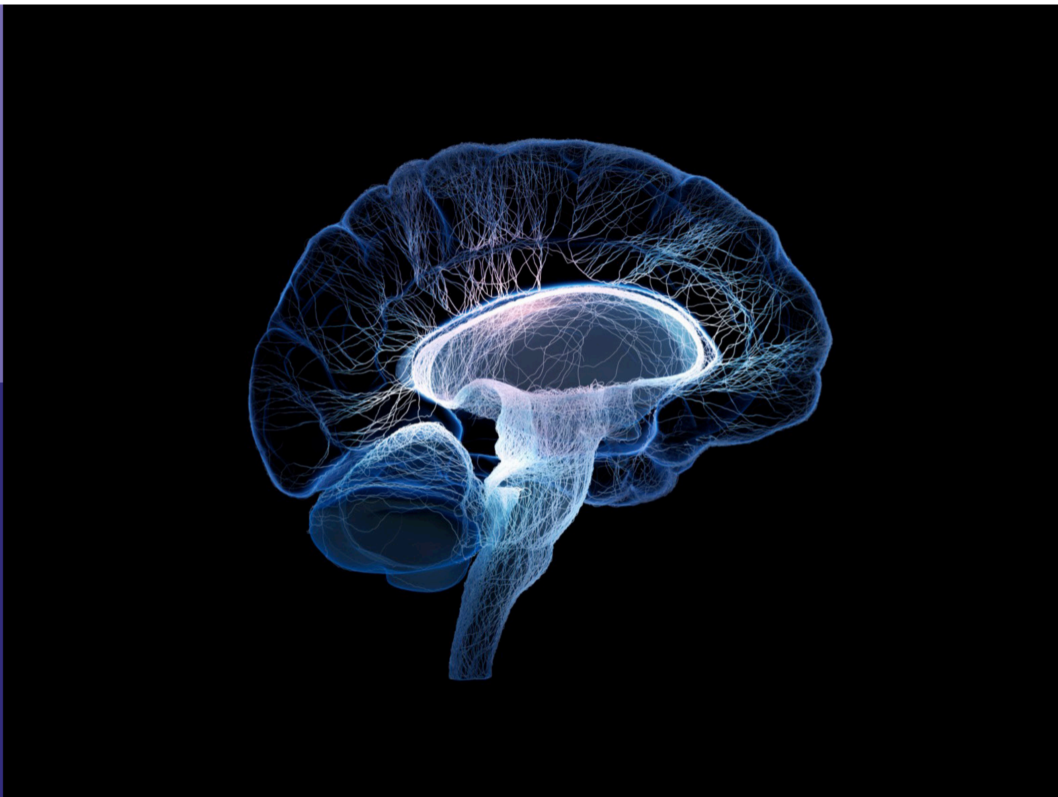
frontiers | Research Topics

Cognitive hearing science: Investigating the relationship between selective attention and brain activity

Edited by
Jerker Rönnberg, Anu Sharma, Tom A. Campbell, Patrik Sörqvist
and Carine Signoret

Published in
Frontiers in Neuroscience
Frontiers in Psychology
Frontiers in Human Neuroscience



ISSN 1664-8714
ISBN 978-2-83251-167-1
DOI 10.3389/978-2-83251-167-1

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

Cognitive hearing science: Investigating the relationship between selective attention and brain activity

Topic editors
Jerker Rönnberg — Linköping University, Sweden
Anu Sharma — University of Colorado Boulder, United States
Tom A. Campbell — Tampere University, Finland
Patrik Sörqvist — University of Gävle, Sweden
Carine Signoret — Linköping University, Sweden

Citation
Rönnberg, J., Sharma, A., Campbell, T. A., Sörqvist, P., Signoret, C., eds. (2023). *Cognitive hearing science: Investigating the relationship between selective attention and brain activity*. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-83251-167-1

Table of contents

05 **Editorial: Cognitive hearing science: Investigating the relationship between selective attention and brain activity**
Jerker Rönnberg, Anu Sharma, Carine Signoret, Tom A. Campbell and Patrik Sörqvist

11 **Brain activity during divided and selective attention to auditory and visual sentence comprehension tasks**
Mona Moisala, Viljami Salmela, Emma Salo, Synnöve Carlson, Virve Vuontela, Oili Salonen and Kimmo Alho

26 **Brain activations during bimodal dual tasks depend on the nature and combination of component tasks**
Emma Salo, Teemu Rinne, Oili Salonen and Kimmo Alho

36 **Phonological Task Enhances the Frequency-Following Response to Deviant Task-Irrelevant Speech Sounds**
Kimmo Alho, Katarzyna Żarnowiec, Natàlia Gorina-Careta and Carles Escera

46 **Modulation of Brain Activity by Selective Attention to Audiovisual Dialogues**
Alina Leminen, Maxime Verwoert, Mona Moisala, Viljami Salmela, Patrik Wikman and Kimmo Alho

62 **Decoding the Attended Speaker From EEG Using Adaptive Evaluation Intervals Captures Fluctuations in Attentional Listening**
Manuela Jaeger, Bojana Mirkovic, Martin G. Bleichner and Stefan Debener

78 **Short-Term Audiovisual Spatial Training Enhances Electrophysiological Correlates of Auditory Selective Spatial Attention**
Christina Hanenberg, Michael-Christian Schlüter, Stephan Getzmann and Jörg Lewald

96 **The Effects of Working Memory Load on Auditory Distraction in Adults With Attention Deficit Hyperactivity Disorder**
Rina Blomberg, Andrea Johansson Capusan, Carine Signoret, Henrik Danielsson and Jerker Rönnberg

111 **Acoustic Stimuli Can Improve and Impair Somatosensory Perception**
Matthias Nuernberger, Denise Schaller, Carsten Klingner, Otto Witte and Stefan Brodoehl

121 **Age-Related Differences in Early Cortical Representations of Target Speech Masked by Either Steady-State Noise or Competing Speech**
Bruce A. Schneider, Cristina Rabaglia, Meital Avivi-Reich, Dena Krieger, Stephen R. Arnott and Claude Alain

137 **Speech to noise ratio improvement induces nonlinear parietal phase synchrony in hearing aid users**
Payam Shahsavari Baboukani, Carina Graversen, Emina Alickovic and Jan Østergaard

151 **Aberrant resting-state connectivity of auditory, ventral attention/salience and default-mode networks in adults with attention deficit hyperactivity disorder**
Rina Blomberg, Carine Signoret, Henrik Danielsson, Irene Perini, Jerker Rönnberg and Andrea Johansson Capusan

163 **Intention-based predictive information modulates auditory deviance processing**
Andreas Widmann and Erich Schröger

frontiers | Frontiers in Neuroscience

TYPE Editorial
PUBLISHED 13 December 2022
DOI 10.3389/fnins.2022.1098340

OPEN ACCESS

EDITED AND REVIEWED BY
Isabelle Peretz,
Université de Montréal, Canada

*CORRESPONDENCE
Jerker Rönnberg
jerker.ronnberg@liu.se

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

RECEIVED 14 November 2022
ACCEPTED 22 November 2022
PUBLISHED 13 December 2022

CITATION
Rönnberg J, Sharma A, Signoret C, Campbell TA and Sörqvist P (2022) Editorial: Cognitive hearing science: Investigating the relationship between selective attention and brain activity. *Front. Neurosci.* 16:1098340. doi: 10.3389/fnins.2022.1098340

Editorial: Cognitive hearing science: Investigating the relationship between selective attention and brain activity

Jerker Rönnberg[1]*, Anu Sharma[2], Carine Signoret[1], Tom A. Campbell[3] and Patrik Sörqvist[4]

[1]Department of Behavioral Sciences and Learning, Linnaeus Centre HEAD, Linköping University, Linköping, Sweden, [2]Department of Speech, Language and Hearing Sciences University of Colorado at Boulder, Boulder, CO, United States, [3]Faculty of Information Technology and Communication Sciences, Tampere University, Tampere, Finland, [4]Department of Building Engineering, Energy Systems and Sustainability Science, University of Gävle, Gävle, Sweden

KEYWORDS

selective attention, divided attention, cognitive hearing science, brain activity, multi-modality, working memory, aging, brain imaging methods

Editorial on the Research Topic
Cognitive hearing science: Investigating the relationship between selective attention and brain activity

Introduction

"Everyone knows what attention is. It is the taking possession by the mind in clear and vivid form, of one out of what seem several simultaneously possible objects or trains of thought. Focalization, concentration, of consciousness are of its essence. It implies withdrawal from some things in order to deal effectively with others, and is a condition which has a real opposite in the confused, dazed, scatterbrained state which in French is called distraction, and Zerstreutheit in German," (James, 1890, p. 403–404). How does such a psychological concept relate to human brain activity? An influential model in clinical neuropsychology (Sohlberg and Mateer, 2001) differentiates five dissociable components that are focused, sustained, alternating, selective, and divided attention. Sustained attention concerns focusing attention on stimulation for an extended period. Selective attention concerns concentrating on one source of information in exclusion of another, in the service of some task. Divided attention concerns attending to one task when there are other attentional demands, such as another parallel task. From a cognitive hearing science perspective, attention has been a somewhat nebulous concept that depends partially on working memory (Barrouillet and Camos, 2020; Rönnberg et al., 2022a,b) and upon related executive control mechanisms (Badre, 2021).

In the current topic, we include multi-modal attention studies employing a plethora of measures from several brain-imaging and behavioral techniques. The topic reveals how the field is developing, maturing, and diversifying. This collection assembles world-leading researchers' more exciting developments from groundbreaking experiments spanning the last seven years. This research not only provides us with new knowledge about attentional processes but also about the intricacies of perceptual-cognitive interactions.

Sörqvist et al. (2012) make a case for cognitive hearing and the early attentional-steering of auditory input: Visual working memory load dampens Wave V of the auditory brainstem response. Such a visual working memory load also targets auditory cortex (Sörqvist et al., 2016). Accordingly, conscious and intentional processing of stimuli, presented cross-modally, can penetrate modular brain functions performing auditory processing within the first few milliseconds of the onset of a sound (see also Ikeda and Campbell, 2021). Generally speaking, sensory and cognitive processing blend to a much larger extent than previously acknowledged (Rönnberg et al., 2022a,b). Cognitive hearing science's new early filter model explains the top-down influences upon early sensory processing in relation to existing corticopetal-corticofugal loops (Marsh and Campbell, 2016; Campbell and Marsh, 2018, 2019).

Selectively attending to perceived dialogue or sounds is vital for smooth communication processes. Factors that affect selective attention not only include the source of speech or nonspeech sound, hearing status, and the listener's motivation for attending to the sound, but also effort and listener fatigue (see Pichora-Fuller et al., 2016). Further, recent research investigating the attentional processing of speech is revealing other key factors that affect our selective attention. These new factors include: the quality of attended speech, semantic predictability, grammatical complexity, and the number of competing sources of speech, as well as whether the masker speech is in the listener's mother tongue or not. This Research Topic, therefore, gathers together studies investigating the effects of what Sohlberg and Mateer (2001) term sustained, selective, and divided attention upon brain regions that relate to the aforementioned cognitive and perceptual factors.

To foreshadow the ensuing editorial, on a simple level of analysis, supramodal cortical regions during audiovisual divided attention are not the neural equivalent of capacity-limited bottlenecks. However, selective attention to visual phonological material exhibits an intermodal character that affects the brain's representation of the auditory stimulus. Turning from a simple to a more complex communicative level, hemodynamic investigations characterize the different kinds of activation when selecting or dividing attention concerning auditory and visual modalities. Intriguing is how that division, particularly under adverse conditions, can compromise the activation of the social brain network. On this complex level, we bring you new EEG approaches to indexing listening effort and fluctuations in sustained attention, as has a future in brain-computer interfaces to dynamically steer the signal processing in hearing-assistive devices according to transient neurocognitive state. We then introduce you to how cognitive training can relatively rapidly re-calibrate the perceptual systems dealing with speech. The editorial then evaluates the successes of investigations that psychologically characterize inter-individual differences in attentional effects on hemodynamic measures of brain activity in special populations. These populations are not only of elderly individuals but also of adults with attention deficit hyperactivity disorder (ADHD). We conclude with how explicit intention can limit the cortical processing of predictable pitch change, arguably *via* brain processes relating to the top-down influence of auditory selective attention. The consensus in the field hitherto considered such cortical processing of pitch deviance as largely task-independent, if not preattentive and impenetrable to volition.

Levels of analysis

Simple levels

In a relatively early functional magnetic resonance imaging (fMRI) dual-task study centering upon divided attention (Salo et al.), participants concurrently attended to a time-varying series of spoken and written letters. Participants perform one of nine bimodal discrimination tasks, which had a visual and an auditory task component. Either the auditory or visual task component, or both, could concern phonological features, discriminating between whether letters had a *name* starting with a vowel or a *name* starting with a consonant. Either task component, or both, could concern spatial features, discriminating whether the stimulus was on the left or right. Either task component, or both, could also concern simple features, judging the gender of the voice or the font of the letter. Of the nine tasks, the baseline dual task, with which to compare the other dual tasks, involved discriminating the simple auditory feature of gender of voice whilst discriminating the simple visual feature of font. This baseline dual task had no spatial or phonological requirements. A prior study provided the corresponding unimodal single discrimination task data with which to also separately compare the other dual tasks.

Comparison of dual tasks with the baseline dual task revealed different supramodal patterns of activation in the left medial frontal gyrus and right inferior parietal lobule. These findings juxtapose with how these supramodal activations were absent in the comparison with the single task components. The interpretation offered was that supramodal phonological and spatial areas are similarly activated during single tasks requiring phonological or spatial processing in one modality as during dual tasks that require: (i) both auditory and visual phonological processing implicating the left medial frontal gyrus or (ii) both auditory and visual spatial processing implicating the right inferior parietal lobule. These supramodal regions are thus arguably not the seat of some phonological or spatial capacity limitation serving as a bottleneck at the confluence of auditory and visual information.

In an electroencephalographic (EEG) investigation, Alho et al. reveal that the frequency-following response, which to-be-ignored heard distractor syllables elicit, goes relatively unaffected by a primary cross-modal task: Across two different heard syllables, whether that task is either a more challenging phonological task or a non-phonological task, on which performance is faster and more accurate, this response's amplitude does not differ significantly. This frequency-following response phase-locks to the vowel's acoustical fundamental.

As an editorial aside, at first, this null effect of the to-be-attended task seems uncontentious for the notion that any biased competition (Desimone and Duncan, 1995) between distractor and target during selective attention confines to intramodal filtering (Parks et al., 2011). Such a notion thus assumes the independence of modality-specific visual and auditory processing resources. Accordingly, there are no cross-modal effects on frequency-following responses (Szychowska and Wiens, 2021). However, deeper scrutiny of Alho et al.'s data is not so uncontentious for this notion.

The deviance of a rare unexpected syllable, interspersed amidst a sequence of repeated standard syllables in an oddball sequence, can cause the frequency-following response's amplitude to be higher. The presence of this effect of deviance seems to depend upon the acoustical content of the standard-deviant pairing in the oddball sequence of distractors. Crucially, when that pairing is sufficient for an effect of deviance, the extent of the effect proves higher when the primary task is phonological rather than nonphonological. Such a task-dependent influence may well result from the more demanding cross-modal phonological task either augmenting the frequency-following response to the deviant, or suppressing the corresponding response to the standard, or both. In either case, as a further editorial aside, this task-dependent influence is difficult to reconcile theoretically with an independence of visual and auditory processing resources during selective attention such that there are no cross-modal effects on frequency-following responses (Szychowska and Wiens, 2021).

Alho et al. do postulate a top-down modulation of activity in subcortical structures *via* corticofugal connections descending from the auditory cortex, as does cognitive hearing science's new early filter model (Marsh and Campbell, 2016; Campbell and Marsh, 2018, 2019). Alho et al.'s task-dependent influences upon frequency-following response phenomena are in more accord with this model than an independence of visual and auditory processing resources during selective attention. Although Alho et al. do demonstrate that the deviant syllable elicits a mismatch negativity, there is no analogous significant task-dependence of the amplitude of this component that could have functionally unrelated cortical generators. While it could thus be tempting to consider the task-dependent influence on frequency-following responses as purely subcortical, Alho et al.'s stimuli arguably also engage cortical generators (Coffey et al., 2019). These generators are distinct from that of the auditory mismatch negativity and are capable of tracking modulation frequencies upto 200 Hz (Brugge et al., 2009; Nourski et al., 2013).

In a cross-modal study of a slightly different sort, Nuernberger et al. investigate how different forms of noise influence the processing of tactile stimuli during a mechanical detection threshold task. The results show that whereas unpleasant everyday noise, "real noise", leads to an increased tactile sensitivity, white noise impairs such tactile sensitivity. Significant differences in brain activity and connectivity in distributed networks accompany this interaction between acoustic and tactile stimuli. Rather than invoking notions of selective and divided attention, the interpretation that the authors offer is that real noise creates a brain state for enhanced unimodal processing of tactile stimuli as could be favored by "phasic attention" (Schlittmeier et al., 2015). In juxtaposition, white noise increases both activity and connectivity in the auditory and somatosensory cortices, the association cortex, and the thalamus. Such white noise thereby impairs tactile sensitivity cross-modally as could relate to selective attention.

Complex levels

Moisala et al., in a semantic sentence congruency task, compared activations: (i) selectively attending to only the visual modality with (ii) selectively attending to only the auditory modality with (iii) divided attention on a bimodal version of that task. This task activates left prefrontal cortex activity in selective attention conditions, whereas the same areas showed significant activity increases during divided attention. The results suggest that divided attention tasks interfere with each other, stimulating increased activity in the same cortical areas without any compensatory activity in other cortical areas. The cost is therefore lower performance in the divided attention tasks.

The fMRI studies of Salo et al. and Moisala et al. both show that comparing selective with divided attention tasks reveal different brain activation patterns (see also Salo et al., 2017). However, neither study clarifies whether the division of attentional tasks rely upon cross-modal or within-modality interference. Leminen et al., in an ecologically compelling investigation, demonstrate that selective attention to dialogues activates not only areas in brain networks for audiovisual speech processing and understanding, but also a social brain network. What social knowledge that we use and gain in a conversational situation depends not only upon what we can see, hear, and already know, but also upon how we integrate the information from the two modalities, to an extent pre-attentively, when one modality is less informative. As well, this knowledge depends upon our mental flexibility to direct intermodal selective attentional resources to alternate between the auditory and visual modality. Arguably, with hearing-assistive devices that have a limited number of channels, kindred to Leminen et al.'s signal processing, the temporal fine structure of the audio is a valuable source of information in that process that determines audiovisual quality. Reductions in audiovisual quality increase the demands upon selective attention that could relate to Leminen et al.'s observed fronto-parietal activation. Such pathfinding studies pave the future for a cognitive hearing science that shall determine how intermodal brain processes glean socially relevant meanings from heard utterances in audiovisual contexts under adverse conditions.

Along the lines of more complex communicative levels of analysis, Jaeger et al.'s participants attended to one story in the (competing) presence of another. Analysis of intra-individual variation in EEG-decoding performance over time relates to behavioral performance together with subjective ratings of listening effort, motivation, and fatigue. Parameters describing the individual performance indicated significant differences in EEG-decoding performance over time, which closely related to the behavioral performance in the selective listening task. Those fluctuations could have implications for the control of hearing-assistive devices *via* a brain computer interface in multi-talker situations.

Shahsavari Baboukani et al. identified an EEG-based measure of alphaband phase synchrony that arguably indexes listening effort. This investigation measures the EEG of aided listeners with hearing impairments during a continuous speech-in-noise task under conditions of background noise and a competing talker. This study shows that the activation of noise-reduction schemes in hearing aids can non-linearly reduce listening effort in the parietal region-of-interest. Indeed, the authors propose that the investigation of the phase synchrony within regions-of-interest over the scalp can reflect the effects of hearing aids in hearing-impaired individuals under ecological listening conditions.

Training

A 15-min period of audiovisual spatial training (Hanenberg et al.) affects participants' audio-spatial performance on a selective attention task. The task used by Hanenberg et al. requires selecting the auditory target from different positions of three distractor words. Training affected the amplitude of the N2 deflection of the event-related potential (ERP), which is known to index auditory spatial attention. The N2 is significantly higher in amplitude after audiovisual-congruency training compared with other feedback or incongruous training conditions. This finding was apparent for younger, yet not older, participants. These findings arguably offer insights into the cross-modal processes that audiovisual-congruency training alters under "cocktail-party" conditions. This short-term alteration results in enhanced correlates of auditory selective spatial attention. Focusing on the very limited time necessary to improve neural and behavioral performance, the results by Hanenberg et al. are in accord with the independent study of Moradi et al. (2019)—Brief exposure to audiovisual stimulus materials improves performance on auditory perception tasks. This finding generalizes to tasks ranging from the simple auditory gating of vowels and consonants to sentence perception in noise. Moradi et al. (2019) dub this phenomenon "perceptual doping". The brain seems to have the power to rapidly re-calibrate the perceptual systems dealing with speech.

Attention in special populations

Attention ADHD

ADHD per definitionem affects sustained attention and the attention networks of the brain. Although distractibility is not definitively at the core of this syndrome, selective attention can be an issue (Pelletier et al., 2016). Two investigations by Blomberg et al. demonstrate how cognitive and attention networks interact. One of the principal findings of Blomberg, Johansson Capusan, et al. is that under cognitive load, in a visual working memory task, the attention networks tend to become "blended" to a larger extent when working memory is put under stress. Results indicate that adults with ADHD, compared to controls, cannot attenuate auditory cortical responses to the task-irrelevant sound when working memory demands is high (i.e., as in a 2-back version of an n-back task). Further, heightened auditory activity to task-irrelevant sound correlates significantly with both poorer working memory performance and symptomatic inattentiveness. As already shown by Blomberg et al. (2019), a behavioral composite latent working memory capacity measure could predict performance in different kinds of degraded/noise conditions. Finally, in a resting state study (Blomberg, Signoret, et al.), the default mode network still interacts more with the ventral and auditory attention networks for adults with ADHD relative to controls, as arguably compromises selective attention causing higher levels of distraction from auditory stimuli.

Attention and aging

Schneider et al.'s EEG study revealed that, in young adults, there is an increase in the early cortical activity generating the N1 and P2 ERP deflections to a word in babble with longer masker onset delays. This cortical activity in older adults goes unaffected by that delay. These results support the hypothesis that an increase in onset delay improves stream segregation in younger adults in both noise and babble. The results also support the hypothesis that this improvement occurs only in noise for older adults. These influences upon stream segregation are also evident in early cortical processes.

As an editorial comment, one may wonder what a cognitive interpretation of these effects may look like. The ability to segregate competing speech signals is dependent on working memory, while the ability to process a speech signal that competing background (non-speech) noise masks is less dependent upon working memory (Sörqvist and Rönnberg, 2012). Pertinent are the age-related differences in working memory capacity (Wingfield et al., 1988) and the role of that capacity in temporal discrimination and temporal processing (Broadway and Engle, 2011). As such, differences in working memory capacity could thus explain the age-related differences in the ability to make use of the onset delay to support auditory stream segregation in speech noise. One possibility is that high working memory capacity is a prerequisite for detecting, accessing, and utilizing the temporally fine-tuned information necessary for the segregation of the two speech streams.

Intentional-explicit influence on attentional processing

Widmann and Schröger employed an oddball paradigm to investigate intention-based predictive or non-predictive processing of standard and pitch-deviant sounds. The manipulation of predictive processing was that the participant either heard a completely unpredictable oddball sequence or, on-the-fly, had partial control of the oddball sequence: The participant's task was pressing one button occasionally to produce a predicted deviant tone in the sequence, whereas pressing another frequently produces a predicted standard tone, albeit occasionally the system for stimulus presentation randomly produced a mispredicted deviant instead of that predicted standard. Intriguingly, both unpredictable deviants and mispredicted deviants elicited the auditory mismatch negativity ERP component, but predicted deviants did not. This elegant procedure shows that intention-based prediction, which relies upon the top-down influence of the action intention prediction, attenuates this mismatch negativity. Thus, even though the predicted deviant violates an auditory regularity, brain processes that relate to top-down cognitive predictions limit the generation of the mismatch negativity. While the sensory-memory trace hypothesis has fallen from grace and these new findings are difficult to reconcile with the adaptation hypothesis, a role for working memory capacity is not an assumption of the predictive-coding account proliferating extant explanations of the auditory mismatch negativity findings. Widmann and Schröger's findings dovetail with corroborative evidence of a somewhat different sort concerning the investigation of predictability: Semantic cues associated with target sentences prime performance on a speech-perception-in-noise test in which those sentences are stimuli (Zekveld et al., 2013). Whereas working memory capacity predicted the extent of this priming effect, an intriguing open question is how that capacity relates to Widmann and Schröger's influence of prediction.

Closing

In sum, this topic's collection of papers explores the relationship between factors affecting different forms of attention and brain activity, as well as the brain regions that competing audio and audio-visual cues or sources activate. To evaluate, these papers, together, succeed in substantially advancing cognitive hearing science's understanding of human attention and the related brain processes. Having attracted mostly multi-modal investigations, our overwhelming impression from re-reading the articles is that this topic now sets the scene for new avenues in cognitive hearing science: This new avenue shall usefully determine how attention relates to the intermodal brain processes that operate when people extract meaning under adverse conditions.

Author contributions

JR drafted the editorial. AS, CS, TC, and PS contributed to the final submitted version. All authors contributed to the article and approved the submitted version.

Funding

Writing of this editorial by JR was supported by the Swedish Research Council (Grant No. 2017-06092; held by Anders Fridberger).

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

Badre, D. (2021). "Brain networks for cognitive control: Four unresolved questions," in *Intrusive Thinking: From Molecules to Free Will Strûngmann Forum Reports, Kalivas*, P. W., and Paulus, M. P. (Cambridge, MA: The MIT Press) p. 203–228. doi: 10.7551/mitpress/13875.001.0001

Barrouillet, P., and Camos, V. (2020). "The time-based resource-sharing model of working memory," in *Working memory*, Barrouillet, P., and Camos, V. (eds). (Oxford: Oxford University Press). doi: 10.1093/oso/9780198842286.003.0004

Blomberg, R., Danielsson, H., Rudner, M., Söderlund, G. B. W., and Rönnberg, J. (2019). Speech processing difficulties in attention deficit hyperactivity disorder. *Front Psychol*, 10:1536. doi: 10.3389/fpsyg.2019.01536

Broadway, J. M., and Engle, R. (2011). Individual differences in working memory capacity and temporal discrimination. *PLoS ONE.* 6, e25422. doi: 10.1371/journal.pone.0025422

Brugge, J. F., Nourski, K. V., Oya, H., Reale, R. A., Kawasaki, H., Steinschneider, M., et al. (2009). Coding of repetitive transients by auditory cortex on Heschl's gyrus. *J. Neurophysiol.* 102, 2358–2374. doi: 10.1152/jn.91346.2008

Campbell, T. A., and Marsh, J. E. (2018). Commentary: Donepezil enhances understanding of degraded speech in Alzheimer's disease. *Front. Aging Neurosci.* 10:197. doi: 10.3389/fnagi.2018.00197

Campbell, T. A., and Marsh, J. E. (2019). On corticopetal-corticofugal loops of the new early filter: from cell assemblies to the rostral brainstem. *NeuroReport.* 30, 202–206. doi: 10.1097/wnr.0000000000001184

Coffey, E., Nicol, T., White-Schwoch, T., Chandrasekaran, B., Krizman, J., Skoe, E., et al. (2019). Evolving perspectives on the sources of the frequency-following response. *Nat. Commun.* 10:5036. doi: 10.1038/s41467-019-13003-w

Desimone, R., and Duncan, J. (1995). Neural mechanisms of selective visual attention. *Ann. Rev. Neurosci.* 18, 193–222. doi: 10.1146/annurev.ne.18.030195.001205

Ikeda, K., and Campbell, T. A. (2021). Reinterpreting the human ABR binaural interaction component: Isolating attention from stimulus effects. *Hear. Res.* 410:108350. doi: 10.1016/j.heares.2021.108350

James, W. (1890). *The Principles of Psychology*. New York: H. Holt and Company. doi: 10.1037/11059-000

Marsh, J. E., and Campbell, T. A. (2016). Processing complex sounds passing through the rostral brainstem: the new early filter model. *Front. Neurosci.* 10:136. doi: 10.3389/fnins.2016.00136

Moradi, S., Lidestam, B., Danielsson, H., Ng, E. H. N., and Rönnberg, J. (2019). Perceptual doping: An audiovisual facilitation effect on auditory speech processing, from phonetic feature extraction to sentence identification in noise. *Ear Hear*. 40, 312–327. doi: 10.1097/AUD.0000000000000616

Nourski, K. V., Brugge, J. F., Reale, R. A., Kovach, C. K., Oya, H., Kawasaki, H., et al. (2013). Coding of repetitive transients by auditory cortex on posterolateral superior temporal gyrus in humans: an intracranial electrophysiology study. *J. Neurophysiol.* 109, 1283–1295. doi: 10.1152/jn.00718.2012

Parks, N. A., Hilimire, M. R., and Corballis, P. M. (2011). Steady-state signatures of visual perceptual load, multimodal distractor filtering, and neural competition. *J. Cogn. Neurosci.* 23, 1113–1124. doi: 10.1162/jocn.2010.21460

Pelletier, M. F., Hodgetts, H. M., Lafleur, M. F., Vincent, A., and Tremblay, S. (2016). Vulnerability to the irrelevant sound effect in adult ADHD. *J. Attention Dis.* 20, 306–316. doi: 10.1177/1087054713492563

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

Rönnberg, J., Holmer, E., and Rudner, M. (2022a). "Working memory and the ease of language understanding model," in *The Cambridge Handbook of Working Memory and Language*, Schwieter, W., and Wen, Z. E. (Cambridge: Cambridge University Press) p. 197–218. doi: 10.1017/9781108955638.013

Rönnberg, J., Signoret, C., Andin, J., and Holmer, E. (2022b). The Cognitive Hearing Science perspective on perceiving, understanding, and remembering language: The ELU model. *Front. Psychol.* 13:967260. doi: 10.3389/fpsyg.2022.967260

Salo, E., Salmela, V., Salmi, J., Numminen, J., and Alho, K. (2017). Brain activity associated with selective attention, divided attention and distraction. *Brain Res.* 1664, 25–36. doi: 10.1016/j.brainres.2017.03.021

Schlittmeier, S. J., Feil, A., Liebl, A., and Hellbrück, J. R. (2015). The impact of road traffic noise on cognitive performance in attention-based tasks depends on noise level even within moderate-level ranges. *Noise Health* 17, 148–157. doi: 10.4103/1463-1741.155845

Sohlberg, M. M., and Mateer, C. A. (2001). *Cognitive Rehabilitation: An Integrative Neuropsychological Approach.* New York: Guilford Press.

Sörqvist, P., Dahlström, Ö., Karlsson, T., and Rönnberg, J. (2016). Concentration: The neural underpinnings of how cognitive load shields against distraction. *Front. Hum. Neurosci.* 10:221. doi: 10.3389/fnhum.2016.00221

Sörqvist, P., and Rönnberg, J. (2012). Episodic long-term memory of spoken discourse masked by speech: What is the role of working memory capacity? *J. Speech Lang. Hear. Res.* 55, 210–218. doi: 10.1044/1092-4388(2011/10-0353)

Sörqvist, P., Stenfelt, S., and Rönnberg, J. (2012). Working memory capacity and cognitive load modulate auditory-sensory gating in the brainstem. *J. Cogn. Neurosci.* 24, 2147–2154. doi: 10.1162/jocn_a_00275

Szychowska, M., and Wiens, S. (2021). Visual load effects on the auditory steady-state responses to 20-, 40-, and 80-Hz amplitude-modulated tones. *Physiol. Behav.* 228:113240. doi: 10.1016/j.physbeh.2020.113240

Wingfield, A., Stine, E. A. L., Lahar, C. J., and Aberdeen, J. S. (1988). Does the capacity of working memory change with age? *Exp. Aging Res.* 14, 103–107. doi: 10.1016/j.lmot.2017.01.007

Zekveld, A. A., Rudner, M., Johnsrude, I. S., and Rönnberg, J. (2013). The effects of working memory capacity and semantic cues on the intelligibility of speech in noise. *J. Acoust. Soc. Am.* 134, 2225–2234. doi: 10.1080/03610738808259731

frontiers in
HUMAN NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 19 February 2015
doi: 10.3389/fnhum.2015.00086

Brain activity during divided and selective attention to auditory and visual sentence comprehension tasks

Mona Moisala[1,2]*, Viljami Salmela[1,3], Emma Salo[1], Synnöve Carlson[4,5], Virve Vuontela[5], Oili Salonen[6] and Kimmo Alho[1,3,7,8]

[1] *Division of Cognitive Psychology and Neuropsychology, Institute of Behavioural Sciences, University of Helsinki, Helsinki, Finland*
[2] *Department of Teacher Education, University of Helsinki, Helsinki, Finland*
[3] *Advanced Magnetic Imaging Centre, Aalto NeuroImaging, Aalto University, Espoo, Finland*
[4] *Brain Research Unit, Department of Neuroscience and Biomedical Engineering, Aalto University School of Science, Espoo, Finland*
[5] *Neuroscience Unit, Institute of Biomedicine/Physiology, University of Helsinki, Helsinki, Finland*
[6] *Helsinki Medical Imaging Center, Helsinki University Central Hospital, Helsinki, Finland*
[7] *Helsinki Collegium for Advanced Studies, University of Helsinki, Helsinki, Finland*
[8] *Swedish Collegium for Advanced Study, Uppsala, Sweden*

Edited by:
Shuhei Yamaguchi, Shimane University, Japan

Reviewed by:
Richard Bruce Bolster, University of Winnipeg, Canada
Jennifer Adrienne Johnson, Bloomsburg University of Pennsylvania, USA

***Correspondence:**
Mona Moisala, Division of Cognitive Psychology and Neuropsychology, Institute of Behavioural Sciences, University of Helsinki, P.O. Box 9 (Siltavuorenpenger 1 B), Helsinki, FI 00014, Finland
e-mail: mona.moisala@helsinki.fi

Using functional magnetic resonance imaging (fMRI), we measured brain activity of human participants while they performed a sentence congruence judgment task in either the visual or auditory modality separately, or in both modalities simultaneously. Significant performance decrements were observed when attention was divided between the two modalities compared with when one modality was selectively attended. Compared with selective attention (i.e., single tasking), divided attention (i.e., dual-tasking) did not recruit additional cortical regions, but resulted in increased activity in medial and lateral frontal regions which were also activated by the component tasks when performed separately. Areas involved in semantic language processing were revealed predominantly in the left lateral prefrontal cortex by contrasting incongruent with congruent sentences. These areas also showed significant activity increases during divided attention in relation to selective attention. In the sensory cortices, no crossmodal inhibition was observed during divided attention when compared with selective attention to one modality. Our results suggest that the observed performance decrements during dual-tasking are due to interference of the two tasks because they utilize the same part of the cortex. Moreover, semantic dual-tasking did not appear to recruit additional brain areas in comparison with single tasking, and no crossmodal inhibition was observed during intermodal divided attention.

Keywords: dual-tasking, divided attention, selective attention, functional MRI, semantic processing

INTRODUCTION

Simultaneously performing several tasks is demanding and often leads to decrements in performance speed and accuracy (Pashler, 1994). These decrements may be due to a bottleneck in executive task-coordination systems recruited by multitasking (D'Esposito et al., 1995; Collette et al., 2005). Additional interference may be generated if the component tasks are presented in different sensory modalities and the corresponding sensory cortices have to compete for attentional resources (e.g., Näätänen, 1992). Competition may also occur beyond the sensory cortices in brain areas related to carrying out the component tasks in case these component tasks require similar (e.g., phonological or spatial) processing (Baddeley and Hitch, 1974). In the current study, we compared performance and brain activity in conditions requiring intermodal selective attention to one task with those demanding crossmodal division of attention between two simultaneous tasks requiring similar semantic processing. We asked (i) whether dividing attention recruits specialized executive task-coordinating systems; (ii) how attention modulates activity in the sensory cortices during bimodal linguistic stimulation; and (iii) how brain areas showing attention-related and task-specific activations react when two simultaneous tasks requiring similar processing are performed.

Previous research has suggested that multitasking recruits brain areas specialized in task coordination and managing interfering information from the component tasks (Corbetta et al., 1991; D'Esposito et al., 1995; Yoo et al., 2004; Stelzel et al., 2006). It has been suggested that dual-tasking involves task-coordinating abilities that are distinct from other executive functions such as shifting or inhibition (Miyake et al., 2000). Previous studies have highlighted the importance of frontal and parietal cortical areas as parts of a neural network involved in coordination of multiple parallel tasks. The involved frontal areas include the inferior frontal (Herath et al., 2001; Schubert and Szameitat, 2003; Stelzel et al., 2006) and middle frontal regions (Szameitat et al., 2002; Yoo et al., 2004) and the dorsolateral prefrontal cortex (Corbetta et al., 1991;

D'Esposito et al., 1995; Johnson and Zatorre, 2006). The involved parietal areas, in turn, include the superior parietal lobule (Yoo et al., 2004) and intraparietal sulcus (Szameitat et al., 2002). The existence of specialized multitasking areas has been questioned, however, by studies failing to show multitasking-related activity in areas beyond those activated by the component tasks (Klingberg, 1998; Adcock et al., 2000; Bunge et al., 2000; Nijboer et al., 2014). These studies have shown that the performance of two concurrent tasks results only in a surplus of activation in the regions activated by the component tasks when performed separately, and no additional cortical regions are recruited. The former studies suggest that the main factor limiting performance during multitasking is the involvement of general coordinating or executive functions, whereas the latter studies suggest that limited task-specific resources are responsible for the observed interference during multitasking.

It has been repeatedly shown that when selective attention is directed to one modality, activity elevations in the sensory cortical areas processing attended inputs may be accompanied by diminished activity in the sensory cortical areas processing unattended inputs (Laurienti et al., 2002; Crottaz-Herbette et al., 2004; Shomstein and Yantis, 2004; Johnson and Zatorre, 2005; Mittag et al., 2013; Salo et al., 2013). It is less clear, however, how activity is modulated in the sensory cortices when attention is divided between two modalities. If there is a limited attentional resource allocated to the sensory cortices, sensory activity should decrease during intermodal divided attention when compared with selective attention to one modality. Indeed, there are studies showing such an effect during bimodal attention tasks (Loose et al., 2003; Johnson and Zatorre, 2006).

