 0.606, respectively), indicating overall failures of selective attention. Older adults showed larger failures of selective attention than did young adults (Selective-Attention factors of 0.130 vs. 0.052, respectively). These results are visually presented in **Figure 3**.

DISCUSSION

The current study adopted a psychophysical approach to directly compare young and older adults' sensory thresholds for emotion recognition across two channels of speech: prosody and semantics. We aimed to better understand age-related differences in the processing of spoken emotions, as indicated in the literature, and specifically an age-related decrease in the dominance of prosody over semantics, as found by Ben-David et al. (2019). A total of 29 young adults and 26 older adults listened to 50 spoken sentences presenting different combinations of emotions across prosody and semantics and, in different tasks, were asked to recognize the emotion presented in one of the channels. Sentences were mixed with speech-spectrum noise ranging from SNR of −15 dB (most difficult) to +5 dB (easiest). Individual recognition thresholds were calculated (by fitting psychometric functions), separately for prosodic and for semantic emotion recognition.

Results indicated the following trends, supporting our hypotheses:

1. Recognition thresholds were better for young over older adults (*age-related effects*);
2. Recognition thresholds were better for prosodic over semantic information (*prosodic advantage*);
3. The prosodic advantage in thresholds did not differ between age-groups;
4. However, a significant age-related effect was indicated for *selective attention*, suggesting that older adults were more affected by the irrelevant channel than were young adults.

To the best of our knowledge, this study is the first to directly examine possible age-related differences in the imbalance between thresholds for emotion recognition in different speech channels. To date, only a few studies have tried to directly compare the recognition of the semantic and prosodic channels, mostly in young adults (but see, Dupuis and Pichora-Fuller, 2014). For example, Morgan (2021) showed that sensory thresholds were better for prosodic-emotion recognition than for word recognition in noise for young adults (see also van Zyl and Hanekom, 2011; Ritter and Vongpaisal, 2018; Morgan et al., 2022). However, these studies did not directly measure semantic-emotion recognition, but rather used word/sentence recognition as a placeholder. Clearly, these two processes differ, as semantic-emotion recognition involves both the identification of the spoken words and their integration as the basis for emotional labeling.

As expected, we found lower (better) recognition thresholds for young over older adults. In other words, older adults needed speech to be presented at ∼2.2 dB SNR louder than young adults to reach their recognition threshold in noise. These results are in line with the abundant literature on speech perception in noise (Heinrich et al., 2016). *Semantics:* Age-related changes in semantic emotion recognition follow findings on spoken word recognition. Note, our effects are about half the size of the well-observed 4 dB SNR age-related difference in spoken-word recognition accuracy (Pichora-Fuller et al., 1995; Murphy et al., 2000; Ben-David et al., 2011a). This is probably the outcome of the different tasks used, as we tested emotion recognition thresholds rather than word recognition accuracy. *Prosody:* The current study is the first to directly test age-related changes in recognition thresholds for emotional prosody. Our findings, on an age-related decrease in prosodic recognition thresholds, expand previous findings on age-related diminished prosodic recognition accuracy for speech in noise (Dmitrieva and Gelman, 2012; Dupuis and Pichora-Fuller, 2014). *Maximum asymptotes:* An age-related difference was found for the maximum asymptote of the psychometric functions, indicating that young adults recognize emotions in speech better than older adults, even under very little noise (see Paulmann et al., 2008; Ben-David et al., 2019). Recognition accuracy for older adults did not reach 100% at the maximum asymptotes (easiest SNR). This is not surprising, as the literature suggests that even in quiet older adults are impaired at emotion recognition (Ruffman et al., 2008), speech recognition (Pichora-Fuller and Souza, 2003) and emotional prosody and semantics recognition (Paulmann et al., 2008).

Our results support a sensory prosodic advantage across both age-groups, where recognition thresholds were lower (better) for emotional prosody than for emotional semantics. This suggests that to reach recognition threshold in noise, emotional semantics call for an addition of ∼3.3 dB SNR as compared to emotional prosody. This prosodic advantage across age groups expands previous evidence that focused mainly on an accuracy advantage for prosodic recognition over spoken word recognition (Dupuis and Pichora-Fuller, 2014; Morgan, 2021; Morgan et al., 2022). A noteworthy study by Morgan (2021) reported a 10 dB SNR advantage between emotional prosodic thresholds and spoken word identification thresholds in young adults. This marks a much larger advantage than the 3.3 dB SNR difference we report. This difference possibly stems from the tasks used (word identification vs. emotion recognition in a sentence) and from other methodological differences (such as the different levels of SNRs used in each condition). *Maximum asymptotes:* In contrast to the prosodic advantage in SNR thresholds, it is notable that a small but significant semantic advantage was found for the maximum asymptote of the psychometric functions, indicating that emotional semantics are recognized slightly better than are emotional prosodies under very little noise (see also Ben-David et al., 2019).

How to explain this ease of prosodic detection in noise? As aforementioned, spoken emotional semantic recognition is based on both word identification and context generation as the words unfold in time. These tasks are highly sensitive to noise (Pichora-Fuller et al., 1995), as misapprehension of sound-sharing words might change the emotional meaning of the whole sentence. For example, consider the sentences "I'm so */sad/* right now" versus "I'm so */mad/* right now." Confusing one phoneme for another, a common characteristic of speech-in-noise processing (Ben-David et al., 2011a; Nitsan et al., 2019), shifts the emotional categorization of the sentence from sadness to anger. In contrast, prosodic recognition is based on suprasegmental features that may be less susceptible to noise. Namely, prosodic processing is based on the envelope of speech, speech rate and fundamental frequency fluctuations (Myers et al., 2019). These acoustic features are more immune to interference from energetic masking (Morgan, 2021). Moreover, processing of prosodic features involves several functionally (and anatomically) segregated systems of cortical and sub-cortical networks (Grandjean, 2021). This redundancy might serve to protect from the effects of adverse sensory conditions.

Indeed, prosody has been taken to be a fundamental aspect of speech that scaffolds other aspects of linguistic processing (Myers et al., 2019). Emotional prosody is learned and used already in infancy, before the effective use of semantics in infant-parent interactions (Fernald, 1989). Thus, prosody serves as a basic emotional cue across the life span. Prosody also appears to be a contextualizing marker of verbal interactions that directly leads listeners to the speaker's emotional message (House, 2007). The critical role prosody plays in interpersonal and social situations (Pell and Kotz, 2021) may be generated by its perceptual salience, or may lead to heightened sensitivity to prosodic cues in noise.

Perhaps our most important finding is the lack of interaction between age group and prosodic advantage in sensory thresholds. In other words, the prosodic advantage was similar in extent for older and young adults (around a 33% advantage in both groups). Our data do not support suggestions in the literature that older adults might have specific impairments in prosodic processing as compared to young adults (Mitchell, 2007; Orbelo et al., 2005). Rather, they are in line with a general age-related auditory decline that spans to both segmental and suprasegmental features (Paulmann et al., 2008). Results could also support a general age-related decrease in emotional perception and processing (Ruffman et al., 2008; but see Castro and Isaacowitz, 2019) across the two speech channels.

In contrast to the preserved prosodic advantage in recognition thresholds, we observed significant age-related differences in selective attention. When asked to focus on one speech channel, older adults were affected to a larger extent by the content of the other, irrelevant channel. This finding could be taken to support the age-related inhibitory deficit hypothesis (Hasher and Zacks, 1988; Ben-David et al., 2014), with older adults experiencing larger difficulties in inhibiting irrelevant information. Alternatively, our results could be based on an information degradation hypothesis (Schneider and Pichora-Fuller, 2000; Ben-David and Schneider, 2009, 2010), whereby age-related sensory changes lead to performance changes. In the current study, information in the prosodic and semantic channels was degraded to a similar extent due to auditory sensory degradation in aging. Clearly, pure-tone thresholds for older adults were significantly worse than for young adults (see **Supplementary Appendix A**). These and other age-related audiological changes (e.g., frequency selectivity and loudness recruitment; Füllgrabe, 2020) are likely to have had an impact on age-related sensory degradation of speech perception. Consequently, older adults in our study might have adopted a wider processing strategy and integrated information from both speech channels to form a clearer picture of the speaker's intent (Hess, 2005, 2006, 2014). Whereas this strategy improves processing in congruent prosody-semantic sentences, it leads to failures in selective attention in incongruent sentences.

It is notable that older adults in our sample experienced a larger extent of hearing loss in the higher frequencies (4,000 and 8,000 Hz, see **Supplementary Appendix A**). This high-frequency hearing loss is common for older adults with clinically normal hearing (in the lower frequency ranges) recruited for speech processing studies (Dupuis and Pichora-Fuller, 2014, 2015; Nagar et al., 2022). It has been suggested that this age-related difference may have a specific effect on semantic processing, as many speech cues are available in a range around 4,000 Hz (Vinay and Moore, 2010); whereas prosodic cues, such as f0 and the envelope of speech, might still be preserved. Our findings do not necessarily support this option, as we found an equivalent SNR prosodic advantage for older and young adults. In other words, age-related sensory degradation appears to have had a similar impact on semantic and prosodic emotional processing in the current study. Thus, our results follow the literature indicating that age-related sensory changes are not the sole source of difficulties older adults experience when speech is presented in adverse listening conditions (Roberts and Allen, 2016). For example, Füllgrabe et al. (2015) found age-related deficits in speech-in-noise identification to persist even when audiograms for older and young adults were matched (see also Grassi and Borella, 2013). Following Cardin (2016), listening in adverse conditions becomes effortful in aging and demands more cognitive resources, thus speech processing is affected by age-related changes in both sensory and cognitive factors.

Caveats, Future Directions, and Implications

Limitations of the current study include relatively small numbers of participants in each age group. However, this number is not different than that found in the pertinent literature (e.g., 20 participants, Morgan, 2021). Even though the range of SNR used was large enough to include individual thresholds, future studies may increase the range to improve the assessment's accuracy. In addition, the current study used speech-spectrum noise, a standard noise type widely used in age-related comparisons (Ben-David et al., 2011a, 2012). Future studies may wish to test further types of auditory distortions (e.g., Ritter and Vongpaisal, 2018; Dor et al., 2020). Future studies may also test the effects of individual audiometric thresholds (see Grassi and Borella, 2013), demographic characteristics (e.g., gender, socio-economic status and education), as well as emotional traits and mental health (e.g., empathy and alexithymia, see Leshem et al., 2019) on emotion recognition thresholds. Indeed, mental health was also found to affect the recognition of negative and positive emotions differently (e.g., detection of emotionally negative words was related to PTSD and forensic schizophrenia; Cisler et al., 2011; Leshem et al., 2020). Finally, this study used a unique set of validated and standardized spoken sentences that present emotional content in both semantics and prosody. Future studies may wish to expand the scope of this study's findings by using different sets of sentences.

In sum, the current study is the first to directly compare emotion recognition thresholds for spoken semantics and prosody in young and older adults. Mainly, we found a recognition threshold advantage for young over older adults, an advantage for prosody over semantics that was not affected by age group, and larger failures of selective attention for older adults. Previous studies indicate that older adults assign different relative weights to prosodic and semantic spoken emotions than do young adults, possibly resulting in an inter-generational communication breakdown (Dupuis and Pichora-Fuller, 2010; Ben-David et al., 2019). The current study does not support a sensory source for this age-related difference in speech processing, hinting to a possible cognitive source. Future studies should directly test whether processing of prosodic and semantic emotions demands a different extent of cognitive resources for young and older adults.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics Committee of Reichamn University, Herzliya, Israel. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

YD and BB-D wrote the manuscript, they are responsible for the design of the paradigm, the analysis and interpretation of the data. DA and BB-D supervised the research project, DA and VS made invaluable contributions to the conceptualizing the research question and the final manuscript. BB-D is the corresponding author and the study was conducted in his lab. All authors had a prominent intellectual contribution to the study, are accountable for the data and approved the final version of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2022.846117/full#supplementary-material

REFERENCES

Ben-David, B. M., Eidels, A., and Donkin, C. (2014). Effects of aging and distractors on detection of redundant visual targets and capacity: do older adults integrate visual targets differently than younger adults? *PLoS One* 9:e113551. doi: 10.1371/journal.pone.0113551

Ben-David, B. M., Erel, H., Goy, H., and Schneider B. A. (2015). 'Older is always better:' age-related differences in vocabulary scores across 16 years. *Psychol. Aging* 30, 856–862 doi: 10.1037/pag0000051

Ben-David, B. M., Gal-Rosenblum, S., van Lieshout, P. H., and Shakuf, V. (2019). Age-related differences in the perception of emotion in spoken language: the relative roles of prosody and semantics. *J. Speech Lang. Hear. Res.* 62, 1188–1202. doi: 10.1044/2018_JSLHR-H-ASCC7-18-0166

Ben-David, B. M., and Icht, M. (2018). The effect of practice and visual feedback on oral-diadochokinetic rates for younger and older adults. *Lang. Speech* 61, 113–134. doi: 10.1177/0023830917708808

Ben-David, B. M., Mentzel, M., Icht, M., Gilad, M., Dor, Y. I., Ben-David, S., et al. (2021). Challenges and opportunities for telehealth assessment during COVID-19: iT-RES, adapting a remote version of the test for rating emotions in speech. *Int. J. Audiol.* 60, 319–321. doi: 10.1080/14992027.2020.1833255

Ben-David, B. M., Multani, N., Shakuf, V., Rudzicz, F., and van Lieshout, P. H. (2016). Prosody and semantics are separate but not separable channels in the perception of emotional speech: test for rating of emotions in speech. *J. Speech Lang. Hear. Res.* 59, 72–89. doi: 10.1044/2015_JSLHR-H-14-0323

Ben-David, B. M., and Schneider, B. A. (2009). A sensory origin for color-word Stroop effects in aging: a meta-analysis. *Neuropsychol. Dev. Cogn. B Aging Neuropsychol. Cogn.* 16, 505–534. doi: 10.1080/13825580902855862

Ben-David, B. M., and Schneider, B. A. (2010). A sensory origin for color-word Stroop effects in aging: simulating age-related changes in color-vision mimics age-related changes in Stroop. *Neuropsychol. Dev. Cogn. B Aging Neuropsychol. Cogn.* 17, 730–746. doi: 10.1080/13825585.2010.510553

Ben-David, B. M., Thayapararajah, A., and van Lieshout, P. H. (2013). A resource of validated digital audio recordings to assess identification of emotion in spoken language after a brain injury. *Brain Inj.* 27, 248–250. doi: 10.3109/02699052.2012.740648

Ben-David, B. M., Tse, V. Y., and Schneider, B. A. (2012). Does it take older adults longer than younger adults to perceptually segregate a speech target from a background masker? *Hear. Res.* 290, 55–63. doi: 10.1016/j.heares.2012.04.022

Ben-David, B. M., Chambers, C. G., Daneman, M., Pichora-Fuller, M. K., Reingold, E. M., and Schneider, B. A. (2011a). Effects of aging and noise on real-time spoken word recognition: evidence from eye movements. *J. Speech Lang. Hear. Res.* 54, 243–262. doi: 10.1044/1092-4388(2010/09-0233)

Ben-David, B. M., van Lieshout, P. H., and Leszcz, T. (2011b). A resource of validated affective and neutral sentences to assess identification of emotion in spoken language after a brain injury. *Brain Inj.* 25, 206–220. doi: 10.3109/02699052.2010.536197

Bilger, R. C., Nuetzel, J. M., Rabinowitz, W. M., and Rzeczkowski, C. (1984). Standardization of a test of speech perception in noise. *J. Speech Lang. Hear. Res.* 27, 32–48. doi: 10.1044/jshr.2701.32

Boersma, P., and Weenink, D. (2019). *Praat: Doing Phonetics by Computer [Computer Program]. Version 6.1.* Available online at: http://www.praat.org/ (accessed July 2019).

Cardin, V. (2016). Effects of aging and adult-onset hearing loss on cortical auditory regions. *Front. Neurosci.* 10:199. doi: 10.3389/fnins.2016.00199

Carl, M., Icht, M., and Ben-David, B. M. (2022). A cross-linguistic validation of the test for rating emotions in speech (T-RES): acoustic analyses of emotional sentences in English, German, and Hebrew. *J. Speech Lang. Hear. Res.* 65, 991–1000. doi: 10.1044/2021_JSLHR-21-00205

Castro, V. L., and Isaacowitz, D. M. (2019). The same with age: evidence for age-related similarities in interpersonal accuracy. *J. Exp. Psychol. Gen.* 148, 1517–1537. doi: 10.1037/xge0000540

Chan, E. K. H. (2014). "Standards and guidelines for validation practices: development and evaluation of measurement instruments," in *Validity and Validation in Social, Behavioral, and Health Sciences. Social Indicators Research Series*, Vol. 54, B. Zumbo and E. Chan (Cham: Springer), 9–24. doi: 10.1007/978-3-319-07794-9_2

Cisler, J. M., Wolitzky-Taylor, K. B., Adams T. G. Jr., Babson, K. A., Badour, C. L., and Willems, J. L. (2011). The emotional Stroop task and posttraumatic stress disorder: a meta-analysis. *Clin. Psychol. Rev.* 31, 817–828. doi: 10.1016/j.cpr.2011.03.007

Dmitrieva, E. S., and Gelman, V. Y. (2012). The relationship between the perception of emotional intonation of speech in conditions of interference and the acoustic parameters of speech signals in adults of different gender and age. *Neurosci. Behav. Physiol.* 42, 920–928. doi: 10.1007/s11055-012-9658-z

Dor, Y., Rosenblum, M., Kenet, D., Shakuf, V., Algom, D., and Ben-David, B. M. (2020). "Can you hear what I feel? Simulating high-frequency hearing loss mimics effects of aging and tinnitus in emotional speech perception," in *Proceedings of the 36th Annual Meeting of the International Society for Psychophysics Fechner Day 2020*, eds J. R. Schoenherr, T. Hubbard, W. Stine, and C. Leth-Steensen (International Society for Psychophysics), 13–16.

Dupuis, K., and Pichora-Fuller, M. K. (2010). Use of affective prosody by young and older adults. *Psychol. Aging* 25, 16–29. doi: 10.1037/a0018777

Dupuis, K., and Pichora-Fuller, M. K. (2014). Intelligibility of emotional speech in younger and older adults. *Ear Hear.* 35, 695–707. doi: 10.1097/AUD.0000000000000082

Dupuis, K., and Pichora-Fuller, M. K. (2015). Aging affects identification of vocal emotions in semantically neutral sentences. *J. Speech Lang. Hear. Res.* 58, 1061–1076. doi: 10.1044/2015_JSLHR-H-14-0256

Fernald, A. (1989). Intonation and communicative intent in mothers' speech to infants: is the melody the message? *Child Dev.* 60, 1497–1510. doi: 10.2307/1130938

Füllgrabe, C. (2020). On the possible overestimation of cognitive decline: the impact of age-related hearing loss on cognitive-test performance. *Front. Neurosci.* 14:454. doi: 10.3389/fnins.2020.00454

Füllgrabe, C., Moore, B. C., and Stone, M. A. (2015). Age-group differences in speech identification despite matched audiometrically normal hearing: contributions from auditory temporal processing and cognition. *Front. Aging Neurosci.* 6:347. doi: 10.3389/fnagi.2014.00347

Giroud, N., Keller, M., Hirsiger, S., Dellwo, V., and Meyer, M. (2019). Bridging the brain structure—brain function gap in prosodic speech processing in older adults. *Neurobiol. Aging* 80, 116–126. doi: 10.1016/j.neurobiolaging.2019.04.017

Goodman, L. (2001). *Translation of WAIS-III - Wechsler Adult Intelligence Scale*. Jerusalem: Psych tech.

Grandjean, D. (2021). Brain networks of emotional prosody processing. *Emot. Rev.* 13, 34–43. doi: 10.1177/1754073919898522

Grassi, M., and Borella, E. (2013). The role of auditory abilities in basic mechanisms of cognition in older adults. *Front. Aging Neurosci.* 5:59. doi: 10.3389/fnagi.2013.00059

Grunwald, I. S., Borod, J. C., Obler, L. K., Erhan, H. M., Pick, L. H., Welkowitz, J., et al. (1999). The effects of age and gender on the perception of lexical emotion. *Appl. Neuropsychol.* 6, 226–238. doi: 10.1207/s15324826an0604_5

Harel-Arbeli, T., Wingfield, A., Palgi, Y., and Ben-David, B. M. (2021). Age-related differences in the online processing of spoken semantic context and the effect of semantic competition: evidence from eye gaze. *J. Speech Lang. Hear. Res.* 64, 315–327. doi: 10.1044/2020_JSLHR-20-00142

Hasher, L., and Zacks, R. T. (1988). Working memory, comprehension, and aging: a review and a new view. *Psychol. Learn. Motiv.* 22, 193–225. doi: 10.1016/S0079-7421(08)60041-9

Heinrich, A., Gagne, J. P., Viljanen, A., Levy, D. A., Ben-David, B. M., and Schneider, B. A. (2016). Effective communication as a fundamental aspect of active aging and well-being: paying attention to the challenges older adults face in noisy environments. *Soc. Inq. Well Being* 2, 51–69. doi: 10.13165/SIIW-16-2-1-05

Helfer, K. S., Merchant, G. R., and Wasiuk, P. A. (2017). Age-related changes in objective and subjective speech perception in complex listening environments. *J. Speech Lang. Hear. Res.* 60, 3009–3018. doi: 10.1044/2017_JSLHR-H-17-0030

Henry, J. D., and Crawford, J. R. (2005). The short-form version of the Depression Anxiety Stress Scales (DASS-21): construct validity and normative data in a large non-clinical sample. *Br. J. Clin. Psychol.* 44, 227–239. doi: 10.1348/014466505X29657

Hess, T. M. (2005). Memory and aging in context. *Psychol. Bull.* 131, 383–406. doi: 10.1037/0033-2909.131.3.383

Hess, T. M. (2006). Adaptive aspects of social cognitive functioning in adulthood: age–related goal and knowledge influences. *Soc. Cogn.* 24, 279–309. doi: 10.1521/soco.2006.24.3.279

Hess, T. M. (2014). Selective engagement of cognitive resources: motivational influences on older adults' cognitive functioning. *Perspect. Psychol. Sci.* 9, 388–407. doi: 10.1177/1745691614527465

House, J. (2007). The role of prosody in constraining context selection: a procedural approach. *Nouv. Cah. Linguist. Fr.* 28, 369–383.

IBM Corp (2011). *IBM SPSS Statistics for Windows, Version 20.0*. Armonk, NY: IBM Corp.

Isaacowitz, D. M., Löckenhoff, C. E., Lane, R. D., Wright, R., Sechrest, L., Riedel, R., et al. (2007). Age differences in recognition of emotion in lexical stimuli and facial expressions. *Psychol. Aging* 22, 147–159. doi: 10.1037/0882-7974.22.1.147

Kaufman, A. S., Reynolds, C. R., and McLean, J. E. (1989). Age and WAIS–R intelligence in a national sample of adults in the 20- to 74-year age range: a cross-sectional analysis with educational level controlled. *Intelligence* 13, 235–253. doi: 10.1016/0160-2896(89)90020-2

Lambrecht, L., Kreifelts, B., and Wildgruber, D. (2012). Age-related decrease in recognition of emotional facial and prosodic expressions. *Emotion* 12, 529–539. doi: 10.1037/a0026827

Leshem, R., Icht, M., Bentzur, R., and Ben-David, B. M. (2020). Processing of emotions in speech in forensic patients with schizophrenia: impairments in identification, selective attention, and integration of speech channels. *Front. Psychiatry* 11:601763. doi: 10.3389/fpsyt.2020.601763

Leshem, R., van Lieshout, P. H. H. M., Ben-David, S., and Ben-David, B. M. (2019). Does emotion matter? The role of alexithymia in violent recidivism: a systematic literature review. *Crim. Behav. Ment. Health* 29, 94–110. doi: 10.1002/cbm.2110

Livingston, G., Sommerlad, A., Orgeta, V., Costafreda, S. G., Huntley, J., Ames, D., et al. (2017). Dementia prevention, intervention, and care. *Lancet* 390, 2673–2734. doi: 10.1016/S0140-6736(17)31363-6

Mather, M. (2016). The affective neuroscience of aging. *Annu. Rev. Psychol.* 67, 213–238. doi: 10.1146/annurev-psych-122414-033540

McMurray, B. (2017). *Nonlinear Curvefitting for Psycholinguistics (Version 13)*. Available online at: https://osf.io/4atgv/ (accessed November 21, 2021).

Melara, R. D., and Algom, D. (2003). Driven by information: a tectonic theory of Stroop effects. *Psychol. Rev.* 110, 422–471. doi: 10.1037/0033-295X.110.3.422

Mitchell, R. L. (2007). Age-related decline in the ability to decode emotional prosody: primary or secondary phenomenon? *Cogn. Emot.* 21, 1435–1454. doi: 10.1080/02699930601133994

Mitchell, R. L., and Kingston, R. A. (2011). Is age-related decline in vocal emotion identification an artefact of labelling cognitions? *Int. J. Psychol. Stud.* 3, 156–163. doi: 10.5539/ijps.v3n2p156

Morgan, S. D. (2021). Comparing emotion recognition and word recognition in background noise. *J. Speech Lang. Hear. Res.* 64, 1758–1772. doi: 10.1044/2021_JSLHR-20-00153

Morgan, S. D., Garrard, S., and Hoskins, T. (2022). Emotion and word recognition for unprocessed and vocoded speech stimuli. *Ear Hear.* 43, 398–407. doi: 10.1097/AUD.0000000000001100

Murphy, D. R., Craik, F. I., Li, K. Z., and Schneider, B. A. (2000). Comparing the effects of aging and background noise on short-term memory performance. *Psychol. Aging* 15, 323–334. doi: 10.1037/0882-7974.15.2.323

Myers, B. R., Lense, M. D., and Gordon, R. L. (2019). Pushing the envelope: developments in neural entrainment to speech and the biological underpinnings of prosody perception. *Brain Sci.* 9:70. doi: 10.3390/brainsci9030070

Nagar, S., Mikulincer, M., Nitsan, G., and Ben-David, B. M. (2022). Safe and sound: the effects of experimentally priming the sense of attachment security on pure-tone audiometric thresholds among young and older adults. *Psychol. Sci.* 33, 424–432. doi: 10.1177/09567976211042008

Nitsan, G., Wingfield, A., Lavie, L., and Ben-David, B. M. (2019). Differences in working memory capacity affect online spoken word recognition: evidence from eye movements. *Trends Hear.* 23, 1–12. doi: 10.1177/2331216519839624

Orbelo, D. M., Grim, M. A., Talbott, R. E., and Ross, E. D. (2005). Impaired comprehension of affective prosody in elderly subjects is not predicted by age-related hearing loss or age-related cognitive decline. *J. Geriatr. Psychiatry Neurol.* 18, 25–32. doi: 10.1177/0891988704272214

Paulmann, S., Pell, M. D., and Kotz, S. A. (2008). How aging affects the recognition of emotional speech. *Brain Lang.* 104, 262–269. doi: 10.1016/j.bandl.2007.03.002

Pell, M. D., and Kotz, S. A. (2021). Comment: the next frontier: prosody research gets interpersonal. *Emot. Rev.* 13, 51–56. doi: 10.1177/1754073920954288

Phillips, L. H., MacLean, R. D., and Allen, R. (2002). Age and the understanding of emotions: neuropsychological and sociocognitive perspectives. *J. Gerontol. B Psychol. Sci. Soc. Sci.* 57, 526–530. doi: 10.1093/geronb/57.6.P526

Pichora-Fuller, M. K., Schneider, B. A., and Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *J. Acoust. Soc. Am.* 97, 593–608. doi: 10.1121/1.412282

Pichora-Fuller, M. K., and Souza, P. E. (2003). Effects of aging on auditory processing of speech. *Int. J. Audiol.* 42(Suppl. 2), 11–16. doi: 10.3109/14992020309074638

Ritter, C., and Vongpaisal, T. (2018). Multimodal and spectral degradation effects on speech and emotion recognition in adult listeners. *Trends Hear.* 22, 1–17. doi: 10.1177/2331216518804966

Roberts, K. L., and Allen, H. A. (2016). Perception and cognition in the aging brain: a brief review of the short-and long-term links between perceptual and cognitive decline. *Front. Aging Neurosci.* 8:39. doi: 10.3389/fnagi.2016.00039

Ruffman, T., Henry, J. D., Livingstone, V., and Phillips, L. H. (2008). A meta-analytic review of emotion recognition and aging: implications for neuropsychological models of aging. *Neurosci. Biobehav. Rev.* 32, 863–881. doi: 10.1016/j.neubiorev.2008.01.001

Schneider, B. A., and Pichora-Fuller, M. K. (2000). "Implications of perceptual deterioration for cognitive aging research," in *The Handbook of Aging and Cognition*, eds F. I. M. Craik and T. A. Salthouse (London: Lawrence Erlbaum Associates Publishers), 155–219

Taitelbaum-Swead, R., Icht, M., and Ben-David, B. M. (2022). More than words: the relative roles of prosody and semantics in the perception of emotions in spoken language by postlingual cochlear implant recipients. *Ear Hear.* [Epub ahead of print]. doi: 10.1097/AUD.0000000000001199

van Zyl, M., and Hanekom, J. J. (2011). Speech perception in noise: a comparison between sentence and prosody recognition. *J. Hear. Sci.* 1, 54–56.

Vinay, and Moore, B. C. (2010). Psychophysical tuning curves and recognition of highpass and lowpass filtered speech for a person with an inverted V-shaped audiogram. *J. Acoust. Soc. Am.* 127, 660–663. doi: 10.1121/1.3277218

Wingfield, A., and Tun, P. A. (2001). Spoken language comprehension in older adults: interactions between sensory and cognitive change in normal aging. *Semin. Hear.* 22, 287–302. doi: 10.1055/s-2001-15632

Zalsman, G., Aizenberg, D., Sigler, M., Nahshoni, E., and Weizman, A. (1998). Geriatric depression scale-short form–validity and reliability of the Hebrew version. *Clin. Gerontol.* 18, 3–9. doi: 10.1300/J018v18n03_02

Conflict of Interest: 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 Psychology

ORIGINAL RESEARCH
published: 28 April 2022
doi: 10.3389/fpsyg.2022.859722

Relationships Between Health-Related Quality of Life and Speech Perception in Bimodal and Bilateral Cochlear Implant Users

Nadav Brumer*[1], *Elizabeth Elkins*[1], *Jake Hillyer*[2], *Chantel Hazlewood*[1] and *Alexandra Parbery-Clark*[1]

[1]Auditory Research Laboratory, Swedish Neuroscience Institute, Seattle, WA, United States, [2]College of Medicine, University of Arizona, Phoenix, AZ, United States

Edited by:
Leah Fostick,
Ariel University, Israel

Reviewed by:
Michel Hoen,
Oticon Medical, Sweden
Aaron Moberly,
The Ohio State University, United States

****Correspondence:***
Alexandra Parbery-Clark
alexandra.parbery-clark@swedish.org

Specialty section:
This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Psychology

Received: *24 January 2022*
Accepted: *28 March 2022*
Published: *28 April 2022*

Citation:
Brumer N, Elkins E, Hillyer J, Hazlewood C and Parbery-Clark A (2022) Relationships Between Health-Related Quality of Life and Speech Perception in Bimodal and Bilateral Cochlear Implant Users. Front. Psychol. 13:859722. doi: 10.3389/fpsyg.2022.859722

Purpose: Previous studies examining the relationship between health-related quality of life (HRQoL) and speech perception ability in cochlear implant (CI) users have yielded variable results, due to a range of factors, such as a variety of different HRQoL questionnaires and CI speech testing materials in addition to CI configuration. In order to decrease inherent variability and better understand the relationship between these measures in CI users, we administered a commonly used clinical CI speech testing battery as well as two popular HRQoL questionnaires in bimodal and bilateral CI users.

Methods: The Glasgow Benefit Inventory (GBI), a modified five-factor version of the GBI (GBI-5F), and the Nijmegen Cochlear Implant Questionnaire (NCIQ) were administered to 25 CI users (17 bimodal and 8 bilateral). Speech perception abilities were measured with the AzBio sentence test in several conditions (e.g., quiet and noise, binaural, and first-ear CI only).

Results: Higher performance scores on the GBI general subscore were related to greater binaural speech perception ability in noise. There were no other relationships between the GBI or NCIQ and speech perception ability under any condition. Scores on many of the GBI-5F factors were substantially skewed and asymmetrical; therefore, correlational analyses could not be applied. Across all participants, binaural speech perception scores were greater than first-ear CI only scores.

Conclusion: The GBI general subscore was related to binaural speech perception, which is considered the everyday listening condition of bimodal and bilateral CI users, in noise; while the more CI-specific NCIQ did not relate to speech perception ability in any listening condition. Future research exploring the relationships between the GBI, GBI-5F, and NCIQ considering bimodal and bilateral CI configurations separately is warranted.

Keywords: cochlear implant, health related quality of life, speech perception, hearing loss, bimodal, bilateral, Glasgow Benefit Inventory, Nijmegen Cochlear Implant Questionnaire

INTRODUCTION

Cochlear implants (CIs) significantly improve quality of life (QoL) and speech perception abilities for individuals with severe to profound hearing loss (Gaylor et al., 2013; Mosnier et al., 2015). Currently, speech perception scores are the primary outcome measures utilized for quantifying CI benefit in adult users; however, there is a growing movement to further quantify CI benefit with QoL measures. This is because while objective measures of CI benefit, such as speech perception tests, are important for evaluating CI performance, CIs also influence other aspects of a patient's life, such as self-esteem and socializing, that are not always captured by traditional objective measures. In general, QoL measures capture important information regarding the subjective wellbeing of a patient at a given point in time. On the other hand, health-related quality of life (HRQoL) measures are specific to certain aspects of QoL affected by health conditions, such as hearing loss, or medical procedures, such as cochlear implantation. As such, HRQoL measures may be more sensitive to differences in CI benefit compared to general QoL measures (Krabbe et al., 2000; Hirschfelder et al., 2008; Sladen et al., 2017). As HRQoL measures are increasingly used with CI recipients in clinical settings, it is important to understand any relationships between these HRQoL metrics and the traditional metric of CI-aided speech perception scores.

The use of HRQoL measures in addition to speech perception scores to monitor CI benefit is growing in popularity (Services, U.D.O.H.A.H., 2011). HRQoL questionnaires that are commonly used with CI patients are the Glasgow Benefit Inventory (GBI; Robinson et al., 1996) adapted for CI users (Ho et al., 2009) and the Nijmegen Cochlear Implant Questionnaire (NCIQ; Hinderink et al., 2000). The GBI was initially developed to be a post-intervention outcome measure for medical treatments including surgical procedures, and is scored on a scale from −100 to +100, meaning that clinicians can identify if there has been an overall improvement (closer to +100) or worsening (closer to −100) of QoL post-intervention. The GBI consists of 18 questions, and scoring includes a total score as well as three subscores: general, social support, and physical health. One strength of the GBI is that it addresses the direct success of CI implantation with questions regarding whether or not an individual would undergo the procedure again or recommend it to others, providing additional information for capturing subjective benefit of a CI. However, recent work highlights the need to explore the construct validity of the GBI subscores; specifically, which questions are designated to each subscore (Browning et al., 2021). A confirmatory factor analysis performed by Browning et al. (2021) found that the original three subscore model of the GBI was a poor fit for data from 4,799 otolaryngologic patient responses and that the total score and general subscore contained a large number of heterogeneous questions that do not converge on any one construct. Browning et al. (2021) further identified three questions that were either redundant or not pertinent to otolaryngologic intervention in this group (e.g., question nine centered around job opportunities and was the most frequently unanswered of the 18 total questions in the GBI in this population). As such, Browning recommended that, for an otolaryngologic patient population, the general subscore of the original GBI be split into three additional subscores (QoL, self-confidence, and social involvement) and that three less relevant questions be removed from the original 18 questions. This modified GBI questionnaire was renamed the five-factor Glasgow Benefit Inventory (GBI-5F), which has 15 rather than 18 questions and contains a total of five subscores.

A second popular clinical questionnaire is the NCIQ (Hinderink et al., 2000). Unlike the GBI and GBI-5F, the NCIQ was specifically created for CI users and is scored on a Likert scale from 1–5 with transformed scores ranging from 0 (very poor) to 100 (optimal). In addition to the 5-point Likert system, the NCIQ includes a sixth "not applicable" option to all questions, which may be helpful for distinguishing between what domains are not affected by cochlear implantation and which are simply less relevant to a certain population of CI users. Unlike the brief 18 question GBI, the NCIQ consists of 60 questions centered around several physical, psychological, and social domains related to CI use. While the NCIQ is considerably longer than the GBI, this allows for a theoretically more comprehensive assessment of CI outcomes which may increase sensitivity to various clinical changes (Hinderink et al., 2000). Indeed, the NCIQ has been shown to be sensitive to pre- to post-implantation performance change which distinguishes itself from retrospective HRQoL questionnaires, such as the GBI (Straatman et al., 2014; McRackan et al., 2018).

Relating subjective HRQoL questionnaire responses to clinical objective speech perception ability for CI recipients has yielded mixed results. The GBI total and general subscores have been shown to relate to CI-aided speech perception ability in specific instances, such as listening for sentences in quiet (Palmer et al., 1999; Hillyer et al., 2019); however, similar relationships are not apparent when different CI age groups or speech perception materials are included (i.e., sentences versus words). For example, correlations between GBI scores and speech perception of monosyllabic words in quiet existed for younger (<55 years) but not older (≥55 years) CI users (Vermeire et al., 2005). Conversely, Sorrentino et al. (2020) found the opposite effect, with relationships observed between GBI scores and speech perception ability in quiet for three different test stimuli (i.e., disyllabic words, sentences, and question comprehension) in an older (≥65 years) but not in a younger (≤50 years) group of CI users. Meanwhile, Forli et al. (2019) observed no correlations between GBI score and speech perception ability of Italian disyllabic words in quiet or noise, regardless of age group (42–80 years).

Inconsistent relationships between speech perception ability and NCIQ scores have also been observed. CI benefit measured with pre-operative NCIQ scores and 12 months post-operative NCIQ scores related to gains in speech perception of disyllabic words (Mosnier et al., 2015) and monosyllabic words (Sladen et al., 2017). This finding has been further supported by associations between NCIQ subdomains and speech perception abilities of words in quiet (Capretta and Moberly, 2016) and sentences in noise (Olze et al., 2012; Capretta and Moberly, 2016).

Conversely, Hirschfelder et al. (2008) found no relationship between CI-aided speech perception ability of monosyllabic words in noise and NCIQ scores. Similarly, our previous research found no association between NCIQ scores and speech perception ability of sentences in quiet (Hillyer et al., 2019), possibly because participants were considered high-performing CI users, thus limiting variability in CI-aided speech perception performance. Consistent with these variable results, a meta-analysis of 13 studies examining HRQoL questionnaires, including the NCIQ, found negligible to weak but significant correlations between HRQoL measures and speech perception ability of sentences in quiet and noise (McRackan et al., 2018).

In addition to the large variation of participant's age groups and speech stimuli used, another potential explanation for mixed relationships observed between HRQoL and speech perception ability is that previous studies have included participants using a variety of CI configurations. Previous research has often included a variety of combinations of CI users, including unilateral [i.e., one CI with no contralateral amplification], bimodal [i.e., one CI (electric signal) with a contralateral hearing aid (acoustic signal) and/or bilateral [i.e., two CIs (two electric signals)] users, with users experiencing conditions in both quiet and noise. Indeed, Olze et al. (2012) evaluated unilateral CI users only, while other works evaluated primarily unilateral CI and bimodal users (Mosnier et al., 2015; Forli et al., 2019) or a combination of unilateral CI, bimodal and bilateral users (Sanchez-Cuadrado et al., 2015; Capretta and Moberly, 2016; McRackan et al., 2018; Hillyer et al., 2019; Sorrentino et al., 2020). Additionally, some studies do not delineate between bimodal or bilateral CI users (Vermeire et al., 2005; Hirschfelder et al., 2008; Sladen et al., 2017). Therefore, it is challenging to generalize how HRQoL measures may relate to speech perception abilities in CI users when CI configurations included in studies are variable.