Many of the previous studies examining multitasking effects have used component tasks that do not necessarily rely on the same cortical areas, such as a semantic categorization task and a face recognition task (Adcock et al., 2000), or a spatial rotation and semantic judgment task (D'Esposito et al., 1995). It is therefore still unclear how task-related cortical activations are affected when several tasks competing for the use of those areas are performed simultaneously. In the current study, our participants performed two simultaneous sentence congruence judgment tasks. This type of task, when performed separately as a single task, has been shown to activate cortical areas related to semantic processing (e.g., Kiehl et al., 2002; Humphries et al., 2007). Functional magnetic resonance imaging (fMRI) studies using semantic congruence manipulations have consistently observed greater hemodynamic activity for incongruent than congruent sentences. The most commonly found areas to show this effect are the left superior temporal and left inferior frontal gyri, both when the sentences are presented as written text (Baumgaertner et al., 2002; Kuperberg et al., 2003; Service et al., 2007) and when they are presented as speech (Ni et al., 2000; Cardillo et al., 2004). These activations might be related to the N400 event-related potential (ERP) response elicited, for example, by an incongruent last word of sentence (e.g., "the pizza was too hot to *sing*"; Kutas and Hillyard, 1980).

In the present fMRI study, participants performed a sentence comprehension task involving spoken or written sentences, or both. The participants' task was to rate the sentences as congruent or incongruent in only one modality at a time, or in both modalities simultaneously. This experimental setup allowed us to address three separate research questions related to multitasking. First, we investigated brain activity during simultaneous performance of two tasks in comparison with brain activity during the same tasks when performed separately. This allowed us to determine whether any additional cortical areas would be recruited during the divided attention condition. Second, the issue of crossmodal suppression of sensory cortices during selective attention to one modality was addressed. By performing a separate analysis in the auditory and visual cortices, we expected to see crossmodal suppression in the auditory cortex during selective attention to the visual modality, and vice versa. Moreover, in accordance with the hypothesis of limited resources, we expected to observe smaller attention-related activity in the visual and auditory cortices during division of attention between the two modalities than during intermodal selective attention to the written and spoken sentences, respectively. Third, we examined modulation of brain activity associated with linguistic processing when participants perform two simultaneous sentence comprehension tasks. This was accomplished by comparing activity elicited by incongruent sentences with activity elicited by congruent sentences during selective attention, thereby presumably isolating brain areas related specifically to semantic processing, and then examining activity modulations in these areas during divided attention. We hypothesized that as the number of tasks requiring semantic judgments is increased from one to two, activity in semantic processing areas increases. We expected to see that this increase would be non-additive due to limited processing capacity, leading to deficits in performing two simultaneous semantic judgments.

MATERIALS AND METHODS

STIMULI

Visual stimuli

Visual stimuli used in the experiment were written sentences and sentence-like nonsense text. They were projected onto a mirror mounted on the head coil and presented in the middle of the screen (font: Arial, size 14). The size of the sentences at the viewing distance of ∼40 cm was ∼1.4° vertically and ∼24° horizontally.

Text. Written sentences were either semantically congruent or incongruent sentences in Finnish. The incongruent sentences were created by taking a subset of the congruent sentences (e.g., "*This morning I ate a bowl of cereal*") and replacing the last word of the sentences with a semantically incongruent (but syntactically plausible) word (e.g., "*This morning I ate a bowl of shoes*"). Each participant saw a total of 192 congruent sentences and 144 incongruent sentences, because in the dual-task conditions more congruent sentences were needed (for details see Section Procedure).

Nonsense text. Sentence-like nonsense text was created by randomly selecting a subset of the congruent written sentences and replacing each vowel in those sentences with a different vowel. This procedure resulted in nonsensical sentences with word lengths and letter frequencies similar to the Finnish language. Forty eight different nonsense written sentences were used.

Auditory stimuli

Auditory stimuli used in the experiment consisted of speech, nonsense speech, and music. All auditory stimuli were presented binaurally through insert earphones (Sensimetrics model S14; Sensimetrics, Malden, MA, USA). All auditory stimuli were broadband stimuli high-pass filtered with a cut-off at 100 Hz and low-pass filtered with a cut-off at 7000 Hz. The intensity of auditory stimuli was scaled so that their total power in RMS units, the square root of the mean of the squared signal, was similar (0.1). The intensity of the sounds was individually set to a loud, but pleasant level, and was ~80 dB SPL as measured from the tip of the earphones. All adjustments to the auditory stimuli were made using Audacity [1] and Matlab (Mathworks Inc., Natick, MA, USA) softwares.

Speech. Spoken sentences were semantically congruent or incongruent Finnish sentences spoken by a female native Finnish speaker. The incongruent sentences were created in a similar way as the incongruent written sentences, that is, by replacing the last word in the congruent sentences. Each participant heard a total of 192 congruent sentences and 144 incongruent sentences, because in the dual-task conditions more congruent sentences were needed (see Section Procedure for details).

Nonsense speech. The nonsense speech stimuli consisted of recorded nonsensical sentences (see above) spoken by a female native Finnish speaker. Each participant heard a total of 112 nonsense speech sentences. The lengths of the sentences were adjusted so that each sentence had a duration of 2.5 s.

Music. 2.5-s excerpts of instrumental music were obtained from a free-source online music website. The music excerpts represented various genres from hip-hop to classical music. Forty eight music clips were used.

Functional localizers

Functional localizers were used in order to accurately localize the auditory and visual sensory cortices of each participant. The auditory functional localizer was created by phase-scrambling spoken sentences by chopping the signal into short (10 ms) time-windows and shuffling the segments (Ellis, 2010). The visual functional localizer was a contrast-reversing checkerboard flickering at 8 Hz. The size of the checkerboard was similar to written sentences (~1.4° × ~24°), and it was centered at the middle of the screen. The auditory and visual localizers were presented simultaneously for 2.5 s, followed by a 1-s fixation cross (~1.4° × ~1.4°) at the center of screen.

PARTICIPANTS

Participants were 18 healthy volunteering adults (9 females), all right handed and native Finnish speakers between 21 and 34 years of age (mean age 26 years) with normal hearing, normal or corrected-to-normal vision, and no history of psychiatric or neurological illnesses. An informed written consent was obtained from each participant before the experiment. The experimental protocol was approved by the Coordinating Ethics Committee of The Hospital District of Helsinki and Uusimaa, Finland.

fMRI/MRI DATA ACQUISITION

Functional brain imaging was carried out with 3 T MAGNETOM Skyra whole-body scanner (Siemens Healthcare, Erlangen, Germany) using a 20-channel head coil. The functional echo planar (EPI) images were acquired with an imaging area consisting of 43 contiguous oblique axial slices (TR 2500 ms, TE 32 ms, flip angle 75°, voxel matrix 64 × 64, field of view 20 cm, slice thickness 3.0 mm, in-plane resolution 3.1 mm × 3.1 mm × 3.0 mm). Image acquisition was performed at a constant rate, but was asynchronized with stimulus onsets. Four functional runs of 240 volumes were measured for each participant. A total of 960 functional volumes were obtained in one session (session duration approximately 37 min).

High-resolution anatomical images (voxel matrix 256 × 256, in-plane resolution 1 mm × 1 mm × 1 mm) were acquired from each participant between the third and fourth functional runs.

PROCEDURE

A total of ten experimental task blocks (each consisting of the nine experimental conditions with the divided attention condition repeated twice), one rest block, and one functional localizer block were included in each functional run. In the beginning of each block, instructions for the current task type were shown for 3.5 s. During the rest and localizer blocks, the participants were asked to look at the fixation cross. In subsequent task blocks, 12 sentences (visual or auditory) or sentence pairs (visual and auditory) were presented, each with a duration of 2.5 s. Each sentence was followed by a 1-s response window during which the participants were instructed to respond with an appropriate button press whether the attended sentence was congruent or not (or during the divided attention task whether both attended sentences were congruent or not) using their right index and middle finger, respectively. During the response window, a question mark (size 1.4° × 1.0°) was presented at the center of screen. The fixation cross preceded each written sentence for 500 ms on the left side of the screen where the first letter of the sentence subsequently appeared. When only speech stimuli were presented, the fixation cross was shown at the center of screen during the entire trial. At the end of each block, the participant was shown the percentage of correct responses in that block. The score was shown for 2 s, and followed by 4 s of rest before the next block.

A total of nine different experimental conditions were used. In the single-task conditions, the participants were instructed to attend to the sentences in just one modality (auditory or visual). There was either no stimuli presented in the other modality (the *unimodal condition*, two blocks), or distractor stimuli were

---

[1] http://audacity.sourceforge.net

present in the other modality and the participants were instructed to ignore them (the *selective attention condition*, two blocks). Auditory distractors were spoken sentences (one block), music (one block) or nonsense speech (one block). The visual distractors were written sentences (one block), which the participants were instructed to ignore by holding a steady fixation on a fixation cross presented in the middle of the screen. Two additional visual distractor conditions were included in order to control for eye movements: a moving fixation cross (one block) and the participants were instructed to follow it while attending to speech; nonsense written sentences (one block) and the participants were instructed to scan through the nonsense text while attending to speech. These two control conditions did not differ from the condition including written sentences as distractors and were therefore discarded from further analyses. In the *divided attention condition* (two blocks), the participants were presented with simultaneous spoken and written sentences and instructed to attend to both modalities, and asked to decide whether or not both sentences were congruent (both sentences were never incongruent).

There were four functional runs, 12 blocks in each run, and 12 trials (i.e., sentences, sentence pairs, or functional localizers) in each block. Each run included one block of each task type, except the divided attention task, which was repeated twice. This was done in order to ensure an equal amount of trials where the incongruent sentence was in the visual/auditory modality between the divided attention and the unimodal and selective attention condition blocks, since in the divided attention condition only half of the incongruent trials had an incongruent sentence in the visual/auditory modality. This resulted in a total of 96 trials for the divided attention task (4 × 2 × 12), and 48 trials for all the other task types (4 × 1 × 12). The order of tasks within the run was random, except that the rest block was always in the middle of the run between the 6th and 7th task block. All stimuli (sentences and distractors) were presented in randomized order. The sentences were randomized in the following way. First, the sentences were divided randomly into 4 sets (1 per run) that were identical for all participants. Then the order of sentences within a set was randomized, and the presentation order of these 4 sets was randomized and counterbalanced across participants. Each sentence was presented only once to each participant. The congruent and incongruent versions of the same sentence were never presented within the same run.

fMRI DATA ANALYSIS

Image preprocessing and statistical analysis was performed using Statistical Parametric Mapping (SPM8) analysis package (Wellcome Department of Cognitive Neurology, London, UK; Friston et al., 1994a) as implemented in Matlab. In order to allow for initial stabilization of the fMRI signal, the first four dummy volumes were excluded from analysis. In pre-processing, the slice timing was corrected, data were motion corrected, high-pass filtered (cut-off at 1/128 Hz), and spatially smoothed with a 6 mm Gaussian kernel. The EPI images were intra-individually realigned to the middle image in each time series and un-warping was performed to correct for the interaction of susceptibility artifacts and head movements.

For the first-level statistical analysis, the general linear model was set up including a regressor for incongruent and congruent sentences in each of the 9 different and analyzed experimental conditions, resulting in 18 regressors. Separate regressors for the responses of the participants and for instructions (2.5-s periods between the blocks and a 6-s period at the beginning of each run) were also included. 6 movement parameters were added to the model as nuisance regressors. The regressors were convoluted with the canonical hemodynamic response function.

In the second-level analysis, the anatomical images were normalized to a canonical T1 template (MNI standard space) provided by SPM8 and then used as a template to normalize the contrast images for each participant (tri-linear interpolation, 3 mm × 3 mm × 3 mm using 16 nonlinear iterations). Statistical parametric maps of individual contrasts between task types and between tasks and rest were then averaged across participants. A voxel-wise height t-value threshold and a cluster size threshold were set depending on the contrast type (the specific values are stated below each contrast image). The statistical images were cluster corrected at $p < 0.005$ (Friston et al., 1994b). Anatomical regions corresponding to the activity foci were identified using the xjView toolbox for SPM.[2]

REGION OF INTEREST ANALYSIS

To study activity modulations in areas specifically related to dual-tasking, the divided attention condition was contrasted separately with the selective attention to text condition and the selective attention to speech condition. *Dual-tasking regions of interest* (ROIs) were then drawn manually using Freesurfer software to cover areas showing overlap between these two contrasts. Further statistical analyses were conducted using repeated-measures analyses of variance (ANOVAs) for voxels within these ROIs. Activity modulations between task conditions were compared by conducting an ANOVA with the factor Condition (9 levels) for each ROI separately and for data averaged across the ROIs. To compare activity modulations between the dual-tasking ROIs in the different task conditions, a 5 (Dual-tasking ROI) × 9 (Condition) ANOVA was conducted. Laterality effects in the dual-tasking ROIs were studied using a 2 (Hemisphere of the dual-tasking ROI) × 9 (Condition) ANOVA. To study the effects of attention in the unimodal, selective attention and divided attention conditions, a 5 (Dual-tasking ROI) × 3 (Task type) ANOVA was carried out. Finally, the effect of attended modality (irrespective of task type) was examined using a 5 (Dual-tasking ROI) × 3 (Attended modality) ANOVA. Note that the nine task conditions used in the ANOVAs also include the three conditions which were used to select dual-tasking ROIs.

ROI analyses were also conducted to examine activity modulations in the sensory cortices. To this end, voxels activated by the functional localizer (family-wise error corrected $p < 0.05$) were used as visual- and auditory-cortex ROIs of each individual participant. The mean percentage of blood-oxygen-level dependent (BOLD) signal change within the ROIs was calculated per voxel and normalized by dividing it by the overall

---

[2] http://www.alivelearn.net/xjview

average BOLD signal amplitude within a participant, and then averaged within each contrast of interest. To address the issue of possible crossmodal inhibition of the sensory cortices during selective attention, a 2 (Sensory cortex) × 2 (Attended modality) ANOVA was carried out for selective attention condition. To study dual-tasking effects, the divided attention condition was compared with the unimodal and selective attention conditions using an ANOVA with a 5 (Condition) × 2 (Hemisphere) × 2 (Sensory cortex) ANOVA.

ROI analysis was also used to study modulations of activity during divided attention in areas involved in semantic processing of sentences. In this analysis, contrasts between incongruent and congruent sentences in the second-level analysis were used to map areas of enhanced activity separately for written and spoken sentences. The incongruence contrast for speech sentences was created by summing together separately for attended incongruent and congruent sentences all the conditions where attention was directed to speech sentences (attention to speech in the unimodal condition, selective attention to speech with a text distractor, and the two additional visual distractor conditions) and all the conditions where attention was directed to written sentences (attention to text in the unimodal condition, selective attention to text presented together with speech, music or nonspeech distractors), and then contrasting the incongruent vs. congruent sentences within each modality. *Semantic* ROIs were drawn manually using Freesurfer software so that they covered areas showing overlap between the incongruence contrasts for spoken and written sentences. The mean percentage of BOLD signal change within each semantic ROI was calculated and normalized across the nine experimental conditions, and then averaged for each contrast of interest. To study the effects of the different task types on activity in the semantic ROIs, a 2 (Hemisphere of the ROI) × 3 (Task type) × 2 (Semantic congruence) ANOVA was carried out. Note that the three tasks types for an ANOVA were created by averaging the nine conditions used to select semantic ROIs.

ANALYSIS OF BEHAVIORAL DATA

The total percentage of correct responses per task type was calculated. The difference in the number of correct responses between task types was analyzed using a repeated-measures ANOVA with three Task levels (unimodal conditions vs. selective attention conditions vs. divided attention condition), where the two unimodal conditions where averaged together, and the six selective attention conditions were averaged together. An ANOVA was conducted on the three selective attention to text conditions (attention to text with a speech, nonsense speech or music distractor) in order to determine the effect of Auditory distractor type, and a similar ANOVA was conducted for the three selective attention to speech conditions (attention to speech with a text, nonsense text or moving fixation cross distractor) to study the effects of Visual distractor type. The effect of Attended modality was analyzed using an ANOVA with three levels (conditions where attention was targeted to written sentences vs. speech sentences vs. both written and speech sentences). The effect of the modality of incongruent sentences during the divided attention condition was analyzed using a paired sample t-test.

For all conducted ANOVAs the Greenhouse-Geisser p-value was used (as indicated by the correction value ε) if the Mauchly's test of sphericity showed a significant result for a variable with more than two levels. However, original degrees of freedom will be reported with the F-value even in these cases. A 95% confidence interval was used in all ANOVAs. When an ANOVA yielded a significant result, Bonferroni *post hoc* tests were conducted. IBM SPSS Statistics 21 for Windows (IBM SPSS, Armonk, NY, USA) was used for statistical analyses.

RESULTS

BEHAVIORAL RESULTS

The mean percentage of correct responses (± standard error of the mean, SEM) was 97.6% ± 0.6% for the unimodal conditions, 95.3% ± 0.95% for the selective attention conditions, and 90.2% ± 1.6% for divided attention condition (**Figure 1**). The ANOVA with three Task levels showed a main effect of Task type ($F_{(2,32)} = 23.69$, $p < 0.001$) and subsequent *post hoc* tests revealed that the percentage of correct responses was significantly lower during divided attention than during attention in the unimodal condition ($p < 0.001$) or intermodal selective attention ($p < 0.005$) conditions, and significantly lower during selective attention than during attention in the unimodal condition ($p < 0.05$). The modality of the attended sentences did not affect the percentage of correct responses in single tasks ($p = 0.24$). The nature of the auditory distractor during selective attention to written sentences did not affect performance ($p = 0.78$). The ANOVA for auditory selective attention conditions showed a significant main effect of Visual distractor type ($F_{(2,32)} = 6.31$, $p < 0.005$, $\varepsilon = 0.58$) and *post hoc* tests indicated that significantly fewer correct responses were given when the visual distractor was regular written text than when it was a moving fixation cross or nonsense text (in both cases, $p < 0.05$).

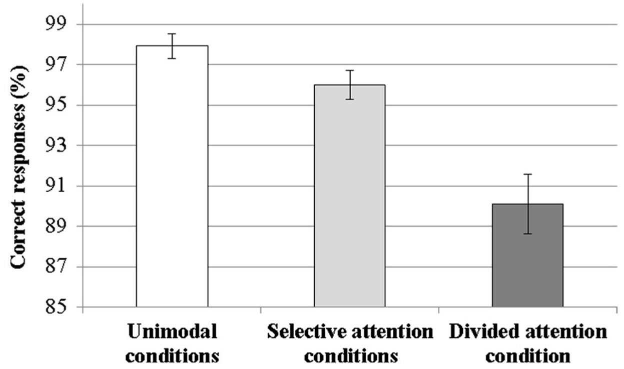


FIGURE 1 | Percentages of correct responses in the unimodal conditions (data combined across auditory and visual conditions), selective attention conditions (data combined across 6 conditions: selective attention to speech with a text, nonsense text or moving fixation cross distractor and selective attention to text with a speech, nonsense speech or music distractor) and divided attention condition. Error bars indicate SEMs. Percentages of correct responses in the three condition types differed significantly from each other (in all cases, $p < 0.05$).

Further analyses were conducted for the divided attention condition to study possible task strategy biases. These analyses showed that the percentage of correctly identified incongruent sentences did not depend on whether the incongruent sentence was in the visual or auditory modality ($p = 0.24$). Half of the participants had slightly higher accuracy rates (max. 12.5%) when the incongruent sentence was visual, and the other half when the incongruent sentence was auditory. Two participants performed at about chance level (46%) for auditory incongruent sentences, but the remaining participants had high response accuracies (>70%, mean 89%) for incongruent sentences in both modalities.

BRAIN ACTIVITY DURING DIVIDED VS. NON-DIVIDED ATTENTION

Cortical networks recruited by selective attention to text with a speech distractor and selective attention to speech with a text distractor are shown in **Figure 2**. Activity during the selective attention tasks was compared with activity in the rest blocks. For the selective attention to text condition (**Figure 2A**), activity enhancement was seen bilaterally in the visual and auditory sensory cortices (BA 17/18/19, BA 41/42/22), and in the medial supplementary motor area (SMA; BA 6), precentral gyrus (BA 4/6), and inferior and middle frontal gyri (IFG and MFG; BA 44 and BA 46/9, respectively), and in the left superior and inferior parietal lobule (BA 7 and BA 40, respectively). A similar cortical network was activated by the selective attention to speech condition, with the exception of no significant activations in the visual sensory cortices (**Figure 2B**). **Figure 2C** shows comparisons between the areas recruited by the two selective attention conditions combined and the divided attention condition, demonstrating that these two networks largely overlap with each other. The activation map from the combined selective attention conditions compared with rest are denoted with red, and the activation map from the divided attention condition compared with rest is denoted with yellow. Areas showing overlap between these two contrasts are denoted with orange. This overlapping network includes bilaterally the visual and auditory cortices, medial SMA extending to more anterior regions of the medial superior frontal gyrus (BA 8/32), and the IFG and MFG, as well as the left precentral gyrus and superior and inferior parietal lobules.

Next, we contrasted the divided attention condition separately with the visual and auditory selective attention conditions with similar stimulation as during divided attention (i.e., selective attention to text with a speech distractor and selective attention to speech with a text distractor). The resulting contrast images were then overlaid on top of each other (**Figure 3**). Areas showing overlap between these two contrasts (orange areas in **Figure 3**)

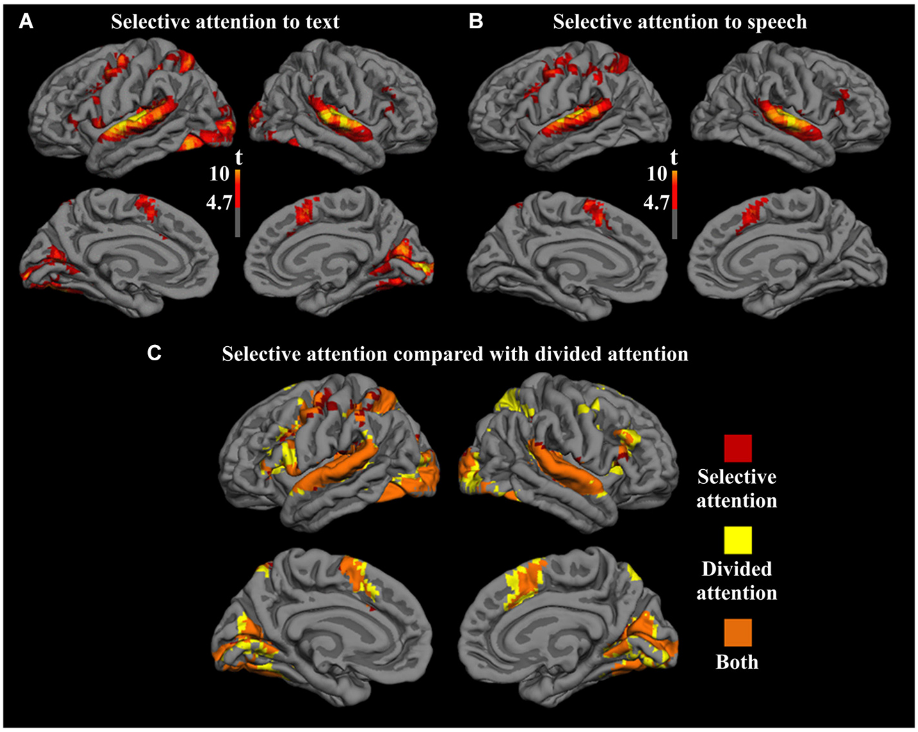


FIGURE 2 | Significant activity enhancements in relation to the rest blocks in the conditions (A) selective attention to text with a speech distractor and (B) selective attention to speech with a text distractor. **(C)** A combination of these two contrasts is overlaid with the contrast showing activity enhancements during divided attention compared to rest. Areas showing significant activation enhancements only in the selective attention conditions are denoted with red and areas showing activation enhancements only in the divided attention condition are denoted with yellow. Areas showing overlap between these enhancements are denoted with orange. Voxel-wise height threshold $t = 4.7$, cluster size > 250, cluster-corrected $p < 0.001$.

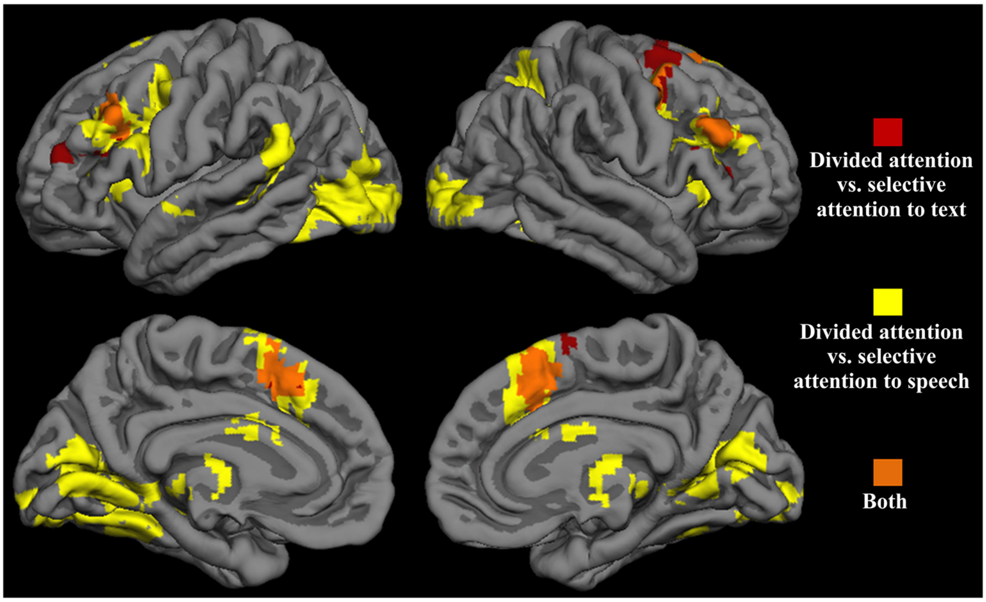


FIGURE 3 | Significant activity enhancements during divided attention in relation to selective attention to text with a speech distractor (red), selective attention to speech with a text distractor (yellow), and both (orange). Voxel-wise height threshold $t = 2.5$, cluster size > 250, cluster corrected $p < 0.001$.

included clusters in the dorsolateral and medial portions of the frontal lobe. More specifically, clusters in the MFG (BA 9/6) and medial SMA (BA 6) showed greater activity bilaterally during divided attention than in either selective attention condition. Five dual-tasking ROIs were subsequently drawn to cover these regions showing overlap: the left and right anterior middle frontal gyrus (aMFG) ROIs, the left and right SMA ROIs, and the right posterior middle frontal gyrus (pMFG) ROI. Subsequent analyses were performed for voxels within these ROIs.

Figure 4 shows mean signal changes in the dual-tasking ROIs for each task condition. A significant main effect of Condition was observed for all five ROIs (for all ROIs, $p < 0.001$). There were no significant effects of Hemisphere for the MFG dual-tasking ROIs or SMA ROIs, or Condition × Hemisphere interactions. Since the five dual-tasking ROIs displayed a similar general pattern of activation for the different conditions (**Figure 4**, top and middle), and because no other main effects were observed, the data were averaged across the five dual-tasking ROIs in further analyses (**Figure 4**, bottom).

An ANOVA including Task type (unimodal vs. selective attention vs. divided attention) as the factor indicated a main effect of Task type ($F_{(2,34)} = 47.72$, $p < 0.001$). *Post hoc* tests revealed that the selective attention conditions resulted in significantly larger BOLD signal increases in the dual-tasking ROIs than the unimodal conditions ($p < 0.05$) and that divided attention was associated with greater signal increases than unimodal and selective attention conditions (in both cases, $p < 0.001$), which was expected since the dual-tasking ROIs were defined as areas showing greater activity during divided attention than selective attention. Another ANOVA with Attended modality (visual vs. auditory vs. both) as the factor indicated a significant main effect of Attended modality ($F_{(2,34)} = 61.19$, $p < 0.001$). Subsequent *post hoc* tests showed that BOLD signal increases in the dual-tasking ROIs were smallest in conditions where speech sentences were attended, followed by conditions where the text sentences were attended, and greatest when attention was divided between text and speech (in all cases, $p < 0.001$). When an ANOVA was conducted for data that were averaged across the dual-tasking ROIs (**Figure 4**, bottom), attending to text with a nonsense speech distractor caused greater BOLD signal increases than when no auditory distractor was present ($p < 0.05$). When speech was attended, both text and nonsense text distractors caused a greater signal increase than when no visual distractor was present ($p < 0.05$ in both). When the distractor was a moving fixation cross, signal increases did not differ from the condition with no visual distractor ($p = 0.13$), but were smaller than when a text distractor was present ($p < 0.05$).

ATTENTION EFFECTS ON ACTIVITY IN THE SENSORY CORTICES

When the activity in sensory cortices during the selective attention conditions (attention to text with a speech distractor and attention to speech with a text distractor) was examined, the interaction Sensory cortex (visual vs. auditory) × Attended modality (visual vs. auditory) was significant ($F_{(1,17)} = 15.85$, $p < 0.001$), that is, the visual cortex showed greater activity when attention was selectively directed to text than when it was directed to speech while the auditory cortex showed an opposite pattern.

The results from the ANOVA including the factors Condition (attention to text in a unimodal condition vs. attention to speech in a unimodal condition vs. selective attention to text with a speech distractor vs. selective attention to speech with a text distractor vs. divided attention), Sensory cortex, and Hemisphere

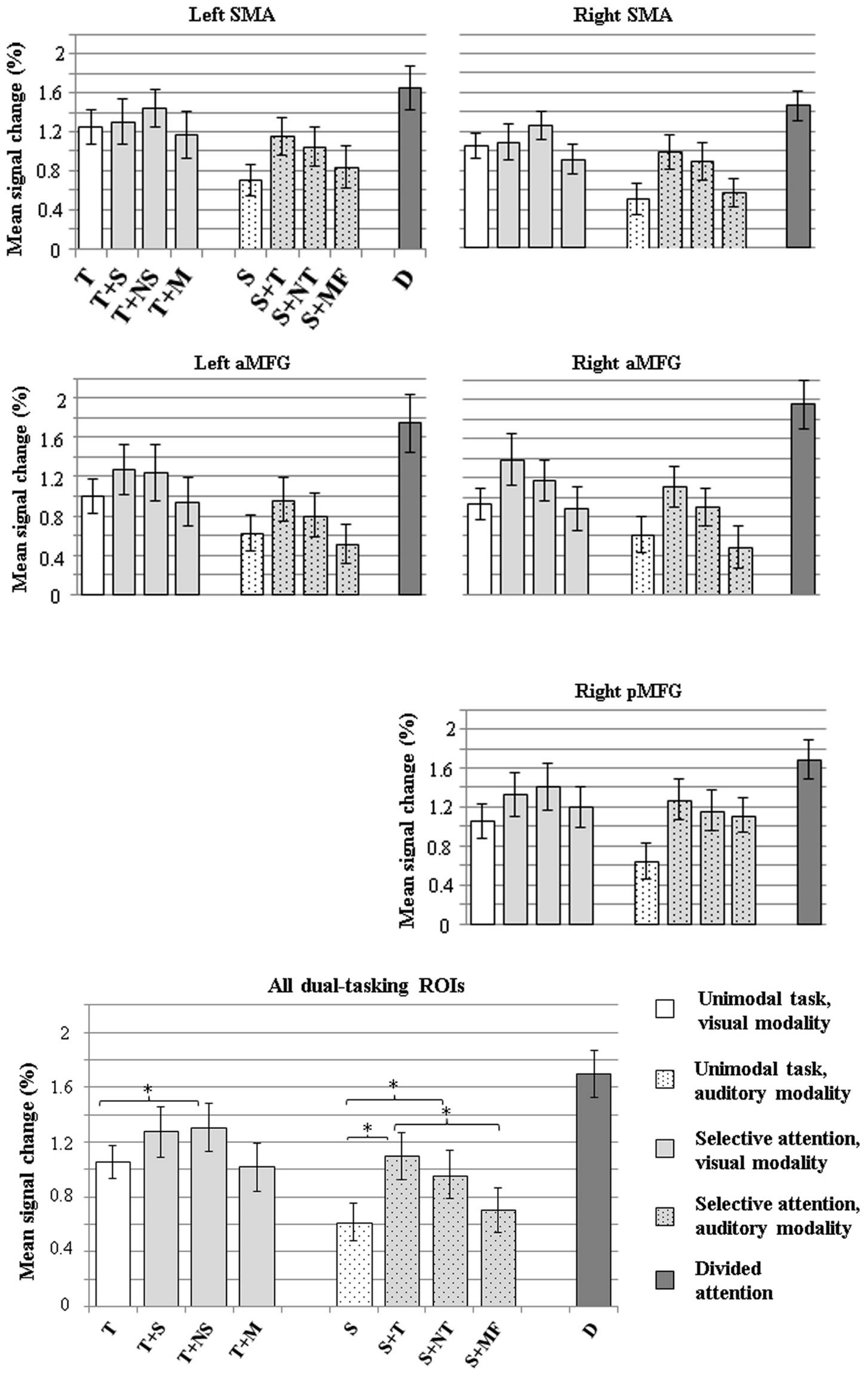


FIGURE 4 | Mean signal changes (%) compared with rest in the five dual-tasking ROIs during the nine experimental conditions. Top: mean signal increases in the left and right SMA ROIs. Middle: mean signal increases in the left and right anterior MFG ROIs and the right posterior MFG ROI. Bottom: mean signal increases averaged across all five dual-tasking ROIs. The conditions in each graph are grouped based on the attended modality (left bar cluster: visual modality attended, middle bar cluster: auditory modality attended, rightmost bar: both modalities attended). Error bars indicate SEMs. Conditions differing significantly from each other are indicated with asterisks (*$p < 0.05$). Note that the nine conditions include the three conditions which were used to select dual-tasking ROIs. (T = attention to text in a unimodal condition, T+S = attention to text with a speech distractor, T+NS = attention to text with a nonsense speech distractor, T+M = attention to text with a music distractor, S = attention to speech in a unimodal condition, S+T = attention to speech with a text distractor, S+NT = attention to speech with a nonsense text distractor, S+MF = attention to speech with a moving fixation cross distractor, D = divided attention).

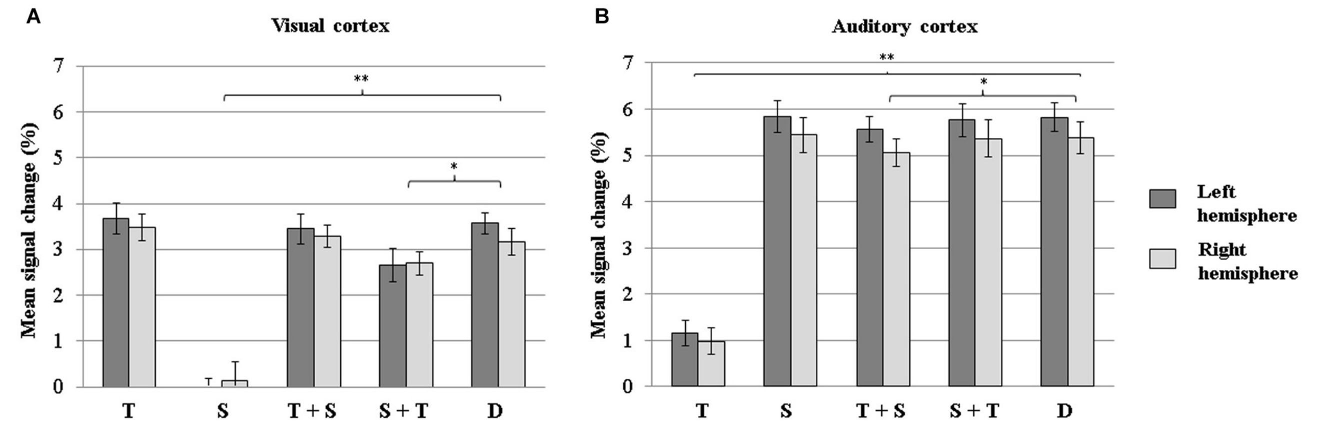


FIGURE 5 | Mean signal changes (%) in the visual (A) and auditory (B) cortices in the left and right hemispheres (dark gray and light gray bars, respectively) during attention to text in the unimodal condition, attention to speech in the unimodal condition, selective attention to text (with a speech distractor), selective attention to speech (with a text distractor), and divided attention. Error bars indicate SEMs. Conditions differing significantly from the divided attention condition are indicated with asterisks (*$p < 0.05$; **$p < 0.005$). (T = attention to text in a unimodal condition, S = attention to speech in a unimodal condition, T+S = attention to text with a speech distractor, S+T = attention to speech with a text distractor, D = divided attention).

are illustrated in **Figure 5**. A significant main effect for Sensory cortex was observed ($F_{(1,17)} = 43.53$, $p < 0.001$), demonstrating that, overall, mean signal changes were greater in the auditory cortex than in the visual cortex. There was no significant main effect of Hemisphere (although there was some insignificant tendency for the left-hemisphere activity being higher than the right-hemisphere activity, $p = 0.12$). However, the main effect of Condition was significant ($F_{(4,68)} = 63.04$, $p < 0.001$, $\varepsilon = 0.85$). Subsequent pairwise comparisons revealed that the BOLD signal change was greatest during divided attention, followed by selective attention to text with a speech distractor, then by selective attention to speech with a text distractor, and lastly by attention to speech and attention to text in the unimodal conditions. Also, a significant interaction Condition × Sensory cortex was found ($F_{(4,68)} = 190.12$, $p < 0.001$, $\varepsilon = 0.51$). Pairwise comparisons revealed that in the visual cortex, the mean signal change during divided attention did not differ significantly from that during attention to the visual modality in the unimodal ($p = 0.27$) or selective attention condition ($p = 0.98$), but was significantly greater than that during auditory attention in the unimodal ($p < 0.001$) or selective attention condition ($p < 0.005$). Activity in the visual cortex during visual attention did not depend significantly on the presence of an auditory (speech) distractor ($p = 0.27$ for attention to text during selective attention vs. during the unimodal condition). In the auditory cortex, the mean signal change during divided attention did not differ significanly from that during attention to the auditory modality in the unimodal ($p = 0.84$) or selective attention condition ($p = 0.83$), but was signifcantly higher than that during visual attention in the unimodal ($p < 0.001$) or selective attention condition ($p < 0.05$). Activity in the auditory cortex during auditory attention did not depend significantly on the presence of a visual (text) distractor ($p = 0.70$ for attention to speech during selective attention vs. unimodal condition).

BRAIN ACTIVITY RELATED TO SEMANTIC PROCESSING

As seen in **Figure 6**, analysis across the auditory single-task conditions showed that attended spoken incongruent sentences elicited a greater hemodynamic response than attended spoken congruent sentences bilaterally in the IFG (BA 44) extending to the MFG (BA 9/6), and in the superior temporal gyrus (BA 41/42/22). A similar comparison for attended written sentences in the visual single-task conditions showed activity enhancements for written incongruent sentences in relation to written congruent sentences bilaterally in the IFG (BA 44) extending to the MFG (BA 9/6), and in the posterior part of the left middle temporal gyrus (BA 21/37). When these two contrasts were overlaid (orange areas in **Figure 6**), two clusters corresponding roughly to the left and right IFG (BA 44) showed overlap between the two contrasts. In the left hemisphere, the overlap region covered both the pars opercularis and pars triangularis, and in the right hemisphere, the region was smaller and extended to the inferior frontal sulcus. Areas showing overlap were used as semantic ROIs and subsequent analyses were performed for voxels within these ROIs.