The first goal of the present study was to examine how HRQoL score relates to speech perception ability in bimodal and bilateral CI users. By focusing on bimodal and bilateral CI users, we decrease some of the inherent variability in group performance that is observed when unilateral CI users are included. Our second goal was to examine speech perception ability with a first-ear CI only and binaural configuration across all participants. We administered the GBI, the GBI-5F, and the NCIQ, and had participants complete CI-aided speech perception tasks in four conditions: (1) first-ear CI only configuration in quiet, (2) first-ear CI only configuration in noise, (3) binaural (i.e., two bilateral CI's or HA and CI) configuration in quiet, and (4) binaural configuration in noise. Clinical speech scores were used in this study as the overarching goal of this research was to improve understanding of the relationship between clinical CI speech understanding and HRQoL measures. Given that similar data is collected by audiologists across the United States as standard of care, this work has the potential for meaningful clinical translation and interpretation by audiologists providing care to this patient population. We predicted that higher speech perception scores would relate to higher HRQoL scores for the GBI, GBI-5F, and NCIQ. However, we also predicted that HRQoL domains less impacted by speech perception abilities (e.g., physical health) would be less related to speech performance. We also predicted that participants would have higher speech perception scores in the binaural condition relative to the first-ear CI only condition.

MATERIALS AND METHODS

Participants

Twenty-five (14 females, 11 males) experienced CI users (>6 months CI listening experience, $M = 63.4$ months, SD = 31.86 months, range of 14–145 months) with bimodal ($n = 17$) and bilateral ($n = 8$) configurations, between the ages of 52 and 82 years ($M = 67.28$, SD = 10.09) were recruited from the patient pool at the Center for Hearing and Skull Base Surgery at The Swedish Neuroscience Institute in Seattle, Washington. Experienced CI users were recruited because maximum comfortable levels and threshold levels are optimally achieved after 6 months of use and programming (Gajadeera et al., 2017). Inclusion criteria required participants to have no recorded symptoms or diagnosis of dementia, no report of cognitive decline, and no history of congenital or pre-lingual hearing loss. All participants were native speakers of English, had at least a high school education, and demonstrated normal IQ scores ($M = 107.24.11$, SD = 7.85), as measured by the Test of Non-verbal Intelligence—4th Edition (TONI-4; Brown et al., 2010). All testing procedures were approved by the Swedish Medical Center Institutional Review Board (#SWD56152-14) and participants provided informed written consent. All speech testing was conducted in a booth, and all questionnaires were completed in a clinic room at the Swedish Neuroscience Institute in Seattle, WA, United States. All subjects completed all measures except for five who did not complete the NCIQ ($n = 20$; 15 bimodal, 5 bilateral for this measure).

CI-Aided Speech Perception Testing (AzBio)

All participants completed speech perception testing in both quiet and noise conditions, with both first-ear CI only (i.e., no HA or second CI) and binaural configurations (i.e., either CI + HA or CI + CI) in a randomized manner. The speech perception test material chosen was the AzBio Sentence Test (Spahr et al., 2012), comprised of recordings of 20 sentences spoken by two male and two female talkers. Sentences range from 4 to 10 words, spoken by one talker at a time in a conversational style with minimal contextual cues (e.g., "She missed a week of work and nobody noticed"). All words presented are keywords for scoring purposes. Speech testing was administered in a sound-proof booth, with internal dimensions of 2.74 m x 2.82 m. Speech stimuli were presented at 60 dB SPL from a loudspeaker (GN Otometrics Astera Sound Field Speakers) at 0 degrees azimuth, 2 m from the participant, who was instructed to repeat back what they heard. The noise condition, presented at a signal-to-noise ratio (SNR) of +8 dB, included additional 10-talker babble from the same loudspeaker. All speech testing and scoring was performed by a CI audiologist as part of each participant's routine audiologic care and represents the data that was used for clinical decision making for programming and treatment. AzBio speech scores were reported as a percentage (%) of total words correctly repeated, with higher scores indicating better performance.

Health-Related Quality of Life Measures

The Glasgow Benefit Inventory (GBI) is based on a five-point Likert scale that ranges from one, signifying a large change for the worse, to five, signifying a large change for the better, with a score of three signifying no change. Total scores (i.e., the sum responses to 18 questions which is then scaled and averaged) range from −100 (i.e., maximum worsening of overall health status post-intervention) to +100 (maximum improvement of overall health status post-intervention). The total composite score and general subscore were also calculated with question 9 excluded given the evidence of question 9 (i.e., "job opportunities") being potentially less relevant to an otolaryngological population (Browning et al., 2021) or older population as in our study.

The GBI-5F (Browning et al., 2021) is a revised version of the original GBI, with five subscores or factors instead of three. These five subscores are: QoL, self-confidence, support, social involvement, and general health, as well as a sixth total score. The general health and support subscores are identical to the original GBI physical and social support subscores, respectively. The GBI-5F removed questions 9, 10, and 14 from the original GBI based on relative importance to otolaryngologic intervention, redundancy, and the fact that questions 10 and 14 did not fit into any of the new constructs or factors created in the new GBI-5F.

The Nijmegen Cochlear Implant Questionnaire (NCIQ) is a HRQoL questionnaire specific to CI users (Hinderink et al., 2000). The 60-item scale consists of three domains: physical, psychological, and social. Within each domain are various subdomains consisting of 10 questions each. The physical domain consists of three subdomains: *basic sound perception*, *advanced sound perception,* and *speech production*; the psychological domain has one subdomain: *self-esteem*, and the social domain consists of two subdomains: *activity limitations* and *social interactions*. The NCIQ is scored on a Likert scale from 1 to 5 and transformed so that $1 = 0$, $2 = 25$, $3 = 50$, $4 = 75$, and $5 = 100$ These scores were then summed together and then divided by the number of completed questions, with scores ranging from 0 (poor) to 100 (optimal). Higher scores indicate better overall health-related quality of life.

Statistical Analyses

Statistical analyses were completed using SPSS Version 28 (IBM Corp, 2021). Prior to analysis, normality of data was evaluated using Shapiro–Wilk tests. HRQoL and speech perception scores were analyzed using Pearson correlations, as well as paired sample t-test comparisons to assess the effects of CI configuration (i.e., first-ear CI only vs. binaural). For Pearson correlations between HRQoL scores and speech perception performance, a power analysis conducted utilizing G * Power indicated that there was an 80% chance of detecting a medium to large effect for a sample size that ranged between 9 and 29 ($\alpha = 0.05$). For paired sample *t*-tests, G * Power indicated an 80% chance of detecting a medium to large effect with a sample size that ranged from 15 to 34 ($\alpha = 0.05$). Spearman Rho correlations were used when needed for measures that were not normally distributed. Semi-partial correlations were employed to examine the relationship between speech perception ability and HRQoL measures given that age was related to speech perception ability in all conditions, except for first-ear CI only in quiet, but not any HRQoL scores. Semi-partial correlations are similar to partial correlations but are used to examine the relationship between two variables while taking into account a covariate that is related to only one of these variables, such as age. Several GBI-5F scores were significantly asymmetrical and skewed (i.e., the distribution of the data was skewed toward the maximum possible score). Skewness was defined as any point beyond the established Fisher's skewedness coefficient (SK_F) range of −1.96 and + 1.96 (Pett, 2015), meaning the GBI-5F could not be subjected to correlational analyses and thus only descriptive statistics are reported. Ceiling and floor effects for the GBI-5F were defined as significant if ≥15% of participants scored the highest or lowest possible score for a given subscore (Gulledge et al., 2019). Bonferroni corrections were applied where appropriate. All reported statistics reflect two-tailed significance values.

RESULTS

HRQoL Descriptive Statistics

The GBI total score and general subscore (calculated with and without question 9) were interrelated (all $r \leq 0.991$, $p \leq 0.001$), but the GBI social support and physical health subscores were not related to the GBI total score or other subscores (all $\rho \leq 0.211$, $r \leq 0.965$; Bonferroni adjusted $\alpha = 0.008$; see **Table 1**). For the GBI-5F, three out of five of the subscores were substantially asymmetrical, with QoL ($kurtosis = 2.90$, $skewness = -1.69$, $SK_F = -3.64$), social involvement ($kurtosis = 1.87$, $skewness = -1.25$, $SK_F = -2.70$), and the total score ($kurtosis = 2.29$, $skewness = -1.358$, $SK_F = -2.93$) being significantly skewed. Additionally, 56% and 28% of scores on the QoL and self-confidence subscores, respectively, were significantly at ceiling. In comparison, 4% of support scores, 0% of general health scores, 8% of social involvement scores, and 0% of total scores were at ceiling. No floor effects were observed for any subscores. On the general health subscore, 80% of participants indicated no change or had a "0" score since the CI surgery (see **Figure 1**). NCIQ total score was related to all NCIQ subscores (advanced sound perception, speech production, self-esteem, activity limitation, and social interaction; all $\rho \leq 0.757$, $p \leq 0.002$) except for basic sound perception ($\rho = 0.560$, $p = 0.010$; Bonferroni adjusted $\alpha = 0.007$). Only NCIQ activity limitation and social interaction subscores of the NCIQ were interrelated ($r = 0.737$, $p < 0.001$); no other relationships between NCIQ subdomains were observed (all $r \leq 0.493$, $p \leq 0.951$; see **Table 2**).

Speech Perception and HRQoL Questionnaires

No relationships between speech perception ability with a first-ear CI only configuration in quiet and any GBI scores

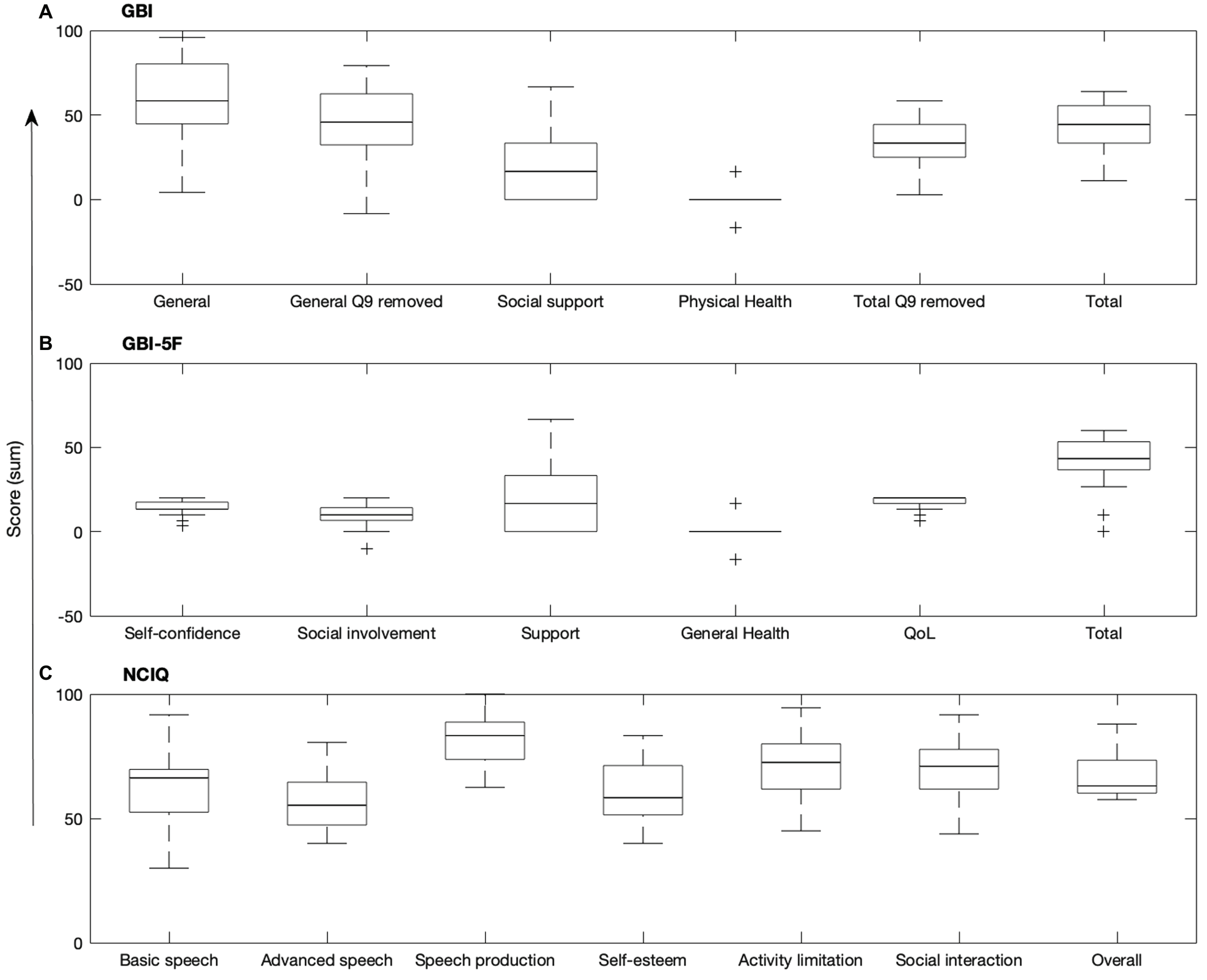


FIGURE 1 | Boxplots of the **(A)** GBI, **(B)** GBI-5F, and **(C)** NCIQ. Central mark indicates median, bottom and top edges of box indicate 25th and 75th percentiles, whiskers extend to minimum and maximum points, and plus signs indicate outliers. *Q9, question 9*.

were observed (all $\rho \leq 0.482$, $p \leq 0.846$). There were significant semi-partial correlations between binaural speech perception performance in noise and the GBI general subscore with question 9 included ($r = 0.463$, $p = 0.016$) and question 9 excluded ($r = 0.442$, $p = 0.024$; see **Table 3**). No semi-partial relationships between speech perception scores and any other GBI scores were observed (all $r \leq 0.428$, $p \leq 0.591$; see **Table 3**). GBI-5F scores were not subjected to correlational analyses due to asymmetry and skewedness. There were no relationships between speech perception performance scores and any NCIQ scores (all $r \leq 0.190$, $p \leq 0.972$; see **Table 4**).

First-Ear CI Only Versus Binaural Speech Perception Analyses

Across all participants, a paired sample *t*-test comparing speech perception ability in quiet with the first-ear CI only ($M = 84.36$, SD = 13.59) versus the binaural configuration in quiet ($M = 92.44$, SD = 7.43) demonstrated better performance scores with a binaural configuration [$t(24) = -3.58$, $p = 0.002$; see **Figure 2**]. Speech perception ability in noise with a binaural configuration ($M = 59.36$, SD = 19.84) was better than with a first-ear CI only configuration [$M = 45.84$, SD = 20.19; $t(24) = -4.28$, $p < 0.001$]. The average increase in speech perception score going from a first-ear CI only to a binaural configuration was 8.1% points in quiet (SD = 11.28) and 13.52% points in noise (SD = 15.81).

DISCUSSION

The primary goal of this study was to explore how subjective ratings of CI benefit measured with the GBI, GBI-5F, and NCIQ related with objective CI outcomes measures of speech

TABLE 1 | Intercorrelations between the GBI.

	Total	Total Q9 removed	General subscore	General subscore Q9 removed	Social support	Physical health
Total	–	**$r = 0.970$, $p < 0.001$**	**$r = 0.969$, $p < 0.001$**	**$r = 0.962$, $p < 0.001$**	$\rho = -0.145$, $p = 0.490$	$\rho = 0.145$, $p = 0.489$
Total Q9 removed	–	–	**$r = 0.936$, $p < 0.001$**	**$r = 0.940$, $p < 0.001$**	$\rho = -0.104$, $p = 0.621$	$\rho = 0.211$, $p = 0.311$
General subscore	–	–	**–**	**$r = 0.991$, $p < 0.001$**	$\rho = -0.326$, $p = 0.112$	$\rho = 0.033$, $p = 0.875$
General subscore Q9 removed	–	–	–	–	$\rho = -0.308$, $p = 0.134$	$\rho = 0.014$, $p = 0.948$
Social support	–	–	–	–	–	$\rho = -0.009$, $p = 0.965$
Physical health	–	–	–	–	–	–

Q9, question 9; Bold = significant with Bonferroni correction.

TABLE 2 | Intercorrelations between the NCIQ.

	Basic sound perception	Advanced sound perception	Speech production	Self-esteem	Activity limitation	Social interaction	Overall
Basic sound perception	–	$r = 0.387$, $p = 0.083$	$r = 0.208$, $p = 0.379$	$r = 0.447$, $p = 0.048$	$r = 0.195$, $p = 0.410$	$r = 0.348$, $p = 0.133$	$\rho = 0.560$, $p = 0.010$
Advanced sound perception	–	–	$r = 0.099$, $p = 0.679$	$r = 0.335$, $p = 0.149$	$r = 0.493$, $p = 0.027$	$r = 258$, $p = 0.271$	**$\rho = 0.700$, $p < 0.001$**
Speech production	–	–	–	$r = 0.147$, $p = 0.536$	$r = 0.048$, $p = 0.840$	$r = -0.015$, $p = 0.951$	**$\rho = 0.309$, $p = 0.185$**
Self-esteem	–	–	–	–	$r = 0.557$, $p = 0.011$	$r = 0.391$, $p = 0.088$	**$\rho = 0.722$, $p < 0.001$**
Activity limitation	–	–	–	–	–	**$r = 0.737$, $p < 0.001$**	**$\rho = 0.757$, $p = 0.001$**
Social interaction	–	–	–	–	–	–	**$\rho = 0.659$, $p = 0.002$**
Overall	–	–	–	–	–	–	–

Bold = significant with Bonferroni correction.

TABLE 3 | Relationships between AzBio and GBI scores.

GBI measures	AzBio first-ear CI only in quiet	AzBio first-ear CI only in noise	AzBio binaural quiet	AzBio binaural noise
Total	$\rho = 0.407$, $p = 0.044$	$r = 0.380$, $p = 0.055$	$r = 0.203$, $p = 0.321$	$r = 0.371$, $p = 0.062$
Total Q9 removed	$\rho = 0.385$, $p = 0.057$	$r = 0.370$, $p = 0.063$	$r = 0.132$, $p = 0.522$	$r = 0.293$, $p = 0.148$
General	$\rho = 0.337$, $p = 0.099$	$r = 0.406$, $p = 0.037$	$r = 0.268$, $p = 0.181$	**$r = 0.463$, $p = 0.016$**
General Q9 removed	$\rho = 0.354$, $p = 0.082$	$r = 0.428$, $p = 0.029$	$r = 0.249$, $p = 0.221$	**$r = 0.442$, $p = 0.024$**
Social support	$\rho = 0.041$, $p = 0.846$	$r = -0.110$, $p = 0.601$	$r = -0.291$, $p = 0.146$	$r = -0.405$, $p = 0.038$
Physical Health	$\rho = 0.132$, $p = 0.530$	$r = -0.201$, $p = 0.336$	$r = -0.085$, $p = 0.671$	$r = -0.166$, $p = 0.405$

Q9, question 9; Bold = significant with Bonferroni correction.

TABLE 4 | Relationships between AzBio and NCIQ scores.

NCIQ measures	AzBio first-ear CI only in quiet	AzBio first-ear CI only in noise	AzBio binaural in quiet	AzBio binaural in noise
Basic sound perception	$\rho = 0.380$, $p = 0.187$	$r = -0.038$, $p = 0.758$	$r = -0.156$ $p = 0.530$	$r = -0.087$, $p = 0.768$
Advanced speech perception	$\rho = 0.130$, $p = 0.585$	$r = 0.201$, $p = 0.482$	$r = 0.011$, $p = 0.940$	$r = 0.176$, $p = 0.411$
Speech production	$\rho = -0.070$, p = 0.768	$r = 0.035$ $p = 0.822$	$r = -0.068$, $p = 0.769$	$r = -0.1488$, $p = 0.515$
Self-esteem	$\rho = -0.356$, $p = 0.123$	$r = -0.464$, $p = 0.056$	$r = -0.237$, $p = 0.277$	$r = -0.257$, $p = 0.213$
Activity limitation	$\rho = 0.037$, $p = 0.878$	$r = -0.030$, $p = 0.877$	$r = -0.033$, $p = 0.897$	$r = -0.113$, $p = 0.656$
Social interaction	$\rho = 0.180$, $p = -0.045$	$r = -0.082$, $p = 0.703$	$r = 0.108$, $p = 0.653$	$r = -0.063$, $p = 0.812$
Overall	$\rho = -0.045$, $p = 0.850$	$r = -0.096$, $p = 0.677$	$r = -0.104$, $p = 0.674$	$r = -0.116$, $p = 0.642$

perception ability in bimodal and bilateral CI users. With respect to our first goal, the general subscore of the GBI related with speech perception ability in noise with CI users in a binaural configuration. However, no other relationships with the GBI, GBI-5F, or NCIQ were observed. Regarding our second goal, across all participants (bimodal

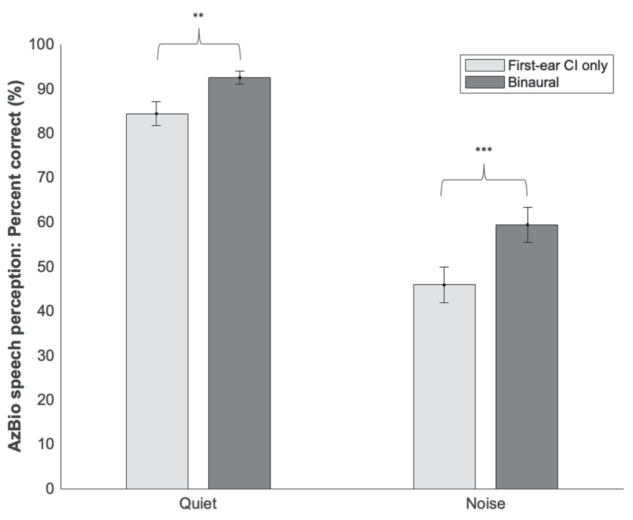


FIGURE 2 | Means and SD's of first-ear CI only and binaural speech perception scores in quiet and noise. Significant differences were noted between a first-ear CI only and binaural configuration in both quiet and noise. **≤0.01, ***≤0.001.

and bilateral combined) higher speech perception ability was observed with the respective binaural configuration over a first-ear CI only configuration.

In the current study, we found that the GBI general subscore related with CI-aided speech perception ability in noise, only when a binaural configuration was used. Indeed, speech perception in noise is generally considered more reflective of everyday living conditions given that there is a certain level of noise present in our daily listening environment (Schafer et al., 2007; Wu et al., 2018). While we saw a positive correlation between CI speech in noise perception and the GBI general subscore, the GBI also contains a variety of questions aimed at various domains of health, such as physical health. In line with our second prediction regarding domains less relevant to speech perception ability, we found that GBI subscores calculated from questions not related to speech perception ability demonstrated no relationship with CI-aided speech perception performance. For example, the physical health subscore, which contains questions centered around changes in medications or frequency of illnesses that have occurred since cochlear implantation, was not correlated with any speech measures, indicating physical health may not be a strong indicator of CI benefit in this population. Indeed, in our study, 93% of participants reported no change on the three questions contributing to the physical health subscore, with 100% of participants reporting no change for question 12: *Since you had cochlear implant surgery, do you catch colds or infection more or less often?* Similarly, on average, 68% of participants reported no change on the three questions in the social support subscore, with 96% of participants reporting no change for question 11: *Since your cochlear implant surgery are there more or fewer people who really care about you?* The physical health subscore, followed by the social support subscore, were the subscores closest to an overall average of zero, further indicating that participants experienced the least amount of change in these domains after receiving a CI. These results suggest that these subscores and the questions included in them were less relevant to our group of CI users in terms of overall benefit in HRQoL from their CI. These results are consistent with previous studies where the least amount of benefit post-implantation was measured *via* the physical health subscore followed by the social support subscore (Lassaletta et al., 2006; Straatman et al., 2014; Sanchez-Cuadrado et al., 2015; Amin et al., 2021).

While the GBI general subscore did relate to speech perception abilities in this study, the total composite score did not, perhaps due to questions included that may have been less relevant to the patient population in this study. The GBI general subscore does not include questions from either the physical or social subscore, whereas the total composite score includes questions from both. In our study, it could be argued that speech perception ability may not have related to the total subscore for precisely this reason, in that it contained a larger number of questions less pertinent to CI outcomes. Similar to Browning et al. (2021), four subjects in our study indicated that question 9 was not relevant to them and therefore chose "no change" but would have preferred a "not applicable" option. This is understandable due to the nature of question 9 which discusses employment opportunities which may be less of a consideration for older individuals who were retired, did not work or participants who had not experienced recent job transitions. To assess the impact of this question we created an alternate score for any subscore that included this question (i.e., the total and general subscore). As expected, when removing this question from the total and general subscore, we did find significant differences between the average scores with and without this question removed. However, removal of question 9 did not alter the correlational relationships with CI speech perception, indicating that this question alone did not have a significant impact on the GBI's general subscores sensitivity to CI speech ability. Browning et al. (2021) found that removal of question 9 from 3,436 participants that had completed the question made no material difference in terms of the average total score, although the N of the study was much greater and the differences between the general subscore with and without question 9 removed were not reported.

Another method to reorganize the GBI into potentially more meaningful constructs by grouping more homogenous questions and removing those less pertinent was developed by Browning et al. (2021). By employing this scoring method named the GBI-5F, we were able to explore whether these new constructs which are embedded within the original GBI may be more reflective of HRQoL for CI users. However, our results demonstrated that in this CI population, GBI-5F subscores were substantially skewed compared to the original GBI (see **Figure 1**). For example, 78% of participants reported a 5 (i.e., much better) in response to questions regarding change in QoL, suggesting that QoL greatly increased following cochlear implantation, but also that this subset of questions may not be specific enough for differentiating between degrees of benefit for this patient population if a majority of participants chose the maximum value. Given that the original GBI and GBI-5F general health and support subscores are identical with both scoring methods, no change to the restricted range was noted in the GBI-5F. The ceiling effects and restricted range observed within the new GBI-5F domains may be because the GBI-5F removed several questions from the original GBI (9, 10, and 14), thus reducing its sensitivity in this population. However, it may also be because the GBI-5F was developed for a broader range of otolaryngologic patients rather than a specific subset of that population, (i.e., CI users). Given the unique nature of CI users within the sphere of otolaryngologic intervention, additional work is needed before any clinical recommendations for GBI-5F use in this population can be determined.

While we saw a positive correlation between the GBI general subscores and CI speech measures in noise, this was not apparent for CI speech measures and the NCIQ. These results, however, are consistent with previous studies (Capretta and Moberly, 2016; Hillyer et al., 2019) and a meta-analysis which indicated low to negligible correlations between the NCIQ and speech perception abilities in both quiet and noise (McRackan et al., 2018). This may be because the NCIQ was developed to identify CI benefit by comparing pre-operative and post-operative scores (Hinderink et al., 2000), whereas our study examined post-implantation scores only. Indeed, studies that have demonstrated a relationship with NCIQ speech domains (i.e., basic sound perception, advanced sound perception, and speech production) and speech perception, have analyzed the change in speech perception in fixed pre- and post-implantation time ranges (Hirschfelder et al., 2008; Olze et al., 2012; Häußler et al., 2019). As such, the NCIQ appears to be more clinically applicable to CI-aided speech perception abilities when evaluating benefit through pre-post implantation scores rather than relating them to post-implantation scores alone.

The second goal of our study explored the differences within bimodal and bilateral CI user groups in terms of CI-aided speech perception ability. As expected in our third prediction, we found that across all participants (bimodal and bilateral combined) CI-aided speech perception scores in quiet and noise were higher with a binaural configuration versus a first-ear CI only configuration. This was evidenced by an 8.1% point increase on average in quiet and a 13.5% point increase in noise after adding a second CI for bilateral users or a HA for bimodal users, suggesting that binaural amplification was beneficial for speech perception performance. These results align with previous research demonstrating a speech perception benefit when moving from unilateral CI to a bimodal (Ching et al., 2004, 2008; Iwaka et al., 2004; Morera et al., 2005; Schafer et al., 2007; Illg et al., 2014; Farinetti et al., 2015; Hua et al., 2017) or bilateral CI (Gantz et al., 2002; Ramsden et al., 2005; Litovsky et al., 2006; Schafer et al., 2007; Buss et al., 2008).

One limitation of our study was a small sample size, specifically with regards to bilateral CI users (8 and 5 for the GBI and NCIQ respectively). While our sample size is not atypical of research surrounding bilateral users (Potts and Litovsky, 2014; Gifford et al., 2015; Moberly et al., 2018) comparing speech and HRQoL measures between bimodal and bilateral CI users can only be considered preliminary. In this study, these preliminary results indicate that bimodal and bilateral CI demonstrated essentially equivalent performance on CI-aided speech perception ability with a first-ear CI only configuration. However, bilateral CI users had an average binaural speech perception score of 97.13% in quiet and 69.88% in noise, while bimodal CI users had an average binaural speech perception score of 90.24% in quiet and 54.41% in noise. These results would appear in line with the meta-analysis of bimodal and bilateral CI users by Schafer et al. (2011), in which bilateral CI users had a slight but significant advantage in binaural performance over bimodal users in noise. It may be that more difficult speech perception tests, such as speech in background noise, are more likely to reveal a binaural benefit in bilateral CI users as evidenced by Wackym et al. (2007) in which the greatest and most consistent binaural benefit was observed with sentences presented in noise, followed by words presented in quiet. These results are also consistent with previous research that has seen more binaural benefit for bilateral CI users relative to a unilateral configuration, in noise than in quiet, for both word and sentence-level materials (Ramsden et al., 2005; Schafer et al., 2007). Future work with a larger sample size is needed to address the potential differences in binaural benefit between bimodal and bilateral CI.

CONCLUSION

Our work demonstrates that the GBI general subscore related to speech perception ability in bimodal and bilateral everyday listening conditions unlike the total, physical or social support subscores, while the CI-specific NCIQ did not relate to speech perception ability in any domains. The GBI-5F had significant limitations when applied to this patient population due to skewedness, and therefore, recommendations for clinical applicability in CI users would be premature. Given the variability in the current literature due to the wide variety of speech testing materials and HRQoL questionnaires used, future research should aim to explore the relationships between clinical measures and these HRQoL questionnaires in each of the CI configurations separately.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because the Auditory Research Laboratory is part of a hospital system that does not allow for data sharing due to patient privacy requirements. Requests to access the datasets should be directed to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Swedish Medical Center Institutional Review Board

(#SWD56152-14). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

NB, AP-C, and JH participated in data collection. NB and AP-C created figures and tables and completed statistical analysis of data. NB, EE, and AP-C wrote manuscript. CH and JH contributed to editing the manuscript. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

The authors wish to thank the patients who donated their valuable time and the Swedish Neuroscience Institute for supporting this study.

REFERENCES

Amin, N., Wong, G., Nunn, T., Jiang, D., and Pai, I. (2021). The outcomes of Cochlear implantation in elderly patients: A single United Kingdom center experience. *Ear Nose Throat J.* 100(Suppl. 5), 842S–847S. doi:10.1177/0145561320910662

Brown, L., Sherbenou, R. J., and Johnsen, S. K. (2010). *Test of Nonverbal Intelligence: TONI-4*. TX: Pro-Ed, Inc.

Browning, G. G., Kubba, H., and Whitmer, W. M. (2021). Revised 15-item Glasgow benefit inventory with five factors based on analysis of a large population study of medical and surgical otorhinolaryngological interventions. *Clin. Otolaryngol.* 46, 213–221. doi: 10.1111/coa.13649

Buss, E., Pillsbury, H. C., Buchman, C. A., Pillsbury, C. H., Clark, M. S., Haynes, D. S., et al. (2008). Multicenter US bilateral MED-EL Cochlear implantation study: speech perception over the first year of use. *Ear Hear.* 29, 20–32. doi: 10.1097/AUD.0b013e31815d7467

Capretta, N. R., and Moberly, A. C. (2016). Does quality of life depend on speech recognition performance for adult cochlear implant users? *Laryngoscope* 126, 699–706. doi: 10.1002/lary.25525

Ching, T. Y., Incerti, P., and Hill, M. (2004). Binaural benefits for adults who use hearing aids and cochlear implants in opposite ears. *Ear Hear.* 25, 9–21. doi: 10.1097/01.AUD.0000111261.84611.C8

Ching, T. Y. C., Massie, R., Van Wanrooy, E., Rushbrooke, E., and Psarros, C. (2008). Bimodal fitting or bilateral implantation? *Cochlear Implants Int.* 10, 23–27. doi: 10.1002/cii.381

Farinetti, A., Roman, S., Mancini, J., Baumstarck-Barrau, K., Meller, R., Lavieille, J. P., et al. (2015). Quality of life in bimodal hearing users (unilateral cochlear implants and contralateral hearing aids). *Eur. Arch. Otorhinolaryngol.* 272, 3209–3215. doi: 10.1007/s00405-014-3377-8

Forli, F., Lazzerini, F., Fortunato, S., Bruschini, L., and Berrettini, S. (2019). Cochlear implant in the elderly: results in terms of speech perception and quality of life. *Audiol. Neurootol.* 24, 77–83. doi: 10.1159/000499176

Gajadeera, E. A., Galvin, K. L., Dowell, R. C., and Busby, P. A. (2017). The change in electrical stimulation levels During 24 months Postimplantation for a large cohort of adults using the nucleus® Cochlear implant. *Ear Hear.* 38, 357–367. doi: 10.1097/aud.0000000000000405

Gantz, B. J., Tyler, R. S., Rubinstein, J. T., Wolaver, A., Lowder, M., Abbas, P., et al. (2002). Binaural Cochlear implants placed during the same operation. *Otol. Neurotol.* 23, 169–180. doi: 10.1097/00129492-200203000-00012

Gaylor, J. M., Raman, G., Chung, M., Lee, J., Rao, M., Lau, J., et al. (2013). Cochlear implantation in adults: a systematic review and meta-analysis. *JAMA Otolaryngol. Head Neck Surg.* 139, 265–272. doi: 10.1001/jamaoto.2013.1744

Gifford, R. H., Driscoll, C. L., Davis, T. J., Fiebig, P., Micco, A., and Dorman, M. F. (2015). A within-subjects comparison of bimodal hearing, bilateral cochlear implantation, and bilateral cochlear implantation with bilateral hearing preservation: high-performing patients. Otology and neurotology: official publication of the American Otological Society, American Neurotology Society [and] European academy of. *Otol. Neurotol.* 36, 1331–1337. doi: 10.1097/MAO.0000000000000804

Gulledge, C. M., Smith, D. G., Ziedas, A., Muh, S. J., Moutzouros, V., and Makhni, E. C. (2019). Floor and ceiling effects, time to completion, and question burden of PROMIS CAT domains Among shoulder and knee patients undergoing nonoperative and operative treatment. *JBJS Open Access* 4:15. doi: 10.2106/JBJS.OA.19.00015

Häußler, S. M., Knopke, S., Wiltner, P., Ketterer, M., Gräbel, S., and Olze, H. (2019). Long-term benefit of unilateral Cochlear implantation on quality of life and speech perception in bilaterally deafened patients. *Otol. Neurotol.* 40, e430–e440. doi: 10.1097/mao.0000000000002008

Hillyer, J., Elkins, E., Hazlewood, C., Watson, S. D., Arenberg, J. G., and Parbery-Clark, A. (2019). Assessing cognitive abilities in high-performing Cochlear implant users. *Front. Neurosci.* 12:1056. doi: 10.3389/fnins.2018.01056

Hinderink, J. B., Krabbe, P. F., and Van Den Broek, P. (2000). Development and application of a health-related quality-of-life instrument for adults with cochlear implants: the Nijmegen cochlear implant questionnaire. *Otolaryngol. Head Neck Surg.* 123, 756–765. doi: 10.1067/mhn.2000.108203

Hirschfelder, A., Grabel, S., and Olze, H. (2008). The impact of cochlear implantation on quality of life: the role of audiologic performance and variables. *Otolaryngol. Head Neck Surg.* 138, 357–362. doi: 10.1016/j.otohns.2007.10.019

Ho, E. C., Monksfield, P., Egan, E., Reid, A., and Proops, D. (2009). Bilateral bone-anchored hearing aid: impact on quality of life measured with the Glasgow benefit inventory. *Otol. Neurotol.* 30, 891–896. doi: 10.1097/MAO.0b013e3181b4ec6f

Hua, H., Johansson, B., Magnusson, L., Lyxell, B., and Ellis, R. J. (2017). Speech recognition and cognitive skills in bimodal Cochlear implant users. *J. Speech Lang. Hear. Res.* 60, 2752–2763. doi: 10.1044/2017_JSLHR-H-16-0276

IBM Corp (2021). *IBM SPSS Statistics for Macintosh, Version 28.0.* Armonk, NY: IBM Corp.

Illg, A., Bojanowicz, M., Lesinski-Schiedat, A., Lenarz, T., and Büchner, A. (2014). Evaluation of the bimodal benefit in a large cohort of Cochlear implant subjects using a contralateral hearing aid. *Otol. Neurotol.* 35, e240–e244. doi: 10.1097/mao.0000000000000529

Iwaka, T., Matsushiro, N., Mah, S.-R., Sato, T., Yasuoka, E., Yamamoto, K.-I., et al. (2004). Comparison of speech perception between monaural and binaural hearing in cochlear implant patients. *Acta Otolaryngol.* 124, 358–362. doi: 10.1080/00016480310000548a

Krabbe, P. F., Hinderink, J. B., and van den Broek, P. (2000). The effect of cochlear implant use in postlingually deaf adults. *Int. J. Technol. Assess. Health Care* 16, 864–873. doi: 10.1017/s0266462300102132

Lassaletta, L., Castro, A., Bastarrica, M., de Sarria, M. J., and Gavilan, J. (2006). Quality of life in postlingually deaf patients following cochlear implantation. *Eur. Arch. Otorhinolaryngol.* 263, 267–270. doi: 10.1007/s00405-005-0987-1

Litovsky, R., Parkinson, A., Arcaroli, J., and Sammeth, C. (2006). Simultaneous bilateral Cochlear implantation in adults: A multicenter clinical study. *Ear Hear.* 27, 714–731. doi: 10.1097/01.aud.0000246816.50820.42

McRackan, T. R., Bauschard, M., Hatch, J. L., Franko-Tobin, E., Droghini, H. R., Nguyen, S. A., et al. (2018). Meta-analysis of quality-of-life improvement after cochlear implantation and associations with speech recognition abilities. *Laryngoscope* 128, 982–990. doi: 10.1002/lary.26738

Moberly, A. C., Harris, M. S., Boyce, L., Vasil, K., Wucinich, T., Pisoni, D. B., et al. (2018). Relating quality of life to outcomes and predictors in adult cochlear implant users: are we measuring the right things? *Laryngoscope* 128, 959–966. doi: 10.1002/lary.26791

Morera, C., Manrique, M., Ramos, A., Garcia-Ibanez, L., Cavalle, L., Huarte, A., et al. (2005). Advantages of binaural hearing provided through bimodal stimulation via a cochlear implant and a conventional hearing aid: A 6-month comparative study. *Acta Otolaryngol.* 125, 596–606. doi: 10.1080/00016480510027493

Mosnier, I., Bebear, J. P., Marx, M., Fraysse, B., Truy, E., Lina-Granade, G., et al. (2015). Improvement of cognitive function after cochlear implantation in elderly patients. *JAMA Otolaryngol. Head Neck Surg.* 141, 442–450. doi: 10.1001/jamaoto.2015.129

Olze, H., Grabel, S., Forster, U., Zirke, N., Huhnd, L. E., Haupt, H., et al. (2012). Elderly patients benefit from cochlear implantation regarding auditory rehabilitation, quality of life, tinnitus, and stress. *Laryngoscope* 122, 196–203. doi: 10.1002/lary.22356

Palmer, C. S., Niparko, J. K., Wyatt, J. R., Rothman, M., and de Lissovoy, G. (1999). A prospective study of the cost-utility of the multichannel cochlear implant. *Arch. Otolaryngol. Head Neck Surg.* 125, 1221–1228. doi: 10.1001/archotol.125.11.1221

Pett, M. A. (2015). *Nonparametric Statistics for Health Care Research: Statistics for Small Samples and Unusual Distributions*. United States: Sage Publications.