Mean signal changes in the semantic ROIs for congruent and incongruent sentences in the different task condition types are shown in **Figure 7**. As expected, the significant main effect of sentence congruence ($F_{(1,17)} = 34.32$, $p < 0.001$) confirmed that incongruent sentences caused greater increases in the BOLD signal than congruent sentences in both the left- and right-hemisphere semantic ROI. A main effect of Task type ($F_{(2,34)} = 22.41$, $p < 0.001$) revealed a greater increase in overall signal change during the divided attention condition than during the unimodal or selective attention conditions ($p < 0.001$ in both), and a greater increase during the selective attention conditions than unimodal conditions ($p < 0.05$) in the semantic ROI of each hemisphere. Also a main effect of Laterality was observed ($F_{(1,17)} = 7.97$, $p < 0.05$), demonstrating a greater signal change

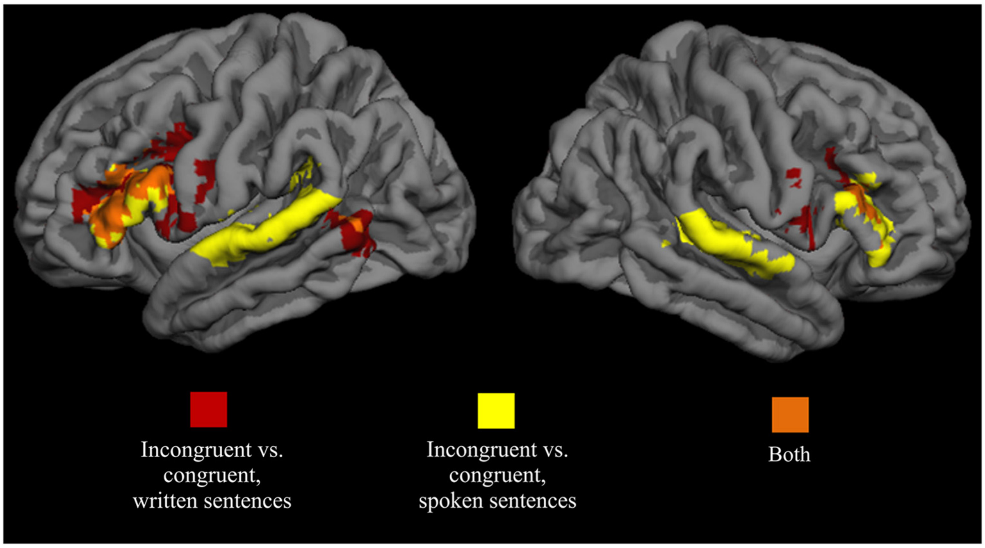


FIGURE 6 | Brain areas showing significant activity enhancements for attended incongruent written (red) and spoken (yellow) sentences (area overlaps shown in orange) in relation to respective congruent sentences. Data combined across all single-task conditions for each modality. Voxel-wise height threshold $t = 2.5$, cluster size > 250, cluster-corrected $p < 0.001$.

in the left-hemisphere semantic ROI than in the right-hemisphere semantic ROI. There were no significant interactions between the factors.

DISCUSSION

TASK PERFORMANCE

The behavioral results indicate that even though task performance was significantly worse during divided than during selective attention, the participants were still able to reach a high level of performance accuracy while attending to two stimuli simultaneously (even during divided attention mean response accuracy was over 90%).

DIVIDED ATTENTION VS. FOCUSED AND SELECTIVE ATTENTION

The difference between the selective attention and divided attention conditions was examined in order to determine whether any cortical activity was specifically related to dividing attention. Because in both conditions stimuli were presented in both modalities, the effect of sensory stimulation was controlled for in the contrast between these conditions. The results showed that divided attention recruited a very similar cortical network as the component tasks performed alone, since the activation maps showed a high degree of overlap.

When a direct comparison was made between the divided attention and the selective attention conditions, bilateral clusters both on the medial and dorsolateral frontal cortex showed significantly greater BOLD signal increases in the divided attention condition compared to the selective attention conditions. More specifically, these clusters were situated in the medial SMA and MFG of both hemispheres. The MFG has been implicated in memory rehearsal processes (Awh et al., 1996), rapid adaptation and coordination of actions required in dual-tasking (Szameitat et al., 2002), and detection of unexpected relevant stimuli (Corbetta and Shulman, 2002). The medial SMA, in turn, has been associated with performance monitoring, pre-response conflict, decision uncertainty, response errors, and processing of negative feedback (for a review, see Ridderinkhof et al., 2004). The need to inhibit a response to one sentence when it conflicts with the response to the other sentence, or the overall increase in difficulty in choosing the correct response in the divided attention condition might therefore explain the increase in SMA activity.

Areas showing higher activity during divided attention than during both selective attention to text and selective attention to speech were defined as dual-tasking ROIs. These ROIs were located in the medial SMA and MFG bilaterally. The smallest BOLD signal increases in these ROIs were seen during the unimodal conditions. The selective attention conditions activated these regions to a greater degree, with some activation differences that depended on the nature of the distractor stimuli. More specifically, nonsense speech as an auditory distractor and text and nonsense text as visual distractors caused greater activity increases than when no distractors were present. Since divided attention activated these ROIs the most, this might mean that these distractors were the most effective in drawing attention away from the actual task and creating a situation where attention was unintentionally divided between the attended and to-be-ignored modality.

Taking into account the high degree of overlap between the cortical networks activated by selective and divided attention, and the fact that dual-tasking ROIs showed a graded activation increase related to task difficulty (unimodal condition < selective attention < divided attention), our results suggest that at least semantic dual-tasking does not recruit new cortical areas, but

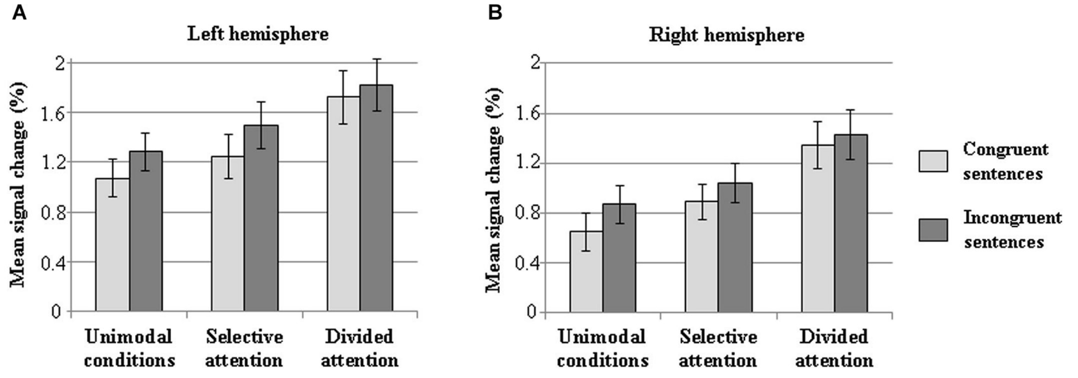


FIGURE 7 | Mean signal changes (%) in the semantic ROIs for attended incongruent and congruent sentences compared with rest during the unimodal (data combined across the unimodal auditory and visual conditions), selective attention (data combined across all auditory and visual selective attention conditions), and divided attention conditions in the left (A) and right (B) hemisphere. Error bars indicate SEMs.

places more demands on the brain regions already in use by the component tasks. This finding is in accordance with several previous studies showing that no additional neural regions are activated when interfering information needs to be coordinated (Klingberg, 1998; Adcock et al., 2000; Bunge et al., 2000; Nijboer et al., 2014), but rather that the component tasks compete for resources in a "global neuronal workspace" most likely located in frontoparietal regions (Hein et al., 2007). Some studies have reported opposite results, however, showing that frontal regions are recruited only during divided attention (Corbetta et al., 1991; D'Esposito et al., 1995; Herath et al., 2001; Szameitat et al., 2002; Schubert and Szameitat, 2003; Yoo et al., 2004; Johnson and Zatorre, 2006; Stelzel et al., 2006). These conflicting results may be explained more by the nature of the single tasks used in the individual studies than by the need to divide attention *per se*. Frontal recruitment may depend on the specific task demands of the single-tasks and vary from one task combination to the other. In our study, there are several possible explanations for the observed frontal recruitment during the component tasks. First, it could be related to inhibiting the processing of irrelevant information from the unattended modality. Frontal regions have been shown to be involved in gating sensory information according to task-specific demands (Miller and Cohen, 2001; Staines et al., 2002). Another possible explanation is catching of attention by stimuli in the unattended modality. It has been shown that a distributed network including frontal and parietal areas is activated when attention is involuntarily shifted to events in the sensory environment (Downar et al., 2000; Corbetta and Shulman, 2002; Salmi et al., 2009). The sentences in the unattended modality might therefore have caused an involuntary shift of attention to the unattended modality, resulting in frontal and parietal activity increases. Finally, our results might be explained by the difficulty of the component tasks used in the study. It could be argued that since our component tasks were complex sentence comprehension tasks, performing them required central executive functions to a great degree even in the absence of distracting stimuli or a need to divide attention between two modalities.

The frontoparietal cortical network observed in our selective attention and divided attention conditions bares a close resemblance to the multiple-demand (MD) network described by Duncan (2010). This general-purpose network includes cortex in and around the inferior frontal sulcus, the pre-SMA and the intraparietal sulcus, and it is activated by a variety of demanding cognitive tasks that require the formation of a series of subtasks. The tasks employed in our experiment can indeed be broken down into a succession of subtasks: internalizing the task instructions, evaluating the meaning of the presented sentence, choosing the correct response option, forming a motor response, reorienting to the next task instruction, etc. In the case of the present selective attention conditions, an additional subtask of inhibiting processing of the unattended stimulus is introduced. When two streams of stimuli have to be attended simultaneously, the amount of subtasks is even further increased even though the time given to complete these subtasks remains unchanged, adding to the demands placed on the MD network. It might therefore be that the observed BOLD signal increases in dorsolateral and medial frontal areas are a result of the task becoming more complex (i.e., involving more subtasks) and requiring quicker shifts from one subtask to the next, and not a result of a need to divide attention between two sensory streams.

The dorsolateral frontal activity increases during divided attention could also be explained by the recruitment of working memory when two tasks need to be performed simultaneously (Johnson et al., 2007). In our divided attention condition, the participants most likely had to maintain one sentence in a working memory buffer while making a congruence judgment concerning the other simultaneously presented sentence, whereas in the single-task condition no such demands were placed on working memory. In other words, the participants, at least some of them, may have adopted a rehearsal strategy during the divided attention task but not during the single-tasks. This could have led to the observed frontal activity increase, since the role of the dorsolateral prefrontal cortex in working memory (D'Esposito et al., 1995; Petrides, 2000) and more specifically in subvocal rehearsal (Awh et al., 1996) is well known. An experimental design specifically aimed at teasing apart the effects of increasing working memory load, divided attention, and overall task difficulty would be needed in order to determine the primary role of the dorsolateral prefrontal cortex in dual-tasking paradigms.

When interpreting our results with regards to dual tasking, it is important to note that the participants may not have been performing the divided attention task as the experimenters intended.

For example, the participants may have been attending to just one modality in the divided attention condition. Our behavioral results indicate that this is most likely not the case, however, because attending selectively to only one modality and performing at guess level for the other modality would have resulted in a response accuracy of 63–75%, a rate which our participants far surpassed. In addition, almost all participants performed at a high level of accuracy irrespective of the presentation modality of the incongruent sentence, demonstrating that there was no clear tendency to attend to just one modality. Another strategy used by our participants might have been to first attend to the written text and then switch to the unattended spoken sentence stored in a short-term memory (Norman, 1969), thus not really dual tasking but switching between the two tasks. The use of such a strategy might explain the increased parietal activity during dual tasking, as parietal regions have been shown to be involved in the voluntary shifting of attention between vision and audition (Shomstein and Yantis, 2004). On the other hand, it seems unlikely that such a strategy could have been used successfully in our experiment due to the fast pace of stimulus presentation. Participants had a total of 3.5 s per trial. The average length of the text sentences was 55 characters, which takes around 2.5 s to read at the average reading speed of Finnish text (Hahn et al., 2006). It is therefore unlikely that participants had had enough time to read out the spoken sentence from a short-term memory buffer after reading the text sentence, as subvocal rehearsal of auditory phonological material occurs in real time (Baddeley, 1992). In addition, if the text sentence was evaluated first, the participants would have likely detected incongruent written sentences significantly more accurately than incongruent spoken sentences, but according to our behavioral results this was not the case. As a final possible task strategy during dual tasking, our participants may have converted the written sentences they read into subvocalized speech rehearsed in the articulatory-phonological loop (Baddeley, 1992). If this indeed were the case, our divided attention task would not have been truly a bimodal one. However, even if this were the case, our main findings regarding brain activity associated with dual tasking would not be undermined, because our participants would still have been performing two tasks simultaneously albeit mainly in the same (auditory) modality.

ATTENTION EFFECTS IN THE SENSORY CORTICES

During bimodal stimulation when the participants were attending to just one sensory modality, the sensory cortical areas subserving the attended modality showed increased activity and the ignored sensory cortices showed a decrease in activity compared with when attention was directed to the other modality. This result is in accordance with previous studies showing a similar interaction between the attended modality and activity in the relevant sensory cortices (Shomstein and Yantis, 2004; Johnson and Zatorre, 2005, 2006; Salo et al., 2013).

The visual cortex was shown to be activated to the same extent during divided attention as during attention to visual stimuli in both the unimodal and selective attention conditions, and this activity was greater than when attention was directed to the auditory modality. An analogous pattern of results was observed for the auditory cortex. This result is in contrast to our initial hypothesis: Since several previous studies suggest that a common attentional resource is shared between the sensory modalities (Just et al., 2001; Loose et al., 2003; Johnson and Zatorre, 2006) we expected to see a decrease in sensory-cortex activations during divided attention in relation to auditory or visual selective attention. Our results also indicated that the addition of a distractor stimulus to the unattended modality did not affect activity in the sensory cortical areas subserving the attended modality, even though activity in the cortical areas processing the unattended stimuli increased significantly. If a common attentional resource were indeed shared among the different modalities, this would mean that no resources were allotted to the unattended modality. This would, however, make it difficult to account for the performance accuracy decrease seen in the selective attention condition compared with the unimodal condition. Therefore our results do not support the notion of a constrained total amount of attentional resources being spread out to all recruited sensory cortices.

ACTIVITY RELATED TO SEMANTIC PROCESSING DURING DIVIDED ATTENTION

When only single-task conditions were examined, contrasting incongruent sentences with congruent sentences revealed an increase of activity in bilateral inferior frontal clusters for the written sentences, and in inferior frontal and temporal clusters for the spoken sentences. These foci of activity are well in line with the existing literature describing the role temporal and frontal areas (especially in the left hemisphere) in both semantic and syntactic language-related processing (Friederici et al., 2003; Hickok and Poeppel, 2004; for a review, see Vigneau et al., 2006). The increased activity in these areas in response to incongruent sentence endings is possibly due to the difficulty of integrating the unexpected last word to the preceding information, resulting in increased processing costs (Kutas and Hillyard, 1980). In electrophysiological studies, semantic integration was reflected as an increase in the amplitude of a specific ERP component, the N400 (Kutas and Hillyard, 1980; for reviews, see Kutas and Federmeier, 2000; Lau et al., 2008). The temporal activity clusters observed for the spoken sentences in our study is a likely candidate source for the N400 component (Humphries et al., 2006). The observed temporal activity could also be related to another ERP component, the phonological mismatch negativity (PMN; Connolly and Phillips, 1994), which is elicited when the initial phoneme of the last word in a sentence does not match the phoneme of the expected word (as was the case in our experiment). This component is elicited only when sentences are presented in the auditory modality, and it has been localized to the anterior superior temporal cortex predominantly in the left hemisphere (Kujala et al., 2004), and would therefore explain why we observed the temporal activity clusters only for the spoken sentences.

The IFG was activated bilaterally by both written and spoken incongruent sentences, this effect being stronger in the left than the right hemisphere. The important role of the IFG in processing the semantic content of linguistic stimuli has been demonstrated in previous studies (Baumgaertner et al., 2002; Kiehl et al., 2002). The IFG does not seem to contribute to the N400 component, however, as lesions to frontal areas including the IFG do not affect the N400 component (Friederici et al., 1999). Our results therefore add to the discrepancy between hemodynamic and electrophysiological studies describing the contribution of the IFG to semantic processing. Our study makes a valuable contribution to this debate, since we used both written and spoken sentence stimuli in the same study, and show that the IFG was activated for incongruent sentences irrespective of the presentation modality.

When two tasks that occupy a common part of the cortex are performed simultaneously, interference can occur at the level of these common regions (Roland and Zilles, 1998). In the case of our experiment, ROI analyses were conducted in the semantic ROIs (i.e., bilaterally in the IFG) during divided attention in order to study task interference more carefully. During divided attention, participants had to make two simultaneous or consecutive congruence judgments, presumably both relying on the same amodal semantic processing areas. When the overall activity in the semantic ROIs was examined, our results pointed to an increase in activity during divided attention when compared with the unimodal and selective attention conditions. This suggests that more demands were placed on semantic processing areas when two semantic tasks were performed in parallel, which possibly contributed to the observed performance decrements.

It is important to take into consideration the possibility that incongruent sentences elicited more IFG activity due to other cognitive functions than semantic processing. For example, it has been shown that the IFG is activated when prepotent responses are inhibited (Menon et al., 2001; Aron et al., 2004). Reading or listening to sentences where an anomaly occurs at the very end may create a situation where a response that the sentence is congruent is always chosen first, but then has to be inhibited and replaced by a new response when an anomaly is detected. This may explain the observed IFG activity enhancements. Yet another possible explanation relates to the observation that the IFG is involved with the detection of salient stimuli irrespective of task type (Hampshire et al., 2010). Sentences with semantic violations may represent such an unexpected and salient stimulus, thus involving the IFG.

CONCLUSIONS

The participants of our study performed significantly more errors when they had to make two simultaneous sentence congruence judgments in separate modalities than when they performed just one such judgment in one modality. This dual-task interference could potentially be caused by mutual inhibition of the sensory cortices, or by the recruitment of additional cortical areas responsible for additional cognitive operations related to dual-tasking, or by interference of the two tasks because they utilize the same part of the cortex. Our results indicate that crossmodal inhibition of the sensory cortices is not responsible for the observed performance decrements, and that no dual-task-specific areas are recruited when attention is divided between two simultaneous semantic tasks involving parallel attention to speech and written text. Competition for resources in cortical areas used by both component tasks most likely contributes to dual-tasking interference.

ACKNOWLEDGMENTS

This study was supported by the Academy of Finland grants #260054, #259752 and #273147, the aivoAALTO project, Riksbankens Jubileumsfond, the Academy of Finland Mind Program project “Mind the Gap between digital natives and educational practices” (grant #265528), and the Päivikki and Sakari Sohlberg Foundation.

REFERENCES

Adcock, R. A. R., Constable, T., Gore, J. C., and Goldman-Rakic, P. S. (2000). Functional neuroanatomy of executive processes involved in dual-task performance. *Proc. Natl. Acad. Sci. U S A* 97, 3567–3572. doi: 10.1073/pnas.97.7.3567

Aron, A. R., Robbins, T. W., and Poldrack, R. A. (2004). Inhibition and the right inferior frontal cortex. *Trends Cogn. Sci.* 8, 170–177. doi: 10.1016/j.tics.2004.02.010

Awh, E., Jonides, J., Smith, E. E., Schumacher, E. H., Koeppe, R., and Katz, S. (1996). Dissociation of storage and rehearsal in verbal working memory: evidence from positron emission tomography. *Psychol. Sci.* 7, 25–31. doi: 10.1111/j.1467-9280.1996.tb00662.x

Baddeley, A. (1992). Working memory: the interface between memory and cognition. *J. Cogn. Neurosci.* 4, 281–288. doi: 10.1162/jocn.1992.4.3.281

Baddeley, A. D., and Hitch, G. (1974). “Working memory,” in *The Psychology of Learning and Motivation: Advances in Research and Theory*, ed G. Bower (New York, NY: Academic Press), 47–90.

Baumgaertner, A., Weiller, C., and Büchel, C. (2002). Event-related fMRI reveals cortical sites involved in contextual sentence integration. *Neuroimage* 16, 736–745. doi: 10.1006/nimg.2002.1134

Bunge, S. A., Klingberg, T., Jacobsen, R. B., and Gabrieli, J. D. (2000). A resource model of the neural basis of executive working memory. *Proc. Natl. Acad. Sci. U S A* 97, 3573–3578. doi: 10.1073/pnas.97.7.3573

Cardillo, E. R., Aydelott, J., Matthews, P. M., and Devlin, J. T. (2004). Left inferior prefrontal cortex activity reflects inhibitory rather than facilitatory priming. *J. Cogn. Neurosci.* 16, 1552–1561. doi: 10.1162/0898929042568523

Collette, F., Olivier, L., Van der Linden, M., Laureys, S., Delfiore, G., Luxen, A., et al. (2005). Involvement of both prefrontal and inferior parietal cortex in dual-task performance. *Brain Res. Cogn. Brain Res.* 24, 237–251. doi: 10.1016/j.cogbrainres.2005.01.023

Connolly, J., and Phillips, N. (1994). Event-related potential components reflect phonological and semantic processing of the terminal word of spoken sentences. *J. Cogn. Neurosci.* 6, 256–266. doi: 10.1162/jocn.1994.6.3.256

Corbetta, M., Miezin, F. M., Dobmeyer, S., Shulman, G. L., and Petersen, S. E. (1991). Selective and divided attention during visual discriminations of shape, color and speed: functional anatomy by positron emission tomography. *J. Neurosci.* 11, 2383–2402.

Corbetta, M., and Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nat. Rev. Neurosci.* 3, 201–215. doi: 10.1038/nrn755

Crottaz-Herbette, S., Anagnoson, R. T., and Menon, V. (2004). Modality effects in verbal working memory: differential prefrontal and parietal responses to auditory and visual stimuli. *Neuroimage* 21, 340–351. doi: 10.1016/j.neuroimage.2003.09.019

D'Esposito, M., Detre, J. A., Alsop, D. C., Shin, R. K., Atlas, S., and Grossman, M. (1995). The neural basis of the central executive system of working memory. *Nature* 378, 279–281. doi: 10.1038/378279a0

Downar, J., Crawley, A. P., Mikulis, D. J., and Davis, K. D. (2000). A multimodal cortical network for the detection of changes in the sensory environment. *Nat. Neurosci.* 3, 277–283. doi: 10.1038/72991

Duncan, J. (2010). The multiple-demand (MD) system of the primate brain: mental programs for intelligent behaviour. *Trends Cogn. Sci.* 14, 172–179. doi: 10.1016/j.tics.2010.01.004

Ellis, D. P. W. (2010). "Time-domain scrambling of audio signals in Matlab". Available online at: http://www.ee.columbia.edu/~dpwe/resources/matlab/scramble/

Friederici, A. D., Rüschemeyer, S. A., Hahne, A., and Fiebach, C. J. (2003). The role of left inferior frontal and superior temporal cortex in sentence comprehension: localizing syntactic and semantic processes. *Cereb. Cortex* 13, 170–177. doi: 10.1093/cercor/13.2.170

Friederici, A. D., von Cramon, D. Y., and Kotz, S. A. (1999). Language related brain potentials in patients with cortical and subcortical left hemisphere lesions. *Brain* 122, 1033–1047. doi: 10.1093/brain/122.6.1033

Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J. P., Frith, C. D., and Frackowiak, R. S. (1994a). Statistical parametric maps in functional imaging: a general linear approach. *Hum. Brain Mapp.* 2, 189–210. doi: 10.1002/hbm.460020402

Friston, K. J., Worsley, K. J., Frackowiak, R. S. J., Mazziotta, J. C., and Evans, A. C. (1994b). Assessing the significance of focal activations using their spatial extent. *Hum. Brain Mapp.* 1, 210–220. doi: 10.1002/hbm.460010306

Hahn, G. A., Penka, D., Gehrlich, C., Messias, A., Weismann, M., Hyvärinen, L., et al. (2006). New standardised texts for assessing reading performance in four European languages. *Br. J. Ophthalmol.* 90, 480–484. doi: 10.1136/bjo.2005.087379

Hampshire, A., Chamberlain, S. R., Monti, M. M., Duncan, J., and Owen, A. M. (2010). The role of the right inferior frontal gyrus: inhibition and attentional control. *Neuroimage* 50, 1313–1319. doi: 10.1016/j.neuroimage.2009.12.109

Hein, G., Alink, A., Kleinschmidt, A., and Müller, N. G. (2007). Competing neural responses for auditory and visual decisions. *PLoS One* 2:e320. doi: 10.1371/journal.pone.0000320

Herath, P., Klingberg, T., Young, J., Amunts, K., and Roland, P. (2001). Neural correlates of dual task interference can be dissociated from those of divided attention: an fMRI study. *Cereb. Cortex* 11, 796–805. doi: 10.1093/cercor/11.9.796

Hickok, G., and Poeppel, D. (2004). Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition* 92, 67–99. doi: 10.1016/j.cognition.2003.10.011

Humphries, C., Binder, J. R., Medler, D. A., and Liebenthal, E. (2006). Syntactic and semantic modulation of neural activity during auditory sentence comprehension. *J. Cogn. Neurosci.* 18, 665–679. doi: 10.1162/jocn.2006.18.4.665

Humphries, C., Binder, J. R., Medler, D. A., and Liebenthal, E. (2007). Time course of semantic processes during sentence comprehension: an fMRI study. *Neuroimage* 36, 924–932. doi: 10.1016/j.neuroimage.2007.03.059

Johnson, J. A., Strafella, A. P., and Zatorre, R. J. (2007). The role of the dorsolateral prefrontal cortex in bimodal divided attention: two transcranial magnetic stimulation studies. *J. Cogn. Neurosci.* 19, 907–920. doi: 10.1162/jocn.2007.19.6.907

Johnson, J. A., and Zatorre, R. J. (2005). Attention to simultaneous unrelated auditory and visual events: behavioral and neural correlates. *Cereb. Cortex* 15, 1609–1620. doi: 10.1093/cercor/bhi039

Johnson, J. A., and Zatorre, R. J. (2006). Neural substrates for dividing and focusing attention between simultaneous auditory and visual events. *Neuroimage* 31, 1673–1681. doi: 10.1016/j.neuroimage.2006.02.026

Just, M. A., Carpenter, P. A., Keller, T. A., Emery, L., Zajac, H., and Thulborn, K. R. (2001). Interdependence of non-overlapping cortical systems in dual cognitive tasks. *Neuroimage* 14, 417–426. doi: 10.1006/nimg.2001.0826

Kiehl, K. A., Laurens, K. R., and Liddle, P. F. (2002). Reading anomalous sentences: an event-related fMRI study of semantic processing. *Neuroimage* 17, 842–850. doi: 10.1006/nimg.2002.1244

Klingberg, T. (1998). Concurrent performance of two working memory tasks: potential mechanisms of interference. *Cereb. Cortex* 8, 593–601. doi: 10.1093/cercor/8.7.593

Kujala, A., Alho, K., Service, E., Ilmoniemi, R. J., and Connolly, J. F. (2004). Activation in the anterior left auditory cortex associated with phonological analysis of speech input: localization of the phonological mismatch negativity response with MEG. *Brain Res. Cogn. Brain Res.* 21, 106–113. doi: 10.1016/j.cogbrainres.2004.05.011

Kuperberg, G. R., Holcomb, P. J., Sitnikova, T., Greve, D., Dale, A. M., and Caplan, D. (2003). Distinct patterns of neural modulation during the processing of conceptual and syntactic anomalies. *J. Cogn. Neurosci.* 15, 272–293. doi: 10.1162/089892903321208204

Kutas, M., and Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends Cogn. Sci.* 4, 463–470. doi: 10.1016/s1364-6613(00)01560-6

Kutas, M., and Hillyard, S. A. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science* 207, 203–205. doi: 10.1126/science.7350657

Lau, E. F., Phillips, C., and Poeppel, D. (2008). A cortical network for semantics: (de)constructing the N400. *Nat. Rev. Neurosci.* 9, 920–933. doi: 10.1038/nrn2532

Laurienti, P. J., Burdette, J. H., Wallace, M. T., Yen, Y. F., Field, A. S., and Stein, B. E. (2002). Deactivation of sensory-specific cortex by cross-modal stimuli. *J. Cogn. Neurosci.* 14, 420–429. doi: 10.1162/089892902317361930

Loose, R., Kaufmann, C., Auer, D. P., and Lange, K. W. (2003). Human prefrontal and sensory cortical activity during divided attention tasks. *Hum. Brain Mapp.* 18, 249–259. doi: 10.1002/hbm.10082

Menon, V., Adleman, N. E., White, C. D., Glover, G. H., and Reiss, A. L. (2001). Error-related brain activation during a Go/NoGo response inhibition task. *Hum. Brain Mapp.* 12, 131–143. doi: 10.1002/1097-0193(200103)12:3<131::aid-hbm1010>3.0.co;2-c

Miller, E. K., and Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.* 24, 167–202. doi: 10.1146/annurev.neuro.24.1.167

Mittag, M., Inauri, K., Huovilainen, T., Leminen, M., Salo, E., Rinne, T., et al. (2013). Attention effects on the processing of task-relevant and task-irrelevant speech sounds and letters. *Front. Neurosci.* 7:231. doi: 10.3389/fnins.2013.00231

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., and Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: a latent variable analysis. *Cogn. Psychol.* 41, 49–100. doi: 10.1006/cogp.1999.0734

Näätänen, R. (1992). *Attention and Brain Function.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Ni, W., Constable, R., Mencl, W., Pugh, K., Fulbright, R., Shaywitz, S., et al. (2000). An event-related neuroimaging study distinguishing form and content in sentence processing. *J. Cogn. Neurosci.* 12, 120–133. doi: 10.1162/08989290051137648

Nijboer, M., Borst, J., van Rijn, H., and Taatgen, N. (2014). Single-task fMRI overlap predicts concurrent multitasking interference. *Neuroimage* 100, 60–74. doi: 10.1016/j.neuroimage.2014.05.082

Norman, D. A. (1969). Memory while shadowing. *Q. J. Exp. Psychol.* 21, 85–93. doi: 10.1080/14640746908400200

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

Petrides, M. (2000). The role of the mid-dorsolateral prefrontal cortex in working memory. *Exp. Brain Res.* 133, 44–54. doi: 10.1007/s002210000399

Ridderinkhof, K. R., Ullsperger, M., Crone, E. A., and Nieuwenhuis, S. (2004). The role of the medial frontal cortex in cognitive control. *Science* 306, 443–447. doi: 10.1126/science.1100301

Roland, P. E., and Zilles, K. (1998). Structural divisions and functional fields in the human cerebral cortex. *Brain Res. Brain Res. Rev.* 26, 87–105. doi: 10.1016/s0165-0173(97)00058-1

Salmi, J., Rinne, T., Koistinen, S., Salonen, O., and Alho, K. (2009). Brain networks of bottom-up triggered and top-down controlled shifting of auditory attention. *Brain Res.* 1286, 155–164. doi: 10.1016/j.brainres.2009.06.083

Salo, E., Rinne, T., Salonen, O., and Alho, K. (2013). Brain activity during auditory and visual phonological, spatial and simple discrimination tasks. *Brain Res.* 1496, 55–69. doi: 10.1016/j.brainres.2012.12.013

Schubert, T., and Szameitat, A. J. (2003). Functional neuroanatomy of interference in overlapping dual tasks: an fMRI study. *Brain Res. Cogn. Brain Res.* 17, 733–746. doi: 10.1016/s0926-6410(03)00198-8

Service, E., Helenius, P., Maury, S., and Salmelin, R. (2007). Localization of syntactic and semantic brain responses using magnetoencephalography. *J. Cogn. Neurosci.* 19, 1193–1205. doi: 10.1162/jocn.2007.19.7.1193

Shomstein, S., and Yantis, S. (2004). Control of attention shifts between vision and audition in human cortex. *J. Neurosci.* 24, 10702–10706. doi: 10.1523/jneurosci.2939-04.2004

Staines, W. R., Graham, S. J., Black, S. E., and McIlroy, W. E. (2002). Task-relevant modulation of contralateral and ipsilateral primary somatosensory cortex and the role of a prefrontal-cortical sensory gating system. *Neuroimage* 15, 190–199. doi: 10.1006/nimg.2001.0953

Stelzel, C., Schumacher, E. H., Schubert, T., and D'Esposito, M. (2006). The neural effect of stimulus-response modality compatibility on dual-task performance: an fMRI study. *Psychol. Res.* 70, 514–525. doi: 10.1007/s00426-005-0013-7

Szameitat, A. J., Schubert, T., Müller, K., and Von Cramon, D. Y. (2002). Localization of executive functions in dual-task performance with fMRI. *J. Cogn. Neurosci.* 14, 1184–1199. doi: 10.1162/089892902760807195

Vigneau, M., Beaucousin, V., Herve, P. Y., Duffau, H., Crivello, F., Houde, O., et al. (2006). Meta-analyzing left hemisphere language areas: phonology, semantics and sentence processing. *Neuroimage* 30, 1414–1432. doi: 10.1016/j.neuroimage.2005.11.002

Yoo, S. S., Paralkar, G., and Panych, L. P. (2004). Neural substrates associated with the concurrent performance of dual working memory tasks. *Int. J. Neurosci.* 114, 613–631. doi: 10.1080/00207450490430561

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

Received: 17 October 2014; paper pending published: 17 January 2015; accepted: 03 February 2015; published online: 19 February 2015.

Citation: Moisala M, Salmela V, Salo E, Carlson S, Vuontela V, Salonen O and Alho K (2015) Brain activity during divided and selective attention to auditory and visual sentence comprehension tasks. Front. Hum. Neurosci. ***9****:86. doi: 10.3389/fnhum.2015.00086*

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

frontiers in
HUMAN NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 26 February 2015
doi: 10.3389/fnhum.2015.00102

Brain activations during bimodal dual tasks depend on the nature and combination of component tasks

Emma Salo[1,2]*, Teemu Rinne[1], Oili Salonen[3] and Kimmo Alho[1,4,5]

[1] *Division of Cognitive Psychology and Neuropsychology, Institute of Behavioural Sciences, University of Helsinki, Helsinki, Finland*
[2] *Advanced Magnetic Imaging Centre, Aalto University School of Science and Technology, Espoo, Finland*
[3] *Helsinki Medical Imaging Center, Helsinki University Central Hospital, Helsinki, Finland*
[4] *Helsinki Collegium for Advanced Studies, University of Helsinki, Helsinki, Finland*
[5] *Swedish Collegium for Advanced Study, Uppsala, Sweden*

Edited by:
John J. Foxe, Albert Einstein College of Medicine, USA

Reviewed by:
Jessica A. Bernard, University of Colorado Boulder, USA
Christian Beste, Ruhr Universität Bochum, Germany
Joaquin Alberto Anguera, University of California, San Francisco, USA

****Correspondence:***
Emma Salo, Division of Cognitive and Neuropsychology, Institute of Behavioural Sciences, University of Helsinki, PO Box 9, Helsinki FI 00014, Finland
e-mail: emma.salo@helsinki.fi

We used functional magnetic resonance imaging to investigate brain activations during nine different dual tasks in which the participants were required to simultaneously attend to concurrent streams of spoken syllables and written letters. They performed a phonological, spatial or "simple" (speaker-gender or font-shade) discrimination task within each modality. We expected to find activations associated specifically with dual tasking especially in the frontal and parietal cortices. However, no brain areas showed systematic dual task enhancements common for all dual tasks. Further analysis revealed that dual tasks including component tasks that were according to Baddeley's model "modality atypical," that is, the auditory spatial task or the visual phonological task, were not associated with enhanced frontal activity. In contrast, for other dual tasks, activity specifically associated with dual tasking was found in the left or bilateral frontal cortices. Enhanced activation in parietal areas, however, appeared not to be specifically associated with dual tasking *per se*, but rather with intermodal attention switching. We also expected effects of dual tasking in left frontal supramodal phonological processing areas when both component tasks required phonological processing and in right parietal supramodal spatial processing areas when both tasks required spatial processing. However, no such effects were found during these dual tasks compared with their component tasks performed separately. Taken together, the current results indicate that activations during dual tasks depend in a complex manner on specific demands of component tasks.

Keywords: dual task, divided attention, fMRI, phonological processing, spatial processing

INTRODUCTION

Performing two or more cognitive tasks simultaneously is assumed to require executive functions such as coordination of cognitive resources (Alvarez and Emory, 2006). It has been proposed that brain activity during dual tasks that cannot be associated with either of the component tasks would reflect such functions. Previous functional magnetic resonance imaging (fMRI) studies suggest that dual tasking would activate prefrontal cortical areas involved in coordination of limited processing resources (Corbetta and Shulman, 2002; Schubert and Szameitat, 2003; Johnson and Zatorre, 2006; Stelzel et al., 2006; Johnson et al., 2007) and posterior parietal cortical areas involved in control (e.g., shifting) of attention (Corbetta and Shulman, 2002; Shomstein and Yantis, 2004; Corbetta et al., 2008).

However, performing two cognitively demanding tasks simultaneously may deteriorate performance in either task or in both tasks. It is generally assumed that dual task performance deteriorates when the component tasks require the same limited sensory, cognitive or cortical resources (Welford, 1952; Mowbray, 1953; Pashler, 1994; Roland and Zilles, 1998; Alais et al., 2006). Furthermore, due to limitations in dividing attention between two sensory modalities, task related activity in the auditory and visual cortex is lower during intermodal divided attention than during auditory or visual selective attention, respectively (Johnson and Zatorre, 2006). Taken together, these results suggest that dual tasking comprises several processes that are not yet fully understood.

Previous studies have used a limited number of task combinations (e.g., two unimodal single tasks and one bimodal dual task) in order to identify activations associated specifically with dual tasking. Therefore, in the present study, we examined the effects of three auditory and three visual component tasks and their nine combinations on brain activity during dual tasking. The dual tasks comprised an auditory phonological, spatial or simple (speaker-gender) discrimination task and a visual phonological, spatial or simple (font-shade) discrimination task. All tasks were performed on identical stimuli and required identical motor responses to targets. The auditory and visual phonological tasks (A_{Phon} and V_{Phon}, respectively), as well as the auditory and visual spatial tasks (A_{Spat} and V_{Spat}, respectively), were designed to be as similar as possible to each other in terms of task requirements. The auditory and visual simple tasks (A_{Simp} and V_{Simp}, respectively), in turn, were designed to require modality specific processing (voice and luminance contrast discrimination, respectively). This design allowed us to study the functional significance of activations associated with different dual tasks. Our previous study (Salo et al., 2013), using identical stimuli and the same participants as the present study, investigated activations associated with the three auditory component tasks and the three visual component tasks when performed as the only task. To evaluate whether all dual tasks activate some common brain areas, we compared the present dual task data with single task data from our previous study.