Potts, L. G., and Litovsky, R. Y. (2014). Transitioning From bimodal to bilateral Cochlear implant listening: speech recognition and localization in four individuals. *Am. J. Audiol.* 23, 79–92. doi: 10.1044/1059-0889(2013/11-0031)

Ramsden, R., Greenham, P., O'Driscoll, M., Mawman, D., Proops, D., Craddock, L., et al. (2005). Evaluation of bilaterally implanted adult subjects with the nucleus 24 Cochlear implant system. *Otol. Neurotol.* 26, 988–998. doi: 10.1097/01.mao.0000185075.58199.22

Robinson, K., Gatehouse, S., and Browning, G. G. (1996). Measuring patient benefit from otorhinolaryngological surgery and therapy. *Ann. Otol. Rhinol. Laryngol.* 105, 415–422. doi: 10.1177/000348949610500601

Sanchez-Cuadrado, I., Lassaletta, L., Perez-Mora, R., Muñoz, E., and Gavilan, J. (2015). Reliability and validity of the Spanish Glasgow benefit inventory after cochlear implant surgery in adults. *Eur. Arch. Otorhinolaryngol.* 272, 333–336. doi: 10.1007/s00405-013-2844-y

Schafer, E. C., Amlani, A. M., Paiva, D., Nozari, L., and Verret, S. (2011). A meta-analysis to compare speech recognition in noise with bilateral cochlear implants and bimodal stimulation. *Int. J. Audiol.* 50, 871–880. doi: 10.3109/14992027.2011.622300

Schafer, E. C., Amlani, A. M., Seibold, A., and Shattuck, P. L. (2007). A Meta-analytic comparison of binaural benefits between bilateral cochlear implants and bimodal stimulation. *J. Am. Acad. Audiol.* 18, 760–776. doi: 10.3766/jaaa.18.9.5

Services, U. D. O. H. A. H (2011). Foundation health measure report: health-related quality of life and well-being.

Sladen, D. P., Peterson, A., Schmitt, M., Olund, A., Teece, K., Dowling, B., et al. (2017). Health-related quality of life outcomes following adult cochlear implantation: A prospective cohort study. *Cochlear Implants Int.* 18, 130–135. doi: 10.1080/14670100.2017.1293203

Sorrentino, T., Donati, G., Nassif, N., Pasini, S., and Redaelli de Zinis, L. O. (2020). Cognitive function and quality of life in older adult patients with cochlear implants. *Int. J. Audiol.* 59, 316–322. doi: 10.1080/14992027.2019.1696993

Spahr, A. J., Dorman, M. F., Litvak, L. M., Van Wie, S., Gifford, R. H., Loizou, P. C., et al. (2012). Development and validation of the AzBio sentence lists. *Ear Hear.* 33, 112–117. doi: 10.1097/AUD.0b013e31822c2549

Straatman, L. V., Huinck, W. J., Langereis, M. C., Snik, A. F. M., and Mulder, J. J. (2014). Cochlear implantation in late-implanted Prelingually deafened adults: changes in quality of life. *Otol. Neurotol.* 35:253. doi: 10.1097/MAO.0b013e3182a4758e

Vermeire, K., Brokx, J. P., Wuyts, F. L., Cochet, E., Hofkens, A., and Van de Heyning, P. H. (2005). Quality-of-life benefit from cochlear implantation in the elderly. *Otol. Neurotol.* 26, 188–195. doi: 10.1097/00129492-200503000-00010

Wackym, P. A., Runge-Samuelson, C. L., Firszt, J. B., Alkaf, F. M., and Burg, L. S. (2007). More challenging speech-perception tasks demonstrate binaural benefit in bilateral Cochlear implant users. *Ear Hear.* 28, 80S–85S. doi: 10.1097/AUD.0b013e3180315117

Wu, Y.-H., Stangl, E., Chipara, O., Hasan, S. S., Welhaven, A., and Oleson, J. (2018). Characteristics of real-world signal to noise ratios and speech listening situations of older adults with mild to moderate hearing loss. *Ear Hear.* 39, 293–304. doi: 10.1097/AUD.0000000000000486

Conflict of Interest: 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 | Frontiers in Psychology

ORIGINAL RESEARCH
published: 16 May 2022
doi: 10.3389/fpsyg.2022.821044

Using Eye-Tracking to Investigate an Activation-Based Account of False Hearing in Younger and Older Adults

*Eric Failes and Mitchell S. Sommers**

Department of Psychological and Brain Sciences, Washington University in St. Louis, St. Louis, MO, United States

Several recent studies have demonstrated context-based, high-confidence misperceptions in hearing, referred to as *false hearing*. These studies have unanimously found that older adults are more susceptible to false hearing than are younger adults, which the authors have attributed to an age-related decline in the ability to inhibit the activation of a contextually predicted (but incorrect) response. However, no published work has investigated this activation-based account of false hearing. In the present study, younger and older adults listened to sentences in which the semantic context provided by the sentence was either unpredictive, highly predictive and valid, or highly predictive and misleading with relation to a sentence-final word in noise. Participants were tasked with clicking on one of four images to indicate which image depicted the sentence-final word in noise. We used eye-tracking to investigate how activation, as revealed in patterns of fixations, of different response options changed in real-time over the course of sentences. We found that both younger and older adults exhibited anticipatory activation of the target word when highly predictive contextual cues were available. When these contextual cues were misleading, younger adults were able to suppress the activation of the contextually predicted word to a greater extent than older adults. These findings are interpreted as evidence for an activation-based model of speech perception and for the role of inhibitory control in false hearing.

Keywords: false hearing, speech perception, aging, eye-tracking, inhibition

Edited by:
Leah Fostick,
Ariel University, Israel

Reviewed by:
Gal Nitsan,
University of Haifa, Israel
Arthur Wingfield,
Brandeis University, United States

****Correspondence:***
Mitchell S. Sommers
msommers@wustl.edu

Specialty section:
This article was submitted to Psychology of Aging, a section of the journal Frontiers in Psychology

Received: 23 November 2021
Accepted: 28 March 2022
Published: 16 May 2022

Citation:
Failes E and Sommers MS (2022) Using Eye-Tracking to Investigate an Activation-Based Account of False Hearing in Younger and Older Adults. Front. Psychol. 13:821044. doi: 10.3389/fpsyg.2022.821044

INTRODUCTION

What a listener reports hearing is influenced by what they expect to hear. Indeed, there are many studies that demonstrate how speech perception is facilitated by the presence of valid semantic contexts, which allow the listener to anticipate what will be said (Hutchinson, 1989; Nittrouer and Boothroyd, 1990; Wingfield et al., 1991; Pichora-Fuller et al., 1995; Sommers and Danielson, 1999; Dubno et al., 2000; Benichov et al., 2012; Rogers et al., 2012; Sommers et al., 2015; Failes et al., 2020). The availability of contextual cues may be especially important when the speech signal is degraded—such as when speech is presented in noise or when the listener suffers from hearing loss—as the semantic cues may allow the listener to infer what was said in cases where acoustic information was missed.

A compelling demonstration of the influence of context in speech perception is *false hearing*, instances in which listeners erroneously report hearing a contextually predicted word when a similar sounding but unpredicted word is presented (Rogers et al., 2012; Sommers et al., 2015; Failes et al., 2020). For example, Sommers et al. (2015) had younger and older adults identify sentence-final words in three different conditions: (1) sentences providing no context for predicting the sentence-final word (baseline condition: *He was thinking about the* sheep); (2) sentences providing a valid context for predicting the sentence-final word (congruent condition: *The shepherd watched his* sheep); and (3) sentences in which the sentence-final word was a phonological neighbor of the predicted sentence-final word (incongruent condition: *The shepherd watched his* sheath). In all cases, the participant's task was to report the sentence-final word, which was presented in background noise, and to judge their confidence in the accuracy of their response on a 0–100 point scale. To account for differences in hearing acuity across younger and older adults (see Morrell et al., 1996; Sommers et al., 2011 for hearing acuity trends across the lifespan), signal-to-noise ratios (SNRs) were set individually to obtain approximately 50% accuracy for all participants in the baseline condition. Importantly, all participants were warned that sentences would sometimes be misleading, and there were three times as many incongruent sentences as congruent sentences. Both the warning instructions and the disproportionate use of incongruent sentences should have discouraged a context-based response strategy. Despite conditions that discouraged using context, Sommers et al. found that both younger and older adults experienced false hearing in the incongruent condition (e.g., reported hearing *sheep* when presented *The shepherd watched his* sheath). However, older adults were more susceptible to and more confident in their false hearing responses than were younger adults (0.50 vs. 0.39). Older adults were also much more likely (0.16 vs. 0.04) than younger adults to report maximum confidence in cases of false hearing (100% confidence in the incorrect, but semantically predicted response). Participants' continued use of contextual cues in conditions that discouraged use of context lends further support to the argument that both age groups—but especially older adults—relied on context as a basis for responding.

False hearing is often described as resulting from an inability to suppress an expected response, reflecting a failure of inhibitory control (Rogers et al., 2012; Sommers et al., 2015; Failes et al., 2020). Using the stimuli from Sommers et al. (2015) as an example, the sentence "*The shepherd watched his…*" creates a strong expectation of what word should follow (*sheep*). This expectation acts a source of increased activation of the expected lexical item. In the case of incongruent sentences, the participant must then suppress the highly activated word *sheep* to correctly hear the presented (but unpredicted) word *sheath*.

Although the role of inhibitory control in false hearing has yet to be tested directly, there is evidence that inhibitory control influences veridical speech perception. Sommers and Danielson (1999, Experiment 2), for example, found that better inhibitory control—as assessed by a composite of a selective attention paradigm developed by Garner (1974) and an auditory Stroop task (Stroop, 1935)—was associated with improved ability to identify lexically hard words (i.e., words with many, high-frequency phonological neighbors) but not lexically easy words (i.e., words with fewer, lower-frequency phonological neighbors) in noise. Additionally, the correlation between inhibitory control and the ability to identify lexically hard words was reduced when the target word was preceded by a high-predictability sentence (e.g., *She was walking along the* path) relative to when preceded by a low-predictability sentence (e.g., *She was thinking about the* path). The authors interpreted their findings within the Neighborhood Activation Model (NAM; Luce and Pisoni, 1998), which suggests that when listening to a spoken word, both the word and its phonological neighbors become activated in the mental lexicon and compete for perception. Sommers and Danielson suggested that inhibitory control might be used to suppress the activation of competing phonological neighbors, and that the availability of highly predictive (and valid) contextual cues may reduce the need to suppress competitors by selectively increasing the activation of the contextually congruent target word.

The proposal by Sommers and Danielson (1999, Experiment 2) that inhibitory control is needed to reduce activation of a target word's phonological neighbors and that context increases the activation of semantically viable words is highly pertinent to the studies of false hearing. Recall that the target words in the incongruent condition of the study by Sommers et al. (2015) were phonological neighbors of the word predicted by context. Based on the NAM (Luce and Pisoni, 1998) and the findings of Sommers and Danielson, we would expect that hearing the incongruent target word (e.g., *sheath* in the sentence *The shepard watched his* sheath) would activate the contextually predicted phonological neighbor of the target word (e.g., *sheep*), and that the activation of this phonological neighbor would be boosted due to its compatibility with available contextual cues. This set of conditions should result in an increased probability that participants would mistakenly "hear" the contextually predicted phonological neighbor, a case of false hearing. Therefore, if the role of inhibitory control is to decrease activation of phonological neighbors of the target word as suggested by Sommers and Danielson, then we might expect that individuals with better inhibitory control would be less susceptible to false hearing than those with poorer inhibitory control. Given the well-established age-related decline in inhibitory control (Cohn et al., 1984; Hasher and Zacks, 1988; MacLeod, 1991; Hasher et al., 1997; Sommers and Danielson, 1999; Jacoby et al., 2005, Experiment 2; Sommers and Huff, 2003, Experiment 2; West and Alain, 2000), older adults' increased susceptibility to false hearing may result, at least in part, from declines in the ability to inhibit activation on phonological neighbors activated by both semantic context and phonological similarity.

One experimental method that may be particularly useful for testing the inhibitory control account of false hearing is eye-tracking. Eye-tracking has been increasingly used to study language processing because it allows the researcher to observe changes in attentional focus over time, providing a real-time assessment of the processing that occurs before a response is made. In speech perception research, eye-tracking is often used within a visual world paradigm (Allopenna et al., 1998; Dahan and Gaskell, 2007; Revill and Spieler, 2012; Mishra and Singh, 2014; Ito et al., 2018; Kukona, 2020), in which different response options are depicted in the form of written words or pictures. It has been argued that changes in the proportion of fixations on the written words or pictures in the visual world paradigm can be used as an index of changes in activation, with more highly activated response options receiving a greater proportion of fixations than less highly activated options (Tanenhaus et al., 2000). Supporting this claim, images depicting high-frequency words, which are assumed to gain more activation than low-frequency words within speech perception models such as the NAM (Luce and Pisoni, 1998), tend to receive a greater proportion of fixations than images depicting low-frequency words (Dahan et al., 2001; Dahan and Gaskell, 2007; Revill and Spieler, 2012). Similarly, following from claim of Sommers and Danielson (1999) that words gain activation when supported by contextual cues, images depicting words that are congruent with available semantic context tend to receive a greater proportion of fixations than images depicting words that do not fit with context (Ito et al., 2018; Kukona, 2020). Therefore, the eye-tracking methodology could allow for a test of the inhibitory control account of false hearing. Misleading sentences, as in the incongruent condition of the study by Sommers et al. (2015), should lead to increased fixations on an image depicting the contextually predicted (but incorrect) word. To the extent that participants are able to suppress activation of the contextually predicted word when the unpredicted target word is presented, they should be able to fixate more on the image depicting the correct, but not predicted, target item.

Two recent studies (Ayasse and Wingfield, 2020; Harel-Arbeli et al., 2021) used eye-tracking to compare activation of semantic, rather than phonological, competitors in older and younger adults. Ayasse and Wingfield (2020) compared the time-course of gaze fixations for older and younger adults on sentence-final target items when a semantically plausible competitor was either present or absent as well as for a control condition in which context was not predictive of the sentence final item. For young adults, growth curve analyses indicated similar slopes of gaze fixations whether or not a semantically related competitor was presented. In contrast, older adults showed shallower slopes for target fixations (i.e., slower time-course of target fixations) when a semantic competitor was present compared to when an unrelated word was used as the response alternative. Moreover, the age-related difference in target fixations in the presence of a semantic competitor was driven in part by individual differences in inhibitory control as evidenced by greater interference from a semantic competitor for older adults with lower versus higher measures of inhibitory control. Similar results were reported by Harel-Arbeli et al. (2021) who found that older adults were slower than younger listeners to look at a target picture when a semantic competitor was present, suggesting an age-related decline in the ability to inhibit semantic competition.

In the present study, we used eye-tracking with the visual world paradigm to examine age-related changes in inhibiting phonological competition as revealed by fixations in a false hearing paradigm. Specifically, the current study investigated how activation of different response options change over the course of neutral, valid, and misleading sentence contexts. The visual world task was similar to the one used by Harel-Arbeli et al. (2021) but used the speech perception in noise (SPIN) task described earlier (Sommers et al., 2015) in which context preceding a sentence-final word in noise was either non-predictive (baseline condition), predictive and valid (congruent condition) or predictive but invalid (incongruent condition). As each sentence was played, four pictures were presented on the computer screen. The pictures depicted the target word (e.g., box), a phonological neighbor of the target word (e.g., fox), and two words that did not sound like the target word and were not predicted by the sentence context (e.g., key and paw). Using eye-tracking, we were able to determine how the proportion of fixations on each of the images changed over the course of the sentence and after the target word was presented.

We formed specific hypotheses regarding how younger and older adults' fixation patterns would change over the course of baseline, congruent, and incongruent sentences. In the baseline condition, we predicted that the proportion of fixations on each of the images should remain approximately equal until the target word was presented since there were no contextual cues upon which to base an expectation. After the target word was presented in the baseline condition, fixations on the target image should increase in accordance with participants' ability to accurately hear the target word. In congruent and incongruent sentences, we hypothesized that both younger and older adults would become increasingly fixated on the contextually predicted image leading up to presentation of the target word, demonstrating increasing anticipatory activation of the word supported by context. We predicted that this increased focus on the contextually predicted image might be greater for older than younger adults, reflecting older adults' increased context-based responding demonstrated in previous studies (Rogers et al., 2012; Sommers et al., 2015; Failes et al., 2020). Whereas fixations on the target image should continue to increase for both age groups once the target word was presented in the congruent condition, we predicted that the age groups would differ in their reaction to presentation of the target word in the incongruent condition. Specifically, we predicted that younger adults would decrease their proportion of fixations on the contextually predicted (but not presented) visual image and increase their fixations on the unpredicted (but correct) target image after the target word was presented, reflecting their ability to suppress the activation of the expected word. Older adults, on the other hand, were expected to maintain or even increase their fixations on the contextually predicted image after the target word was presented in incongruent sentences, reflecting an inability to suppress the activation of the expected word. This would align with the theory that false hearing reflects a failure to inhibit a highly activated response and the findings of previous studies suggesting that older adults have poorer inhibitory control than younger adults (Cohn et al., 1984; Hasher and Zacks, 1988; MacLeod, 1991; Hasher et al., 1997; Sommers and Danielson, 1999; Jacoby et al., 2005, Experiment 2; Sommers and Huff, 2003, Experiment 2; West and Alain, 2000).

MATERIALS AND METHODS

Participants

Participants were 23 younger adults ages 18–29 ($M = 21.0$, SD = 2.68) and 19 older adults ages 66–81 ($M = 73.31$, SD = 4.45). All participants were native English speakers who did not require the use of a hearing aid and self-reported normal or corrected-to-normal vision. To assess English language competency, participants completed the Shipley Vocabulary Test (Shipley, 1940) in which participants decided which of four words was most similar in meaning to 40 distinct target words. Good vocabulary knowledge was exhibited by both younger ($M = 33.04$, SD = 3.04) and older adults ($M = 34.42$, SD = 3.91), and the two groups did not differ in vocabulary knowledge, $t(33.61) = -1.26$, $p > 0.05$. Hearing thresholds were assessed for octave frequencies from 250 to 8,000 Hz in a sound-attenuating booth using standard audiometry. As would be expected due to age-related hearing loss (Morrell et al., 1996; Sommers et al., 2011), older adults ($M = 23.68$, SD = 11.78) had poorer best-ear PTAs (500/1,000/2,000 Hz frequencies) than younger adults ($M = 4.42$, SD = 3.39), $t(20.47) = -6.90$, $p < 0.001$.

Stimuli and Materials

Stimuli in the SPIN task were 31 carrier sentences (baseline sentences: *The word is page*) and 62 high-predictability sentences (congruent sentences: *She put the toys in a box*) selected from the Revised Speech Perception in Noise Test (Bilger et al., 1984) or created specifically for this study. For half of the congruent sentences, the first or last sound in the sentence-final word was changed to form an alternative word that was not predicted by the sentence context (incongruent sentences: *She put the toys in the fox*). For half of these changed items, the onset was altered and for the remaining half we changed the offset. The length, frequency, phonological neighborhood density, and concreteness of target words were collected from the English Lexicon Project (Balota et al., 2007), and averages across sentence conditions are presented in **Table 1**. Target words did not differ significantly in terms of any of these lexical characteristics across sentence conditions, all ps > 0.05.

All sentences were recorded at 44,100 Hz and 16-bit resolution in a double-walled, sound-attenuating booth, and were spoken at a normal rate by a male with a Midwestern American accent. Periods of silence of different lengths were inserted at the start of each sentence so that the onset of the target word began at the same time on each trial to facilitate eye-tracking analyses. All sentences were played at an average amplitude of 64 dB sound pressure level (SPL).