We expected that especially dual tasks requiring parallel phonological or spatial processing would show strong activation modulations. Our previous study showed that, when performed separately, the A_{Phon} and V_{Phon} tasks activate the same area in the left prefrontal cortex involved in phonological processing and that the A_{Spat} and V_{Spat} tasks activate the same area in the right inferior parietal cortex involved in spatial processing (Salo et al., 2013). Thus, we expected to find strong modulation of activity especially in these areas. Finally, we hypothesized that all dual tasks would show activity enhancements in the same areas of dorsolateral prefrontal cortex and posterior parietal cortex involved in task coordination and control of attention, in addition to some activity decrements in the primary sensory cortices due to intermodally divided attentional resources.

METHODS

PARTICIPANTS

Participants ($N = 15$, 8 female) were native Finnish speakers, between 20 and 35 years of age (mean 25 years). All participants were right handed, had normal hearing, normal or corrected-to-normal vision, and no history of psychiatric or neurological illnesses (all self-reported). An informed written consent was obtained from each participant before the experiment. The experimental protocol was approved by the Ethical Committee of the Hospital District of Helsinki and Uusimaa, Finland. One to three weeks before the present dual task session, all participants had participated in an fMRI session where all component tasks of the present study were performed separately in single task conditions (Salo et al., 2013). In addition, 1–7 days prior the present session the participants took part in a short practice session to familiarize them with dual task instructions.

STIMULI

The exact stimulus parameters are reported in our previous study (Salo et al., 2013). In brief, auditory stimuli consisted of 17 meaningless consonant-vowel and vowel-consonant syllables each having a duration of 250 ms. Seven syllables started with a vowel (ab, ad, ag, ah, ak, ap, at) and 10 started with a consonant (du, fu, ku, lu, mu, nu, pu, ru, su, vu). The syllables were uttered by four female and four male native Finnish speakers. Interaural time difference was used to produce eight spatial locations organized in two spatial categories: central (four locations, ca. 2.5° or 5° to the left or right from midline) and peripheral (four locations, ca. 20° or 25° to the left or right from midline). Visual stimuli (duration 250 ms) consisted of 17 consonant letters (height ca. 0.018°). Participants were required to discriminate between letters with a *name* starting with a vowel and letters with a *name* starting with a consonant (e.g., in English, the name of letter R is pronounced like “are” and thus starts with a vowel, while the name of letter T is pronounced like “tea” and starts with a consonant). The Finnish names of seven chosen consonant letters started with a vowel and ended in a consonant (F, L, M, N, R, S, X) and names of 10 chosen consonant letters started with a consonant and ended in a vowel (B, C, D, G, H, J, K, P, T, V). The letters were presented on a gray background (Red, $R = 128$, Green, $G = 128$, Blue, $B = 128$) in either darker gray (four shades) or a lighter gray (four shades). Moreover, they occurred in eight locations either centrally near the fixation asterisk (four diagonal locations ca. 0.029° from fixation) or more peripherally (four diagonal locations ca. 0.075° from fixation). Asynchronous auditory and visual sequences were presented in bimodal 30 s blocks that alternated with 15.3 s breaks. Within each modality, stimulus onset-to-onset intervals varied randomly between 375 and 625 ms in 10 ms steps (rectangular distribution).

PROCEDURE

The participants were presented with concurrent asynchronous streams of spoken syllables and written letters that varied in their phonological, spatial and modality specific (voice or font shade) features (**Figure 1**). For both modalities, there were three different component tasks. In the auditory phonological, spatial and simple tasks (A_{Phon}, A_{Spat}, and A_{Simp}, respectively), targets were syllables starting with a vowel, syllables presented at more peripheral (left or right) locations, and syllables uttered by a female speaker, respectively. In the visual phonological, spatial and simple tasks (V_{Phon}, V_{Spat}, and V_{Simp}, respectively), targets were letters with their name beginning with a vowel, letters at more peripheral locations, and letters presented with a darker gray than the background, respectively. The component tasks were combined to make nine bimodal dual tasks ($A_{Phon}V_{Phon}$, $A_{Phon}V_{Spat}$, $A_{Phon}V_{Simp}$, $A_{Spat}V_{Phon}$, $A_{Spat}V_{Spat}$, $A_{Spat}V_{Simp}$, $A_{Simp}V_{Phon}$, $A_{Simp}V_{Spat}$, and $A_{Simp}V_{Simp}$). During all dual tasks, the participants were required to focus on a black fixation asterisk constantly shown at the center of the screen and to press a button with their right index finger to the auditory and visual targets as fast as possible. During the breaks, they focused on the fixation asterisk and waited for the next task. Eye position was monitored with an iView X MEyetrack LR long range camera and a matching iView X MEyetrack mirror box (SensoMotoric Instruments, Teltow, Germany).

An instruction chart (including the fixation asterisk) was shown in the middle of the screen for 5 s before the onset of the next block. The chart consisted of two rows and four columns of text (in Finnish). The upper and lower rows of the first column contained a black letter A (for Auditory tasks) and V (for Visual tasks), respectively. The rows of the second column contained “female voice” and “dark letter.” The third column had “vowel beginning” and the fourth column “peripheral” on both rows. The columns of the chart were identical for all dual task combinations, except that the target feature for each modality was indicated with black letters, the text in black on the first row indicating target feature in the auditory modality and the text in black on the second row indicating target feature in the visual modality.

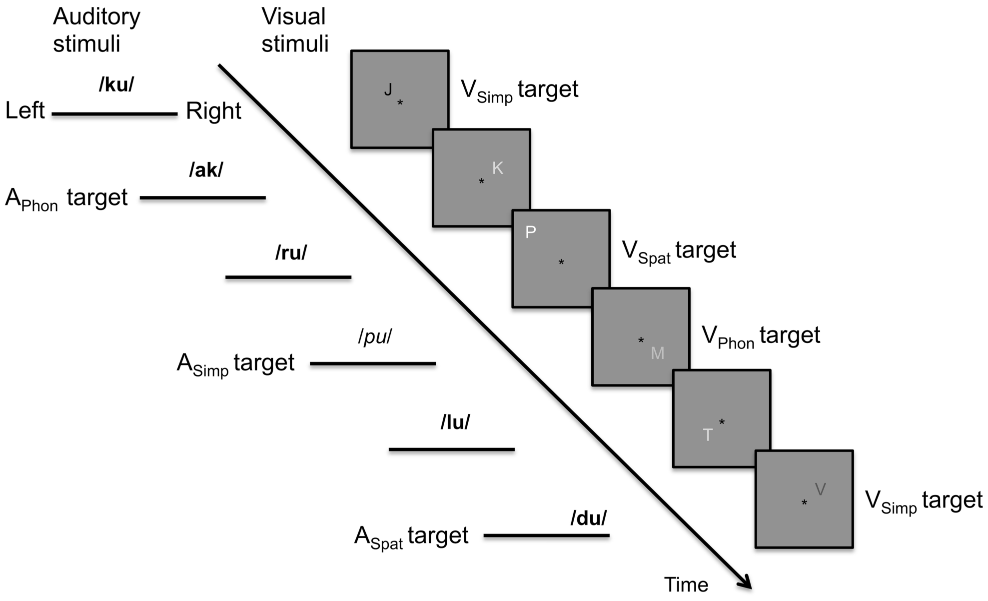


FIGURE 1 | Schematic illustration of the auditory and visual stimulus streams in the present study. In the auditory stream, the targets in phonological, spatial, and simple tasks were syllables starting (vs. ending) with a vowel, presented at the left or right lateral (vs. central) loci or spoken by a female (vs. male) voices. In the figure, female and male voices are indicated in italic and bold font, respectively. In the visual stream, the targets in phonological, spatial, and simple tasks were letters with their name beginning (vs. ending) with a vowel, letters presented farther off (vs. closer to) the fixation asterisk or letters written with a font darker (vs. lighter) than the background.

The other texts were written in gray letters that were darker than the background.

For each of the nine dual task combinations, there were seven blocks and thus altogether 63 blocks were presented. All participants were presented with the same series of 63 stimulus blocks. However, the order of the tasks to be performed in these blocks was randomized separately for each participant. Each block contained 60–80 auditory and 60–80 visual stimuli with a target probability of 0.2 per modality. In both auditory and visual stimulus sequences, stimulus features (17 syllables or letters, 8 auditory and 8 visual locations, 8 font shades and 8 voices, 4 male and 4 female) varied randomly, except that any feature that was used as a target in one of the tasks (i.e., syllables starting with a vowel, peripheral auditory location, female voice, letters with their name starting with a vowel, peripheral visual location, and darker-than-background font) had an independent probability of 0.2. Therefore, a stimulus could contain 0–3 target features, although only one feature was relevant to the task at hand. This allowed us to present similar stimulus blocks during all tasks. The auditory and visual stimuli often overlapped partly in time, but a total overlap was very improbable. The target features of auditory and visual stimulus sequences were randomized independently and thus it was possible that also auditory and visual targets overlapped. In cases where two targets would be presented virtually simultaneously, the participants were instructed to press the response button twice.

ANALYSIS OF THE BEHAVIORAL DATA

In order to minimize effects due to response selection (i.e., due to a response selection bottleneck; Pashler, 1994), responses to auditory and visual targets were given with the same button. Targets were considered in temporal order. The first response occurring within 200–1000 ms from target onset was labeled as a hit. Each response was classified only once. Hit rate (HR) was defined as the number of hits divided by the number of targets. False alarm rate (FaR), in turn, was defined as the number responses given outside the hit response window divided by the overall number of responses.

To compare task performance between single and dual task conditions, mean RTs to auditory targets were calculated for each participant across the dual tasks including the A_{Phon} task ($A_{Phon}V_{Phon}$, $A_{Phon}V_{Spat}$, and $A_{Phon}V_{Simp}$), across the dual tasks including the A_{Spat} task ($A_{Spat}V_{Phon}$, $A_{Spat}V_{Spat}$, and $A_{Spat}V_{Simp}$) and across the dual tasks including the A_{Simp} task ($A_{Simp}V_{Phon}$, $A_{Simp}V_{Spat}$, and $A_{Simp}V_{Simp}$). These mean RTs were then compared with the RT for the corresponding auditory component task performed as a single task in our previous study (Salo et al., 2013). Similarly, mean RTs to visual targets were calculated across the dual tasks including the V_{Phon} task ($A_{Phon}V_{Phon}$, $A_{Spat}V_{Phon}$, $A_{Simp}V_{Phon}$), across the dual tasks including the V_{Spat} task ($A_{Phon}V_{Spat}$, $A_{Spat}V_{Spat}$, $A_{Simp}V_{Spat}$) and across the dual tasks including the V_{Simp} task ($A_{Phon}V_{Simp}$, $A_{Spat}V_{Simp}$, $A_{Simp}V_{Simp}$) and then compared with the RT for the corresponding visual component task performed as a single task. Similar comparisons were made for each participant's HRs and FaRs.

In the ANOVAs, the degrees of freedom were Greenhouse-Geisser corrected when needed. However, the original degrees of freedom will be reported below together with the corrected *P*-value. The reported correction term ε implicates corrections.

fMRI DATA ACQUISITION AND ANALYSIS

Functional brain imaging was carried out with a 3.0 T GE Signa MRI scanner (GE Medical Systems, USA) using an eight channel head coil. The functional echo planar (EPI) images were acquired with an imaging area consisting of 31 contiguous oblique axial slices (TR 2000 ms, TE 32 ms, flip angle 90°, voxel matrix 64 × 64, field of view 22 cm, slice thickness 3.0 mm, in-plane resolution 3.4 × 3.4 × 3.0 mm). Image acquisition was independent of stimulation, that is, jittered acquisition was used.

A total of 1436 functional volumes were obtained in one 48 min session. Immediately after the functional scan, a fluid attenuated inversion recovery image using the same image slices but with a denser in-plane resolution was acquired for anatomical co-alignment (FLAIR; TR 10000 ms, TE 120 ms, voxel matrix 320 × 192, field of view 22 cm, slice thickness 3.0 mm, in-plane resolution 0.7 × 1.1 mm). High-resolution anatomical images (voxel matrix 256 × 256, slice thickness 1.0 mm, in-plane resolution 1 × 1 mm) were acquired in a preceding session (Salo et al., 2013).

The data were analyzed with FSL (4.1.0, www.fmrib.ox.ac.uk/fsl) using one general linear model (GLM) with 10 explanatory variables (nine different tasks and instruction). The first four volumes of the session were excluded from analysis. The data were motion corrected, spatially smoothed (7 mm full width half maximum), and high pass filtered (cutoff 100 s). The hemodynamic response was modeled using a gamma function (mean lag 6 s, SD 3 s) and its temporal derivative. Several contrasts were defined to compare activations during dual tasks with those during the dual task baseline. For group (mixed effects) analysis, the results of lower level analyses were transformed into a standard space (MNI152; Montreal Neurological Institute). Z-statistic images were thresholded using clusters determined by $Z > 2.3$ and a (corrected) cluster significance threshold of $P < 0.05$ (using Gaussian random field theory).

COMPARISON OF DUAL TASK AND SINGLE TASK DATA

Activity increments and decrements associated with dual tasks were investigated by comparing activity during the present dual tasks with activity during the corresponding single tasks measured in our previous study (Salo et al., 2013). These comparisons (fixed effects) were conducted in the space of each participant's high resolution anatomical image followed by group analysis (mixed effects) in the MNI152 space. First we contrasted brain activity during each bimodal dual task with activations during the corresponding auditory single task. We assumed that these contrasts would reveal a combination of activations associated with dual tasking and the visual component of dual task (because the visual stimuli were ignored in the auditory single tasks). Then each dual task was contrasted with the corresponding visual single task to reveal activations associated with dual tasking and with the auditory component of the dual task. The resulting statistic images were then entered into nine conjunction analyses (using the easythresh script) to reveal significant activation enhancements ($Z > 2.3$, cluster corrected $P < 0.05$) specific to dual tasking.

RESULTS

DUAL TASK PERFORMANCE

Reaction times (RTs) and hit rates (HRs) for auditory and visual targets in dual tasks were averaged across dual tasks with a particular auditory or visual component task, respectively. The RTs in the dual tasks were comparable to those in our previous study where each component task was performed as a single task by the same participants (Salo et al., 2013). For auditory RTs, a repeated-measures analysis of variance (ANOVA) with factors Condition (Single task, Dual task) and Task (Phonological, Spatial, Simple) revealed significant main effect of Task $F_{(2,\ 28)} = 7.02$, $P < 0.01$, the RTs being higher for $\mathrm{A_{Spat}}$ than for $\mathrm{A_{Phon}}$ or $\mathrm{A_{Simp}}$, but no significant effect of Condition. Likewise, a similar ANOVA for visual RTs revealed a significant main effect of Task $F_{(2,\ 28)} = 103.56$, $P < 0.001$, but no significant effect of Condition. However, there was a significant Condition × Task interaction $F_{(2,\ 28)} = 38.53$, $P < 0.001$. As seen in **Figure 2** the RTs for $\mathrm{V_{Spat}}$ or $\mathrm{V_{Simp}}$ were longer during dual task conditions than during single task conditions, while the opposite was true for RTs during $\mathrm{V_{Phon}}$.

The HRs, in turn, were lower during the present dual tasks than during the previous single tasks clearly indicating costs of dual tasking. For auditory targets, an ANOVA with factors Condition and Task revealed significant main effects of Condition $F_{(1,\ 14)} = 17.37$, $P < 0.001$ and Task $F_{(2,\ 28)} = 237.88$, $P < 0.001$ and a significant Condition × Task interaction $F_{(2,\ 28)} = 16.01$, $P < 0.001$. The auditory HRs were lower during dual task conditions than during single task conditions. Within both dual and single task conditions, the HRs were lowest for $\mathrm{A_{Spat}}$ and similar for $\mathrm{A_{Phon}}$ and $\mathrm{A_{Simp}}$ and the HRs for $\mathrm{A_{Spat}}$ differed the least between the dual and single task conditions. For visual targets, a similar ANOVA revealed significant main effects of Condition $F_{(1,\ 14)} = 39.94$, $P < 0.001$) and Task $F_{(2,\ 28)} = 21.11$, $P < 0.001$ and a significant Condition × Task interaction $F_{(2,\ 28)} = 4.91$, $P < 0.05$. The visual HRs were lower during dual tasking than during single tasking and highest for $\mathrm{V_{Simp}}$ and lowest for $\mathrm{V_{Phon}}$, and the HR for $\mathrm{V_{Phon}}$ differed the least the dual and single task conditions.

The false alarm rate (FaR) was defined as the number responses given outside the hit response window divided by the overall number of responses. For each participant, the FaR in each dual task was only 7% at the highest and the mean FaR for the nine dual tasks varied between 2% (±0.5%) and 3% (±0.3%). For auditory false alarms, an ANOVA with factors Condition (Single task, Dual task) and Task (Phonological, Spatial, Simple) revealed significant main effect of Task $F_{(2,\ 28)} = 11.61$, $P < 0.01$, $\varepsilon = 0.57$ and a significant Condition × Task interaction $F_{(2,\ 28)} = 14.21$, $P < 0.01$, $\varepsilon = 0.61$. The FaRs were higher for $\mathrm{A_{Spat}}$ component task than for $\mathrm{A_{Phon}}$ or $\mathrm{A_{Simp}}$, this effect being stronger in single task conditions than in dual task conditions.

For visual false alarms, a similar ANOVA revealed significant main effects of Condition $F_{(1,\ 14)} = 40.33$, $P < 0.001$ and Task $F_{(2,\ 28)} = 39.21$, $P < 0.001$, and a significant Condition × Task interaction $F_{(2,\ 28)} = 9.25$, $P < 0.01$, $\varepsilon = 0.63$. The visual FaRs were higher during dual task conditions than during single task conditions. In single task conditions, the FaRs were highest for $\mathrm{V_{Phon}}$, intermediate for $\mathrm{V_{Spat}}$ and lowest for $\mathrm{V_{Simp}}$, whereas in

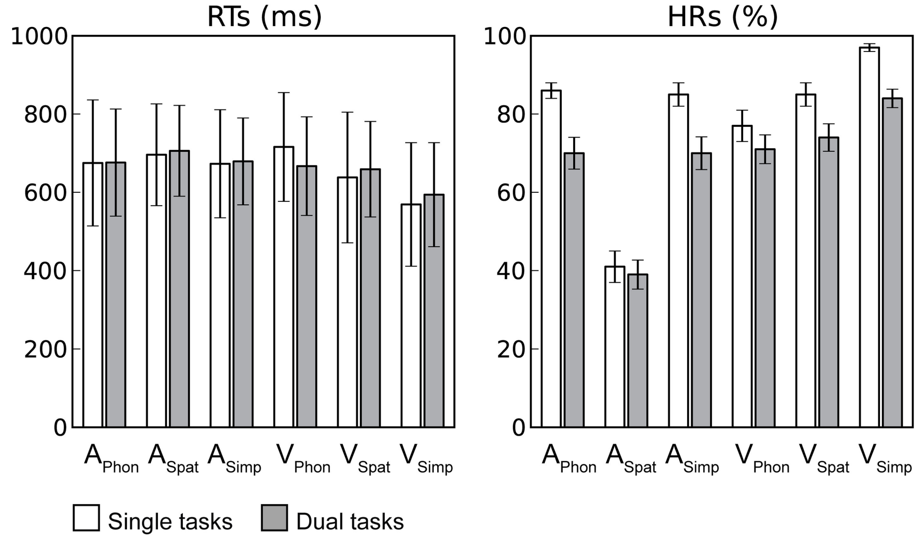


FIGURE 2 | Task performance. Mean reaction times (RTs; includes only hit responses) and hit rates (HRs) during single (white columns) and dual (gray columns) task conditions for each auditory and visual component task. Error bars indicate SEMs (A_{Phon}, A_{Spat}, and A_{Simp} = Auditory phonological, spatial and simple component tasks, respectively and V_{Phon}, V_{Spat}, and V_{Simp} = Visual phonological, spatial and simple component tasks, respectively).

dual task conditions, the FaRs for V_{Phon} and V_{Spat} did not differ and the FaR for V_{Simp} remained lowest.

BRAIN ACTIVITY DURING DUAL TASKS IN RELATION TO THE $A_{Simp}V_{Simp}$ DUAL TASK

$A_{Simp}V_{Simp}$ was used as a baseline dual task with which the other dual tasks were compared. Brain activity enhancements during this baseline dual task in relation to resting periods with no stimuli and during the other dual tasks in relation to this baseline are shown in **Figure 3** ($Z > 2.3$, cluster corrected $P < 0.05$, see also Supplementary Tables 1, 2).

As seen in **Figure 3**, activity enhancements during the other dual tasks in comparison with the baseline dual task showed large variation. In brief, during the $A_{Phon}V_{Phon}$ dual task enhanced activity was detected in a small area of the left middle frontal gyrus (MFG; **Figure 3A**, top row) close to the area activated by both A_{Phon} and V_{Phon} in the single task conditions (Salo et al., 2013). The $A_{Spat}V_{Spat}$ dual task, in turn, showed enhanced activity in the right inferior parietal lobule (IPL; **Figure 3B**, middle row, red areas) activated by both A_{Spat} and V_{Spat} in the single task conditions. In addition, several dual tasks were associated with activity enhancements in the right IPL and in the left or bilateral superior parietal lobule (SPL; **Figure 3**, red areas, see also Supplementary Table 1). Activity decrements, in turn, were observed especially in dual tasks including the V_{Phon} component task in the superior temporal gyrus (STG) and pre- and postcentral gyri in one or both hemispheres depending on the auditory component task (**Figure 3**, top row, blue areas, see also Supplementary Table 2).

To reveal activity enhancements associated systematically with dual tasks including a certain component tasks, additional comparisons were implemented. Mean activations across the dual tasks including the A_{Phon} (i.e., across $A_{Phon}V_{Phon}$,$A_{Phon}V_{Spat}$ and $A_{Phon}V_{Simp}$), A_{Spat}($A_{Spat}V_{Phon}$, $A_{Spat}V_{Spat}$ and $A_{Spat}V_{Simp}$), A_{Simp}($A_{Simp}V_{Phon}$ and $A_{Simp}V_{Spat}$), V_{Phon} ($A_{Phon}V_{Phon}$, $A_{Spat}V_{Phon}$ and $A_{Simp}V_{Phon}$), V_{Spat} ($A_{Phon}V_{Spat}$, $A_{Spat}V_{Spat}$ and $A_{Simp}V_{Spat}$), and V_{Simp} ($A_{Phon}V_{Simp}$ and $A_{Spat}V_{Simp}$) component task were separately contrasted with $A_{Simp}V_{Simp}$, the baseline dual task. The results of these contrasts are shown in **Figure 4**. In brief, all component tasks, except V_{Simp}, were associated with enhanced activity ($Z > 2.3$, cluster corrected $P < 0.05$) in the left posterior parietal cortex. Dual tasks including the A_{Spat} and V_{Spat} component tasks showed enhanced activity bilaterally in SPL and in large areas in the right IPL (**Figure 4**, middle row). Dual tasks including the V_{Phon} task, in turn, showed enhanced activity also in the left MFG and were associated with decreased activity bilaterally in the pre- and postcentral gyri, left SPL and IPL, left posterior STG, and in the right pars opercularis and right middle STG.

ACTIVITY ENHANCEMENTS DURING DUAL TASKS IN RELATION TO THE COMPONENT TASKS

We contrasted the dual tasks with their component tasks performed separately in our previous study (Salo et al., 2013). These comparisons revealed enhanced activations not directly associated with either of the component tasks or with bimodal stimulus presentation. Thus, these activations might be specific to dual tasking. Interestingly, not all dual tasks were associated with such activity enhancements.

Four dual tasks, namely $A_{Phon}V_{Spat}$, $A_{Phon}V_{Simp}$, $A_{Simp}V_{Spat}$, and $A_{Simp}V_{Simp}$, each showed enhanced activity in relation to *both* its auditory component task *and* its visual component task (conjunction analysis, $Z > 2.3$, cluster corrected $P < 0.05$, see also Supplementary Table 3). As seen in **Figure 5**, all these dual tasks were associated with such enhanced activity in the left superior precentral gyrus (for $A_{Simp}V_{Spat}$ and $A_{Simp}V_{Simp}$ there were even two left precentral enhancement clusters). In addition, $A_{Phon}V_{Spat}$ and $A_{Simp}V_{Simp}$ showed enhanced activity in relation to both of their component tasks in the left MFG and $A_{Phon}V_{Simp}$ in

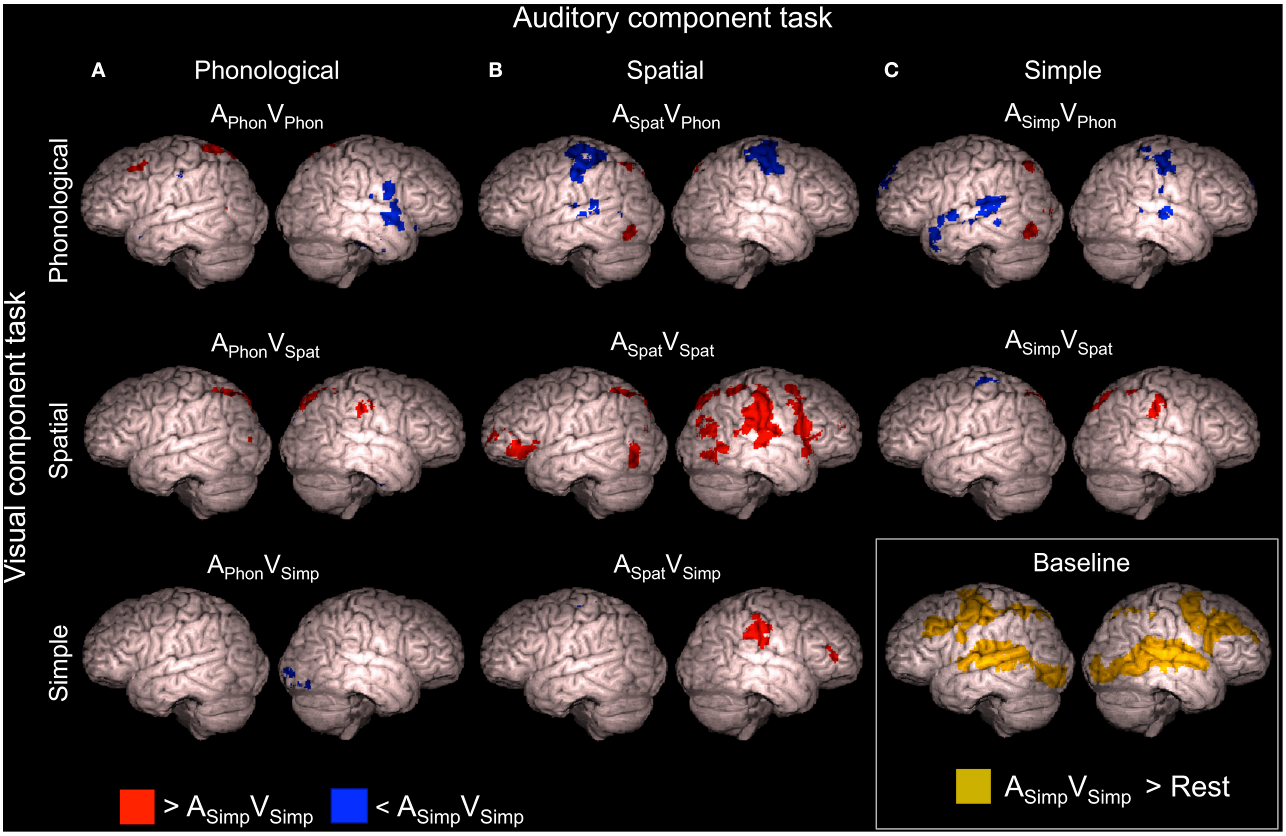


FIGURE 3 | Brain activity during dual tasks in relation to the baseline dual task. Areas showing significant ($Z > 2.3$, cluster corrected $P < 0.05$) activity in relation to the $A_{Simp}V_{Simp}$ dual task (simultaneous speaker-gender and font-shade discrimination task) used as the baseline. Dual tasks including the **(A)** auditory phonological, **(B)** auditory spatial, and **(C)** auditory simple component tasks compared with the baseline dual task. Dual tasks including the visual phonological, visual spatial and visual simple component tasks compared with the baseline dual task are shown in top, middle and bottom rows, respectively. Brain activity during the $A_{Simp}V_{Simp}$ dual task in relation to brain activity during the resting periods is shown in the right bottom corner. Cortical activations are superimposed from 10 mm under the cortex on surface of rendered brain images.

the bilateral MFG. Finally, for both $A_{Phon}V_{Simp}$ and $A_{Simp}V_{Simp}$ there was such activity enhancement even in the right superior precentral gyrus.

ACTIVITY DECREMENTS DURING DUAL TASKS IN RELATION TO THE COMPONENT TASKS

To study activation decrements associated with dual tasking, mean activations across dual tasks including the A_{Phon} task ($A_{Phon}V_{Phon}$, $A_{Phon}V_{Spat}$, and $A_{Phon}V_{Simp}$), A_{Spat} task ($A_{Spat}V_{Phon}$, $A_{Spat}V_{Spat}$, and $A_{Spat}V_{Simp}$) and A_{Simp} task ($A_{Simp}V_{Phon}$, $A_{Simp}V_{Spat}$, and $A_{Simp}V_{Simp}$) were separately contrasted with activity during the corresponding auditory task performed as a single task in our previous study (Salo et al., 2013). As seen in **Figure 6A** (see also Supplementary Table 4), all these comparisons showed significantly decreased activity ($Z > 2.3$, cluster corrected $P < 0.05$) during dual tasking than during auditory single tasking in the left posterior STG. In addition, for the dual tasks including the A_{Simp} component task, activity decreased in relation to A_{Simp} performed as a single task in the right posterior STG and in the ventromedial prefrontal cortex (VMPC; **Figure 6A**, bottom row).

Correspondingly, dual tasks including the V_{Phon} task ($A_{Phon}V_{Phon}$, $A_{Spat}V_{Phon}$, and $A_{Simp}V_{Phon}$), V_{Spat} task ($A_{Phon}V_{Spat}$, $A_{Spat}V_{Spat}$, and $A_{Simp}V_{Spat}$) and V_{Simp} task ($A_{Phon}V_{Simp}$, $A_{Spat}V_{Simp}$, and $A_{Simp}V_{Simp}$) were contrasted with the visual component task performed as a single task. All these comparisons showed significantly lower activity ($Z > 2.3$, cluster corrected $P < 0.05$) during dual tasking than during visual single tasking in the left posterior STG (**Figure 6B** and Supplementary Table 4). For the dual tasks including the V_{Spat} or V_{Simp} component task, activity decreased significantly also in the VMPC (**Figure 6B**, middle and bottom rows).

TASK DIFFICULTY AS COVARIATE

A separate analysis was performed to investigate the possibility that task difficulty as such would explain activity observed during different dual tasks. A behavioral covariate of each participants HR was included as an additional explanatory variable in the general linear model (GLM; for other variables and details, see fMRI data acquisition and analysis). Only the $A_{Spat}V_{Phon}$ dual task showed significant activity enhancements associated with higher HRs ($Z > 2.3$, cluster corrected $P < 0.05$) and only in

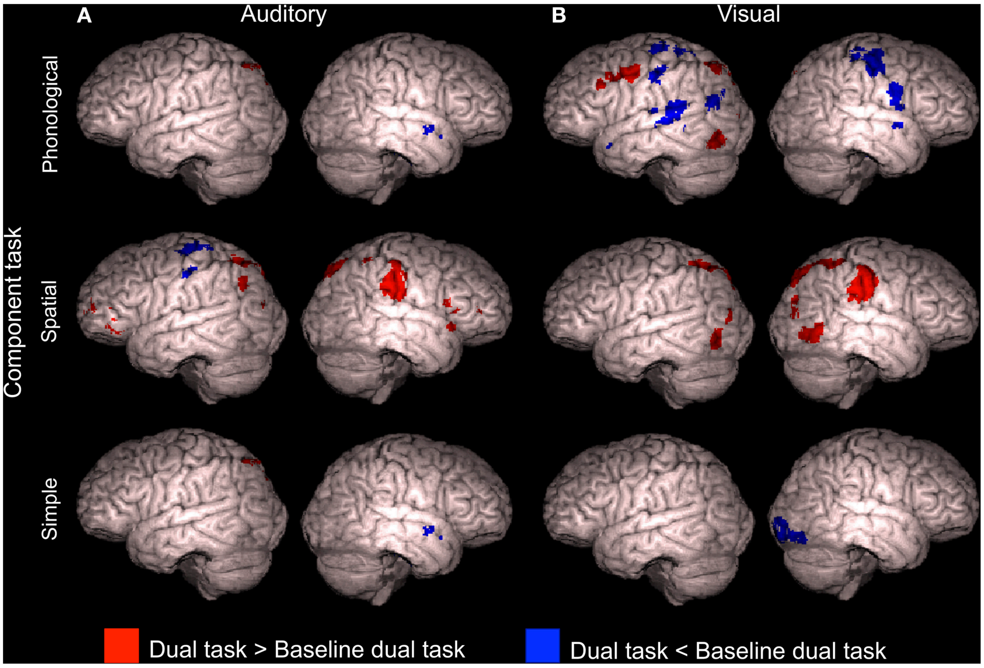


FIGURE 4 | Brain activity during dual tasks including a certain component tasks. Areas showing significant ($Z > 2.3$, cluster corrected $P < 0.05$) activity during dual tasks including **(A)** auditory phonological ($\mathrm{A_{Phon}V_{Phon}}$, $\mathrm{A_{Phon}V_{Spat}}$, and $\mathrm{A_{Phon}V_{Simp}}$), spatial ($\mathrm{A_{Spat}V_{Phon}}$, $\mathrm{A_{Spat}V_{Spat}}$, and $\mathrm{A_{Spat}V_{Simp}}$) or simple ($\mathrm{A_{Simp}V_{Phon}}$ and $\mathrm{A_{Simp}V_{Spat}}$) component task and **(B)** dual tasks including the visual phonological ($\mathrm{A_{Phon}V_{Phon}}$, $\mathrm{A_{Spat}V_{Phon}}$, and $\mathrm{A_{Simp}V_{Phon}}$), spatial ($\mathrm{A_{Phon}V_{Spat}}$, $\mathrm{A_{Spat}V_{Spat}}$, and $\mathrm{A_{Simp}V_{Spat}}$) or simple ($\mathrm{A_{Phon}V_{Simp}}$ and $\mathrm{A_{Spat}V_{Simp}}$) component task compared with $\mathrm{A_{Simp}V_{Simp}}$, the baseline dual task.

the anterior cingulate cortex. However, dual task effects in this model were nearly identical to those in the original analysis. Thus, variation in task difficulty did not systematically affect brain activations during dual tasking.

DISCUSSION

ACTIVITY ENHANCEMENTS DURING DUAL TASKING

We hypothesized that dual task performance is challenging especially when the two tasks require processing in the same brain areas. In particular, we assumed that supramodal phonological and spatial processing areas in the left frontal and right inferior parietal cortex (cf. Salo et al., 2013), respectively, would either show enhanced activation reflecting the double effort needed in the dual tasks where both component tasks are phonological or spatial, or decreased activation reflecting interference of simultaneous auditory and visual phonological or spatial processing. We found that in relation to $\mathrm{A_{Simp}V_{Simp}}$ used as the baseline dual task, both dual tasks requiring overlapping processing showed specific activation enhancements not found for the other dual tasks: $\mathrm{A_{Phon}V_{Phon}}$ was associated with enhanced activity in the left MFG (**Figure 3A**, top row) and $\mathrm{A_{Spat}V_{Spat}}$ with enhanced activity in the right IPL (**Figure 3B**, middle row). These results appear to support the idea that the left frontal phonological areas are the bottleneck for two simultaneous phonological tasks and the right parietal spatial areas are the bottleneck for two simultaneous spatial tasks, and that activity in these bottleneck areas is enhanced when they are recruited by parallel phonological or spatial tasks, respectively.

However, when activations during $\mathrm{A_{Phon}V_{Phon}}$ and $\mathrm{A_{Spat}V_{Spat}}$ were compared with activations during their component tasks, no activity enhancements were found for $\mathrm{A_{Phon}V_{Phon}}$ in the left frontal cortex or for $\mathrm{A_{Spat}V_{Spat}}$ in the right inferior parietal cortex. These results suggest that enhanced left MFG activity during $\mathrm{A_{Phon}V_{Phon}}$ and enhanced right IPL activity during $\mathrm{A_{Spat}V_{Spat}}$ in relation to $\mathrm{A_{Simp}V_{Simp}}$ were simply due to more intensive phonological processing during $\mathrm{A_{Phon}V_{Phon}}$ and more intensive spatial processing during $\mathrm{A_{Spat}V_{Spat}}$ than during $\mathrm{A_{Simp}V_{Simp}}$ where the component tasks were nonphonological and nonspatial.

Additional comparisons investigating activity associated with dual tasks including a certain component task revealed that all dual tasks, except those including the $\mathrm{V_{Simp}}$ component task, were associated with enhanced activity in the left superior parietal cortex (**Figure 4**). Such activations might be explained by processing of the spatially varying auditory and visual stimuli, or by dual tasking in general, since SPL activity is also implicated in cross-modal shifting of attention (Corbetta and Shulman, 2002; Shomstein and Yantis, 2004; Corbetta et al., 2008; Salmi et al., 2009).

The comparisons between dual tasks and their component tasks performed separately as single tasks revealed no activity enhancements that were common for all nine dual tasks. Thus, the present results do not support the assumption that all dual tasks rely on some specific cortical areas. However, based on the present results, it is also clear that not all dual tasks are alike and that activations during a particular dual task depend on the task combination.

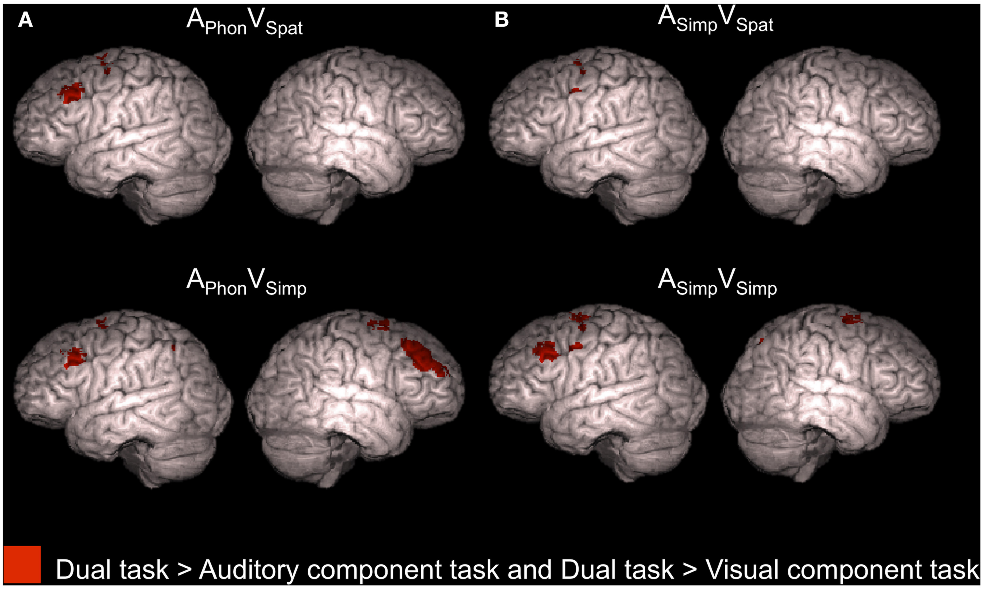


FIGURE 5 | Activity enhancements during dual tasks in relation to single tasks. According to conjunction analyses, the colored areas showed higher activity ($Z > 2.3$, cluster corrected $P < 0.05$) for each of the four dual tasks of the present study in relation to *both* its auditory component task *and* its visual component task when performed separately in our previous study (Salo et al., 2013).

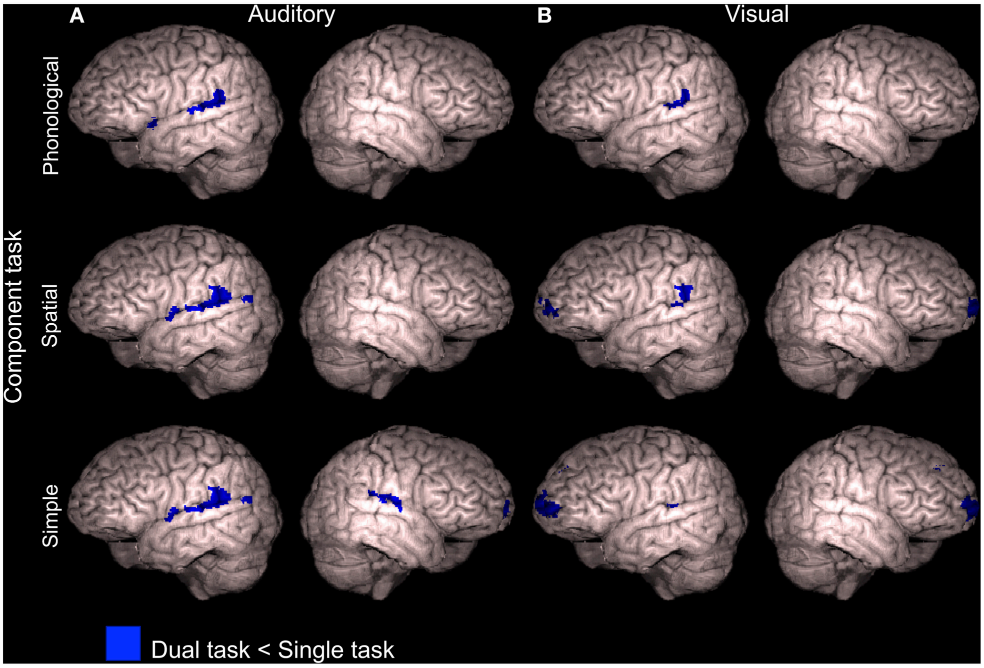


FIGURE 6 | Activity decrements during dual tasks in relation to single tasks. Areas showing lower activity ($Z > 2.3$, cluster corrected $P < 0.05$) during dual tasks than during the component tasks performed separately in our previous study (Salo et al., 2013). **(A)** Dual tasks including the auditory phonological ($A_{Phon}V_{Phon}$, $A_{Phon}V_{Spat}$, and $A_{Phon}V_{Simp}$), spatial ($A_{Spat}V_{Phon}$, $A_{Spat}V_{Spat}$, and $A_{Spat}V_{Simp}$) or simple ($A_{Simp}V_{Phon}$, $A_{Simp}V_{Spat}$, and $A_{Simp}V_{Simp}$) component task compared with corresponding auditory component tasks (A_{Phon}, A_{Spat}, and A_{Simp}, respectively). **(B)** Dual tasks including the visual phonological ($A_{Phon}V_{Phon}$, $A_{Spat}V_{Phon}$, and $A_{Simp}V_{Phon}$), spatial ($A_{Phon}V_{Spat}$, $A_{Spat}V_{Spat}$, and $A_{Simp}V_{Spat}$) or simple ($A_{Phon}V_{Simp}$, $A_{Spat}V_{Simp}$, and $A_{Simp}V_{Simp}$) component task compared with corresponding visual component tasks (V_{Phon}, V_{Spat},and V_{Simp}, respectively). Note that the brain images are tilted 20° to the left or right to reveal ventromedial brain areas.

Previous studies have shown enhanced activity in the dorsolateral prefrontal cortex during dual tasking (Corbetta and Shulman, 2002; Schubert and Szameitat, 2003; Johnson and Zatorre, 2006; Stelzel et al., 2006; Johnson et al., 2007). Consistently, in the present study, we detected activity enhancements in the dorsolateral prefrontal cortex associated with dual tasking during four dual tasks (**Figure 5**). However, even with a more lenient threshold ($Z > 1.6$, nonsignificant), we found no such frontal activations when the dual task included either the A_{Spat} or V_{Phon} component task, or both. According to Baddeley and Hitch (1974), the A_{Spat} task and the V_{Phon} task require *mental modality change*. If prefrontal activity is related to integration of two parallel tasks (Johnson and Zatorre, 2006), then the lack of prefrontal activity enhancements in relation to single tasking during dual tasks including the A_{Spat} or V_{Phon} task, or both, suggests that mental modality change required by these tasks complicated such integration. The complexity and bimodal nature of the modality atypical A_{Spat} and V_{Phon} tasks might have made them highly demanding even during single tasking, since only in these tasks, the hit rates did not markedly decrease during dual tasking in relation to single tasking (see **Figure 2**).

Activation in the superior parietal cortex has been associated with cross-modal and within-modality attention shifts (Corbetta et al., 1993; Bushara et al., 1999; Weeks et al., 1999; Yantis et al., 2002; Giesbrecht et al., 2003; Shomstein and Yantis, 2004, 2006; Salmi et al., 2007, 2009), as well as with goal-oriented attention (Corbetta et al., 2008). Although all present dual tasks presumably required vigorous cross-modal shifting of attention between the auditory and visual tasks, we found no systematic parietal activity enhancements during dual tasks, when compared with the component tasks performed as single tasks (**Figure 5**). Perhaps spatial variation of stimuli in both modalities required shifting of spatial attention in every task and therefore parietal areas were activated already in all single tasks, resulting in weak or no parietal activation differences between the single and dual tasks.

Four dual tasks showed enhanced activation in the left superior precentral gyrus. It is probable, that this activation is related to motor responses. In both single and dual conditions, the participants were instructed to respond with their right hand index finger to targets. In the dual task conditions, the participants were required to attend both modalities simultaneously, and thus the target amount was double compared to the single task conditions.

ACTIVITY DECREMENTS DURING DUAL TASKING

The present dual tasks showed decrements of activity in relation to their component tasks when performed separately in our previous study (Salo et al., 2013). In relation to auditory component tasks, such decrements were detected mainly in the left or bilateral STG (**Figure 6A**). These decrements may have mainly resulted from stronger auditory attention effects during single tasks requiring selective attention to auditory modality than during dual tasks requiring division of attention between two sensory modalities. However, posterior portions of these decrements in the left hemisphere might be related to active suppression of preattentive phonological change detection in these areas (cf. Alho et al., 1998; Celsis et al., 1999) during all present dual task conditions, since left posterior STG/IPL areas showed decreased activity during dual tasking even in relation to the visual single task conditions with task irrelevant varying spoken syllables in the background (**Figure 6B**).

We also found that activity associated with dual tasks including the A_{Simp}, V_{Spat}, or V_{Simp} component task decreased in VMPC during dual tasks. VMPC has been suggested to be involved in suppressing the processing of irrelevant stimuli. This is supported by enhanced activity in VMPC and adjacent areas in response to distracting stimuli (Corbetta and Shulman, 2002; Shomstein and Yantis, 2004; Corbetta et al., 2008; Salmi et al., 2009) and enhanced electrophysiological responses to such distractors in patients with lesions in these areas (Rule et al., 2002). While in single task conditions of our previous study (Salo et al., 2013) there was probably a need to suppress the processing of stimuli in the unattended modality (see also Mittag et al., 2013), in the present dual task conditions active cross-modal suppression would have deteriorated dual task performance. Perhaps therefore there was less suppression related activity in the VMPC and adjacent areas during some dual tasks than during single tasks.

CONCLUSIONS

The present results suggest that dual tasks including two phonological tasks ($A_{Phon}V_{Phon}$) or two spatial tasks ($A_{Spat}V_{Spat}$) are associated with specific activity enhancements in the left frontal cortex (supramodal phonological processing) and in the right inferior parietal cortex (supramodal spatial processing), respectively. Moreover, in congruence with previous studies, we observed that dual tasking with modality typical component tasks is associated with enhanced frontal activity. However, we found no such frontal activity enhancements during dual tasks including a modality atypical task (A_{Spat} or V_{Phon}) and unlike for the modality typical tasks, the hit rates for the modality atypical tasks did not differ markedly between the dual and single task conditions. These results suggest that (single) tasks requiring mental modality change might be as bimodal as audio-visual dual tasks resulting in similar activations in these conditions. Taken together, our results show that all dual tasks do not simply activate the same cortical areas, but task related activations during dual tasking depend on the combination and nature of the component tasks.

ACKNOWLEDGMENTS

This work was supported by the Academy of Finland (grants #210587 and #260054), the National Doctoral Programme of Psychology, and the Research Funds of the University of Helsinki, and Riksbankens Jubileumsfond, Sweden.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://www.frontiersin.org/journal/10.3389/fnhum.2015.00102/abstract

REFERENCES

Alais, D., Morrone, C., and Burr, D. (2006). Separate attentional resources for vision and audition. *Proc. R. Soc. B.* 273, 1339–1345. doi: 10.1098/rspb.2005.3420

Alho, K., Connolly, J. F., Cheour, M., Lehtokoski, A., Huotilainen, M., Virtanen, J., et al. (1998). Hemispheric lateralization in preattentive processing of speech sounds. *Neurosci. Lett.* 258, 9–12. doi: 10.1016/S0304-3940(98)00836-2

Alvarez, J. A., and Emory, E. (2006). Executive function and the frontal lobes: a meta-analytic review. *Neuropsychol. Rev.* 16, 17–42. doi: 10.1007/s11065-006-9002-x

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 (New York, NY: Academic Press), 47–90.

Bushara, K. O., Weeks, R. A., Ishii, K., Catalan, M. J., Tian, B., Rauschecker, J. P., et al. (1999). Modality-specific frontal and parietal areas for auditory and visual spatial localization in humans. *Nat. Neurosci.* 2, 759–766. doi: 10.1038/11239

Celsis, P., Boulanouar, K., Doyon, B., Ranjeva, J. P., Berry, I., Nespoulous, J. L., et al. (1999). Differential fMRI responses in the left posterior superior temporal gyrus and left supramarginal gyrus to habituation and change detection in syllables and tones. *Neuroimage* 9, 135–144. doi: 10.1006/nimg.1998.0389

Corbetta, M., Miezin, F. M., Shulman, G. L., and Petersen, S. E. (1993). A PET study of visuospatial attention. *J. Neurosci.* 13, 1202–1226.

Corbetta, M., Patel, G., and Shulman, G. L. (2008). The reorienting system of the human brain: from environment to theory of mind. *Neuron* 58, 306–324. doi: 10.1016/j.neuron.2008.04.017

Corbetta, M., and Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nat. Rev. Neurosci.* 3, 201–215. doi: 10.1038/nrn755

Giesbrecht, B., Woldorff, M. G., Song, A. W., and Mangun, G. R. (2003). Neural mechanisms of top-down control during spatial and feature attention. *Neuroimage* 19, 496–512. doi: 10.1016/S1053-8119(03)00162-9

Johnson, J. A., Strafella, A. P., and Zatorre, R. J. (2007). The role of the dorsolateral prefrontal cortex in bimodal divided attention: two transcranial magnetic stimulation studies. *J. Cogn. Neurosci.* 19, 907–920. doi: 10.1162/jocn.2007.19.6.907

Johnson, J. A., and Zatorre, R. J. (2006). Neural substrates for dividing and focusing attention between simultaneous auditory and visual events. *Neuromage* 31, 1673–1681. doi: 10.1016/j.neuroimage.2006.02.026

Mittag, M., Inauri, K., Huovilainen, T., Leminen, M., Salo, E., Rinne, T., et al. (2013). Attention effects on the processing of task-relevant and task-irrelevant speech sounds and letters. *Front. Neurosci.* 231, 1–15. doi: 10.3389/fnins.2013.00231

Mowbray, G. H. (1953). Simultaneous vision and audition—the comprehension of prose passages with varying levels of difficulty. *J. Exp. Psychol.* 46, 365–372. doi: 10.1037/h0054574

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

Roland, P. E., and Zilles, K. (1998). Structural divisions and functional fields in the human cerebral cortex. *Brain Res. Rev.* 26, 87–105. doi: 10.1016/S0165-0173(97)00058-1

Rule, R. R., Shimamura, A. P., and Knight, R. T. (2002). Orbitofrontal cortex and dynamic filtering of emotional stimuli. *Cogn. Affect. Behav. Neurosci.* 2, 264–270. doi: 10.3758/CABN.2.3.264

Salmi, J., Rinne, T., Degerman, A., Salonen, O., and Alho, K. (2007). Orienting and maintenance of spatial attention in audition and vision: multimodal and modality-specific brain activations. *Brain Struct. Funct.* 212, 181–194. doi: 10.1007/s00429-007-0152-2

Salmi, J., Rinne, T., Koistinen, S., Salonen, O., and Alho, K. (2009). Brain networks of bottom-up triggered and top-down controlled shifting of auditory attention. *Brain Res.* 1286, 155–164. doi: 10.1016/j.brainres.2009.06.083

Salo, E., Rinne, T., Salonen, O., and Alho, K. (2013). Brain activity during auditory and visual phonological, spatial and simple discrimination tasks. *Brain Res.* 1496, 55–69. doi: 10.1016/j.brainres.2012.12.013

Schubert, T., and Szameitat, A. J. (2003). Functional neuroanatomy of interference in overlapping dual tasks: an fMRI study. *Cogn. Brain Res.* 17, 733–746. doi: 10.1016/S0926-6410(03)00198-8

Shomstein, S., and Yantis, S. (2004). Control of attention shifts between vision and audition in human cortex. *J. Neurosci.* 24, 10702–10706. doi: 10.1523/JNEUROSCI.2939-04.2004

Shomstein, S., and Yantis, S. (2006). Parietal cortex mediates voluntary control of spatial and nonspatial auditory attention. *J. Neurosci.* 26, 435–439. doi: 10.1523/JNEUROSCI.4408-05.2006

Stelzel, C., Schumacher, E. H., Schubert, T., and D'Esposito, M. (2006). The neural effect of stimulus-response modality compatibility on dual-task performance: an fMRI study. *Psychol. Res.* 70, 514–525. doi: 10.1007/s00426-005-0013-7

Weeks, R. A., Aziz-Sultan, A., Bushara, K. O., Tian, B., Wessinger, C. M., Dang, N., et al. (1999). A PET study of human auditory spatial processing. *Neurosci. Lett.* 262, 155–158. doi: 10.1016/S0304-3940(99)00062-2

Welford, A. T. (1952). The 'psychological refractory period' and the timing of high-speed performance—a review and a theory. *Br. J. Psychol. Gen. Sect.* 43, 2–19. doi: 10.1111/j.2044-8295.1952.tb00322.x

Yantis, S., Schwarzbach, J., Serences, J. T., Carlson, R. L., Steinmetz, M. A., Pekar, J. J., et al. (2002). Transient neural activity in human parietal cortex during spatial attention shifts. *Nat. Neurosci.* 5, 995–1002. doi: 10.1038/nn921

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

Received: 24 September 2014; accepted: 10 February 2015; published online: 26 February 2015.

Citation: Salo E, Rinne T, Salonen O and Alho K (2015) Brain activations during bimodal dual tasks depend on the nature and combination of component tasks. Front. Hum. Neurosci. ***9****:102. doi: 10.3389/fnhum.2015.00102*

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

frontiers
in Human Neuroscience

ORIGINAL RESEARCH
published: 16 July 2019
doi: 10.3389/fnhum.2019.00245

Phonological Task Enhances the Frequency-Following Response to Deviant Task-Irrelevant Speech Sounds

Kimmo Alho[1,2]*, *Katarzyna Żarnowiec*[3], *Natàlia Gorina-Careta*[3,4,5] *and Carles Escera*[3,4,5]

[1] *Department of Psychology and Logopedics, Faculty of Medicine, University of Helsinki, Helsinki, Finland,* [2] *Institute of Biomedicine, Paris Descartes University, Paris, France,* [3] *Brainlab-Cognitive Neuroscience Research Group, Department of Clinical Psychology and Psychobiology, Institute of Neurosciences, University of Barcelona, Barcelona, Spain,* [4] *Institute of Neurosciences, University of Barcelona, Barcelona, Spain,* [5] *Institut de Recerca Sant Joan de Déu, Esplugues de Llobregat, Barcelona, Spain*

OPEN ACCESS

Edited by:
Juliana Yordanova,
Institute of Neurobiology (BAS),
Bulgaria

Reviewed by:
Erich Schröger,
Leipzig University, Germany
Gavin M. Bidelman,
University of Memphis, United States
Maija Sisko Peltola,
University of Turku, Finland
Samira Anderson,
University of Maryland, College Park,
United States
Jason Dunlap,
University of Maryland, College Park,
United States
in collaboration with reviewer SA

***Correspondence:**
Kimmo Alho
kimmo.alho@helsinki.fi

Received: *21 December 2018*
Accepted: *01 July 2019*
Published: *16 July 2019*

Citation:
Alho K, Żarnowiec K, Gorina-Careta N and Escera C (2019) Phonological Task Enhances the Frequency-Following Response to Deviant Task-Irrelevant Speech Sounds. Front. Hum. Neurosci. 13:245. doi: 10.3389/fnhum.2019.00245

In electroencephalography (EEG) measurements, processing of periodic sounds in the ascending auditory pathway generates the frequency-following response (FFR) phase-locked to the fundamental frequency (F0) and its harmonics of a sound. We measured FFRs to the steady-state (vowel) part of syllables /ba/ and /aw/ occurring in binaural rapid streams of speech sounds as frequently repeating *standard* syllables or as infrequent (p = 0.2) *deviant* syllables among standard /wa/ syllables. Our aim was to study whether concurrent active phonological processing affects early processing of irrelevant speech sounds reflected by FFRs to these sounds. To this end, during syllable delivery, our healthy adult participants performed tasks involving written letters delivered on a computer screen in a rapid stream. The stream consisted of vowel letters written in red, infrequently occurring consonant letters written in the same color, and infrequently occurring vowel letters written in blue. In the *phonological task*, the participants were instructed to press a response key to the consonant letters differing phonologically but not in color from the frequently occurring red vowels, whereas in the *non-phonological task*, they were instructed to respond to the vowel letters written in blue differing only in color from the frequently occurring red vowels. We observed that the phonological task enhanced responses to deviant /ba/ syllables but not responses to deviant /aw/ syllables. This suggests that active phonological task performance may enhance processing of such small changes in irrelevant speech sounds as the 30-ms difference in the initial formant-transition time between the otherwise identical syllables /ba/ and /wa/ used in the present study.

Keywords: audition, speech, electroencephalography, frequency-following response, phonological task

INTRODUCTION

Baddeley's influential working-memory model (e.g., Baddeley and Hitch, 1974; Baddeley, 1992) proposes that the so-called articulatory-phonological loop underlies auditory working memory and is also involved in processing of written visual inputs. This model was supported by our recent functional magnetic resonance imaging (fMRI) results (Salo et al., 2013). According to our results, auditory cortex (AC) activity in response to spoken syllables is attenuated during phonological processing of written consonant letters in relation to AC activity elicited by the spoken syllables during non-phonological tasks involving the letters, that is, discriminating their font color or location rather than their phonological content. This pattern of results suggests that phonological processing of written letters occupies same phonological processing systems as processing of speech signals. In addition, our fMRI study showed enhanced AC activity during discrimination tasks involving the spoken syllables in relation to AC activity elicited by the same syllables when they were to be ignored during the visual discrimination tasks involving the written letters. Such attention-related modulation of AC activity is a common finding in fMRI studies on attention to speech or non-speech sounds (for a review and meta-analysis, see Alho et al., 2014), as well as in related studies applying electro- or magnetoencephalography (EEG and MEG, respectively; for reviews, see Näätänen et al., 2002; Fritz et al., 2007; Alain et al., 2013).

Effects of attention on auditory processing have been also found in subcortical structures of the ascending auditory pathway. For example, in their fMRI study, Rinne et al. (2008) found an effect of selective auditory attention on the activity of the inferior colliculus (IC), a brainstem nucleus in the auditory pathway from the inner ear to AC. Participants' selective attention to a stream of tones delivered to one ear while they ignored a stream delivered to the other ear was associated with enhanced activity in the AC (see also Alho et al., 1999) and IC contralateral to the attended input in relation to activity in the ipsilateral AC and IC, respectively. This suggests attention-related facilitation of auditory processing in the AC and IC.

In EEG measurements, processing of periodic sounds in the ascending auditory pathway generates the frequency-following response (FFR) phase-locked to the fundamental frequency (F0) of a sound and its harmonics (H2, H3, etc.; e.g., Skoe and Kraus, 2010). The FFR is assumed to reflect phase-locked activity in subcortical structures of the auditory pathway including the cochlear nucleus, IC and medial geniculate body (MGB) of the thalamus, but it also gets a contribution from the AC at least for frequencies up to 120 Hz (e.g., Chandrasekaran and Kraus, 2010; Coffey et al., 2016; Bidelman, 2018).

According to several studies, attention modulates FFRs. Galbraith et al. (1998) and Lehmann and Schönwiesner (2014) found larger FFR amplitudes for a vowel delivered to one ear and attended by the listeners than for another vowel delivered simultaneously to the opposite ear, suggesting attention-related modulation of auditory processing. Moreover, Hairston et al. (2013) observed attenuated FFRs to task-irrelevant tones when participants performed duration discrimination tasks involving task-relevant auditory or visual stimuli compared with a no-task condition. Attention-related facilitation and suppression of auditory processing reflected by FFRs might be mediated by efferent connections descending from AC to the subcortical auditory nuclei, the so-called corticofugal auditory pathway (Oatman and Anderson, 1980; Galbraith et al., 1998, 2003; Suga et al., 2002; Winer, 2006; Hairston et al., 2013; Lehmann and Schönwiesner, 2014). However, as recent MEG and EEG results indicate that at least up to frequencies of 120 Hz the FFR gets also contribution from the AC (Coffey et al., 2016; Bidelman, 2018), effects of attention on FFR might be partly due to facilitation or suppression of auditory processing in the AC (see Fritz et al., 2007; Alain et al., 2013; Alho et al., 2014). This conclusion is also supported by recent results of Holmes et al. (2018) who found that attention to sounds of ca. 100 Hz may enhance FFRs to them while a similar effect was not observed for sounds of ca. 220 Hz (see also Galbraith and Kane, 1993). In contrast, Galbraith et al. (2003) found an enhancing effect of auditory attention (vs. visual attention) on FFRs elicited by tones of 293 Hz. Yet, it should be noted that many studies found no effects of direction of attention towards sounds or away from them on auditory brainstem potentials to clicks (e.g., Picton et al., 1971; Picton and Hillyard, 1974; Woods and Hillyard, 1978; Hirschhorn and Michie, 1990) or on FFRs around 100 Hz or lower (Okamoto et al., 2011; Varghese et al., 2015).

FFRs are sensitive to infrequent changes in repetitive auditory inputs. Shiga et al. (2015) measured FFRs to amplitude-modulated (AM) tones (tone duration 150 ms, carrier frequency 2,230 Hz) delivered at a constant rate of ca. 3 tones per second. Deviant tones had a higher pitch (AM frequency 410 Hz) and lower probability of occurrence ($p = 0.2$) than standard tones (AM frequency 290 Hz; $p = 0.8$) in tone streams ignored by participants watching a silent film. FFRs to deviant-pitch tones had larger amplitudes than FFRs to AM tones when they were used as standard tones in control tone streams. The pitch-deviant tones elicited also enhanced middle-latency and mismatch negativity (MMN) responses indicating change detection in the AC (for reviews, see Näätänen et al., 2007; Escera et al., 2014).

The FFR results of Shiga et al. (2015) suggesting auditory change detection already at an early processing level are supported by fMRI results of Cacciaglia et al. (2015) showing in addition to AC responses, enhanced IC and MGB responses to deviant higher-pitch noise bursts occurring among lower-pitch standard bursts delivered to participants watching a silent movie. Importantly these response enhancements were observed both in relation to brain activity elicited during stimulus blocks including only lower-pitch tones and in relation to activity elicited by noise burst varying randomly in pitch at five levels. This ruled out the possibility that the enhanced response to deviant-pitch bursts among standard-pitch bursts was simply due to the deviant bursts activating less refractory neuron populations than the standard bursts in the tonotopically organized IC, MGB, and AC.

However, brainstem processing of changes in speech sounds may differ from processing of changes in tones.

Slabu et al. (2012) recorded FFRs to a syllable /ba/ replacing infrequently another syllable /wa/ in a stimulus block delivered to participants watching a silent movie with subtitles. The syllables were produced with a speech synthesizer and they differed only in the duration of transition (20 and 35 ms for /ba/ and /wa/, respectively) in their first and second formant (F1 and F2, respectively) during the initial part of the syllable. To control for effects of stimulus characteristics on FFRs, a reversed block was used where /ba/ and /wa/ swapped their status as a deviant and standard syllable. To control for simple refractoriness/adaptation effects, there was also an additional block including an infrequent /ba/ among equally infrequent four versions of /wa/ differing in their F1 and F2 transition durations. In FFRs, amplitude attenuations were observed in the second and fourth harmonics of F0 of the steady-state (vowel) portion of deviant /ba/ both in comparison to the standard /ba/, and in comparison to the infrequent /ba/ occurring among four versions of /wa/.

As reviewed above, the task performed by participants modulates FFRs (Galbraith et al., 1998; Hairston et al., 2013; Lehmann and Schönwiesner, 2014). Moreover, both speech processing and visual phonological processing have been suggested to involve the same articulatory-phonological loop (Baddeley, 1992). Therefore, the present study examined whether processing of infrequent syllable changes in the ascending auditory pathway reflected by FFRs would be affected by a concurrent phonological task involving visually presented letters.

MATERIALS AND METHODS

Participants

Twenty-three healthy volunteers were recruited among the students of University of Barcelona. Written informed consent was obtained from all participants and they were reimbursed for collaboration with a monetary payment of 7€ per hour. The present study was approved by the Bioethics Committee of the University of Barcelona and conducted in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki).

All participants were native speakers of Catalan or Spanish, or both. They had normal or corrected-to-normal vision, and according to their own report, no personal or familial history of psychiatric disorders, no head injuries or brain surgery, no current use of psychotropic drugs, and no hearing problems. Normal hearing of the participants was verified with a pure tone audiometry (hearing threshold at 250–8,000 Hz below 25 dB SPL for each ear). One participant was excluded from data analysis due to misunderstanding experimental task instructions and another three participants due to over 50% of their collected EEG epochs contaminated by artifacts exceeding rejection criterion (see below). The remaining 19 participants were 20–35 years old (nine males and 10 females; 11 right-handed and eight left-handed according to their own report).

Stimuli and Procedure

Experiments were conducted in a sound-attenuated and electrically shielded room. During recordings, participants were seated comfortably in a reclining chair facing an LCD screen at 155 cm from the participant's head. Independent sequences of written letters and spoken syllables delivered to the participants were programmed and presented using Matlab R2007a, MathWorks, and Psychophysics Toolbox Version 3 (Brainard, 1997; Pelli, 1997; Kleiner et al., 2007).

The letters (Arial font, height 1.3°–2.0°, width 1.1°–1.5°) were flashed for 50 ms in a pseudorandom order with a varying onset asynchrony of 250–500 ms (even distribution) at the center of the screen on a white (R = 255, G = 255, B = 255) background. Eighty percentage of letters were vowels (equiprobably A, E, O, U, a, e, o or u) written in red (R = 255, G = 0, B = 0), 10% of letters were vowels (equiprobably A, E, O, U, a, e, o or u) written in blue (R = 0, G = 0, B = 255), and 10% were consonants written in red (equiprobably B, C, D, F, G, M, N, P, R, S, T, V, Z, b, c, d, f, g, m, n, p, r, s, t, v or z). In separate blocks, the participants were instructed to respond by pressing the Enter key on the keyboard in front of them with their index or middle finger of their preferred hand either to any consonant letter or to any blue vowel. Discriminating the infrequently occurring consonants was regarded as a Phonological Task since they differed from the frequently occurring red vowels in phonology but not in color, whereas discriminating the infrequently occurring blue vowels was regarded as a Non-Phonological Task, since they differed from the frequently occurring red vowels only in a non-phonological feature, namely color. In addition, during both tasks they were instructed to ignore the stream of spoken syllables delivered in parallel with the visual stimulus stream and to keep their gaze on a black (R = 0, G = 0, B = 0) fixation cross (0.9° × 0.9°) visible at the center of the screen when no letter was displayed there.

The spoken syllables were generated with the Klatt speech synthesizer (Klatt, 1980) and delivered binaurally (intensity in each ear 75 dB SPL) in alternating polarities (to minimize contributions of stimulus artifact and cochlear microphonic to FFR; see, e.g., Aiken and Picton, 2008) *via* ER-3A ABR insert earphones (Etymotic Research Inc., Elk Grove Village, IL, USA) in a pseudorandom order with a varying onset asynchrony of 250–500 ms (even distribution). There were three different syllables: /ba/, /wa/, and /aw/. Each syllable had a duration of 170 ms. The syllables /ba/ and /wa/ were also used in a previous FFR study by Slabu et al. (2012). Their fundamental frequency (F0) was 100 Hz and the third (F3), fourth (F4) and fifth (F5) formants were set to 2,900, 3,500 and 4,900 Hz respectively. The first 5 ms of both /ba/ and /wa/ syllables consisted of a rapid glide in their F1 (from 400 to 1,700 Hz) and F2 (from 1,700 to 1,240 Hz), after which there was a 20-ms transition for /ba/ and 50-ms transition for /wa/ in F1 from 125 to 800 Hz and in F2 from 571 to 1,200 Hz. The syllable /aw/ was generated by presenting the syllable /wa/ backwards in time.

In Standard-/wa/ Blocks, 1,000 syllables were delivered in a pseudorandom order (i.e., each block had a duration of ca. 6 min 15 s with the visual task performed throughout the block). The syllable /wa/ was the standard syllable and occurred at a probability of 0.8, while the deviant syllables /ba/ and /aw/ at a probability of 0.1 each. Five Standard-/wa/ Blocks were delivered during the Phonological Task and another five during the Non-Phonological Task. Thus, both deviant /ba/ and deviant /aw/ occurred 500 times in each task condition.

In addition, we presented for each task condition one Standard-/ba/ Block where /ba/ was the standard syllable ($p = 0.8$) and /wa/ and /aw/ were the deviant syllables ($p = 0.1$ for each) and one Standard-/aw/ Block where /aw/ was the standard syllable ($p = 0.8$) and and /wa/ and /ba/ were the deviant syllables ($p = 0.1$ for each). This allowed us to compare FFRs to the deviants /ba/ and /aw/ in the Standard-/wa/ blocks with those to the standard /ba/ and standard /aw/ to control for effects of specific stimulus characteristics on FFRs to the deviant syllables /ba/ and /aw/. In Standard-/ba/ and Standard-/aw/ Blocks, there were 640 syllables in each (i.e., each block had a duration of ca. 4 min) including 512 standard syllables and 64 deviant syllables of each type. Note that the deviant syllables were delivered in Standard-/aw/ and Standard-/ba/ Blocks just to keep the stimulus probability structure within these blocks analogous to that in the Standard-/wa/ Blocks. Due to the small number of deviant syllables in these blocks, FFRs to these deviant syllables were not analyzed.

Thus, altogether 14 blocks were delivered to each participant and the duration of the experiment, including short 1–2 min breaks between the blocks, was about 1 h 30 min. The order of blocks and task conditions was randomized separately for each participant.

EEG Data Acquisition, Processing and Analysis

To obtain FFRs, EEG (0.05–3,000 Hz; sampling rate 20 kHz) was recorded during the experimental blocks with SynAmps RT amplifier (Compumedics Neuroscan, Charlotte, NC, USA) and Neuroscan 4.4 acquisition software as a voltage between the fronto-central midline (FCz) Ag/AgCl scalp electrode in the Neuroscan Quik-Cap electrode system and an Ag/AgCl electrode attached to the right earlobe (A2). The default Quick-Cap ground electrode was located between the frontal (Fz) and fronto-polar (Fpz) midline sites. All electrode impedances were kept below 5 kΩ.

Analysis of EEG data was performed using Matlab R2012a, MathWorks, and EEGLAB, an open source toolbox for analysis of single-trial EEG dynamics (Delorme and Makeig, 2004). First, frequencies between 70 Hz and 1,500 Hz were filtered from the EEG data with a Kaiser finite impulse response (FIR) bandpass filter (transition bandwidth 15 Hz, passband ripple 0.001). Then FFRs were obtained for the deviant /ba/ and /aw/ syllables of the Standard-/wa/ blocks and for the standard /ba/ and /aw/ syllables of the Standard-/ba/ and Standard-/aw/ blocks, respectively, separately for each participant and separately for Phonological and Non-Phonological Tasks by averaging EEG epochs starting 40 ms before each syllable onset and ending 180 ms after syllable onset. The 0-μV baseline was set at the mean amplitude during the 40-ms pre-syllable period. Epochs with voltages exceeding ±35 μV were rejected from averaging. As noted above, deviants /wa/ and /aw/ in Standard-/ba/ blocks and deviants /ba/ and /wa/ in Standard-/aw/ blocks were excluded from data analysis due to their small number. Consequently, data for the standard /wa/ were not analyzed as there were not enough data for the deviant /wa/ with which data for the standard /wa/ could be compared.

To analyze FFR amplitudes in the frequency domain during the vowel (steady-state) part of syllables, fast Fourier transform was applied on each participant's demeaned, zero-padded (1-Hz resolution) and Hanning-tapered FFRs. A time window 35–165 ms from syllable onset was used for the FFRs to the deviant and standard /ba/, because /ba/ started with formant transitions due to the initial consonant and ended in a vowel with steady-state formants. For the deviant and standard /aw/, in turn, a time window 10–115 ms from syllable onset was used in the FFR analysis, because /aw/ started with steady-state formants and ended in formant transitions (Note that the FFRs to the syllables /ba/ and /aw/ were not compared statistically with each other and therefore the different time windows used to obtain these FFRs did not affect the statistical results.). The mean FFR amplitude was computed separately for each participant, and separately for the deviant /ba/, standard /ba/, deviant /aw/, and standard /aw/ during Phonological and Non-Phonological Tasks using 10-Hz wide windows centered at the fundamental frequency (F0) of syllables at 100 Hz and at the second (H2) and third harmonic (H3) at 200 and 300 Hz, respectively.

In addition to FFRs, we analyzed long-latency ERPs to deviant and standard syllables in order to study MMN responses elicited by deviant /ba/ and /aw/ syllables. For this analysis, the EEG recorded at the FCz electrode site was resampled at 500 Hz and filtered using a Kaiser FIR filter with a passband of 0.5–20 Hz. ERPs were obtained by averaging EEG epochs starting 40 ms before syllable onset and ending 400 ms after syllable onset. This was done separately for the deviant /ba/ and /aw/ syllables of the Standard-/wa/ blocks and for the standard /ba/ and /aw/ syllables of the Standard-/ba/ and Standard-/aw/ blocks, respectively, separately for each participant, and separately for Phonological and Non-Phonological Tasks. The 0-μV baseline for amplitude measurements was set at the mean amplitude during the 40-ms pre-syllable period. Epochs with voltages exceeding ±35 μV were rejected from averaging.

Statistical Analysis

Button presses given 100–1,100 ms after target-letter onset were regarded as hits. Other responses were classified as false alarms. Hit response times, hit rates (number of hits divided by the number of target letters) and false-alarm rates (number of false alarms divided by the number of non-target letters) were calculated for each participant across the blocks of separately for Phonological and Non-Phonological Tasks and then subjected to one-way repeated-measures analysis of variance (ANOVAs) in order to compare participants' performance speed and accuracy in these tasks. Effects with $p \leq 0.05$ were regarded as significant and effect sizes (η_p^2) were calculated for these cases. It should be noted that since the stimulus onset asynchrony for the letters was randomly 250–500 ms, there is a risk that a delayed hit response to a target letter was classified as a false alarm to a subsequent non-target letter or that false-alarm responses to non-targets given during the 100–1,000 ms time windows following targets were classified as hits. However, probabilities for such misclassifications of responses were similar in Phonological and Non-Phonological Visual Tasks. Therefore, the estimated hit and false-alarm rates are still comparable between these tasks.

Peak amplitudes of frequency spectra of FFRs (measured as mean amplitudes in 10-Hz windows centered at the peak frequency) were analyzed with two-way repeated-measures ANOVAs, performed separately for the FFRs to syllables /ba/ and /aw/, and including factors Task (Phonological vs. Non-Phonological) and Deviance (deviant vs. standard syllable). In case an ANOVA showed significant ($p \leq 0.05$) effects of factors or their significant interaction, effect sizes (η_p^2) were calculated and subsequent *post hoc* *t*-tests (with Bonferroni-corrected significance criterion of $p \leq 0.0125$) were performed for within-condition (Phonological or Non-Phonological Task) comparisons between FFRs to standard and deviant syllables and between-condition comparisons separately for standard syllables and deviant syllables if an ANOVA showed significant effects.

To evaluate significance of differences between long-latency ERPs to deviant and standard /ba/ syllables and between long-latency ERPs to deviant and standard /aw/ syllables due to MMN responses elicited by the deviant syllables, the mean amplitudes of these ERPs during Phonological and Non-Phonological Tasks were measured separately for each participant over consecutive 100-ms periods from syllable onset, that is, over 0–100, 100–200, 200–300, and 300–400 ms. Statistical significance of effects of the factors Deviance (deviant vs. standard syllable) and Task (Phonological vs. Non-Phonological) and their interaction was assessed with a repeated-measures ANOVA performed separately for ERPs to syllables /ba/ and /aw/. Since four dependent ANOVAs were performed for ERPs to each syllable, instead of using a significance criterion of $p \leq 0.05$, Bonferroni-corrected criterion of $p \leq 0.0125$ was applied and effect sizes (η_p^2) were calculated only for effects fulfilling this corrected criterion.

RESULTS

Task Performance

As could be expected, the Phonological Task was somewhat more difficult than the Non-Phonological Task. As seen in **Figure 1**, the participants' response times to target letters were longer in the Phonological Task than in the Non-Phonological Task ($F_{(1,18)} = 246.457$, $p < 0.001$, $\eta_p^2 = 0.93$). Moreover, their hit rates in detecting target letters were lower in the Phonological than Non-Phonological Task ($F_{(1,18)} = 15.108$, $p < 0.002$, $\eta_p^2 = 0.46$) and they made more false alarms in Phonological Task ($F_{(1,18)} = 5.676$, $p < 0.03$, $\eta_p^2 = 0.24$).