Four images were gathered for each sentence for use in the visual world task: one depicting the target word (e.g., *box* for the sentence *She put the toys in the* box), one depicting a semantically unrelated phonological neighbor of the target word that acted as the target word in incongruent sentences (alternative image: *fox*), and two semantically unrelated foil words that did not sound like the target word (e.g., *key* and *paw*). Two foil images were included on each trial as opposed to including a second phonological neighbor or a semantic competitor of the target word so that participants would not divide their fixations across two images tapping the same source of information (semantic or phonological). This allowed for clearer assessment of how fixations were impacted by semantic congruency and phonological similarity. Each image was resized to 300 × 300 pixels. For images that did not have equal width and height, a white border was added to the shorter dimension to achieve the 300 × 300 size.

A pilot test was conducted to ensure that all images to be used in the visual world task were identifiable as the words they were meant to depict. Twenty younger adults participated in this pilot study. First, participants completed a study phase in which they saw each image along with the word the image was meant to depict for 2,000 ms. Participants then completed a test phase wherein each written word was presented at the center of the screen one at a time along with four images. One of the images depicted the written word at the center of the screen, and the other three images were the images paired with the initial image on trials in the SPIN task. For example, when the target word was fox, the four images on screen depicted a fox, a box, a key, and a paw, and the same four images were presented when the target word was box, key, and paw. The images were randomly assigned to one of the four quadrants of the screen. Participants were tasked with clicking on the image that depicted the word at the center of the screen. Average accuracy for identifying the image depicting each target word was 99.42%. In fact, only two images were identified correctly in less than 90% of cases, one of which was only used in the practice trials of the SPIN task (*joker*, Accuracy = 75%) and the other was a foil (*till*, Accuracy = 70%). Given the high identification accuracy of virtually all images in the pilot study and the similarity of the pilot study's procedure to that of the SPIN task, we felt confident that participants would associate each image with the word they were intended to depict in the SPIN task.

TABLE 1 | Means and standard deviations of lexical characteristics of target words across conditions.

	Baseline		**Congruent**		**Incongruent**	
	M	**SD**	***M***	**SD**	***M***	**SD**
Length	4.20	0.76	4.37	0.85	4.23	1.01
Frequency	33,266.53	77,434.77	28,985.50	44,531.67	32,123.90	44,336.55
Phono N	19.07	9.45	18.67	9.89	20.80	11.16
Concreteness	4.66	0.26	4.71	0.41	4.61	0.39

Phono N, Phonological neighborhood density.

A second pilot test was conducted to determine the SNR that would be needed for younger and older adults to achieve approximately 50% accuracy in the baseline condition of the SPIN task. We first tested younger adults at −4 dB SPL and older adults at +1 dB SPL, SNRs used in a similar SPIN task in a prior study (Failes et al., 2020). However, accuracy was at ceiling in the baseline condition for younger adults using the −4 dB SPL SNR ($M = 0.93$), and older adults' performance at the +1 dB SPL SNR was also higher than the desired 0.50 ($M = 0.73$). It was important to ensure that baseline performance was not too high so that participants had room to improve with the addition of congruent context. The high accuracy in the present study using the SNRs from Failes et al. (2020) was unsurprising given that our SPIN task was a four-alternative forced-choice test as opposed to the open-set response format used by Failes et al. It was determined that an SNR of −10 dB SPL for younger adults and −7 dB SPL for older adults would achieve approximately equal performance across groups with accuracy that left room for improvement from the baseline condition to the congruent condition.

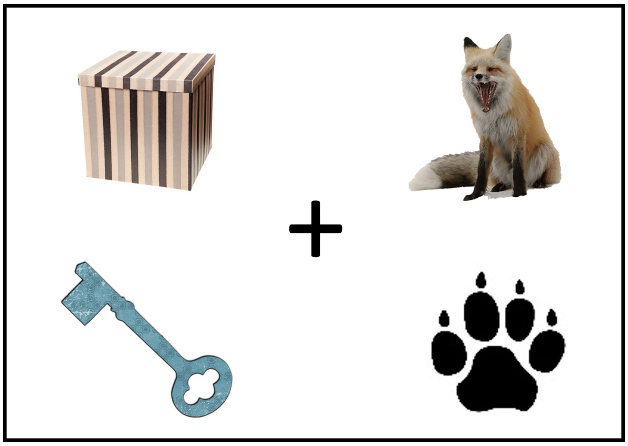


FIGURE 1 | Recreation of the example screen from the visual world task for the congruent sentence She put the toys in the box. For copyright reasons the original figure cannot be reproduced, but is available upon request to the corresponding author. Images sourced from stockvault.net.

Procedure

Participants first completed an audiogram inside a sound-attenuating booth. Following this, participants were seated at an EyeLink 1000 eye-tracking-enabled computer, where they completed the Shipley Vocabulary Test (Shipley, 1940) before beginning the visual world task. Participants placed their chins on a chinrest with their foreheads against a forehead rest to complete the visual world task. The distance from the back of the forehead rest to the eyepiece of the eye-tracker, which was positioned in front and below the computer monitor, was 52.07 cm and the distance from the back of the forehead rest to the center of the computer monitor was 57.78 cm. Participants first completed a study phase wherein each image to be shown in the visual world task was shown with the word the image was meant to depict for 2,000 ms to ensure that participants knew what each image represented. Participants then completed three practice trials, followed by 90 test trials, equally divided between the three sentence types (baseline/congruent/incongruent). Each trial began when the participant clicked on a central fixation cross. On each trial, a sentence was played through headphones with the final word in noise, and the four images associated with that sentence were presented on screen, one randomly assigned to each quadrant (see **Figure 1**). Participants were instructed to look at and click on the image corresponding to the word presented in noise. They were also told that they could move their eyes freely about the screen as long as the images were displayed but would only be able to click on an image once the sentence finished. Participants were specifically instructed that the sentence contexts could sometimes be misleading and were given examples (not included in the main study) of a sentence from each of the three conditions. Images remained on screen until the participant clicked on one of them. After clicking on an image, participants clicked on a number from one to five to indicate their confidence that they had selected the correct image, where one indicated a complete guess and five indicated absolute certainty.

The eye-tracker was calibrated immediately before test trials to ensure accurate eye-tracking. For the calibration task, dots appeared at 13 different locations on the screen and participants were tasked with fixating on each dot when it appeared and continuing to look at the dot until it disappeared. Participants completed the calibration task until it had been rated a "good" calibration by the EyeLink program, then completed an additional validation calibration to ensure that calibration was consistently accurate.

RESULTS

Accuracy

We first analyzed age differences in accuracy across the baseline, congruent, and incongruent conditions with mixed-effects logistic regression using the glmer function from the *lme4* package in R (Bates et al., 2015). The dependent variable in this model was trial-by-trial accuracy. The model included an intercept term corresponding to the odds of an accurate response for younger adults in the baseline condition, two dummy coded variables indicating the change in the odds of an accurate response from the baseline condition to the congruent and incongruent conditions for younger adults, a group variable representing the change in odds of an accurate response from younger to older adults in the baseline condition, and the interaction of group with the congruent and incongruent condition dummy codes to determine whether the change in the odds of an accurate response from the baseline condition to the congruent and incongruent conditions differed between younger and older adults. The odds of an accurate response in the baseline condition and the change in odds of an accurate response from the baseline to the congruent and incongruent conditions were allowed to vary randomly across subjects, and the odds of an accurate response were allowed to vary randomly across items (i.e., target words).

Average accuracy in the baseline, congruent, and incongruent conditions is presented in **Figure 2**. Younger adults displayed better baseline accuracy than older adults [*Odds Ratio* (*OR*) = 0.64, $z = -2.05$, $p < 0.05$], indicating that the SNR manipulation did not successfully equate performance in the baseline condition across groups. As expected, younger adults displayed improved performance in the congruent condition ($OR = 3.17$, $z = 2.67$, $p < 0.01$) and poorer performance in the incongruent condition ($OR = 0.29$, $z = -3.25$, $p < 0.01$) relative to the baseline condition. However, older adults experienced significantly greater benefit in the congruent condition ($OR = 6.62$, $z = 4.15$, $p < 0.001$) and significantly greater detriment to performance in the incongruent condition ($OR = 0.43$, $z = -2.63$, $p < 0.01$) relative to baseline than did younger adults. Although it may be argued that older adults' greater benefit in the congruent condition relative to younger adults could have resulted because younger adults had less room for improvement due to their better baseline performance, this explanation is unlikely since younger adults did not approach ceiling performance in the congruent condition (see **Figure 2**). These results replicate those from past studies of false hearing (Rogers et al., 2012; Sommers et al., 2015; Failes et al., 2020) and support the argument that older adults' performance was influenced more by available contextual cues than was that of younger adults.

False Hearing

To determine whether younger and older adults differed in susceptibility to false hearing, we created another mixed-effects logistic regression model predicting trial-by-trial false hearing on a subset of data that included only incongruent trials. The predictors in this model were an intercept term corresponding to the odds of false hearing in the younger adult group and an age group variable corresponding to the change in the odds of false hearing from the younger adult group to the older adult group. The odds of experiencing false hearing were allowed to vary randomly across subjects and across items.

As shown in **Figure 2**, both younger and older adults experienced false hearing. The odds that a younger adult would experience false hearing on incongruent trials was equivalent to the odds of giving any other possible response on incongruent trials (i.e., a correct response or erroneously choosing one of the foils) combined ($OR = 0.98$, $z = -0.05$, $p > 0.05$). The odds that an older adults would experience false hearing were approximately four times greater than the odds for younger adults, which was significant ($OR = 4.04$, $z = 3.47$, $p < 0.001$).

We also analyzed the odds of experiencing false hearing with maximum confidence—referred to in past studies as *dramatic false hearing* (Rogers et al., 2012; Sommers et al., 2015) —in younger and older adults (see **Figure 2**). The odds that a younger adult would experience dramatic false hearing was far less than the odds of giving any other possible response in the incongruent condition ($OR = 0.03$, $z = -8.14$, $p < 0.001$). Similar to previous studies (Rogers et al., 2012; Sommers et al., 2015), the odds that older adults would experience dramatic false hearing was more than 10 times greater than for younger adults ($OR = 10.46$, $z = 4.97$, $p < 0.001$). Thus, despite having an image depicting the correct target word presented on screen, both younger and older adults incorrectly reported hearing the contextually predicted word in over 50% of incongruent sentences, with older adults doing so more often and being far more likely to report maximum confidence in these errors than younger adults.

Confidence

Accurate Responses

To determine whether sentence condition and age group differences existed for confidence in accurate responses, we created a linear mixed-effects regression model using the lmer function from the *lme4* package in R (Bates et al., 2015) using a subset of data that included only accurate responses. As in the accuracy analyses, the model included an intercept term corresponding to confidence in accurate responses for

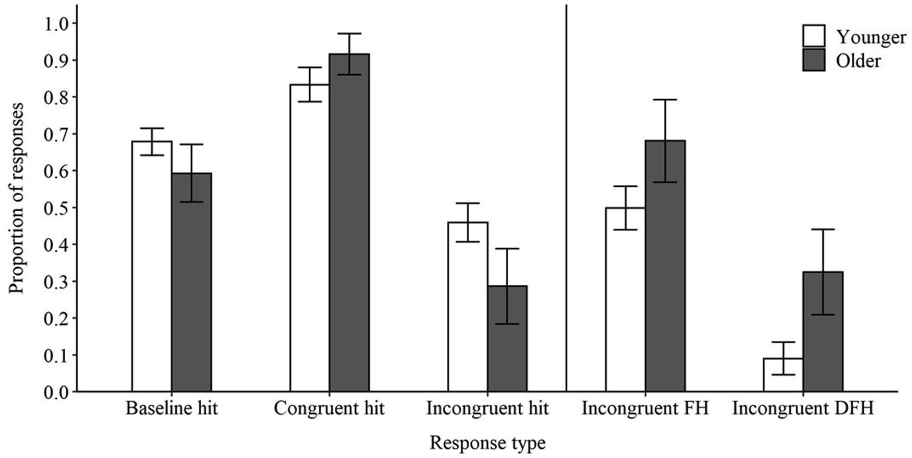


FIGURE 2 | Average proportion of hits (left side) and susceptibility to false hearing (FH; right side) and dramatic false hearing (DFH) for younger and older adults. Error bars represent 95% confidence intervals.

younger adults in the baseline condition, two dummy coded variables indicating the change in confidence from the baseline condition to the congruent and incongruent conditions for younger adults, a group variable representing the change in confidence from younger to older adults in the baseline condition, and the interaction of group with the congruent and incongruent condition dummy codes to determine whether the change in confidence from the baseline condition to the congruent and incongruent conditions differed between younger and older adults. Confidence in the baseline condition and changes in confidence from the baseline condition to the congruent and incongruent conditions were allowed to vary randomly across subjects, and confidence was allowed to vary randomly across items.

Average confidence in accurate responses (hits) in the baseline, congruent, and incongruent conditions is presented in **Figure 3**. Younger adults expressed confidence slightly above a neutral rating for accurate responses in the baseline condition, with the model estimating an average confidence of 3.53 out of 5. Younger and older adults' confidence did not differ in the baseline condition [*Estimated Difference* (*ED*) = 0.06, $t = 0.32$, $p > 0.05$]. Younger adults' confidence for accurate responses did not differ in either the congruent condition ($ED = 0.10$, $t = 0.43$, $p > 0.05$) or the incongruent condition ($ED = -0.23$, $t = -1.10$, $p > 0.05$) relative to the baseline condition. Additionally, the difference in confidence between the baseline and incongruent conditions did not differ in older adults relative to younger adults ($ED = -0.02$, $t = -0.18$, $p > 0.05$). However, there was a significant interaction suggesting that older adults' confidence increased to a greater degree than that of younger adults from the baseline condition to the congruent condition ($ED = 0.39$, $t = 2.18$, $p < 0.05$). Overall, these findings suggest that participants' confidence in accurate responses remained quite stable regardless of the context condition, aside from higher confidence in the congruent condition by older, relative to younger, adults. This differs from in past studies (Rogers et al., 2012; Sommers et al., 2015) where both younger and older adults have demonstrated lower confidence in accurate responses in the baseline condition than in the congruent condition. It is possible that changing to a four-alternative forced-choice paradigm, rather than the open-set response format used in the earlier studies, resulted in the consistently high confidence in accurate responses in the present study. For example, if a participant thought they heard the word *box*, they might be more confident in that response because an image of a box was among the four options on screen. Therefore, the availability of images on the screen may have increased confidence in accurate perceptions.

False Hearing Responses

We then conducted a second linear mixed-effects regression analysis to determine whether younger and older adults differed in their confidence in cases of false hearing. This model was conducted on a subset of data that included only cases of false hearing on incongruent trials. The model included an intercept term corresponding to younger adults' confidence in cases of false hearing and a group variable indicating the change in confidence from younger to older adults. Confidence in cases of false hearing was allowed to vary randomly across subjects and items.

Although there was little difference between age groups for confidence in accurate responses, younger and older adults did differ in their confidence in cases of false hearing (see **Figure 3**). Younger adults expressed approximately neutral confidence in cases of false hearing, with the model estimating an average confidence of 3.12 out of 5. Older adults' estimated average confidence in cases of false hearing was 3.93, which was significantly higher than the confidence displayed by younger adults ($t = 4.08$, $p < 0.001$). Thus, older adults were both more susceptible to and more confident in cases of false hearing than were younger adults.

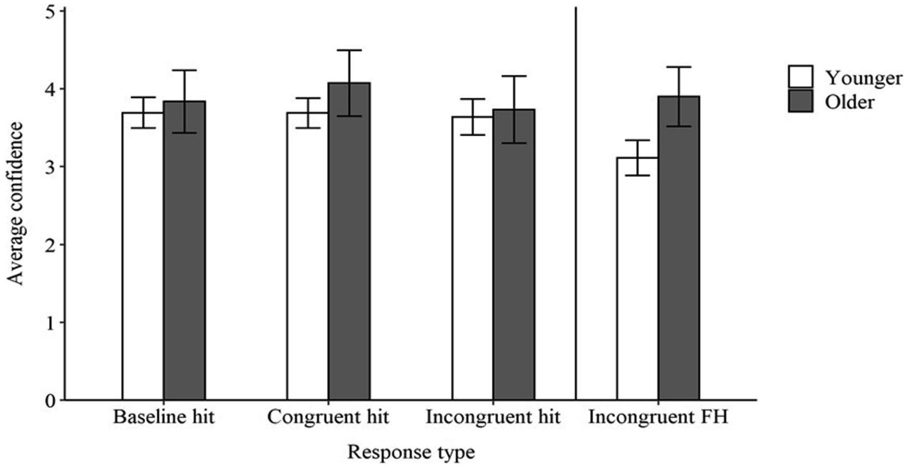


FIGURE 3 | Average confidence in hits in the baseline, congruent, and incongruent conditions, and in cases of false hearing (FH) in the incongruent conditions for younger and older adults. Error bars represent 95% confidence intervals.

Fixation Analyses

To determine changes in the proportion of fixations on each image across time in the visual world task, linear mixed-effects regression was used to analyze the proportion of fixations on each image following the analyses used in a recent eye-tracking study that employed a similar visual world paradigm (Ito et al., 2018). Fixations on locations of the screen other than one of the four images (e.g., on the fixation cross) were not included when calculating the proportion of fixations on each image. Separate analyses were conducted for each sentence type (baseline, congruent, incongruent). Both accurate and inaccurate responses were included in analyses unless otherwise noted. Sentences were divided into three 2,000-ms bins for fixation analyses. The first bin started from 500 ms after the start of the trial since there were very few fixations on any of the images before this time (participants tended to still be looking at the central fixation cross). The second time bin started 2,500 ms into the trial and continued until just before the target word was presented. The third time bin started 4,500 ms into the trial, exactly when the target word onset. These three bins allowed us to determine the proportion of fixations on each image early in the sentence (bin 1), late in the sentence but before the target word was presented (bin 2), and from the presentation of the target word onwards (bin 3). Although other analytic approaches, such as growth-curve analysis, are often used to examine the time course of spoken word recognition in the visual world paradigm, they can sometimes obscure important changes at specific time points, such as the transition between preceding context and target word. We therefore elected to bin the eye-tracking data as described to facilitate comparisons at specific points in the sentence using the raw eye-tracking data. Each mixed-effects model had 24 dummy coded variables corresponding to the four picture types (target image, alternative image, and two foil images) within each age group (younger/older adults) at each time bin as fixed effects predicting the proportion of fixations. Proportion of fixations was allowed to vary randomly across subjects and items for each image type within each time bin. Linear combinations of the fixed effects were tested using the *multcomp* package in R (Hothorn et al., 2008) to determine how the proportion of fixations on each image changed over time, whether the proportion of fixations differed across image types within time bins, and whether these effects differed across age groups. Since the target image and the alternative image (i.e., the contextually predicted image in incongruent sentences) were of primary interest to our hypotheses, we will focus on fixation trends for these image types in the Results section. Analyses pertinent to the two foil images are presented in the Supplementary Material.

Baseline Sentences

The proportion of fixations over time for baseline sentences is presented in **Figure 4**. For baseline sentences, we predicted that fixations on the target image would not increase until time bin 3 since there were no contextual cues in baseline sentences to afford anticipatory activation to any particular response. This prediction was supported by the fixation analysis. In baseline sentences, there was no difference in fixations on the target image from time bin 1 to time bin 2 ($ED = 0.03$, $z = 1.48$, $p > 0.05$), but fixations on the target image increased from time bin 2 to time bin 3 ($ED = 0.27$, $z = 12.80$, $p < 0.001$). There was no interaction with age group for the difference from time bin 1 to time bin 2 ($ED = 0.00$, $z = 0.08$, $p > 0.05$) or from time bin 2 to time bin 3 ($ED = 0.02$, $z = 0.74$, $p > 0.05$). Therefore, as predicted, both younger and older adults only increased fixations on the target image after the target word had been presented in baseline sentences, demonstrating that neither group experienced anticipatory activation of the target word when no contextual cues were present.