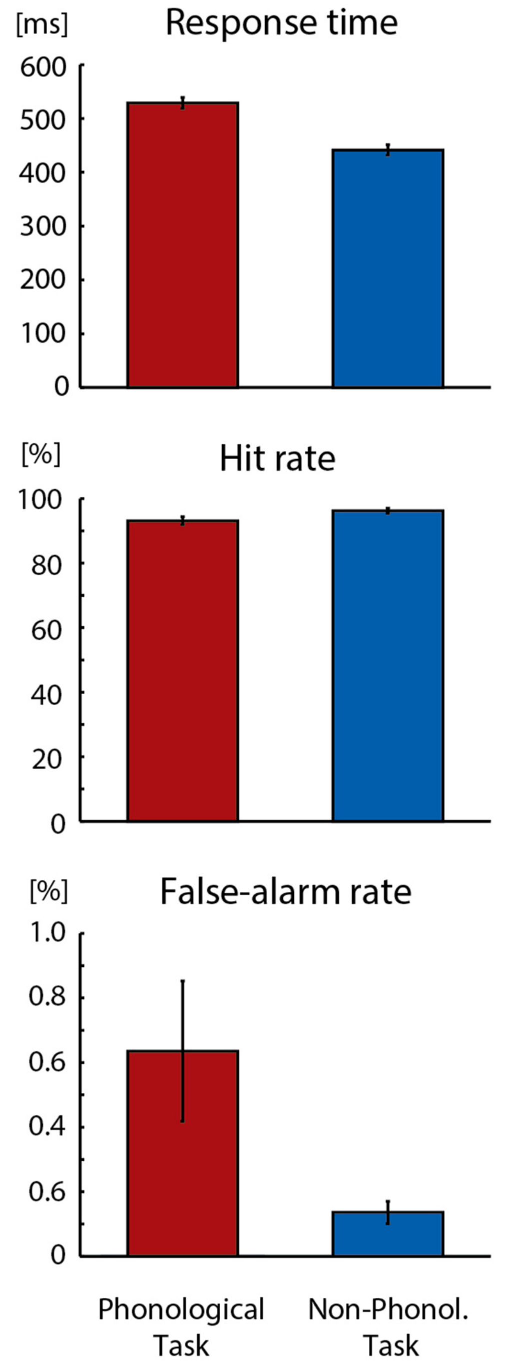


FIGURE 1 | Mean response times, hit rates and false-alarm rates in detecting visual target stimuli in Phonological and Non-Phonological Tasks. Error bars indicate standard errors of the mean.

FFRs

FFRs averaged for deviant and standard /ba/ and /aw/ syllables are shown in **Figures 2**, **3** depict frequency spectra of these FFRs and mean amplitudes in these around F0 during Phonological and Non-Phonological Tasks.

A two-way ANOVA for FFR amplitudes over 95–105 Hz (i.e., around the F0 frequency of 100 Hz) to syllable /ba/ indicated a significant effect of Deviance ($F_{(1,18)} = 6.158$, $p < 0.025$, $\eta_p^2 = 0.25$), but no significant effect of Task. However, there was a significant Task × Deviance interaction ($F_{(1,18)} = 4.733$, $p < 0.05$, $\eta_p^2 = 0.21$). As seen in **Figure 3**, FFRs were larger to deviant /ba/ than to standard /ba/ during Phonological Task but not during Non-Phonological task. Four *post hoc* *t*-tests (with a Bonferroni-corrected significance criterion: $p \leq 0.0125$) comparing FFR amplitudes to deviants vs. standards within each

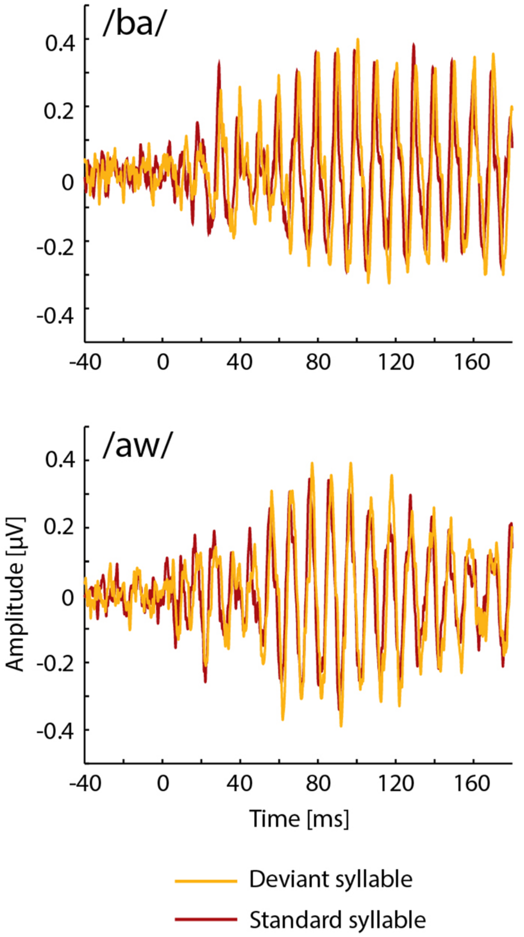


FIGURE 2 | Across 19 participants averaged frequency-following responses (FFRs) to deviant (orange lines) and standard (red lines) syllables /ba/ (top panel) and /aw/ (bottom panel) delivered binaurally *via* earphones during the visual Phonological Task involving letters displayed on a computer screen.

task and to deviants or standards across the tasks indicated that during Phonological Task, the FFR amplitude around F0 to deviant /ba/ was significantly larger than that to standard /ba/ ($t_{(18)} = 4.802$, $p < 0.0002$) the other differences in these pairwise comparisons being insignificant ($p > 0.25$ in all cases) and the insignificant difference between deviant and standard /ba/ during Non-Phonological Task actually being of opposite polarity (see **Figure 3**). Thus, the significant Task × Deviance interaction indeed resulted from enhanced deviant vs. standard FFR difference during Phonological Task in relation to Non-Phonological Task, but it is not possible to judge from the present data whether attenuation of FFR to standard /ba/ (see **Figure 3**) during Phonological Task in relation to Non-Phonological Task also contributed to this deviant vs. standard FFR difference.

A two-way ANOVA for F0 amplitudes to deviant /aw/ and standard /aw/, in turn, showed no significant effects of Deviance or Task, or significant Task × Deviance interaction. Furthermore, ANOVAs for the amplitudes of H2 and H3 harmonics in FFRs (see **Figure 3**, left column) showed no significant effects of Task or Deviance or significant Task × Deviance interaction for either syllable /ba/ or /aw/.

As seen in **Figure 2**, unexpectedly, the FFRs to deviant syllables /ba/ and /aw/ appeared to be slightly delayed in relation to the FFRs to the respective standard syllables. In order to analyze this in detail, cross correlations were calculated between FFRs to deviant and standard /ba/ and between FFRs to deviant and standard /aw/ (see Russo et al., 2004; Ribas-Prats et al., 2019) separately for Phonological and Non-Phonological Tasks and separately for each participant. According to these cross-correlation analyses, the FFRs to deviant syllables tended to lag in relation to FFRs to standard syllables, this lag being on average 1.7 ms (standard error of the mean ±3.4 ms) and 2.9 ms (±2.0 ms) for the syllable /ba/ during Phonological and Non-Phonological Tasks, respectively, and 0.9 ms (±3.3 ms) and 0.9 ms (±5.4 ms) for the syllable /aw/ during Phonological and Non-Phonological Tasks, respectively. However, subsequent *t*-tests showed that none of these lags differed significantly from 0 ms (in all four cases $t_{(18)} < 1.42$, $p > 0.17$).

MMN

As seen in **Figure 4**, the long-latency ERPs to deviant syllables were negatively displaced in relation to standard syllables. Statistical significance of this difference was evaluated separately for syllable /ba/ and syllable /aw/ at four consecutive 100-ms time windows from syllable onset with two-way ANOVAs (Bonferroni corrected significance criterion: 0.0125) including factors Deviance (deviant vs. standard syllable) and Task (Phonological vs. Non-Phonological). According to these ANOVAs the mean amplitudes over 200–300 ms and over 300–400 ms from syllable onset were significantly more negative in ERPs to deviant /ba/ than in ERPs to standard /ba/ (significant effect of Deviance, 200–300 ms: ($F_{(1,18)} = 12.945$, $p < 0.003$, $\eta_p^2 = 0.42$; 300–400 ms: ($F_{(1,18)} = 19.283$, $p < 0.001$, $\eta_p^2 = 0.52$). For the /aw/ syllable, the ERPs to deviant /aw/ had significantly more negative amplitudes than ERPs to standard /aw/ at 100–200 ms ($F_{(1,18)} = 69.107$, $p < 0.001$, $\eta_p^2 = 0.79$), 200–300 ms ($F_{(1,18)} = 33.349$, $p < 0.001$, $\eta_p^2 = 0.65$), and 300–400 ms ($F_{(1,18)} = 12.374$, $p < 0.003$, $\eta_p^2 = 0.41$). None of these ANOVAs showed significant effects of Task. Neither were there significant Task × Deviance interactions, although **Figure 4** suggests that the differences between ERPs to deviant and standard syllables tended to be slightly smaller during Phonological than Non-Phonological Task.

DISCUSSION

The aim of the present study was to clarify whether processing of infrequent syllable changes in the ascending auditory pathway reflected by FFRs would be affected by a concurrent visual phonological task. While our previous fMRI results suggested suppression of processing of to-be-ignored spoken syllables during a visual phonological task (Salo et al., 2013), the present results showed an enhancement of FFR to a phonetic change in a spoken syllable during the phonological task involving visually presented letters. This unexpected effect was observed

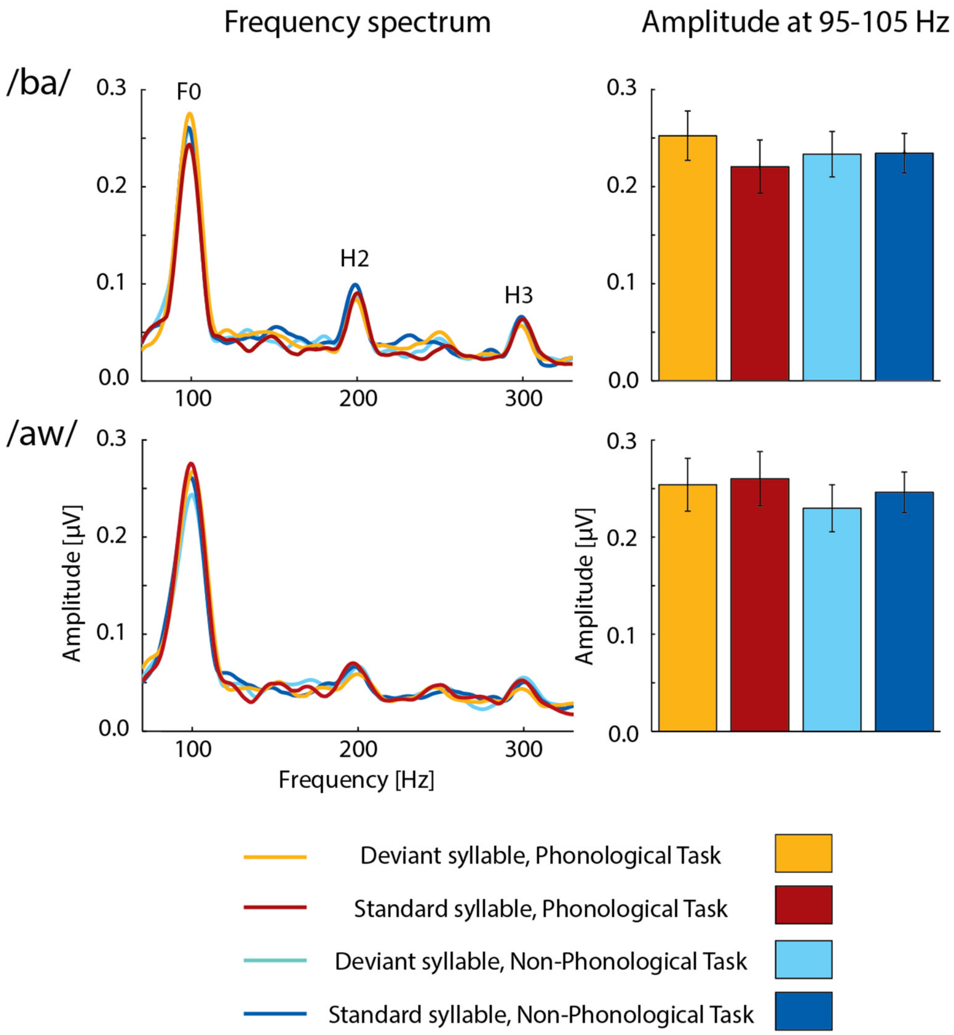


FIGURE 3 | Left column: across 19 participants averaged frequency spectra of FFRs to deviant and standard /ba/ syllables (top row) and deviant and standard /aw/ syllables (bottom row) during the visual Phonological Task (deviants: orange lines, standards: red lines) and Non-Phonological Task (deviants: cyan lines, standards: blue lines). The labels F0, H2, and H3 indicate peaks of fundamental frequency and its second and third harmonics, respectively, in the frequency spectra. Right column: the mean F0 amplitudes (average amplitudes over 95–105 Hz) measured from the frequency spectra for deviant and standard /ba/ and /aw/ syllables during the visual Phonological Task (deviants: orange bars, standards: red bars) and Non-Phonological Task (deviants: cyan bars, standards: blue bars). Error bars indicate standard errors of the mean.

for infrequent changes of a repeating syllable from /wa/ to /ba/. Importantly, this effect for deviant /ba/ syllables was revealed by a comparison of FFRs to deviant /ba/ syllables with FFRs to identical standard /ba/ syllables (delivered in blocks with inverted probabilities of /ba/ and /wa/ syllables) controlling for simple effects of physical stimulus features on FFR.

While the present FFR data suggest enhanced processing of deviant /ba/ syllables during the visual phonological task in comparison with the visual non-phonological task, the nature of visual task was not observed to have any significant effects on the long-latency ERPs to deviant or standard /ba/ or /aw/ syllables. However, these ERPs showed significant effects of syllable deviance: ERPs to deviant syllables were negatively displaced in relation to ERPs to standard syllables, these effects presumably being due to the MMN response elicited by deviant syllables (for reviews, see Näätänen et al., 2007; Escera et al., 2014). Since the MMN has its major generators bilaterally in the AC, the lack of effect of visual phonological task on MMN might be regarded as suggesting functional independence of MMN generator processes from the deviance detection reflected by the present FFRs to deviant /ba/ syllables and enhanced by the visual phonological task. The lack of effect of visual phonological task on the MMN supports the proposal that auditory change detection reflected by the MMN is largely independent of attention (see, e.g., Näätänen et al., 2007). This lack also rules out the possibility that the FFR enhancement observed for the deviant /ba/ syllables during the visual phonological task was due to an effect of this task on the MMN overlapping in time with the

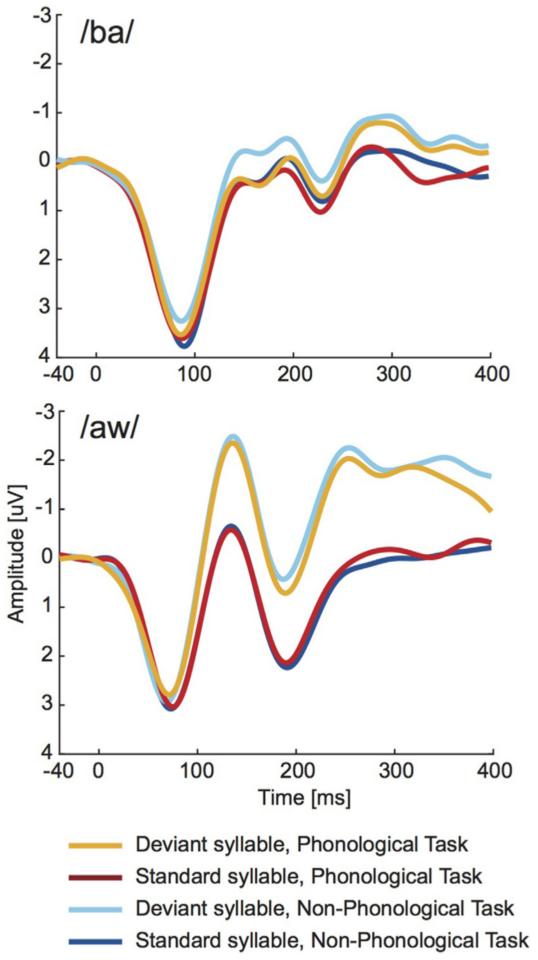


FIGURE 4 | Across 19 participants averaged long-latency ERPs to deviant (orange lines) and standard (red lines) syllables /ba/ (top panel) and /aw/ (bottom panel) delivered during the visual Phonological Task and to respective deviant (cyan lines) and standard (blue lines) syllables during the visual Non-Phonological Task.

late part of FFR to deviant /ba/ syllable and therefore potentially contaminating the FFR results.

Taken the subcortical and cortical generator sources of FRRs (Coffey et al., 2016; Bidelman, 2018), the enhancement of FFRs to deviant /ba/ syllables during visual phonological processing is presumably caused by top-down modulation of speech processing in the AC or hierarchically lower structures of the auditory pathway. Since the processing of deviant /ba/ syllables in the AC reflected by the MMN elicited by these syllables was not affected by the nature of the visual task, it is more likely that the present FFR enhancement observed for the deviant /ba/ syllables during the visual phonological task originated from subcortical structures of the ascending auditory pathway rather than from the AC. However, this enhancement might be due to top-down modulation of activity in these subcortical structures *via* corticofugal connections descending from the AC (see Suga et al., 2002; Galbraith et al., 2003; Winer, 2006).

No difference was observed in the FFRs to deviant /aw/ syllables between the visual phonological and non-phonological tasks. This suggests that the visual phonological task facilitates especially processing of small contrasts between speech sounds, like the 30-ms difference in the frequency transition time between the present /ba/ and /wa/ syllables, rather than processing of large differences, like the difference between the present /aw/ and /wa/ syllables starting without and with a frequency transition, respectively. Yet, it should be borne in mind that all syllables used in the present study contained the same frequencies and therefore it remains to be studied whether FFRs to infrequent small or large changes in the pitch (F0) or frequency structure of spoken syllables or other sounds (see Cacciaglia et al., 2015; Shiga et al., 2015) would be affected by a concurrent phonological task. However, it should be noted that the present /aw/ vs. /wa/ contrast was created by presenting the same syllable /wa/, perceived as a diphthong, forwards or backwards. While such temporal deviancies have been shown to elicit cortical MMN responses (Sams and Näätänen, 1991; Pardo and Sams, 1993), to our knowledge, no previous study has shown that processing of such higher-order stimulus contrasts would be reflected by FFRs.

In their study, Hairston et al. (2013) demonstrated attenuation of FFRs to task-irrelevant sounds during attention-demanding tasks involving other sounds or visual stimuli in comparison with a no-task baseline. Since the present study did not include a no-task condition it is not possible to estimate whether such suppression occurred during the present visual tasks. However, if this suppression were related to attention, one would have expected to see more attenuated FFRs to both standard and deviant syllables during the phonological visual task than during the non-phonological visual task. This is because the present visual phonological task was, according to the present performance speed and accuracy results, more difficult, and thus presumably more attention-demanding, than the visual non-phonological tasks.

In the present study, we observed enhanced amplitudes at the F0 frequency of the FFR in response to deviant /ba/ syllables occurring among standard /wa/ syllables during the visual phonological task in relation to standard /ba/ syllables delivered during a similar visual phonological task. In contrast, Slabu et al. (2012) observed attenuated FFR amplitudes at the second and fourth harmonics of F0 in response to deviant /ba/ syllables in relation to the standard /ba/ syllables in participants watching a silent film with subtitles, that is, performing also a visuo-phonological (reading) task. However, the present difference in the duration of F1 and F2 transition between the deviant /ba/ and standard /wa/ (after an initial common 5-ms transition) was 30 ms (20 ms for /ba/ vs. 50 ms for /wa/), whereas in the study of Slabu et al. (2012), it was only 15 ms (20 ms for /ba/ vs. 35 ms for /wa/) which may not have been large enough to elicit the FFR amplitude enhancement at F0 in response to deviant /ba/. In the present study, in turn, no attenuation of amplitudes at the harmonics of F0 was observed in the FFR to deviant /ba/ syllables differing more from the standard /wa/ syllables than the deviant /ba/ syllables in the study of Slabu et al. (2012).

It might be argued that fatigue or habituation of FFRs (Collet and Duclaux, 1986; Gorina-Careta et al., 2016) during the present ca. 1.5-h experiment or muscle activity and arousal (Dunlop et al., 1965) due to the visual target detection task may have affected the FFRs reported here. However, if there were such effects, they were presumably on average similar for standard and deviant syllables and therefore could not explain the differences between the FFRs elicited by these syllables. Nevertheless, the present visual phonological task was more demanding than the visual non-phonological task. Therefore, contribution of arousal differences between the tasks to the present FFR results cannot be ruled out. Still, also arousal differences between the tasks would be expected to affect similarly the FFRs to deviant and standard syllables and thus are not likely to explain the enhanced FFRs observed specifically for the deviant /ba/ syllables during the visual phonological task.

However, it should be noted that the enhancing effect of visual phonological task on the FFR elicited by a phonological deviance was observed only for one deviance type and that the effect size for the significant Task × Deviance interaction this FFR enhancement caused was rather small ($\eta_p^2 = 0.21$). Therefore, in future studies, the present results need to be replicated with a wider range of phonological deviances. Moreover, our previous studies have shown that deviant sounds eliciting the MMN and subsequent ERP components distract performance in tasks involving subsequent visual target stimuli, this distraction seen as decrease in the speed and accuracy in the visual task (e.g., Alho et al., 1997; Escera et al., 1998). Therefore, in future FFR studies, it would be of interest to clarify whether phonological deviances eliciting enhanced FFRs during a visual phonological task would distract more visual task performance than deviances not eliciting enhanced FFRs. In the present study, with independent sequences of auditorily presented syllables and visually written letters and low rates of both deviant syllables and target letters, there were too few visual target letters immediately following deviant syllables to reliably clarify this issue.

In conclusion, the present observation of enhanced FFR to deviant /ba/ syllables occurring among standard /wa/ syllables during the phonological task involving written letters suggests that active phonological processing and processing of irrelevant speech interfere. While at least up to 100 Hz FFRs get contributions not only from the subcortical ascending auditory pathway, but also from the AC (Coffey et al., 2016; Bidelman, 2018), it is not possible to resolve the origin of the present FFR enhancement. However, since cortical processing of the to-be-ignored syllables reflected by long-latency ERPs, especially the MMN elicited by deviant /ba/ syllables, was not affected by the nature of visual task performed by the participants, it is likely that the present FFR enhancement for deviant /ba/ syllables during phonological processing of written letters originated from subcortical auditory structures. The present results do not allow conclusions about the neural route through which the visual phonological processing enhanced the early processing of deviant /ba/ syllables reflected by the FFR. However, this route appears to bypass subsequent cortical processing of these phoneme deviances reflected by the MMN.

ETHICS STATEMENT

The present study was approved by the Bioethics Committee of the University of Barcelona and conducted in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki).

AUTHOR CONTRIBUTIONS

KA and CE designed the experimental protocol and data collection and analysis. KŻ programmed the experimental design and collected the data. The data were analyzed by KŻ and NG-C in supervision of KA and CE. KA, KŻ, NG-C and CE wrote the manuscript.

FUNDING

This study was supported by the Academy of Finland project grants (260054 and 297848) and a 1-month Visiting Professorship at the Institute of Biomedicine, Paris Descartes University awarded to KA, and by the Catalan Government (Generalitat de Catalunya; SGR2017-974), the Ministerio de Economía y Competitividad (MINECO)/FEDER project PSI2015-63664P and the ICREA Acadèmia Distinguished Professorship awarded to CE.

REFERENCES

Aiken, S. J., and Picton, T. W. (2008). Envelope and spectral frequency-following responses to vowel sounds. *Hear. Res.* 245, 35–47. doi: 10.1016/j.heares.2008.08.004

Alain, C., Arnott, S. R., and Dyson, B. J. (2013). "Varieties of auditory attention," in *The Oxford Handbook of Cognitive Neuroscience Vol. 1: Core Topics*, eds K. N. Ochsner and S. M. Kosslyn (Oxford: Oxford University Press), 215–236.

Alho, K., Escera, C., Díaz, R., Yago, E., and Serra, J. M. (1997). Effects of involuntary auditory attention on visual task performance and brain activity. *Neuroreport* 8, 3233–3237. doi: 10.1097/00001756-199710200-00010

Alho, K., Medvedev, S. V., Pakhomov, S. V., Roudas, M. S., Zeffiro, T., Tervaniemi, M., et al. (1999). Selective tuning of the left and right auditory cortices during spatially directed attention. *Cogn. Brain Res.* 7, 335–341. doi: 10.1016/s0926-6410(98)00036-6

Alho, K., Rinne, T., Herron, T. J., and Woods, D. L. (2014). Stimulus-dependent activations and attention-related modulations in the auditory cortex: a meta-analysis of fMRI studies. *Hear. Res.* 307, 29–41. doi: 10.1016/j.heares.2013.08.001

Baddeley, A. (1992). Working memory: the interface between memory and cognition. *J. Cogn. Neurosci.* 4, 281–288. doi: 10.1162/jocn.1992.4.3.281

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 (New York, NY: Academic Press), 47–90.

Bidelman, G. M. (2018). Subcortical sources dominate the neuroelectric auditory frequency-following response to speech. *Neuroimage* 175, 56–69. doi: 10.1016/j.neuroimage.2018.03.060

Brainard, D. H. (1997). The psychophysics toolbox. *Spat. Vis.* 10, 433–436. doi: 10.1163/156856897X00357

Cacciaglia, R., Escera, C., Slabu, L. S., Grimm, S., Sanjuán, N., Ventura-Campos, A., et al. (2015). Subcortical detection of auditory regularity violations in humans:

direct evidence from functional magnetic resonance imaging. *Neuropsychologia* 68, 51–58. doi: 10.1016/j.neuropsychologia.2015.01.001

Chandrasekaran, B., and Kraus, N. (2010). The scalp-recorded brainstem response to speech: neural origins and plasticity. *Psychophysiology* 47, 236–246. doi: 10.1111/j.1469-8986.2009.00928.x

Coffey, E. B. J., Herholz, S. C., Chepesiuk, A. M. P., Baillet, S., and Zatorre, R. J. (2016). Cortical contributions to the auditory frequency-following response revealed by MEG. *Nat. Commun.* 7:11070. doi: 10.1038/ncomms11070

Collet, L., and Duclaux, R. (1986). Auditory brainstem evoked responses and attention: contribution to a controversial subject. *Acta Otolaryngol.* 101, 439–441. doi: 10.3109/00016488609108629

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

Dunlop, C. W., Webster, W. R., and Simons, L. A. (1965). Effect of attention on evoked responses in the classical auditory pathway. *Nature* 206, 1048–1050. doi: 10.1038/2061048b0

Escera, C., Alho, K., Winkler, I., and Näätänen, R. (1998). Neural mechanisms of involuntary attention switching to novelty and change in the acoustic environment. *J. Cogn. Neurosci.* 10, 590–604. doi: 10.1162/089892998562997

Escera, C., Leung, S., and Grimm, S. (2014). Deviance detection based on regularity encoding along the auditory hierarchy: electrophysiological evidence in humans. *Brain Topogr.* 27, 527–538. doi: 10.1007/s10548-013-0328-4

Fritz, J. B., Elhilali, M., David, S. V., and Shamma, S. A. (2007). Auditory attention—focusing the searchlight on sound. *Curr. Opin. Neurobiol.* 17, 437–455. doi: 10.1016/j.conb.2007.07.011

Galbraith, G. C., Bhuta, S. M., Choate, A. K., Kitahara, J. M., and Mullen, T. A. Jr. (1998). Brain stem frequency-following response to dichotic vowels during attention. *Neuroreport* 9, 1889–1893. doi: 10.1097/00001756-199806010-00041

Galbraith, G. S., and Kane, J. M. (1993). Brainstem frequency-following responses and cortical event-related potentials during attention. *Percept. Mot. Skills* 76, 1231–1241. doi: 10.2466/pms.1993.76.3c.1231

Galbraith, G. C., Olfman, D. M., and Huffman, T. M. (2003). Selective attention affects human brain stem frequency-following response. *NeuroReport* 14, 735–738. doi: 10.1097/01.wnr.0000064983.96259.49

Gorina-Careta, N., Zarnowiec, K., Costa-Faidella, J., and Escera, C. (2016). Timing predictability enhances regularity encoding in the human subcortical auditory pathway. *Sci. Rep.* 6:37405. doi: 10.1038/srep37405

Hairston, W. D., Letowski, T. R., and McDowell, K. (2013). Task-related suppression of the brainstem frequency following response. *PLoS One* 8:e55215. doi: 10.1371/journal.pone.0055215

Hirschhorn, T. N., and Michie, P. T. (1990). Brainstem auditory evoked potentials (BAEPs) and attention revisited. *Psychophysiology* 27, 495–512. doi: 10.1111/j.1469-8986.1990.tb01964.x

Holmes, E., Purcell, D. W., Carlyon, R. P., Gockel, H. E., and Johnsrude, I. S. (2018). Attentional modulation of envelope-following responses at lower (93–109 Hz) but not higher (217–233 Hz) modulation rates. *J. Assoc. Res. Otolaryngol.* 19, 83–97. doi: 10.1007/s10162-017-0641-9

Klatt, D. H. (1980). Software for cascade/parallel formant synthesizer. *J. Acoust. Soc. Am.* 67, 971–975. doi: 10.1121/1.383940

Kleiner, M., Brainard, D., and Pelli, D. (2007). What's new in Psychtoolbox-3? *Perception* 36, 1–16. doi: 10.1068/v070821

Lehmann, A., and Schönwiesner, M. (2014). Selective attention modulates human auditory brainstem responses: relative contributions of frequency and spatial cues. *PLoS One* 9:e85442. doi: 10.1371/journal.pone.0085442

Näätänen, R., Alho, K., and Schröger, E. (2002). "Electrophysiology of attention," in *Stevens' Handbook of Experimental Psychology (Third Edition), Vol. 4: Methodology in Experimental Psychology*, ed. J. Wixted (New York, NY: John Wiley & Sons), 601–653.

Näätänen, R., Paavilainen, P., Rinne, T., and Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clin. Neurophysiol.* 118, 2544–2590. doi: 10.1016/j.clinph.2007.04.026

Oatman, L. C., and Anderson, B. W. (1980). Suppression of the auditory frequency following response during visual attention. *Electroencephalogr. Clin. Neurophysiol.* 49, 314–322. doi: 10.1016/0013-4694(80)90225-4

Okamoto, H., Stracke, H., Bermudez, P., and Pantev, C. (2011). Sound processing hierarchy within human auditory cortex. *J. Cogn. Neurosci.* 23, 1855–1863. doi: 10.1162/jocn.2010.21521

Pardo, P. J., and Sams, M. (1993). Human auditory cortex responses to rising versus falling glides. *Neurosci. Lett.* 159, 43–45. doi: 10.1016/0304-3940(93)90794-l

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat. Vis.* 10, 437–442. doi: 10.1163/156856897x00366

Picton, T. W., and Hillyard, S. A. (1974). Human auditory evoked potentials: II: effects of attention. *Electroencephalogr. Clin. Neurophysiol.* 36, 191–200. doi: 10.1016/0013-4694(74)90156-4

Picton, T. W., Hillyard, S. A., Galambos, R., and Schiff, M. (1971). Human auditory attention: a central or peripheral process? *Science* 173, 351–353. doi: 10.1126/science.173.3994.351

Ribas-Prats, T., Almeida, L., Costa-Faidella, J., Plana, M., Corral, M. J., Gómez-Roig, M. D., et al. (2019). The frequency-following response (FFR) to speech stimuli: a normative dataset in healthy newborns. *Hear. Res.* 371, 28–39. doi: 10.1016/j.heares.2018.11.001

Rinne, T., Balk, M. H., Koistinen, S., Autti, T., Alho, K., and Sams, M. (2008). Auditory selective attention modulates activation of human inferior colliculus. *J. Neurophysiol.* 100, 3323–3327. doi: 10.1152/jn.90607.2008

Russo, N., Nicol, T., Musacchia, G., and Kraus, N. (2004). Brainstem responses to speech syllables. *Clin. Neurophysiol.* 115, 2021–2030. doi: 10.1016/j.clinph.2004.04.003

Salo, E. A., Rinne, T., Salonen, O., and Alho, K. (2013). Brain activity during auditory and visual phonological, spatial and simple discrimination tasks. *Brain Res.* 1496, 55–69. doi: 10.1016/j.brainres.2012.12.013

Sams, M., and Näätänen, R. (1991). Neuromagnetic responses of the human auditory cortex to short frequency glides. *Neurosci. Lett.* 21, 43–46. doi: 10.1016/0304-3940(91)90645-a

Shiga, T., Althen, H., Cornella, M., Zarnowiec, K., Yabe, H., and Escera, C. (2015). Deviance-related responses along the auditory hierarchy: combined FFR, MLR and MMN evidence. *PLoS One* 10:e0136794. doi: 10.1371/journal.pone.0136794

Skoe, E., and Kraus, N. (2010). Auditory brainstem response to complex sounds: a tutorial. *Ear Hear.* 31, 302–324. doi: 10.1097/AUD.0b013e3181cdb272

Slabu, L., Grimm, S., and Escera, C. (2012). Novelty detection in the human auditory brainstem. *J. Neurosci.* 32, 1447–1452. doi: 10.1523/JNEUROSCI.2557-11.2012

Suga, N., Xiao, Z., Ma, X., and Ji, W. (2002). Plasticity and corticofugal modulation for hearing in adult animals. *Neuron* 36, 9–18. doi: 10.1016/s0896-6273(02)00933-9

Varghese, L., Bharadwaj, H. M., and Shinn-Cunningham, B. G. (2015). Evidence against attentional state modulating scalp-recorded auditory brainstem steady-state responses. *Brain Res.* 1626, 146–164. doi: 10.1016/j.brainres.2015.06.038

Winer, J. A. (2006). Decoding the auditory corticofugal systems. *Hear. Res.* 212, 1–8. doi: 10.1016/j.heares.2005.06.014

Woods, D. L., and Hillyard, S. A. (1978). "Attention at the cocktail party: Brainstem evoked responses reveal no peripheral gating," in *Multidisciplinary Perspectives in Event-Related Brain Potential Research*, ed. D. A. Otto (Washington, DC: U.S. Government Printing Office), 230–233.

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

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 12 May 2020
doi: 10.3389/fnins.2020.00436

Modulation of Brain Activity by Selective Attention to Audiovisual Dialogues

Alina Leminen[1,2,3,4]*, *Maxime Verwoert*[1], *Mona Moisala*[1], *Viljami Salmela*[1], *Patrik Wikman*[1] *and Kimmo Alho*[1,5]*

[1] Department of Psychology and Logopedics, Faculty of Medicine, University of Helsinki, Helsinki, Finland, [2] Cognitive Science, Department of Digital Humanities, Helsinki Centre for Digital Humanities (Heldig), University of Helsinki, Helsinki, Finland, [3] Cognitive Brain Research Unit, Department of Psychology and Logopedics, Faculty of Medicine, University of Helsinki, Helsinki, Finland, [4] Center for Cognition and Decision Making, Institute of Cognitive Neuroscience, National Research University – Higher School of Economics, Moscow, Russia, [5] Advanced Magnetic Imaging Centre, Aalto NeuroImaging, Aalto University, Espoo, Finland

OPEN ACCESS

Edited by:
Claude Alain,
Rotman Research Institute (RRI),
Canada

Reviewed by:
Michael S. Beauchamp,
Baylor College of Medicine,
United States
Yang Zhang,
University of Minnesota Twin Cities,
United States

****Correspondence:***
Alina Leminen
alina.leminen@helsinki.fi
Kimmo Alho
kimmo.alho@helsinki.fi

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

Received: *24 September 2019*
Accepted: *09 April 2020*
Published: *12 May 2020*

Citation:
Leminen A, Verwoert M, Moisala M, Salmela V, Wikman P and Alho K (2020) Modulation of Brain Activity by Selective Attention to Audiovisual Dialogues. Front. Neurosci. 14:436. doi: 10.3389/fnins.2020.00436

In real-life noisy situations, we can selectively attend to conversations in the presence of irrelevant voices, but neurocognitive mechanisms in such natural listening situations remain largely unexplored. Previous research has shown distributed activity in the mid superior temporal gyrus (STG) and sulcus (STS) while listening to speech and human voices, in the posterior STS and fusiform gyrus when combining auditory, visual and linguistic information, as well as in left-hemisphere temporal and frontal cortical areas during comprehension. In the present functional magnetic resonance imaging (fMRI) study, we investigated how selective attention modulates neural responses to naturalistic audiovisual dialogues. Our healthy adult participants (N = 15) selectively attended to video-taped dialogues between a man and woman in the presence of irrelevant continuous speech in the background. We modulated the auditory quality of dialogues with noise vocoding and their visual quality by masking speech-related facial movements. Both increased auditory quality and increased visual quality were associated with bilateral activity enhancements in the STG/STS. In addition, decreased audiovisual stimulus quality elicited enhanced fronto-parietal activity, presumably reflecting increased attentional demands. Finally, attention to the dialogues, in relation to a control task where a fixation cross was attended and the dialogue ignored, yielded enhanced activity in the left planum polare, angular gyrus, the right temporal pole, as well as in the orbitofrontal/ventromedial prefrontal cortex and posterior cingulate gyrus. Our findings suggest that naturalistic conversations effectively engage participants and reveal brain networks related to social perception in addition to speech and semantic processing networks.

Keywords: selective attention, noise vocoding, audiovisual integration, social perception, speech, visual speech, fMRI

INTRODUCTION

In everyday life, we are often faced with multiple speaker situations, for instance, when dining in a crowded restaurant or talking to a friend while hearing a radio in the background. Such situations require segregation of speech streams originating from different sources and selection of one of the streams for further processing. The neural mechanisms through which this type of attentional selection is achieved are not yet fully understood (e.g., Rimmele et al., 2015).

A meta-analysis (Alho et al., 2014) of functional magnetic resonance imaging (fMRI) studies on stimulus-dependent sound processing and attention-related modulations in the auditory cortex showed that speech and voice processing activate overlapping areas in the mid superior temporal gyrus and sulcus bilaterally (STG and STS, respectively). Furthermore, selective attention to continuous speech appeared to modulate activity predominantly in the same areas (Alho et al., 2014). Importantly, selectively attending to a particular speaker in a multi-talker situation results in the STG activity that represents the spectral and temporal features of attended speech, as if participants were listening only to that speech stream (Mesgarani and Chang, 2012). In other words, the human auditory system restores the representation of an attended speaker while suppressing irrelevant or competing speech.

In addition to STG/STS, selective attention to non-speech sounds engages prefrontal and parietal cortical areas (Tzourio et al., 1997; Alho et al., 1999; Zatorre et al., 1999; Degerman et al., 2006), which has been associated with top-down control needed to select attended sounds and reject irrelevant sounds. Selective attention to continuous speech, however, does not appear to markedly engage prefrontal and superior parietal areas (Alho et al., 2003, 2006; Scott et al., 2004). This is most probably because selective listening to speech is a highly automatized process, less dependent on fronto-parietal attentional control (Alho et al., 2006; see also Mesgarani and Chang, 2012). Such automaticity might be due to listeners' lifelong experience in listening to speech. However, initial orienting of attention to one of three concurrent speech streams has yielded enhanced activation in the fronto-parietal network, hence, purportedly engaging an attentional top-down control mechanism (Alho et al., 2015; Hill and Miller, 2010).

Natural situations with multiple speakers might not only be complicated by the demand to listen selectively to one speech stream while ignoring competing speech, but also by degraded quality of the attended speech (e.g., when talking in a noisy café on the phone with a poor signal). Studies addressing the comprehension of degraded (e.g., noise-vocoded) speech involving only one speech stream have reported increased activity in the posterior parietal cortex (Obleser et al., 2007) and frontal operculum (Davis and Johnsrude, 2003), as compared to more intelligible speech. Listening to degraded, yet intelligible and highly predictable speech, in turn, elicits activity in the dorsolateral prefrontal cortex, posterior cingulate cortex, and angular gyrus (e.g., Obleser et al., 2007). Moreover, the amount of spectral detail in speech signal was found to correlate with STS and left inferior frontal gyrus (IFG) activity, regardless of semantic predictability (Obleser et al., 2007). McGettigan et al. (2012) observed increasing activity along the length of left dorsolateral temporal cortex, in the right dorsolateral prefrontal cortex and bilateral IFG, but decreasing activation in the middle cingulate, middle frontal, inferior occipital, and parietal cortices associated with increasing auditory quality. Listening to degraded speech has also activated the left IFG, attributed to higher-order linguistic comprehension (Davis and Johnsrude, 2003) and the dorsal fronto-parietal network, related to top-down control of attention (Obleser et al., 2007). Overall, increased speech intelligibility enhances activity in the STS (Scott et al., 2000; Obleser et al., 2007; McGettigan et al., 2012), STG (Davis and Johnsrude, 2003), middle temporal gyrus (MTG; Davis and Johnsrude, 2003), and left IFG (Davis and Johnsrude, 2003; Obleser et al., 2007; McGettigan et al., 2012). Increased activity in these areas may be related to enhanced speech comprehension due to increasing availability of linguistic information.

The studies described above, however, used only single-speaker paradigms. Evans et al. (2016) examined how different masking sounds are processed in the human brain. They used a selective attention paradigm with two speech streams, namely, a masked stream and a target stream. The target speech was always clear, whilst the masked speech was either clear, spectrally rotated or noise-modulated. Increased intelligibility of the masked speech activated the left posterior STG/STS, however, less extensively than a clear single speech alone. This was taken to suggest that syntactic and other higher order properties of masking speech are not actively processed and the masker sounds may be actively suppressed already at early processing stages (see also Mesgarani and Chang, 2012). In contrast, the masked speech yielded increased activation in the frontal (bilateral middle frontal gyrus, left superior orbital gyrus, right IFG), parietal (left inferior and superior parietal lobule) and middle/anterior cingulate cortices, as well as in the frontal operculum and insula. These activations were suggested to reflect increased attentional and control processes. The results corroborate those from earlier positron emission tomography (PET) studies (e.g., Scott et al., 2004) on selective attention to a target speaker in the presence of another speaker (speech-in-speech) or noise (speech-in-noise). More specifically, Scott et al. (2004) found more activity in the bilateral STG for speech-in-speech than speech-in-noise, whereas speech-in-noise elicited more activity in the left prefrontal and right parietal cortex than speech-in-speech. Scott and colleagues suggested that these additional areas might be engaged to facilitate speech comprehension or that they are related to top-down attentional control. Correspondingly, Wild et al. (2012) reported activations in frontal areas (including the left IFG) that were only present when the participants selectively attended to the target speech among non-speech distractors. In contrast to studies reporting increased left IFG activations to increased intelligibility of degraded speech (Davis and Johnsrude, 2003; Obleser et al., 2007; McGettigan et al., 2012), Wild et al. (2012) found greater activity in the left IFG for degraded than for clear target speech. By contrast, STS activity was increased with decreasing speech intelligibility, regardless of attention. Increased activity for attended degraded speech was proposed to reflect "the improvement in intelligibility afforded by explicit, effortful processing, or by additional cognitive processes (such as perceptual learning) that are engaged under directed attention" (Wild et al., 2012, p. 14019). The authors further suggested that top-down influences on early auditory processing might facilitate speech comprehension in difficult listening situations.

The majority of fMRI studies on selective attention to speech have used only auditory speech stimuli (e.g., Alho et al., 2003, 2006; Wild et al., 2012; Evans et al., 2016; Puschmann et al., 2017). However, natural conversations often include also visual speech information. Integrating a voice with mouth movements (i.e., visual speech) facilitates speech understanding in relation to mere listening (Sumby and Pollack, 1954). In accordance, fMRI studies on listening to speech have shown that the presence of visual speech enhances activity in the auditory cortex and higher order speech-processing areas (e.g., Bishop and Miller, 2009; McGettigan et al., 2012). A related magnetoencephalography (MEG) study showed that the presence of visual speech enhances auditory-cortex activity that follows the temporal amplitude envelope of attended speech (Zion Golumbic et al., 2013; for similar electroencephalography (EEG) evidence, see O'Sullivan et al., 2015). Facilitation of speech comprehension by visual speech holds especially true for noisy situations (e.g., Sumby and Pollack, 1954) and degraded quality of attended speech (e.g., McGettigan et al., 2012; Zion Golumbic et al., 2013). Some fMRI studies have suggested maximal facilitation of speech comprehension by visual speech at intermediate signal-to-noise ratios of auditory information (Ross et al., 2007; McGettigan et al., 2012).

Degraded speech increases demands for fronto-parietal top-down control (Davis and Johnsrude, 2003; Evans et al., 2016), whereas adding visual speech appears to facilitate selective attention (Sumby and Pollack, 1954; Zion Golumbic et al., 2013). However, it is still unknown whether fronto-parietal areas are activated during selective attention to visually degraded speech. Moreover, an earlier study that employed a factorial design with different levels of auditory and visual clarity in sentences (McGettigan et al., 2012) did not include an unmodulated (clear) visual and auditory condition. Hence, to our knowledge, brain responses to continuous naturalistic dialogues with varying audio-visual speech quality have not been systematically examined before.

In the current study, we collected whole-head fMRI data in order to identify brain regions critical for selective attention to natural audiovisual speech. More specifically, we examined attention-related modulations in the auditory cortex and associated fronto-parietal activity during selective attention to audiovisual dialogues. In addition, we assessed an interplay between auditory and visual quality manipulations. We also included clear auditory and visual stimulus conditions to investigate brain areas activated during selective attention to naturalistic dialogues in the presence of irrelevant clear speech in the background. Our experimental setup might be regarded as mimicking watching a talk show on a TV while a radio program is playing on the background. Comparing brain activity during attention to the dialogues with activity during control conditions, where the dialogues are ignored and fixation cross is to be attended, allowed us to determine attention-related top-down effects and distinguish them from stimulus-dependent bottom-up effects (Alho et al., 2014).

We predicted that both increased speech intelligibility and increased amount of visual speech information in the attended speech would be associated with stronger stimulus-dependent activity in the STG/STS as well as subsequent activity in brain areas involved in linguistic processing. Moreover, we hypothesized that degrading auditory or visual quality of attended speech might be related to increased fronto-parietal activity due to enhanced attentional demands. Finally, we were interested to see whether attention to audiovisual speech and the quality of this speech would have interactions in some brain areas involved in auditory, visual or linguistic processing, or in the control of attention.

METHODS

Participants

Fifteen healthy right-handed adult volunteers (5 males, age range 20–38 years, mean 25.3 years) participated in the present study. All participants were native Finnish speakers with normal hearing, normal or corrected-to-normal vision, and no history of psychiatric or neurological illnesses. Handedness was verified by the Edinburgh Handedness Inventory (Oldfield, 1971). An informed written consent was obtained from each participant before the experiment. The experimental protocol was approved by the Ethics Review Board in the Humanities and Social and Behavioral Sciences, University of Helsinki.

Stimuli

Stimulus Preparation

The stimuli consisted of 36 video clips showing scripted spoken dialogues (see **Table 1** for an example of a dialogue). The topics of dialogues were neutral, such as weather, vacation, and study plans. The syntactic structure of dialogues was matched as closely as possible. An independent native Finnish speaker subsequently verified the neutrality of dialogues as well as their meaningfulness and grammaticality. Each dialogue always consisted of seven lines spoken alternatingly by two actors, and each line contained 9–13 words.

The stimulus recordings took place in a soundproof studio. The video clips were recorded with a wide angle (23.5 mm G lens) HXR-NX70E digital video camera (SONY Corporation, Tokyo, Japan). Two external microphones were attached to the camera in order to record the left and right audio channels separately (48 kHz sampling frequency, 16-bit quantization).

The actors were two native Finnish speakers (a male and female university student recruited for the recording purposes). They were unaware of the experimental setup and were compensated for their work. The actors memorized the dialogues beforehand but uttered their lines with a natural pace. An external prompter (programmed with Matlab version R2016, Mathworks Inc., Natick, MA, United States) was used to remind each actor to hold a pause before uttering the next line. The pause duration information was used in the subsequent fMRI data processing. The mean duration of dialogues was 60 s (range 55–65 s) with mean line duration of 5.4 s and inter-line pause duration of 3.4 s. Half of the dialogues started with the female speaker and the other half with the male speaker. The speakers sat next to one another with their faces slightly tilted toward each other, making the visual speech setting as natural as possible while maintaining visual speech information visible to a viewer. The video data were then edited with Corel VideoStudio Pro X 8 software (Corel Corporation, Ottawa, ON, Canada) and, finally, with Matlab, see below. The video clips were cut into separate dialogues with 720 ms (18 frames) before the first and after the last spoken words. Thereafter, the videos were split into separate video and audio channels for subsequent editing with Adobe Audition CS6 (Adobe Systems Inc., San Jose, CA, United States) software. The audio channels were then converted to mono, cleaned from all non-voice background sounds, low-pass filtered at 5000 Hz, and scaled to have the same peak sound energy in all dialogues.

In addition to the natural speech, the audio data were noise-vocoded (Shannon et al., 1995; Davis and Johnsrude, 2003) using Praat software (version 6.0.27; Boersma and Weenink, 2001). The audio files were divided into 2 and 4 logarithmically spaced frequency bands between 300 and 5000 Hz (2 band cut-off points: 300, 1385, 5000 and 4 band cut-off points 300, 684, 1385, 2665, 5000). The filter bandwidths were set to represent equal distances along the basilar membrane (according to the Greenwood (1990) equation relating filter position to best frequency). The amplitude envelope from each frequency band was extracted using the standard Praat algorithm. The extracted envelope was then applied to band-pass filtered noise in the same frequency bands. Then, the resulting bands of modulated noise were recombined to produce the distorted speech. Noise vocoded speech sounds like a harsh robotic whisper (Davis and Johnsrude, 2003). Finally, the unchanged F0 (frequencies 0–300 Hz) was added to the noise-vocoded speech in order to maintain the speakers' gender identity clearly perceivable and their voices distinguishable from the irrelevant voice speaking in the background (see below). The speech was perceived to be hardly intelligible with 3 frequency bands (i.e., 2 noise-vocoded bands and the intact F0 band) and quite intelligible with 5 frequency bands (i.e., 4 noise-vocoded bands and the intact F0 band). These two frequency-band manipulations for noise-vocoding were assumed to be optimal for our study on the basis of a behavioral pilot experiment. In this pilot experiment, 5 listeners (not included in the actual fMRI experiment) rated the intelligibility of seven dialogues noise-vocoded across a wide range of frequency bands (2, 4, 6, 8, 10, 12 or 16) with a non-vocoded F0 band. The participants listened to the dialogues one line at a time and provided a typed report on what they could hear. On average, for 2 and 4 noise-vocoded bands, 26.2% (*SD* = 18.6%) and 76.4% (*SD* = 10.3%) of the lines were perceived correctly.

TABLE 1 | Example of one natural speech dialogue by two actors (A and B) used in the experiment.

Dialogue lines	Approximate english translation
A: Pitäisi kohta käydä kaupassa hakemassa välipalaa. Ostanko sinullekin jotain?	A: I should go soon to the store and get something to eat. Should I get something for you as well?
B: Ei kiitos tarvitse, minä pakkasin leipää ja jogurttia tänä aamuna lounaaksi.	B: No thanks, I packed bread and yogurt with me for lunch today.
A: Hyvä on, mutta haluatko tulla mukaan seuraksi kauppaan kuitenkin?	A: Okay, but would you still like to come along with me to the store?
B: Mielelläni, voisin katsoa, jos löytäisin sieltä jotain syötävää myöhemmälle.	B: With pleasure, I could see if I would find something to eat later.
A: Haluatko tulla kanssani puistoon syömään, kun olemme tulleet kaupasta?	A: Would you like to come with me to the park to eat after visiting the store?
B: Ulkona on aika kylmä tänään. Mentäisiinkö mieluummin jonnekin sisälle?	B: It's quite cold outside today. Should we rather go somewhere inside?
A: Totta, voisimme siinä tapauksessa syödä täällä yliopiston kahvihuoneessa.	A: True, in that case we could eat here at the university in the coffee room.

In addition to manipulating auditory information, we parametrically varied the amount of visual speech seen by the participants. This was done by adding different amounts of dynamic white noise onto the region in the videos showing the speakers' faces. Noise was added with Matlab R2016 using built-in functions and custom-made scripts with the following procedure. First, we constructed Gaussian masks for both faces (faces localized from the first frame of the video with Matlab's vision.CascadeObjectDetector). Then we generated two samples (one for each face) of white noise (using Matlab's randn function), and multiplied the noises with the facemasks in order to add noise smoothly only onto the faces. The same sample of noise was added to R, G, and B channels. This was repeated for every frame, and thus the noise was dynamic and it changed in every frame. To get different levels of visual quality, the amount of added noise was scaled so that the root-mean-contrast of the low-quality videos were 20 and 15% lower than the contrast of the highest quality video. Five experienced viewers (not included in the actual fMRI experiment) confirmed that adding the noise reduced the visual quality so that the mouth movements and facial features were only poorly visible at highest noise level (**Figure 1**).

In the final step of stimulus preparation procedure, we recombined the "poor," "medium," and "good" auditory quality sound files (with 2 noise-vocoded bands and an intact F0 band, 4 noise-vocoded bands and an intact F0 band, and clear intact speech, respectively) with the "poor," "medium," and "good" visual quality video files (more masked poorly perceivable visual speech, less masked quite perceivable visual speech, and unmasked clear visual speech, respectively) video files using a custom-made Matlab script. The resulted videos were then compressed using VirtualDub software[1]. Example stimuli of the three experimental conditions (good auditory quality and good visual quality; medium auditory quality and medium visual quality; poor auditory quality and poor visual quality) can be found online[2]. Written informed consent was obtained from the actors to publish identifiable image information.

Taken together, each dialogue had 3 visual and 3 auditory quality variants, which resulted in altogether 9 experimental conditions, one for each quality combination (e.g., poor visual and good auditory quality) with three dialogues in each. All combinations were presented to the participants but each participant saw a different variant of each dialogue.

Furthermore, to increase the attentional load, we added continuous background speech as an auditory distractor. For this purpose, we chose a cultural history audio book (the Finnish

---

[1] http://www.virtualdub.org

[2] https://osf.io/h9er7/

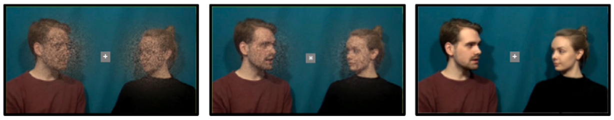


FIGURE 1 | Examples of the visual manipulations. From left to right: Poor, medium and good visual quality. The fixation cross was present during each dialogue and the participants were asked to focus on the cross. Written informed consent was obtained from the actors to publish identifiable image information.

translation of *The Autumn of the Middle Ages* by Johan Huizinga), which is freely distributed online by the Finnish Broadcasting Company (Yleisradio, YLE)[3]. The book was read by a female professional Finnish-native actor. In order for the F0 in this auditory distractor to be perceived approximately equidistant from the F0s of our female (200 Hz) and male (122 Hz) actors, we manipulated the F0 of the reader's voice by using square root of the mean of the female and male voices in the recorded video clips. After some further manipulations based on the estimation of three experienced listeners, the resulting F0 was 156 Hz. The F0 manipulation was performed in Audacity software[4]. The background speech was otherwise presented in its natural form and low-pass filtered at 5000 Hz to match the audio used in the experimental conditions. The audiobook was always presented as clear (i.e., non-vocoded) speech in the background. In addition, loudness differences between attended and unattended speech were kept minimal, as verified by three experienced listeners.

Procedure

Stimulus presentation was controlled through a script written in Presentation 20.0 software (Neurobehavioral Systems Inc., Berkeley, CA, United States). The video clips were projected onto a mirror mounted on the head coil and presented in the middle of the screen. All auditory stimuli were presented binaurally through insert earphones (Sensimetrics model S14; Sensimetrics, Malden, MA, United States). The experiment consisted of 3 functional runs with all 9 experimental conditions (Auditory Quality either poor, medium or good and Visual Quality either poor, medium or good) presented in each run along with 2 visual control conditions. The order of conditions was also randomized; however, the visual control conditions were always presented at the 6th and 7th place within a run. There was a small break of 40 s between these two dialogues. During the rest period, the participants were asked to focus on the fixation cross. Within all three functional runs, the order of the conditions was randomized for each participant. The competing audio distractor (audiobook) was presented 500–2000 ms before video onset and stopped at the offset of the video. The differing durations of dialogues were compensated for by inserting periods with a fixation cross between the instruction and the onset of the dialogue, keeping the overall trial durations constant. The intensity of the sounds was individually set to a loud, but pleasant level, and was approximately 80 dB SPL as measured from the tip of the earphones.

Attention-to-Speech Conditions

In the attention-to-speech conditions, the participants were asked to attentively watch the videos, ignore the background speech, and after each 7-line dialogue answer to seven questions, one question related to each line of the dialogue. More specifically, they were instructed to answer whether a certain topic was discussed in a particular line (see **Table 2**) by pressing the "Yes" or "No" button on a response pad (LUMItouch, Photon control Inc., Richmond, BC, Canada) with their right index or middle finger, respectively. Each written question was presented on the screen for 2 s, during which the participant gave his/her answer. Regardless of the duration of the participant's answer, the next question always started 2 s after the previous one. After the 7 questions, the participants were provided with immediate feedback on their performance (number of the correct answers), and the fixed duration of feedback was 2 s. The next dialogue was presented after a short (2-s) written instruction, telling whether the task was an attention-to-speech or a control (see section Attention-to-the-Fixation-Cross Condition) dialogue, followed by the fixation cross period, presented for 3–13 s to make all the trials equally long. All video clips had a rotating white fixation cross (inside a light gray box), placed in the middle of the screen, with a minimum of 9 and maximum of 15 rotations at a random interval but with minimum of 3 s between rotations. In the attention-to-speech condition, the participants' task was to ignore the cross and concentrate on viewing the people speaking.

Attention-to-the-Fixation-Cross Condition

In addition to the attention-to-speech conditions, we included two control conditions. These consisted of videos with a combination of good auditory and good visual quality and a combination of poor auditory and poor visual quality. Note that only these auditory and visual quality combinations were included into control conditions, because adding the other seven combinations to all three runs would have prolonged the total duration of the experiment by about 30 min and made the experiment too long to be conducted in a single experimental session. The dialogues in these control conditions were the ones not used in the attention-to-speech conditions. Identically to the attention-to-speech conditions, all video clips had a rotating white fixation cross (inside a light gray box), placed in the middle of the screen, with a minimum of 9 and maximum of 15 rotations from "×" to "+", or vice versa, at random intervals with a minimum of 3 s between the rotations. Thus, the attention-to-the-fixation-cross conditions used the same setup as the attention-to-speech conditions, but the task of the participants was to concentrate on counting the number of times the fixation cross rotated and ignore the dialogue and the background voice. After each control block, the participants were presented with seven questions ("Did the cross turn X times?"; the X being 9, 10, 11, 12, 13, 14, or 15 in an ascending order), and they were asked to answer each question with "Yes" or "No" by pressing the corresponding button on the response pad with their right index or middle

---

[3] https://areena.yle.fi/1-3529001

[4] https://sourceforge.net/projects/audacity/

TABLE 2 | Example of quiz questions of a practice dialogue.

Dialogue	English translation	A related quiz question "Did the speakers discuss this topic?"	Correct answer
A: Ostin uuden puhelimen ja siinä on niin paljon toimintoja, että olen sen kanssa ihan hukassa.	A: I bought a new phone and it has so many features that I am completely lost.	Puhuja hukkasi puhelimensa./The speaker lost his/her phone.	No
B: Ai niin joo, sinulla oli ennen sellainen ikivanha kännykkä, joka ei ollut edes älypuhelin.	B: Oh yes, you used to have that ancient phone, which was not even a smart phone.	Puhujan kännykkä oli vanha./The speaker's phone was old.	Yes
A: Joo, ja se oli aivan hyvä puhelin, siihen asti kunnes kissa pudotti sen pöydältä lattialle, se oli sitten siinä.	A: Yes, and it was a perfectly good phone until my cat dropped it on the floor from a table, and that was it.	Koira rikkoi puhelimen./The dog broke the phone.	No
B: Minä kun luulin, että vanhat kännykät kestävät kaiken eivätkä menisi mistään rikki.	B: I thought that old phones take all hits and wouldn't break at all.	Puhuja ihaili uutta puhelinta./The speaker was admiring the new phone.	No
A: No se on kyllä pudonnut monta kertaa, mutta kestänyt kaikki iskut mutta tämä taisi olla sille liikaa.	A: It has indeed fallen many times and always stayed intact but now this was too much for it.	Vanha kännykkä kesti iskut./The old mobile endured all hits.	Yes
B: No, mutta toivotaan, että tässä uudessa kännykässäsi kestää akku hyvin ja olet siihen muutenkin tyytyväinen.	B: Well, let's hope that your new phone has a long-lasting battery and that you are satisfied with it in all aspects.	Puhujalla ei ollut laturia mukanaan./The speaker did not have a charger with him/her.	No
A: Nyt on vielä vähän hankalaa, enkä osaa sitä oikein käyttää mutta kyllä se varmaan tästä!	A: It is still a bit difficult and I really don't know how to use it but I think it will be fine!	Puhuja ei osannut käyttää älypuhelinta./The speaker did not know how to use the smartphone.	Yes

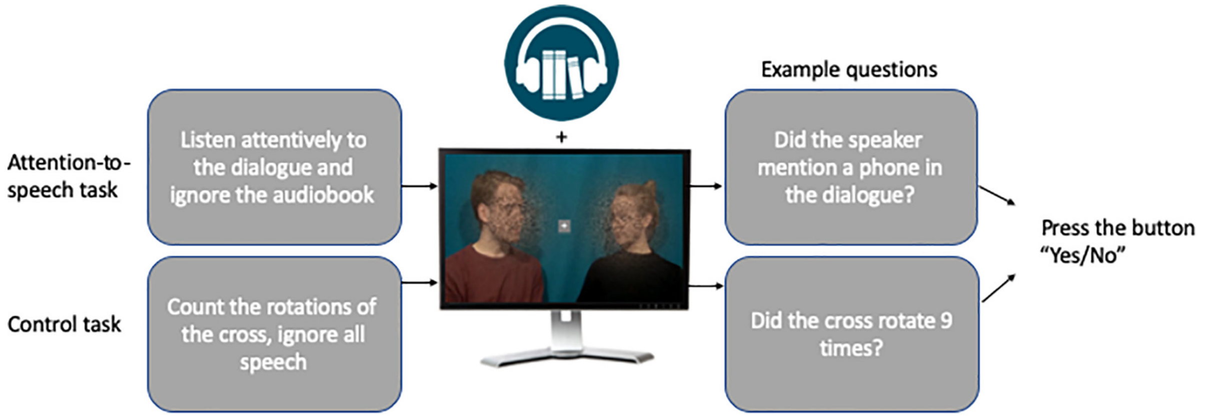


FIGURE 2 | Visualization of the attention-to-speech and control conditions. Written informed consent was obtained from the actors to publish identifiable image information.

finger, respectively. Thereafter, the participants were provided with immediate feedback (e.g., "6/7 correct"). For a schematic presentation of one trial, see **Figure 2**.

Practice Trial

Before the actual fMRI scanning, the participants were familiarized with the task outside the scanner by viewing one practice dialogue with all conditions and answering questions related to its content.

Data Acquisition

Functional brain imaging was carried out with 3T MAGNETOM Skyra whole-body scanner (Siemens Healthcare, Erlangen, Germany) using a 30-channel head coil. The functional echo planar (EPI) images were acquired with an imaging area consisting of 43 contiguous oblique axial slices (TR 2530 ms, TE 32 ms, flip angle 75°, voxel matrix 64 × 64, field of view 20 cm, slice thickness 3.0 mm, in-plane resolution 3.1 mm × 3.1 mm × 3.0 mm). Three functional runs of 368 volumes were measured for each participant. A total of 1158 functional volumes were obtained in one session (session duration approximately 50 min). High-resolution anatomical images (voxel matrix 256 × 256, in-plane resolution 1 mm × 1 mm × 1 mm) were acquired from each participant prior to the functional runs.

Data Analysis

The fMRI data were pre-processed and analyzed in Statistical Parametric Mapping (SPM12; Wellcome Trust Centre for Neuroimaging, London, United Kingdom). The first 4 volumes in each run were dummies and were discarded in further analysis of the data, leaving 382 total volumes per run to be analyzed. The data were slice-time corrected, motion corrected, realigned to the middle image from each run, high-pass filtered (cutoff 1/260 Hz) and spatially smoothed with a Gaussian kernel of 6 mm. The images were normalized to MNI space using a standard pre-processing function in Conn software (Whitfield-Gabrieli and Nieto-Castanon, 2012). For the first-level statistical analysis, the general linear model was created including a regressor for each condition. Separate regressors were also included for (1) the instructions and the responses from the participant and (2) the quiz. This resulted in 13 regressors in total. Additionally, six movement parameters (3 translations, 3 rotations) were included as nuisance regressors. The conditions were modeled using a standard boxcar model. For the second-level analysis, we used the Multivariate and Repeated Measures (MRM) toolbox (McFarquhar et al., 2016). The contrast images of the nine experimental conditions compared to rest from each participant were entered into a 3 × 3 full factorial repeated-measures analysis of variance (ANOVA) with factors Visual Quality (3 levels: poor, medium, good) and Auditory Quality (3 levels: poor, medium, good). Within this model, F-contrasts were computed for the main effects and the interaction effect. A separate 2 × 2 repeated-measures ANOVA was conducted to account for stimulus quality and attentional effects. This additional ANOVA included factors Audiovisual Quality (2 levels: poor auditory and poor visual quality vs. good auditory and good visual quality) and Attention (2 levels: attention to speech vs. attention to the fixation cross). All reported contrasts were thresholded voxel-wise at $p < 0.001$ with a cluster extend threshold of 100 voxels, resulting activity maps that were family-wise error (FWE) corrected at the cluster level, p(FWE) < 0.05.

Statistical analyses of the performance data, that is, responses to the quiz questions during the attention-to-speech condition were submitted to the repeated-measures ANOVA with factors Visual Quality (poor, medium, good) and Auditory Quality (poor, medium, good). Responses to the quiz questions during the attention-to-the-fixation cross condition were submitted to the two-tailed pairwise *t*-test. For all analyses, the Greenhouse-Geisser correction was applied when the assumption of sphericity was violated. IBM SPSS Statistics 24 (IBM SPSS, Armonk, NY, United States) was used for conducting these analyses.

In addition to brain areas showing significant effects of factors included in ANOVAs, we studied brain activity in regions of interest (ROIs) known to be involved in low-level auditory processing and speech processing. These ROIs were located bilaterally in Heschl's gyrus (HG), the anterior, mid and posterior STG, as well as in Broca's area in the left hemisphere and its right hemisphere analogue (Liakakis et al., 2011; Alho et al., 2014; Liebenthal et al., 2014). In addition, due to our focus on visual speech processing, there were additional ROIs in the left and right fusiform face area (FFA; Grill-Spector and Weiner, 2014). These ROIs were based on the Harvard and Oxford cortical structural atlas[5].

RESULTS

Behavioral Results

The mean performance scores for the attention-to-speech conditions are shown in **Figure 3**. Behavioral results demonstrated a significant main effect of Auditory Quality [$F(2, 28) = 57.57$, $p = 0.001$, $\eta_p^2 = 0.80$] and a significant main effect of Visual Quality [$F(2, 28) = 8.2$, $p = 0.002$, $\eta_p^2 = 0.37$]. Although visual quality appeared to have a slightly stronger effect on performance when the auditory quality was poor than when it was medium or good (see **Figure 5**), the interaction between the two factors did not reach significance [$F(4, 56) = 1.64$, $p = 0.176$]. Bonferroni-corrected *post hoc* tests for Auditory Quality revealed significant differences between all Auditory Quality conditions (for all comparisons, $p < 0.001$). *Post hoc* tests for Visual Quality revealed significant differences between the poor and good quality conditions ($p < 0.001$) and between medium and good quality conditions ($p = 0.029$).

The mean performance scores for the attention-to-the-fixation-cross conditions were 6.7/7 (SEM 0.16/7) and 6.5/7 (SEM 0.36/7) for the poor audiovisual quality condition and good audiovisual quality condition, respectively ($p = 0.078$).

fMRI Results

Auditory and Visual Quality

Figure 4 depicts the brain areas where the 3 × 3 ANOVA showed significant main effects of Auditory Quality (3 levels) and **Figure 5** significant main effects of Visual Quality (3 levels) on brain activity measured during attention to speech. **Figures 4** and **5** also depict mean parameter estimates of significant clusters, displaying the direction of the observed cluster effect. No significant interactions between these factors were found with the applied significance threshold.

As seen in **Figure 4A**, Auditory Quality showed a significant effect on brain activity in the STG/STS bilaterally, these effects extending from mid-STG/STS areas to the temporal poles, the left angular gyrus, and the left superior frontal gyrus. **Figure 4B** demonstrates that in all these areas activity was enhanced with increasing auditory quality. **Figures 4C,D** also depict activity in additional ROIs bilaterally: HG, the anterior, mid and posterior STG, and Broca's area in the left hemisphere and its right-hemisphere analogue.

Visual Quality, in turn, had a significant effect on brain activity in the temporal and occipital cortices (**Figure 5A**). As seen in **Figure 5B**, increasing visual quality was associated with enhanced activity in the STS bilaterally, this activity extended in the right hemisphere even to the temporal pole, and in the left hemisphere to the inferior frontal gyrus. However, **Figure 5C** shows that activations increased with decreasing visual quality in the right fusiform gyrus and in the bilateral middle occipital gyrus predominantly in the left hemisphere. **Figure 5D** depicts activity in the bilateral FFA ROIs.

Additional ANOVAs conducted separately for activity in the HG, Broca's area and FFA ROIs included the factors Auditory Quality, Visual Quality, and Hemisphere (to avoid double-dipping, no such ANOVAs were performed for the STG ROIs covered already by the aforementioned 3 × 3 ANOVA). The results indicated a significant main effect of Hemisphere for all ROIs (HG [$F(1, 14) = 19.09$, $p = 0.001$, $\eta_p^2 = 0.58$], Broca's area [$F(1, 14) = 9.53$, $p = 0.008$, $\eta_p^2 = 0.41$] and FFA [$F(1, 14) = 11.81$, $p = 0.004$, $\eta_p^2 = 0.46$]). No other main or interaction effects were found for the HG and FFA ROIs, however, for the Broca's area ROI there was a significant interaction effect between

---

[5] https://identifiers.org/neurovault.collection:262

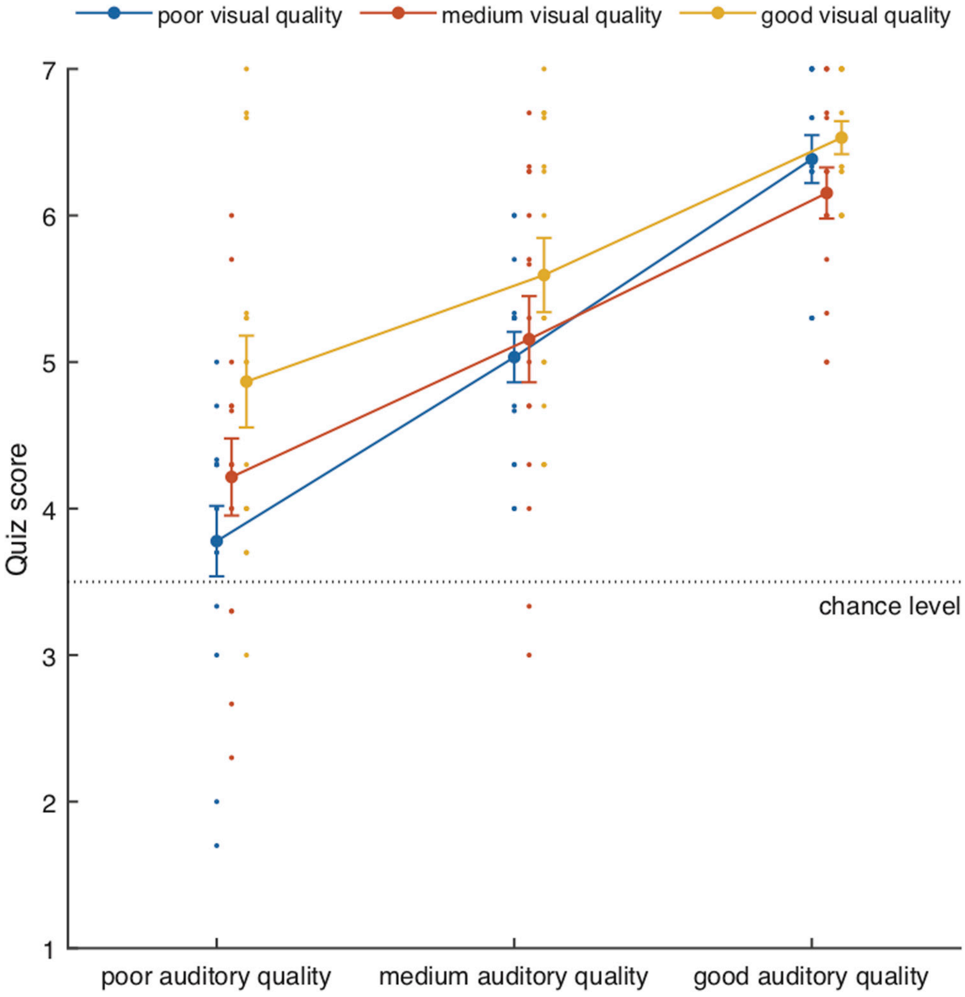


FIGURE 3 | Mean performance scores (± SEM) together with individual data points as a function of Auditory Quality and Visual Quality. Chance level was 3.5.

Hemisphere and Auditory Quality [$F(2, 28) = 7.4$, $p = 0.007$, $\eta_p{}^2 = 0.35$].

Attention-to-Speech vs. Attention-to-the-Fixation Cross

Figure 6 depicts the brain areas where the 2 × 2 ANOVA showed significant main effects of Audiovisual Quality (2 levels: poor auditory and poor visual quality vs. good auditory and good visual quality). **Figure 7** shows the brain areas where the 2 × 2 ANOVA showed significant main effects of Attention (2 levels: attention-to-speech vs. attention-to-the-fixation cross) on brain activity. **Figures 6** and **7** also depict mean parameter estimates of significant clusters, displaying the direction of the observed cluster effect. No significant interactions between the factors Attention and Audiovisual Quality were observed with the applied significance threshold. As seen in **Figure 6A**, in the 2 × 2 ANOVA, there was a significant effect of Audiovisual Quality bilaterally in the STG/STS, these activations extended to the temporal poles, as well as to the left superior parietal lobule, left precuneus, the dorsal part of the right inferior parietal lobule, and in the middle occipital gyrus bilaterally. As seen in **Figure 6B**, in the left and right superior temporal gyri, activity was enhanced with increasing audiovisual quality both during attention-to-speech and attention-to-the-fixation cross. In contrast, in both attention conditions, activity was higher in the left superior parietal lobule, the right inferior parietal, the left precuneus, and bilateral middle occipital gyrus for poorer audiovisual quality (**Figure 6C**). **Figure 6C** shows activity in additional ROIs: the left HG and anterior, mid and, posterior STG, Broca's area, and the right FFA. Results for the additional ANOVAs conducted separately for activity in the left HG, Broca's area and right FFA ROIs are described above.

Figure 7 demonstrates that Attention had a significant effect on brain activity in the left planum polare, the left angular gyrus, the left lingual gyrus, the right temporal pole, the right supramarginal gyrus, the right inferior parietal lobule, as well as in the ventromedial prefrontal cortex/orbitofrontal cortex and posterior cingulate bilaterally. In all these areas, activity

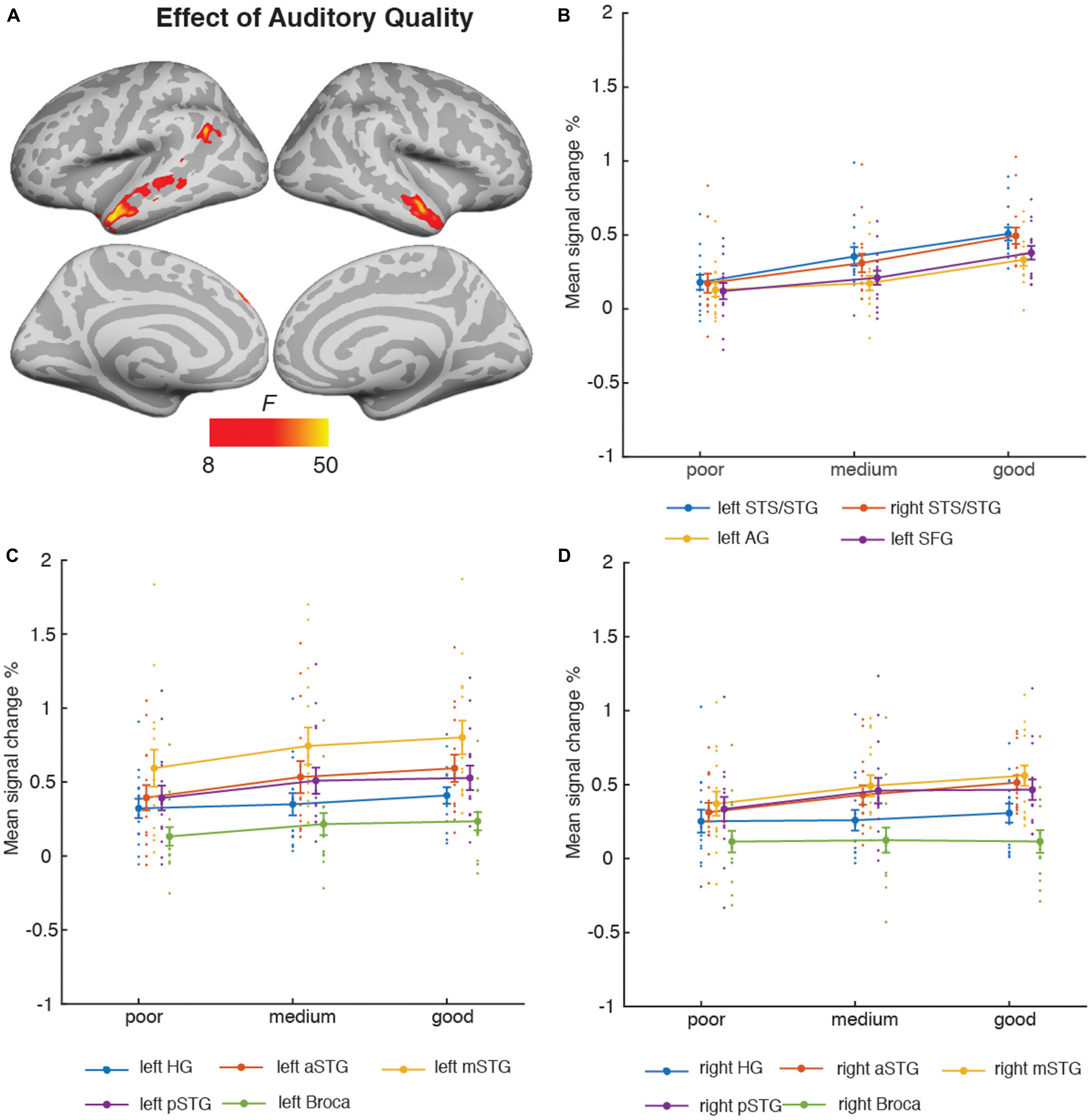


FIGURE 4 | (A) Brain areas showing significant main effects of Auditory Quality of attended audiovisual speech; p(cluster-level FWE) < 0.05, voxel-wise threshold $p < 0.001$, cluster extent $k = 100$. **(B)** Mean signal changes (%) together with individual data points, compared with rest in the clusters showing significant effects of Auditory Quality. **(C)** Mean signal changes (%) together with individual data points compared with rest in the additional left hemisphere ROIs. **(D)** Mean signal changes (%) together with individual data points compared with rest in the additional right hemisphere ROIs. STG, superior temporal gyrus (a, m and p for anterior, mid and posterior, respectively); STS, superior temporal sulcus; AG, angular gyrus; SFG, superior frontal gyrus, HG, Heschl's gyrus.

was higher during attention-to- speech than during attention-to-the-fixation cross.

For additional data, see **Supplementary Material**. **Supplementary Figure S1** presents correlations of the behavioral data with BOLD signal in selected ROIs. **Supplementary Figure S2** depicts raw BOLD activity for a sub-sample of participants in selected ROIs. **Supplementary Figure S3** depicts activity in all conditions for all participants. **Supplementary Figure S4** demonstrates individual participant data for a sub-sample of participants.

DISCUSSION

We investigated brain areas activated during selective attention to audiovisual dialogues. In particular, we expected

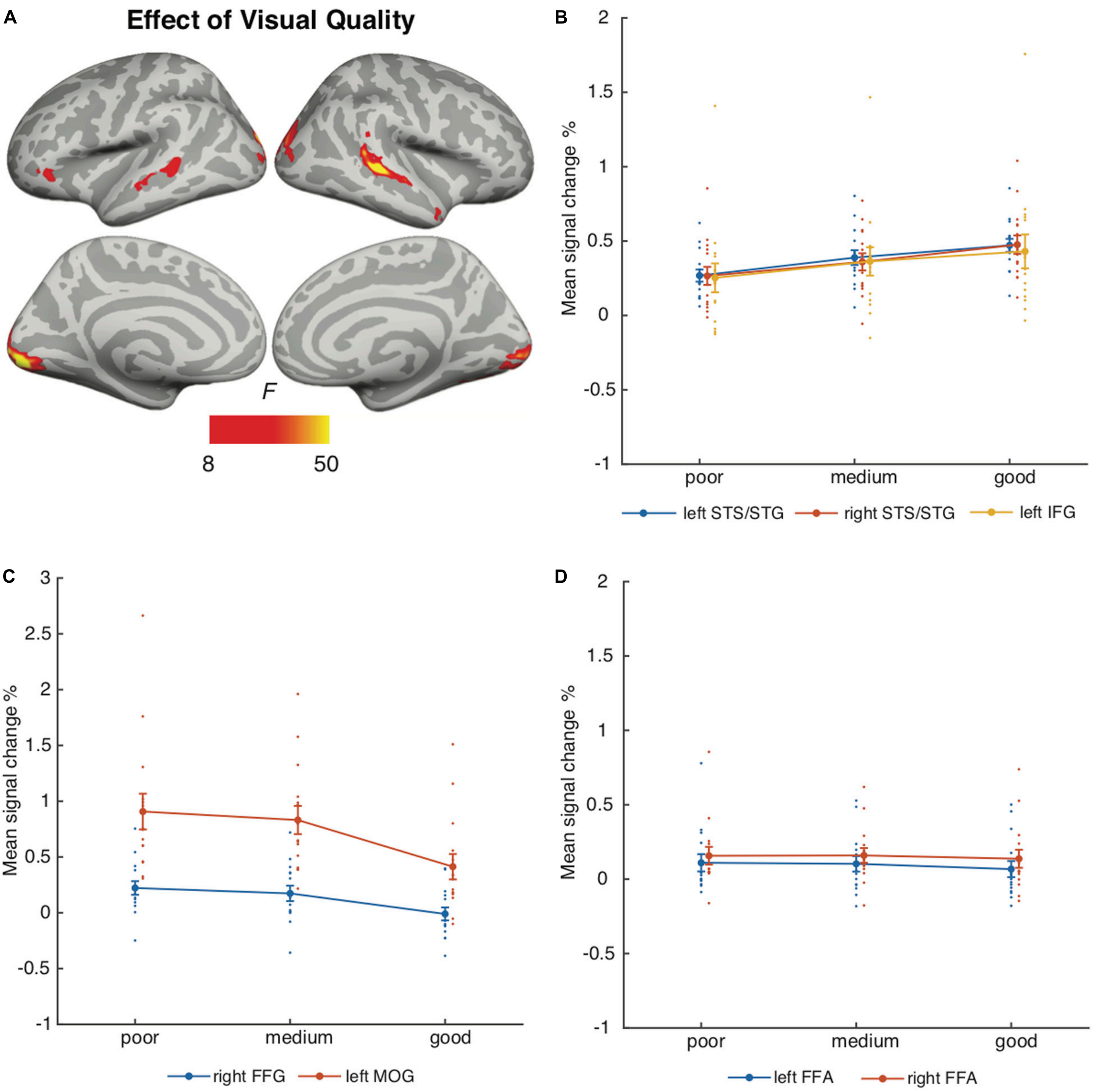


FIGURE 5 | (A) Brain areas showing significant main effects of Visual Quality of attended audiovisual speech; p(cluster-level FWE) < 0.05, voxel-wise threshold $p < 0.001$, cluster extent $k = 100$. **(B)** Mean signal changes (%) together with individual data points compared with rest in the clusters showing significant effects of Visual Quality conditions (clusters showing increasing activity with increasing visual quality). **(C)** Mean signal changes (%) together with individual data points, compared with rest in the clusters showing significant effects of Visual Quality conditions (clusters showing decreasing activity with increasing visual quality). **(D)** Mean signal changes (%) together with individual data points compared with rest in the additional ROIs. STG, superior temporal gyrus; STS, superior temporal sulcus; HG, Heschl's gyrus; IFG, inferior frontal gyrus; FFG, fusiform gyrus; MOG, middle occipital gyrus, FFA, fusiform face area.

attention-related modulations in the auditory cortex and fronto-parietal activity during selective attention to naturalistic dialogues with varying auditory and visual quality. Behaviorally, we observed that increased quality of both auditory and visual information resulted in improved accuracy in answering to the questions related to the content of dialogues. Hence, expectedly, both increased auditory quality (e.g., Davis and Johnsrude, 2003) and increased visual quality (Sumby and Pollack, 1954) facilitated speech comprehension. However, no significant interaction between Auditory Quality and Visual Quality was observed. Thus, our results are not able to give full support to maximal facilitation of speech processing by visual speech at the intermediate signal-to-noise ratio reported, for instance, by McGettigan et al. (2012) and Ross et al. (2007).

In the fMRI analysis, the main effect of Auditory Quality showed that increasing speech quality was associated with

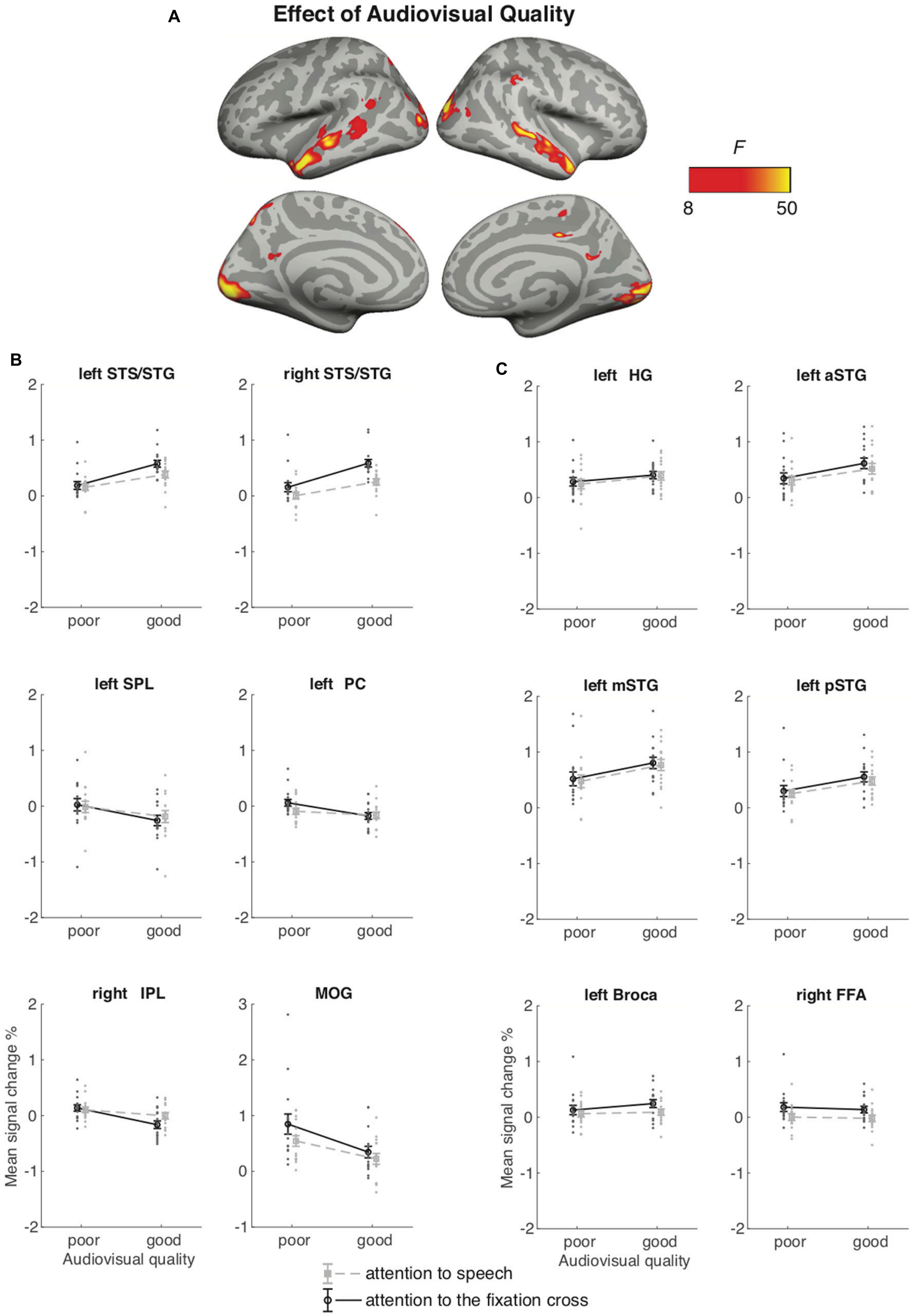


FIGURE 6 | (A) Brain areas showing significant main effects of Audiovisual Quality (good auditory and good visual quality vs. poor auditory and poor visual quality) across conditions with attention to speech and attention to the fixation cross; p(cluster-level FWE) < 0.05, voxel-wise threshold $p < 0.001$, cluster extent $k = 100$. **(B)** Mean signal changes (%) together with individual data points, compared with rest in the clusters showing significant effect of Audiovisual Quality. **(C)** Mean signal changes (%) together with individual data points compared with rest in the additional ROIs. STG, superior temporal gyrus (a, m and p for anterior, mid and posterior, respectively); STS, superior temporal sulcus; HG, Heschl's gyrus; IPL, inferior parietal lobule; SPL, superior parietal lobule; PC, precuneus; MOG, middle occipital gyrus, FFA, fusiform face area. Note the different scale on the y axis for MOG.

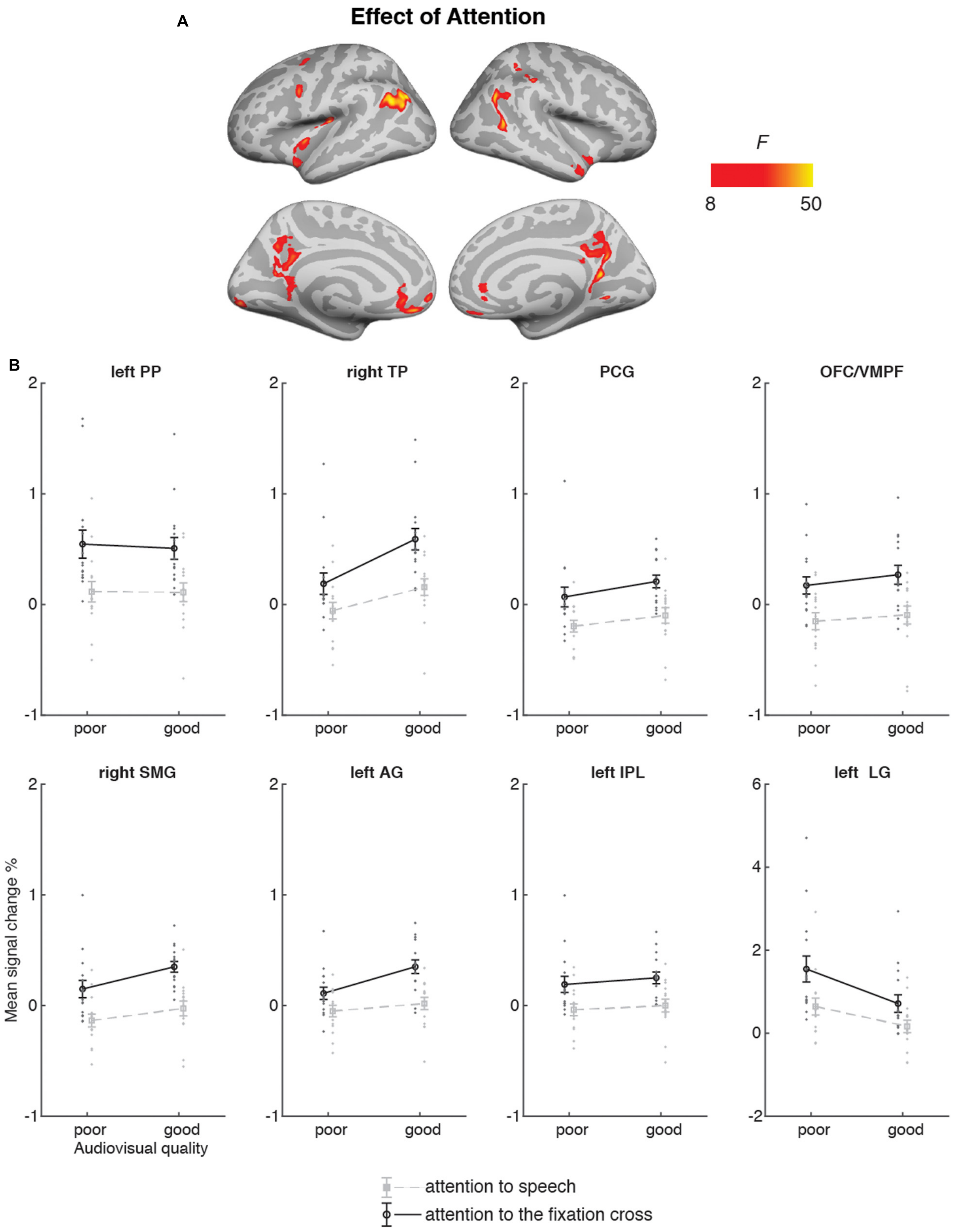


FIGURE 7 | (A) Brain areas showing significant main effects of Attention (attention audiovisual speech vs. attention to speech attention to the fixation cross); p(cluster-level FWE) < 0.05, voxel-wise threshold $p < 0.001$, cluster extent $k = 100$. **(B)** Mean signal changes (%) together with individual data points, compared with rest in the clusters in the showing significantly higher activity during attention to speech than during attention to the fixation cross conditions. No cluster showed an opposite effect. PP, planum polare; TP, temporal pole; PCG, posterior temporal gyrus; OFC/VMPF, orbitofrontal cortex/ventromedial prefrontal cortex; SMG, supramarginal gyrus; AG, angular gyrus; IPL, inferior parietal lobule; LG, lingual gyrus. Note the different scale on the y axis for left LG.

increased activity in the (bilateral) STG/STS, which corroborates previous studies on speech intelligibility (e.g., Scott et al., 2000; Davis and Johnsrude, 2003; Obleser et al., 2007; Okada et al., 2010; McGettigan et al., 2012; Evans et al., 2014, 2016). The bilaterally enhanced activity in the STG/STS is most probably related to increased speech comprehension with increasing availability of linguistic information. The STG/STS activity extended to the temporal pole, which might be associated with enhanced semantic processing with increasing speech quality (Patterson et al., 2007). The right STG/STS activity observed here might also be related to prosodic processing during attentive listening (Alho et al., 2006; McGettigan et al., 2013; Kyong et al., 2014). The right temporal pole, in turn, its most anterior part in particular, has also been associated with social cognition (Olson et al., 2013), which may have been triggered by our naturalistic audiovisual dialogues. In addition, we observed increasing activity in the left angular gyrus and left medial frontal gyrus with increasing speech intelligibility. Enhanced activity in the left angular gyrus may reflect successful speech comprehension, stemming either from increased speech quality or from facilitated semantic processing due to improved speech quality (Humphries et al., 2007; Obleser and Kotz, 2010). The left medial frontal gyrus, in turn, has been attributed to semantic processing as a part of a semantic network (Binder et al., 2009). Hence, an increase in these activations with improving speech quality implies a successful integration of linguistic information onto the existing semantic network and improved comprehension of the spoken input – extending beyond the STG/STS.

The main effect of Visual Quality demonstrated increasing activity in the bilateral occipital cortex and right fusiform gyrus with decreasing visual quality – areas related to object and face recognition, respectively (e.g., Weiner and Zilles, 2016). Enhanced activity in these areas might be due to, for instance, noise-modulation of the videos that contained more random motion on the screen than good quality videos. Visual noise has been shown to activate primary regions in the occipital cortex more than coherent motion (e.g., Braddick et al., 2001). It is, however, also possible that viewing masked visual speech required more visual attention than viewing the unmasked visual speech, perhaps contributing to enhanced activity in the degraded visual conditions especially in the fusiform gyrus. Nevertheless, activity in the middle occipital gyrus was higher for poor visual quality combined with poor auditory quality than for good visual and auditory quality even during attention to the fixation cross (**Figure 6**). This suggests that increased visual cortex activity for poorer visual quality was at least partly caused by random motion of the masker. In a study that used blurring to decrease the reliability of visual information, stronger activity (and connectivity with the STS) was found in the extrastriate visual cortex for more reliable (i.e., less noisy) visual information (Nath and Beauchamp, 2011). It is possible that enhancements in the extrastriate visual cortex were obscured by our noise-modulations, affecting processing already in the primary visual cortex. Activity enhancements with poor (contra good) audiovisual quality were also observed in the left superior parietal lobule, precuneus and right inferior parietal lobule in both attention conditions, implying contribution of random motion in the masker to these effects as well. Increased visual quality was also associated with enhanced activity in the bilateral STG/STS, corroborating other studies reporting these areas being involved in multisensory integration (e.g., Beauchamp et al., 2004a,b, Lee and Noppeney, 2011). Facilitation of speech processing in STG/STS areas by visual speech might be mediated by cortico-cortical connections between brain areas involved in visual and auditory speech processing (Cappe and Barone, 2005; van Wassenhove et al., 2005). We also observed an increase in the left IFG activity with increasing visual quality, an area related to the processing of high-order linguistic information (e.g., Obleser et al., 2007). Activity in the left IFG has been also associated with integration of auditory and visual speech (Lee and Noppeney, 2011), as well as speech and gestures (Willems et al., 2009), suggesting its involvement in multimodal integration also in the current study.

The 2 × 2 ANOVA for brain activity during attention to speech and during attention to the fixation cross showed a main effect of Audiovisual Quality in the bilateral STG/STS (extending to the temporal pole), due to higher activity for good auditory and good visual quality than for poor auditory and poor visual quality. These STG/STS effects were also observed during attention to the fixation cross, implying quite automatic bottom-up speech processing with enhanced audiovisual quality. The 2 × 2 ANOVA indicated also main effects of attention, that is, higher activity during attention to the dialogue than attention to the fixation cross in the left planum polare, angular and lingual gyrus, as well as the right temporal pole. We also observed activity in the dorsal part of the right inferior parietal lobule and supramarginal gyrus, as well as in the oribitofrontal/ventromedial frontal gyrus and posterior cingulate bilaterally. One might wonder why attending to the dialogues in relation to attending to the fixation cross was not associated with activity enhancements in the STG/STS as in some previous studies on selective attention to continuous speech (e.g., Alho et al., 2003, 2006). One possible explanation is the ease of the visual control task (i.e., counting the rotations of the fixation cross), eliminating the need to disregard audiovisual speech in the background altogether. This interpretation is also supported by the STG/STS activations observed even during attention to the fixation cross, at least when the audiovisual quality in the to-be-ignored dialogue was good (see **Figure 6**). Areas in the planum polare have been shown to be associated with task-related manipulations in relation to speech stimuli (Harinen et al., 2013; Wikman and Rinne, 2019).

Auditory attention effects have also been reported outside the STG/STS, for instance, in the middle and superior frontal gyri, precuneus, as well as superior parietal inferior and superior parietal lobule (e.g., Degerman et al., 2006; Salmi et al., 2007). These areas are at least partly involved in the top-down control of auditory cortex during selective attention. Interestingly, even though the participants attended to visual stimuli both during attention to speech and attention to the fixation cross, activity was higher in the lingual gyrus (approximately in areas V2/V3 of the visual cortex) during attention to speech. This effect is presumably explained by differences in visual attention between the tasks (see, e.g., Martínez et al., 1999). In other words, while both tasks demanded visual attention, task-related processing of visual speech was presumably more attention-demanding, especially when the faces were masked, than processing of fixation-cross rotations.

It should be noted that some activity enhancements during attention-to-speech in relation to attention- to-the-fixation-cross might not be associated with attention *per se*, but with higher effort in task performance during attention to the dialogues. However, if this was the case one would expect to see an interaction between Task and Stimulus Quality for the ANOVA depicted in **Figure 6**. That is, higher effort required to process the dialogues with poor auditory quality should modulate activations during the attention-to-speech conditions but not during the attention-to-the-fixation-cross conditions, when participants did not process the dialogues. Moreover, the 3 × 3 ANOVA for the nine auditory-visual quality combinations during the attend-to-speech conditions (**Figures 4** and **5**) did not show enhanced activations due to poorer stimulus quality, except for the bilateral occipital cortex and right fusiform gyrus, where activity increased with decreasing quality of visual speech (see **Figure 4B**). Also, as noted above, enhanced demands for visual attention may have contributed to these effects. Alternatively, these effects might be associated with enhanced effort in perceiving visual speech with decreasing visual quality. However, as discussed above, the 2 × 2 ANOVA for brain activity during attention-to-speech and during attention-to-the-fixation-cross showed a main effect of Audiovisual Quality in the middle occipital gyri, with higher activity for the poor-poor than good-good auditory-visual quality combination, but no significant interaction of Audiovisual Quality and Attention. Since these activity enhancements in the middle occipital gyri and in some parietal areas, associated with decreasing audiovisual quality, were quite similar during the two attention tasks (see **Figure 6B**), it is likely that these effects were due to bottom-up effects associated with visual stimulation differences between conditions with poor and good visual quality, rather than due to differences in effort.

Unfortunately, the relatively small number of participants in the present study does not allow for investigation of behavior-brain relationships. However, it should be noted that the main findings of the present effects of attention and auditory and visual quality of speech on brain activity and performance were replicated in our two subsequent fMRI studies, which are still in preparation (for preliminary results, see Wikman et al., 2018, 2019).

In line with the previous studies on selective attention to continuous speech (Alho et al., 2003, 2006; Scott et al., 2004), attention to audiovisual dialogues did not significantly engage dorsolateral prefrontal and superior parietal areas. This may be due to high automaticity of selective listening to continuous speech, which might, hence, be quite independent of fronto-parietal attentional control (Alho et al., 2006). However, for the present audiovisual attention to speech, we observed activation in the left inferior parietal lobule, which may be related to attentive auditory processing (e.g., Rinne et al., 2009; Alain et al., 2010).

Furthermore, attention to audiovisual speech elicited enhanced activity in the orbitofrontal/ventromedial prefrontal cortex in comparison with attention to the fixation cross. One possible explanation would be that this activity is related to processing of semantic information (e.g., Binder et al., 2009) in attended speech in contrast to visual information in the fixation cross. Alternatively, this effect may be related to the social aspect of the attended dialogues, since the ventromedial frontal area is associated with social cognition, such as theory of mind and moral judgment (Bzdok et al., 2012), as well as evaluation of other persons' traits (Araujo et al., 2013). Moreover, enhanced activity in the posterior cingulate and right superior temporal pole observed here during attention to speech may be related to social perception, as both these areas have been involved in social cognition (Bzdok et al., 2012). To our knowledge, no previous study has shown that attending to emotionally neutral dialogues would enhance activity in these three brain regions related to social perception and cognition.

To summarize, our study is the first to present findings on selective attention to natural audiovisual dialogues. Our results demonstrate that increased auditory and visual quality of speech facilitated selective listening to the dialogues, seen in enhanced brain activity in the bilateral STG/STS and the temporal pole. Enhanced activity in the temporal pole might be related to semantic processing particularly in the left hemisphere, whereas in the right hemisphere, it may index processing of social information activated during attention to the dialogues. The fronto-parietal network was associated with enhanced activity during attention to speech, reflecting top-down attentional control. Attention to audiovisual speech also activated the orbitofrontal/ventromedial prefrontal cortex – a region associated with social and semantic cognition. Hence, our findings on selective attention in realistic audiovisual dialogues emphasize not only involvement of brain networks related to audiovisual speech processing and semantic comprehension but, as a novel observation, the social brain network.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics Review Board in the Humanities and Social and Behavioral Sciences, University of Helsinki. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this manuscript.

AUTHOR CONTRIBUTIONS

AL, KA, MV, and VS designed the fMRI experiment. AL, MV, and VS prepared the stimuli. MV collected the fMRI data.

AL and MV performed the analysis. AL, MV, and KA wrote the manuscript in collaboration with PW, VS, and MM. MM, VS, and PW contributed to the fMRI data analysis.

FUNDING

This study was supported by the Academy of Finland (Grant No. 297848) and by the Russian Science Foundation (grant code RSF 19-18-00534). MV was awarded with Erasmus + scholarship.

ACKNOWLEDGMENTS

We thank Artturi Ylinen for help with stimulus preparation and data collection and Miika Leminen for methodological support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2020.00436/full#supplementary-material

FIGURE S1 | Correlations of the behavioral data with BOLD signal in selected ROIs. Y-axis: behavioral quiz scores (1–7), x-axis: BOLD signal change (%). A1–A3 denote poor, medium and good auditory quality; V1–V3 denote poor, medium and good visual quality. Each experimental condition is depicted in different color. STG, superior temporal gyrus (a, m, and p for anterior, mid and posterior, respectively); STS, superior temporal sulcus; HG, Heschl's gyrus; FFA, fusiform face area.

FIGURE S2 | Raw BOLD activity (first run) for five participants in selected ROIs. Y-axis: time, volume, x-axis: mean BOLD, arbitrary unit. STG, superior temporal gyrus (a, m, and p for anterior, mid and posterior, respectively); STS, superior temporal sulcus; HG, Heschl's gyrus; FFA, fusiform face area.

FIGURE S3 | The average signal changes across all conditions.

FIGURE S4 | Individual participant data for 5 participants across all conditions.

REFERENCES

Alain, C., Shen, D., Yu, H., and Grady, C. (2010). Dissociable memory- and response-related activity in parietal cortex during auditory spatial working memory. *Front. Psychol.* 1:202. doi: 10.3389/fpsyg.2010.00202

Alho, K., Medvedev, S. V., Pakhomov, S. V., Roudas, M. S., Tervaniemi, M., Reinikainen, K., et al. (1999). Selective tuning of the left and right auditory cortices during spatially directed attention. *Cogn. Brain Res.* 7, 335–341. doi: 10.1016/s0926-6410(98)00036-6

Alho, K., Rinne, T., Herron, T. J., and Woods, D. L. (2014). Stimulus-dependent activations and attention-related modulations in the auditory cortex: a meta-analysis of fMRI studies. *Hear. Res.* 307, 29–41. doi: 10.1016/j.heares.2013.08.001

Alho, K., Salmi, J., Koistinen, S., Salonen, O., and Rinne, T. (2015). Top-down controlled and bottom-up triggered orienting of auditory attention to pitch activate overlapping brain networks. *Brain Res.* 1626, 136–145. doi: 10.1016/j.brainres.2014.12.050

Alho, K., Vorobyev, V. A., Medvedev, S. V., Pakhomov, S. V., Roudas, M. S., Tervaniemi, M., et al. (2003). Hemispheric lateralization of cerebral blood-flow changes during selective listening to dichotically presented continuous speech. *Cogn. Brain Res.* 17, 201–211. doi: 10.1016/s0926-6410(03)00091-0

Alho, K., Vorobyev, V. A., Medvedev, S. V., Pakhomov, S. V., Starchenko, M. G., Tervaniemi, M., et al. (2006). Selective attention to human voice enhances brain activity bilaterally in the superior temporal sulcus. *Brain Res.* 1075, 142–150. doi: 10.1016/j.brainres.2005.11.103

Araujo, H. F., Kaplan, J., and Damasio, A. (2013). Cortical midline structures and autobiographical-self processes: an activation-likelihood estimation meta-analysis. *Front. Hum. Neurosci.* 7:548. doi: 10.3389/fnhum.2013.00548

Beauchamp, M. S., Argall, B. D., Bodurka, J., Duyn, J. H., and Martin, A. (2004a). Unraveling multisensory integration: patchy organization within human STS multisensory cortex. *Nat. Neurosci.* 7, 1190–1192. doi: 10.1038/nn1333

Beauchamp, M. S., Lee, K. E., Argall, B. D., and Martin, A. (2004b). Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron* 41, 809–823. doi: 10.1016/s0896-6273(04)00070-4

Binder, J. R., Desai, R. H., Graves, W. W., and Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cereb. Cortex* 19, 2767–2796. doi: 10.1093/cercor/bhp055

Bishop, C. W., and Miller, L. M. (2009). A multisensory cortical network for understanding speech in noise. *J. Cogn. Neurosci.* 21, 1790–1805. doi: 10.1162/jocn.2009.21118

Boersma, P., and Weenink, D. (2001). Praat: doing phonetics by computer. *Glot Int.* 5, 341–345. doi: 10.1016/j.jvoice.2019.07.004

Braddick, O. J., O'Brien, J. M., Wattam-Bell, J., Atkinson, J., Hartley, T., and Turner, R. (2001). Brain areas sensitive to coherent visual motion. *Perception* 30, 61–72. doi: 10.1068/p3048

Bzdok, D., Schilbach, L., Vogeley, K., Schneider, K., Laird, A. R., Langner, R., et al. (2012). Parsing the neural correlates of moral cognition: ALE meta-analysis on morality, theory of mind, and empathy. *Brain Struct. Funct.* 217, 783–796. doi: 10.1007/s00429-012-0380-y

Cappe, C., and Barone, B. (2005). Heteromodal connections supporting multisensory integration at low levels of cortical processing in the monkey. *Eur. J. Neurosci.* 22, 2886–2902. doi: 10.1111/j.1460-9568.2005.04462.x

Davis, M. H., and Johnsrude, I. S. (2003). Hierarchical processing in spoken language comprehension. *J. Neurosci.* 23, 3423–3431. doi: 10.1523/JNEUROSCI.23-08-03423.2003

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

Evans, S., Kyong, J. S., Rosen, S., Golestani, N., Warren, J. E., McGettigan, C., et al. (2014). The pathways for intelligible speech: multivariate and univariate perspectives. *Cereb. Cortex* 24, 2350–2361. doi: 10.1093/cercor/bht083

Evans, S., McGettigan, C., Agnew, Z. K., Rosen, S., and Scott, S. K. (2016). Getting the cocktail party started: masking effects in speech perception. *J. Cogn. Neurosci.* 28, 483–500. doi: 10.1162/jocn_a_00913

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

Grill-Spector, K., and Weiner, K. S. (2014). The functional architecture of the ventral temporal cortex and its role in categorization. *Nat. Rev. Neurosci.* 15, 536–548. doi: 10.1038/nrn3747

Harinen, K., Aaltonen, O., Salo, E., Salonen, O., and Rinne, T. (2013). Task-dependent activations of human auditory cortex to prototypical and nonprototypical vowels. *Hum. Brain Mapp.* 34, 1272–1281. doi: 10.1002/hbm.21506

Hill, K. T., and Miller, L. M. (2010). Auditory attentional control and selection during cocktail party listening. *Cereb. Cortex* 20, 583–590. doi: 10.1093/cercor/bhp124

Humphries, C., Binder, J. R., Medler, D. A., and Liebenthal, E. (2007). Time course of semantic processes during sentence comprehension: an fMRI study. *NeuroImage* 36, 924–932. doi: 10.1016/j.neuroimage.2007.03.059

Kyong, J. S., Scott, S. K., Rosen, S., Howe, T. B., Agnew, Z. K., and McGettigan, C. (2014). Exploring the roles of spectral detail and intonation contour in speech intelligibility: an FMRI study. *J. Cogn. Neurosci.* 26, 1748–1763. doi: 10.1162/jocn_a_00583

Lee, H., and Noppeney, U. (2011). Physical and perceptual factors shape the neural mechanisms that integrate audiovisual signals in speech comprehension. *J. Neurosci.* 31, 11338–11350. doi: 10.1523/JNEUROSCI.6510-10.2011

Liakakis, G., Nickel, J., and Seitz, R. J. (2011). Diversity of the inferior frontal gyrus—a meta-analysis of neuroimaging studies. *Behav. Brain Res.* 225, 341–347. doi: 10.1016/j.bbr.2011.06.022

Liebenthal, E., Desai, R. H., Humphries, C., Sabri, M., and Desai, A. (2014). The functional organization of the left STS: a large scale meta-analysis of PET and fMRI studies of healthy adults. *Front. Neurosci.* 8:289. doi: 10.3389/fnins.2014.00289

Martínez, A., Anllo-Vento, L., Sereno, M. I., Frank, L. R., Buxton, R. B., Dubowitz, D. J., et al. (1999). Involvement of striate and extrastriate visual cortical areas in spatial attention. *Nat. Neurosci.* 2, 364–369. doi: 10.1038/7274

McFarquhar, M., McKie, S., Emsley, R., Suckling, J., Elliott, R., and Williams, S. (2016). Multivariate and repeated measures (MRM): a new toolbox for dependent and multimodal group-level neuroimaging data. *NeuroImage* 132, 373–389. doi: 10.1016/j.neuroimage.2016.02.053

McGettigan, C., Eisner, F., Agnew, Z. K., Manly, T., Wisbey, D., and Scott, S. K. (2013). T'ain't what you say, it's the way that you say it – Left insula and inferior frontal cortex work in interaction with superior temporal regions to control the performance of vocal impersonations. *J. Cogn. Neurosci.* 25, 1875–1886. doi: 10.1162/jocn_a_00427

McGettigan, C., Faulkner, A., Altarelli, I., Obleser, J., Baverstock, H., and Scott, S. K. (2012). Speech comprehension aided by multiple modalities: behavioural and neural interactions. *Neuropsychologia* 50, 762–776. doi: 10.1016/j.neuropsychologia.2012.01.010

Mesgarani, N., and Chang, E. F. (2012). Selective cortical representation of attended speaker in multi-talker speech perception. *Nature* 485, 233–236. doi: 10.1038/nature11020

Nath, A. R., and Beauchamp, M. S. (2011). Dynamic changes in superior temporal sulcus connectivity during perception of noisy audiovisual speech. *J. Neurosci.* 31, 1704–1714. doi: 10.1523/JNEUROSCI.4853-10.2011

Obleser, J., and Kotz, S. A. (2010). Expectancy constraints in degraded speech modulate the language comprehension network. *Cereb. Cortex* 20, 633–640. doi: 10.1093/cercor/bhp128

Obleser, J., Wise, R. J., Dresner, M. A., and Scott, S. K. (2007). Functional integration across brain regions improves speech perception under adverse listening conditions. *J. Neurosci.* 27, 2283–2289. doi: 10.1523/JNEUROSCI.4663-06.2007

Okada, K., Rong, F., Venezia, J., Matchin, W., Hsieh, I. H., Saberi, K., et al. (2010). Hierarchical organization of human auditory cortex: evidence from acoustic invariance in the response to intelligible speech. *Cereb. Cortex* 20, 2486–2495. doi: 10.1093/cercor/bhp318

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9, 97–113. doi: 10.1016/0028-3932(71)90067-4

Olson, I. R., McCoy, D., Klobusicky, E., and Ross, L. A. (2013). Social cognition and the anterior temporal lobes: a review and theoretical framework. *