We next compared the relative proportion of fixations on the target image and the alternative image within each time bin. Participants demonstrated a greater proportion of fixations on the alternative image than the target image in time bins 1 ($ED = -0.18$, $z = -9.31$, $p < 0.001$) and 2 ($ED = -0.10$, $z = -4.84$, $p < 0.001$), but the opposite was true in time bin 3 ($ED = 0.20$, $z = 9.74$, $p < 0.001$). There were no interactions with age group for differences in fixations between the target and alternative images ($ps > 0.05$). This suggests that although the alternative word was activated to a greater degree than the target word before the target word was presented—potentially due to minor differences in word frequency between the target word and alternative word (see Luce and Pisoni, 1998; Dahan et al., 2001; Dahan and Gaskell, 2007; Revill and Spieler, 2012) or differences in characteristics of the target image and alternative image themselves—the target word became more highly activated than the alternative word once the target word was presented.

Congruent Sentences

For congruent sentences, we predicted that both age groups would begin looking toward the contextually predicted target image before the target word was presented. Additionally, we predicted that older adults might increase their fixations on the target image to a greater degree than younger adults, reflecting increased influence of context over responding. As can be seen in **Figure 5**, these predictions were mostly supported by the fixation data. Fixations on the target image increased from time bin 1 to time bin 2 ($ED = 0.25$, $z = 12.98$, $p < 0.001$) and again from time bin 2 to time bin 3 ($ED = 0.26$, $z = 12.81$, $p < 0.001$). Whereas there was no interaction between age group and the change in fixations from time bin 1 to time bin 2 ($ED = -0.01$, $z = -0.27$, $p > 0.05$), there was an interaction with age group for the change in fixations from time bin 2 to time bin 3 ($ED = -0.15$, $z = -7.17$, $p < 0.001$). For younger adults, there was a significant increase in proportion of fixations on the target image from time bin 2 to time bin 3 ($ED = 0.06$, $z = 4.13$, $p < 0.001$), but this increase was significantly greater for older adults ($ED = 0.20$, $z = 13.69$, $p < 0.001$). Thus, both younger and older adults increased fixations on the target image before the target word was presented, demonstrating anticipatory activation of the target word based on available contextual cues. Older adults increased fixations on the target image to a greater degree than younger adults, but only after the target word had been presented. This suggests that younger

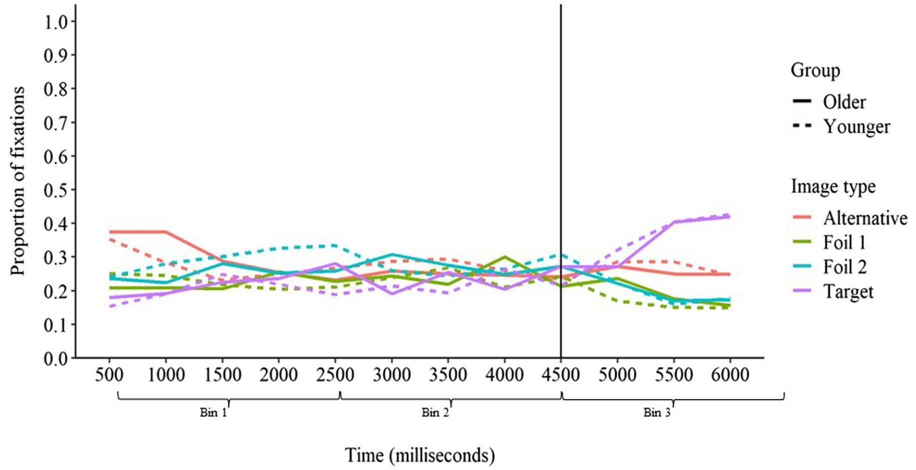


FIGURE 4 | Proportion of fixation on each image type over time in baseline sentences for younger and older adults. The vertical line at 4,500 ms represents the onset of the target word.

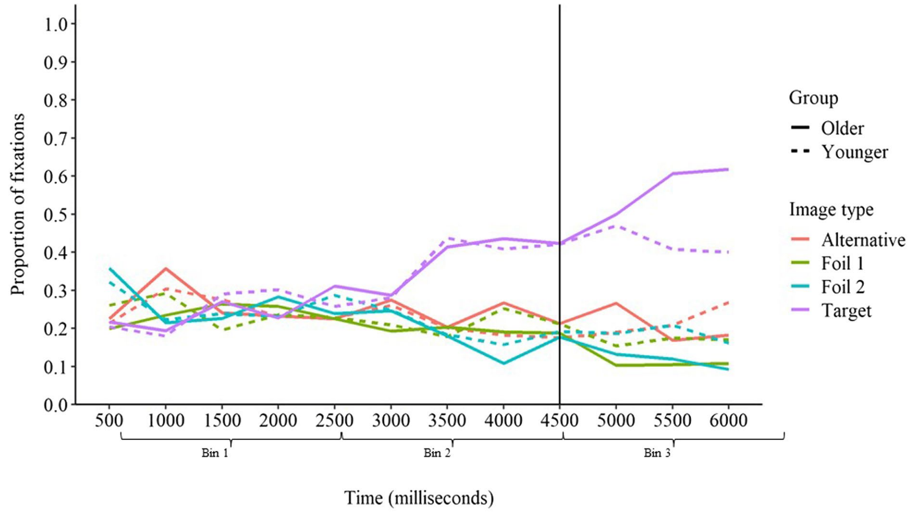


FIGURE 5 | Proportion of fixations on each image type over time in congruent sentences for younger and older adults. The vertical line at 4,500 ms represents the onset of the target word.

and older adults formed context-based expectations at a similar rate, but older adults became more fixated on the contextually predicted response once additional support for this response was provided by presentation of the target word. The greater increase in fixations on the target image after the target word was presented by older, relative to younger, adults suggests that older adults may use the auditory signal to confirm their context-based expectations. When the auditory signal supported the word they expected to hear, older adults become increasingly fixated on that response option, whereas younger adults may have been more cautious and considered alternative options.

As was true in baseline sentences, the alternative image received a greater proportion of fixations than the target image in time bin 1 ($ED = -0.06$, $z = -3.20$, $p < 0.01$). However, as sentence context continued to be introduced in bin 2, the target image began to receive a greater proportion of fixations than the alternative image ($ED = 0.27$, $z = 13.66$, $p < 0.001$). The target image's advantage in terms of fixations increased further in time bin 3 ($ED = 0.55$, $z = 26.89$, $p < 0.001$). Although the difference in fixations between the target and alternative images did not interact with age group for time bin 1 or 2 ($ps > 0.05$), there was a significant interaction for time bin 3 ($ED = -0.21$, $z = -10.02$, $p < 0.001$). This interaction indicated that although younger adults devoted a greater proportion of fixations to the target image relative to the alternative image in time bin 3 ($ED = 0.17$, $z = 12.12$, $p < 0.001$), this difference was substantially greater for older adults ($ED = 0.38$, $z = 25.71$, $p < 0.001$). These findings indicate that although the alternative word was initially more highly activated than the target word, the activation of the target word exceeded that of the alternative word as support for that word was introduced *via* semantic context and the phonological characteristics of the target word.

Incongruent Sentences

Finally, for incongruent sentences, we had again predicted that both younger and older adults would look toward the contextually predicted image before the target word was presented. In incongruent sentences, the alternative word was predicted by context as opposed to the target word, so we predicted that fixations on the alternative image would increase before the target word was presented. Here again we predicted that older adults would increase fixations on the alternative image to a greater degree than younger adults, reflecting greater influence of context over responding. After the target word was presented, we predicted that fixations on the alternative image would decrease and fixations on the target image would increase for younger adults as they realized that the presented target word differed from the contextually predicted word. However, we predicted that older adults would be less likely than younger adults to shift their focus toward the target image, instead either maintaining or increasing fixations on the alternative image, reflecting older adults' poorer ability to suppress the expected response and increased susceptibility to false hearing in incongruent sentences.

Average fixations over time in incongruent sentences can be seen in **Figure 6**. As predicted, the proportion of fixations on the alternative image increased from time bin 1 to time bin 2 ($ED = 0.17$, $z = 8.87$, $p < 0.001$), whereas fixations on the target image did not increase ($ED = -0.02$, $z = -1.07$, $p > 0.05$). There was no interaction with age group for the difference between time bin 1 and time bin 2 for fixations on the alternative image ($ED = -0.01$, $z = -0.70$, $p > 0.05$), but there was a marginally significant interaction for fixations on the target image ($ED = -0.04$, $z = -1.92$, $p = 0.05$). This interaction reflected that younger adults significantly decreased fixations on the target image from time bin 1 to time bin 2 ($ED = -0.03$, $z = -2.24$, $p < 0.05$) whereas older adults' fixations on the target image did not change ($ED = 0.01$, $z = 0.57$, $p > 0.05$). Interestingly, fixations increased from time bin 2 to time bin 3 for both the alternative image ($ED = 0.04$, $z = 2.03$, $p < 0.05$) and the target image ($ED = 0.16$, $z = 7.65$, $p < 0.001$). However, as predicted, there were significant interactions with age group for the difference in fixations from time bin 2 to time bin 3 for both the alternative image ($ED = -0.14$, $z = -6.90$, $p < 0.001$) and the target image ($ED = 0.12$, $z = 5.88$, $p < 0.001$). Younger adults decreased fixations on the alternative image ($ED = -0.05$, $z = -3.51$, $p < 0.001$) and increased fixations on the target image ($ED = 0.14$, $z = 9.75$, $p < 0.001$) from time bin 2 to time bin 3. Older adults, however, increased fixations on the alternative image ($ED = 0.09$, $z = 6.20$, $p < 0.001$) and did not increase fixations on the target image ($ED = 0.02$, $z = 1.23$, $p > 0.05$) from time bin 2 to time bin 3. This supports the conclusion that older adults were less able to suppress the activation of the predicted response in the incongruent condition than were younger adults.

As was true for both baseline and congruent sentences, the alternative image initially received a greater proportion of fixations than the target image ($ED = -0.12$, $z = -6.27$, $p < 0.001$). This difference increased in time bin 2 as contextual cues supporting the alternative image were introduced ($ED = -0.31$, $z = -15.74$, $p < 0.001$). There was no interaction with age group for the difference between the alternative and target images in time bin 1 or 2 ($ps > 0.05$). The alternative image continued to receive a greater proportion of fixations, on average, relative to the target image in time bin 3 ($ED = -0.20$, $z = -9.56$, $p < 0.001$), but this was qualified by a significant interaction with age group ($ED = 0.26$, $z = 12.44$, $p < 0.001$). Although the alternative image did indeed receive a greater proportion of fixations than the target image in time bin 3 for older adults ($ED = -0.23$, $z = -15.61$, $p < 0.001$), the opposite was true for younger adults, with the target image receiving a greater proportion of fixations ($ED = 0.03$, $z = 2.02$, $p < 0.05$). This suggests that younger adults were better able to reduce the activation of the contextually predicted (but incorrect) word below that of the unpredicted target word than were older adults on incongruent trials, which may help to explain why older adults were more susceptible to false hearing.

To better understand how anticipatory activation of the contextually predicted word contributed to the likelihood of false hearing, we next compared changes in fixations on the target and alternative images over time for accurate responses and false hearing responses on incongruent trials. Average fixations on each image type for younger and older adults for accurate and false hearing responses are displayed in **Figure 7**.

The change in fixations on the alternative image from bin 1 to bin 2 differed significantly across accurate and false hearing trials ($ED = -0.19$, $z = -3.17$, $p < 0.01$). Fixations on the alternative image did not change from bin 1 to bin 2 for accurate trials ($ED = -0.01$, $z = -0.28$, $p > 0.05$), whereas fixations on the alternative image increased substantially from bin 1 to bin 2 for false hearing trials ($ED = 0.18$, $z = 4.27$, $p < 0.001$). This finding is interesting, as it suggests that there was relatively little anticipatory activation of the contextually predicted word on incongruent trials in which participants responded accurately relative to those in which false hearing occurred. The proportion of fixations on the target image did not change from bin 1 to bin 2 ($ED = -0.06$, $z = -0.91$, $p > 0.05$) and there was no difference in the change in fixations on the target image from bin 1 to bin 2 across accurate and false hearing trials ($ED = 0.01$, $z = 0.22$, $p > 0.05$).

Unsurprisingly, the change in fixations on the alternative image from bin 2 to bin 3 also differed significantly across accurate and false hearing trials ($ED = -0.46$, $z = -7.50$, $p < 0.001$). Fixations on the alternative image decreased from bin 2 to bin 3 for accurate trials ($ED = -0.20$, $z = -4.48$, $p < 0.001$), whereas fixations on the alternative image increased from bin 2 to bin 3 for false hearing trials ($ED = 0.26$, $z = 6.16$, $p < 0.001$). The change in fixations on the target image from bin 2 to bin 3 also differed significantly across accurate and false hearing trials ($ED = 0.51$, $z = 8.38$, $p < 0.001$). Fixations on the target image increased greatly from bin 2 to bin 3 for accurate trials ($ED = 0.42$, $z = 9.56$, $p < 0.001$), but decreased from bin 2 to

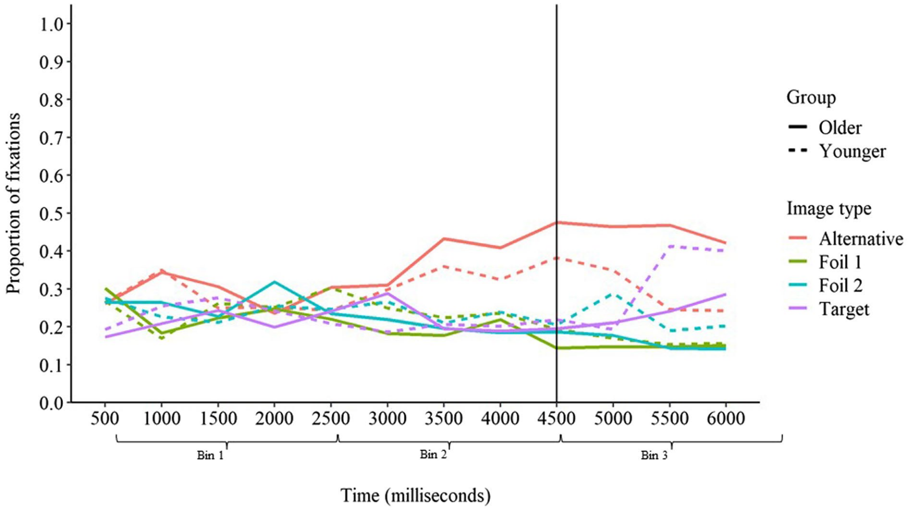


FIGURE 6 | Proportion of fixations on each image type over time in incongruent sentences for younger and older adults. The vertical line at 4,500 ms represents the onset of the target word.

bin 3 for false hearing trials ($ED = -0.09$, $z = -2.19$, $p < 0.05$). All interactions with age group were non-significant when considering accurate and false hearing trials separately ($ps > 0.05$).

It is interesting to note that fixations on the target image did not reach 100% on accurate trials, nor did fixations on the alternative image reach 100% on trials in which false hearing occurred. This may have resulted for several reasons. It is possible that this reflects uncertainty as to the identity of the target word, with participants continuing to consider other response options in time bin 3. It is also possible that participants might approach 100% fixations on the target image if we used shorter time bins. We analyzed our data within 2,500 ms time bins to increase statistical power while still being able to address our research questions, but the extended duration of these bins might capture fixations on other response options leading up to settling on a single image. However, even in **Figure 7**, which presents the fixation data in 500 ms bins, neither age group approaches 100% fixations on the target image. Thus, it is possible that fixations on the image corresponding to the participants' response could have reached 100%, but the short duration of the time bins required to achieve this would have greatly increased the complexity of our models without improving our ability to address our research questions.

For the difference in proportion of fixations between the target image and the alternative image, there were marginally significant interactions indicating differences between accurate and false hearing trials in bins 1 ($ED = -0.10$, $z = -1.70$, $p = 0.09$) and 2 ($ED = 0.10$, $z = 1.68$, $p = 0.09$), and a significant interaction in bin 3 ($ED = 1.08$, $z = 17.76$, $p < 0.001$). Beginning with accurate trials, the alternative image received a greater proportion of fixations in both time bins 1 ($ED = -0.17$, $z = -3.85$, $p < 0.001$) and 2 ($ED = -0.18$, $z = -4.00$, $p < 0.001$), but the opposite was true in time bin 3 ($ED = 0.44$, $z = 10.11$, $p < 0.001$). On incongruent trials in which false hearing occurred, the alternative image did not differ from the target image in fixations in time bin 1 ($ED = -0.06$, $z = -1.53$, $p > 0.05$). The alternative image did, however, receive more fixations than the target image in time bin 2 ($ED = -0.28$, $z = -6.58$, $p < 0.001$), and as indicated by the significant interaction mentioned above, this difference was greater than that in bin 2 of accurate trials. Unlike for accurate trials, the alternative image also received a greater proportion of fixations than the target image in time bin 3 ($ED = -0.64$, $z = -15.08$, $p < 0.001$). Additionally, although there were no interactions with age group for any of the effects mentioned above ($ps > 0.05$), the difference in proportion of fixations between the alternative image and the target image differed in magnitude across age groups for bin 3 of trials in which false hearing occurred ($ED = 0.15$, $z = 3.50$, $p < 0.001$). Specifically, the alternative image received a greater proportion of fixations relative to the target image for younger adults ($ED = -0.24$, $z = -8.35$, $p < 0.001$), but this difference was greater for older adults ($ED = -0.39$, $z = -12.89$, $p < 0.001$). These findings are particularly important, as they suggest that participants were able to reduce the activation of the contextually predicted (but incorrect) word below that of the unpredicted target word on accurate trials but were unable to do so for trials in which false hearing occurred.

DISCUSSION

The current study was designed to provide a more direct test of the inhibitory deficit hypothesis of false hearing using eye-tracking as an online measure of lexical activation. Overall, the findings suggest that false hearing occurs when participants fail to reduce activation on semantically congruent, but incorrect, lexical items that are phonologically similar to target words. Using eye-tracking, we found that fixations, an index of lexical activation, increased on contextually congruent words for both

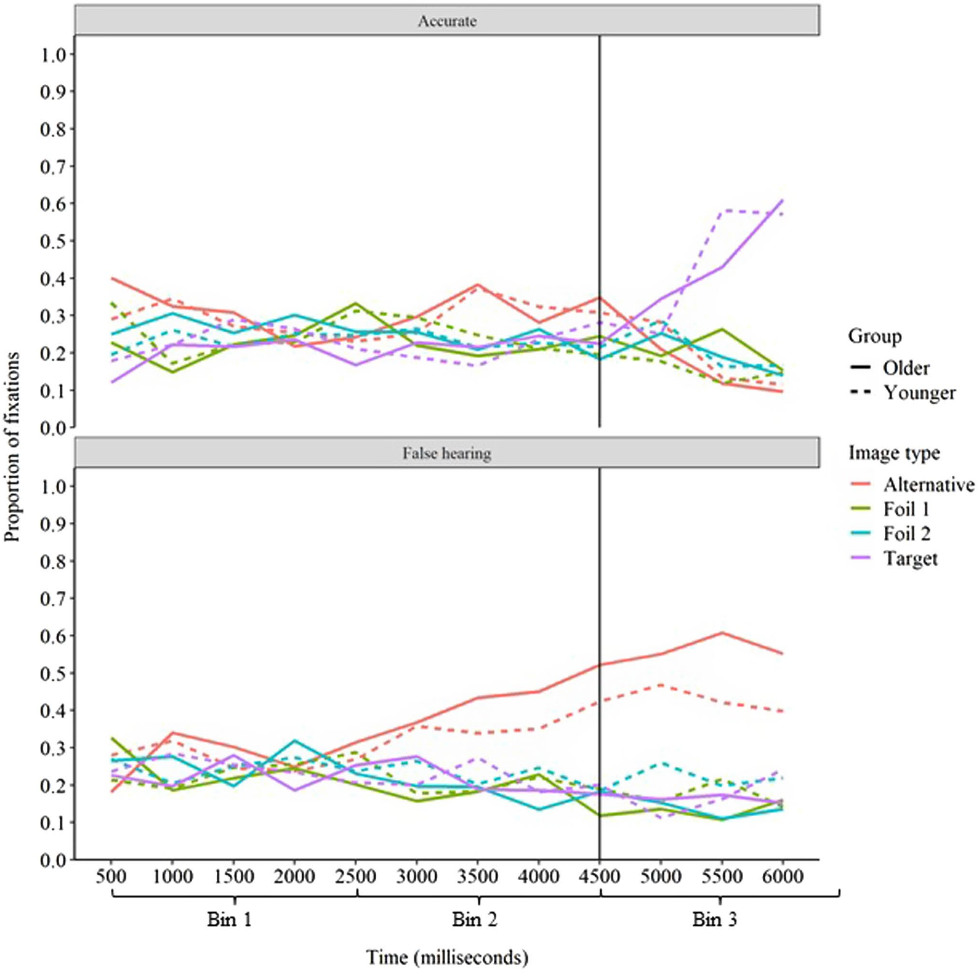


FIGURE 7 | Proportion of fixations on each image type over time in incongruent sentences for younger and older adults on accurate and false hearing. The vertical line at 4,500 ms represents the onset of the target word.

younger and older adults prior to actual target word onset. When semantically congruent items were presented as targets (congruent condition), fixations to the target increased following target word onset and did so to a greater extent for older than for younger adults. Activation of the phonologically similar but semantically incongruent foil in this condition remained low for both older and younger adults. However, when a phonologically similar alternative was the target (incongruent condition), younger adults and older adults again both increased fixations to the semantically congruent (but now incorrect) item prior to target word onset. Once the target word was presented, however, younger adults were more likely than older listeners to switch fixations to the correct, but semantically incongruent target item, whereas older adults maintained high levels of fixation on the incorrect alternative. Furthermore, we found that early activation of the contextually predicted word in incongruent sentences was suppressed below the activation of the target word in cases where participants accurately reported the target word, whereas activation of the contextually predicted, but incorrect, word remained higher than the target word in cases of false hearing. Taken together, the findings suggest that older adults' increased susceptibility to false hearing stems in part from a reduced ability to inhibit the anticipated, but incorrect, response on incongruent trials (Rogers et al., 2012; Sommers et al., 2015; Failes et al., 2020).

To our knowledge, the current study is the first investigation of false hearing with sentence materials to use a closed-set response format (where the target item was presented as one of four alternatives) rather than open-set responding. This change from open- to closed-set responding resulted in a substantially easier task for both older and younger adults as indicated by the necessity of using more difficult SNRs than those used by Failes et al. (2020) to obtain similar baseline accuracy. Nevertheless, the findings largely replicate prior studies of age-related differences in false hearing. Relative to baseline (non-predictive) sentences, older adults were more accurate than younger adults in congruent conditions—a finding that is extremely rare in studies comparing speech in noise perception in younger and older adults—but significantly less accurate than younger adults in the incongruent condition. Taken together, the results demonstrate the robust effects of context on speech perception; even when the right answer was presented to participants in an image, younger adults still chose the contextually predicted (but incorrect) option on approximately 50% of incongruent trials, whereas older adults committed these context-based errors on approximately 70% of incongruent trials.

The findings from the current study provide more direct evidence than in previous studies (Sommers et al., 2015; Failes et al., 2020) that false hearing occurs when listeners are unable to sufficiently inhibit the activation of a contextually predicted, but incorrect, target word. Increased fixations on the contextually predicted image prior to target onset in both congruent and incongruent sentences suggests that the contextually predicted word became activated even before the target word was presented. For congruent sentences, this early target activation led to high accuracy rates for both age groups, with marginally better accuracy for older than for younger adults. Although there is some evidence that older adults benefit more than younger adults from supportive semantic contexts (see Pichora-Fuller, 2008 for review) it is atypical to find better performance for older than for younger adults under conditions of equal audibility. In the present study, this finding in the congruent condition indicates greater reliance on context-based responding for older than for younger adults, a proposal that is further supported by the eye-tracking results showing a significantly greater proportion of fixations to the target image in bin 3 in the congruent condition for older than younger adults.

Support for the role of inhibitory control as a factor contributing to false hearing comes from both accurate and inaccurate responding on the incongruent trials. When participants were correct on incongruent trials, both younger and older adults were able to reduce the activation of the contextually predicted word below that of the target word. In contrast, when participants selected the (inaccurate) alternative on incongruent trials (i.e., exhibited false hearing), fixations on that image did not fall below those for the actual target word, suggesting that participants failed to inhibit activation on the alternative item. Therefore, the eye-tracking data in the present study provide the first direct evidence for the theory that false hearing results when individuals are unable to reduce the activation of a contextually predicted (but incorrect) response below the activation of the correct response.

The eye-tracking data in the present study may also shed light on what aspects of inhibitory control contribute to false hearing. Previous studies have suggested that there are three distinct functions of inhibitory control: controlling what information enters working memory, removing information that is no longer relevant from working memory, and suppressing prepotent, but incorrect, responses (Kramer et al., 1994; Zacks and Hasher, 1997; Lustig et al., 2007). We will refer to these as the information gating, information removal, and response suppression aspects of inhibitory control, respectively. When looking at the eye-tracking data collapsed across all incongruent trials (see **Figure 6**), our data suggest that age differences in susceptibility to false hearing may be influenced primarily by either the response suppression or information removal aspects of inhibitory control. If the information gating aspect of inhibitory control contributed to age differences in susceptibility to false hearing, then we would expect younger and older adults to differ in the degree to which the contextually predicted response became activated before the target word was presented, indexing differences in the ability to control contextual information entering working memory. Instead, we found that the proportion of fixations to the contextually predicted image in the incongruent condition differed between younger and older adults only after the target word was presented. This suggests that contextual information was allowed to enter working memory to a similar degree across age groups, but older adults were either less able to remove this task-irrelevant information or were less able to suppress the contextually predicted response than were younger adults.

Recently, Van Os et al. (2021) have argued against the contribution of cognitive declines, particularly inhibitory abilities, as contributing to either age or individual differences in false hearing. Specifically, using open-set responding and sentence materials similar to those of Sommers et al. (2015), Van Os et al. reported that both older and younger adults misperceived sentence-final items as consistent with a preceding context. In contrast to prior studies of false hearing (Sommers et al., 2015; Failes et al., 2020), however, they did not report high confidence in those misperceptions. Moreover, Van Os et al. reported that the frequency of such misperceptions varied as a function of how well participants could access acoustic information; conditions that increased access to the information in speech signals, such as more favorable SNRs and easier phonetic discriminations, led to fewer misperceptions than those where listeners had less access to such information. Consequently, they argued that responding (incorrectly) on the basis of context should be considered a form of rational language comprehension in which the relative weighting of top-down (context) and bottom-up (acoustic) information varies rationally. When access to acoustic information is limited, listeners place greater emphasis on context. Conversely, when acoustic information is readily available (e.g., listening in quiet), listeners appropriately place greater emphasis on acoustic cues as a basis for responding.

In contrast to claims made by Van Os et al. (2021), however, finding that mishearing (the term used by Van Os et al., 2021) or false hearing (the term used in prior studies; Rogers et al., 2012; Sommers et al., 2015; Failes et al., 2020) varies inversely with access to acoustic cues is entirely consistent with an inhibition-based account of false hearing. Recall that activation-competition models of spoken word recognition (Luce and Pisoni, 1998; Gaskell and Marslen-Wilson, 2002) propose that activation of a particular lexical item is directly proportional to the match between the incoming speech signal and stored lexical representations. When access to the acoustic signal is readily available (e.g., favorable SNRs, normal hearing thresholds), the target item will receive considerably higher activation than other lexical competitors and demands on inhibition will be minimal. In contrast, under more difficult listening conditions, activation of targets and competitors will be more similar, and correct identification will require increased inhibition of competitors. Thus, both rational comprehension and inhibitory deficit accounts would predict increased false hearing (or misperceptions) with less favorable SNRs and more difficult phonetic discriminations –exactly the pattern reported by Van Os et al. (2021) and by Rogers et al. (2012) in their study of false hearing as a function of SNR.

In fact, we see the findings from Van Os et al. (2021) as complementary to the current results. The rational comprehension account suggests that age differences in misperceptions or false hearing result because older adults, rationally, rely on context to a greater extent than younger adults. The eye-tracking results in the current study are consistent with this proposal; despite explicit warnings that context could be misleading and a greater percentage of trials where it was misleading, older adults fixated on the semantically consistent item more than younger adults in both the congruent and incongruent conditions. Our proposal regarding the role of inhibition simply extends this account by proposing that younger adults are better able than older adults to inhibit this activation on incorrect, but semantically congruent, items than are older adults, resulting in a greater frequency of false hearing.

One limitation of the current study is that it did not include direct measures of inhibition or other cognitive abilities that may be important contributors to individual differences in the costs and benefits of semantic context. Huettig and Janse (2016), for example, found that individual differences in both working memory and processing speed were predictive of anticipatory eye movements in a visual world paradigm examining the benefits of gender marking of articles that preceded a target word in Dutch. An important direction for future studies examining the contributions of different types of supportive contexts, including semantic information and gender marking, is to include a battery of cognitive measures that can be used to establish factors that contribute to both age and individual differences in the ability to benefit from such information.

CONCLUSION

The present study shed light on the processing that underlies false hearing and differences in this processing across younger and older adults. Specifically, we presented evidence that false hearing occurs when participants fail to inhibit an incorrect response that is highly activated due to its congruence with available contextual cues. Furthermore, our results advance our understanding of sentence processing, generally. We have shown that both younger and older adults form expectations regarding what will be said in the future based on preceding semantic context, expectations which are used to fill in the blanks in speech perception caused by difficult listening conditions (e.g., background noise, hearing loss). Although contextual cues in natural speech are rarely misleading—making responding based on context an efficient speech perception strategy—our demonstration of robust rates of false hearing in both younger and older adults highlight the importance of carefully attending to how information is presented. When important information is framed by valid contextual cues, both younger and older adults may more easily hear the important information. If preceding contextual cues are misleading, however, both younger and older adults –although especially older adults– may be led astray.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Washington University School of Medicine. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

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

ACKNOWLEDGMENTS

We would like to acknowledge the hard work of the research assistants in the Speech & Hearing Lab at Washington University in St. Louis who assisted with this project: Liam Gibbs, who recorded the auditory stimuli used in the SPIN task, and Adam Isaacs, Aryn Lyke, Audrey Ulfers, Davis Holmes, and Forrest Whiting, who aided in data collection.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2022.821044/full#supplementary-material

REFERENCES

Allopenna, P. D., Magnuson, J. S., and Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: evidence for continuous mapping models. *J. Mem. Lang.* 38, 419–439. doi: 10.1006/jmla.1997.2558

Ayasse, N. D., and Wingfield, A. (2020). The two sides of linguistic context: eye-tracking as a measure of semantic competition in spoken word recognition among younger and older adults. *Front. Hum. Neurosci.* 14:132. doi: 10.3389/fnhum.2020.00132

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., et al. (2007). The English lexicon project. *Behav. Res. Methods* 39, 445–459. doi: 10.3758/BF03193014

Bates, D., Maechler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01

Benichov, J., Cox, L. C., Tun, P. A., and Wingfield, A. (2012). Word recognition within a linguistic context: effects of age, hearing acuity, verbal ability and cognitive function. *Ear Hear.* 33, 250–256. doi: 10.1097/AUD.0b013e31822f680f

Bilger, R. C., Nuetzel, J. M., Rabinowitz, W. M., and Rzeczkowski, C. (1984). Standardization of a test of speech perception in noise. *J. Speech Hear. Res.* 27, 32–48. doi: 10.1121/1.2017541

Cohn, N. B., Dustman, R. E., and Bradford, D. C. (1984). Age-related decrements in Stroop color test performance. *J. Clin. Psychol.* 40, 1244–1250. doi: 10.1002/1097-4679(198409)40:5<1244::AID-JCLP2270400521>3.0.CO;2-D

Dahan, D., and Gaskell, M. G. (2007). The temporal dynamics of ambiguity resolution: evidence from spoken-word recognition. *J. Mem. Lang.* 57, 483–501. doi: 10.1016/j.jml.2007.01.001

Dahan, D., Magnuson, J. S., and Tanenhaus, M. K. (2001). Time course of frequency effects in spoken-word recognition: evidence from eye movements. *Cogn. Psychol.* 42, 317–367. doi: 10.1006/cogp.2001.0750

Dubno, J. R., Ahlstrom, J. B., and Horwitz, A. R. (2000). Use of context by younger and aged adults with normal hearing. *J. Acoust. Soc. Am.* 107, 538–546. doi: 10.1121/1.428322

Failes, E., Sommers, M. S., and Jacoby, L. L. (2020). Blurring past and present: using false memory to better understand false hearing in younger and older adults. *Mem. Cogn.* 48, 1403–1416. doi: 10.3758/s13421-020-01068-8

Garner, W. R. (1974). *The Processing of Information and Structure*, New York: Erlbaum.

Gaskell, M. G., and Marslen-Wilson, W. D. (2002). Representation and competition in the perception of spoken words. *Cogn. Psychol.* 45, 220–266. doi: 10.1016/S0010-0285(02)00003-8

Harel-Arbeli, T., Wingfield, A., Palgi, Y., and Ben-David, B. M. (2021). Age-related differences in the online processing of spoken semantic context and the effect of semantic competition: evidence from eye gaze. *J. Speech Lang. Hear. Res.* 64, 315–327. doi: 10.1044/2020_JSLHR-20-00142

Hasher, L., Quig, M. B., and May, C. P. (1997). Inhibitory control over no-longer relevant information: adult age differences. *Mem. Cogn.* 25, 286–295. doi: 10.3758/BF03211284

Hasher, L., and Zacks, R. T. (1988). Working memory, comprehension, and aging: a review and a new view. *Psychol. Learn. Motiv.* 22, 193–225. doi: 10.1016/S0079-7421(08)60041-9

Hothorn, T., Bretz, F., and Westfall, P. (2008). Simultaneous inference in general parametric models. *Biom. J.* 50, 346–363. doi: 10.1002/bimj.200810425

Huettig, F., and Janse, E. (2016). Individual differences in working memory and processing speed predict anticipatory spoken language processing in the visual world. *Lang. Cogn. Neurosci.* 31, 80–93. doi: 10.1080/23273798.2015.1047459

Hutchinson, K. M. (1989). Influence of sentence context on speech perception in younger and older adults. *J. Gerontol. Psychol. Sci.* 44, P36–P44. doi: 10.1093/geronj/44.2.p36

Ito, A., Pickering, M. J., and Corley, M. (2018). Investigating the time-course of phonological prediction in native and non-native speakers of English: a visual world eye-tracking study. *J. Mem. Lang.* 98, 1–11. doi: 10.1016/j.jml.2017.09.002

Jacoby, L. L., Bishara, A. J., Hessels, S., and Toth, J. P. (2005). Aging, subjective experience, and cognitive control: dramatic false remembering by older adults. *J. Exp. Psychol. Gen.* 134, 131–148. doi: 10.1037/0096-3445.134.2.131

Kramer, A. F., Humphrey, D. G., Larish, J. F., Logan, G. D., and Strayer, D. L. (1994). Aging and inhibition: beyond a unitary view of inhibitory processing in attention. *Psychol. Aging* 9, 491–512. doi: 10.1037/0882-7974.9.4.491

Kukona, A. (2020). Lexical constraints on the prediction of form: insights from the visual world paradigm. *J. Exp. Psychol. Learn. Mem. Cogn.* 46, 2153–2162. doi: 10.1037/xlm0000935

Luce, P. A., and Pisoni, D. B. (1998). Recognizing spoken words: the Neighborhood Activation Model. *Ear Hear.* 19, 1–36. doi: 10.1097/00003446-199802000-00001

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

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: an integrative review. *Psychol. Bull.* 109, 163–203. doi: 10.1037/0033-2909.109.2.163

Mishra, R. K., and Singh, N. (2014). Language non-selective activation of orthography during spoken word process in Hindi-English sequential bilinguals: an eye tracking visual world study. *Read. Writ. Interdiscip. J.* 27, 129–151. doi: 10.1007/s11145-013-9436-5

Morrell, C. H., Gordon-Salant, S., Pearson, J. D., Brant, L. J., and Fozard, J. L. (1996). Age- and gender-specific reference ranges for hearing level and longitudinal changes in hearing level. *J. Acoust. Soc. Am.* 100, 1949–1967. doi: 10.1121/1.417906

Nittrouer, S., and Boothroyd, A. (1990). Context effects in phoneme and word recognition by younger children and older adults. *J. Acoust. Soc. Am.* 87, 2705–2715. doi: 10.1121/1.399061

Pichora-Fuller, K. M. (2008). Use of supportive context by younger and older adult listeners: balancing bottom-up and top-down information processing. *Int. J. Audiol.* 47(sup2), S72–S82. doi: 10.1080/14992020802307404

Pichora-Fuller, M. K., Schneider, B. A., and Daneman, M. (1995). How younger and old adults listen to and remember speech in noise. *J. Acoust. Soc. Am.* 97, 593–608. doi: 10.1121/1.412282

Revill, K. P., and Spieler, D. H. (2012). The effect of lexical frequency on spoken word recognition in younger and older listeners. *Psychol. Aging* 27, 80–87. doi: 10.1037/a0024113

Rogers, C. S., Jacoby, L. L., and Sommers, M. S. (2012). Frequent false hearing by older adults: the role of age differences in metacognition. *Psychol. Aging* 27, 33–45. doi: 10.1037/a0026231

Shipley, W. C. (1940). A self-administering scale for measuring intellectual impairment and deterioration. *J. Psychol.* 9, 371–377. doi: 10.1080/00223980.1940.9917704

Sommers, M. S., and Danielson, S. M. (1999). Inhibitory processes and spoken word recognition in younger and older adults: the interaction of lexical competition and semantic context. *Psychol. Aging* 14, 458–472. doi: 10.1037/0882-7974.14.3.458

Sommers, M. S., Hale, S., Myerson, J., Rose, N., Tye-Murray, N., and Spehar, B. (2011). Listening comprehension across the adult lifespan. *Ear Hear.* 32, 775–781. doi: 10.1097/aud.0b013e3182234cf6

Sommers, M. S., and Huff, L. M. (2003). The effects of age and dementia of the Alzheimer's type on phonological false memories. *Psychol. Aging* 18, 791–806. doi: 10.1037/0882-7974.18.4.791

Sommers, M. S., Morton, J., and Rogers, C. (2015). "You are not listening to what I said: false hearing in younger and older adults," in *Remembering: Attributions, Processes, and Control in Human Memory (Essays in Honor of Larry Jacoby)*. eds. D. S. Lindsay, C. M. Kelley, A. P. Yonelinas and H. L. Roediger III (New York: Psychology Press), 269–284.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *J. Exp. Psychol.* 18, 643–662. doi: 10.1037/h0054651

Tanenhaus, M. K., Magnuson, J. S., Dahan, D., and Chambers, C. (2000). Eye movements and lexical access in spoken-language comprehension: evaluating a linking hypothesis between fixations and linguistic processing. *J. Psycholinguist. Res.* 29, 557–580. doi: 10.1023/A:1026464108329

Van Os, M., Kray, J., and Demberg, V. (2021). Mishearing as a side effect of rational language comprehension in noise. *Front. Psychol.* 12:679278. doi: 10.3389/fpsyg.2021.679278

West, R., and Alain, C. (2000). Age-related decline in inhibitory control contributes to the increased Stroop effect observed in older adults. *Psychophysiology* 37, 179–189. doi: 10.1111/1469-8986.3720179

Wingfield, A., Aberdeen, J. S., and Stine, E. A. (1991). Word onset gating and linguistic context in spoken word recognition by younger and elderly adults. *J. Gerontol.* 46, P127–P129. doi: 10.1093/geronj/46.3.P127

Zacks, R., and Hasher, L. (1997). Cognitive gerontology and attentional inhibition: a reply to burke and McDowd. *J. Gerontol. B: Psychol. Sci. Soc. Sci.* 52B, P274–P283. doi: 10.1093/geronb/52B.6.P274

Conflict of Interest: 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 and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership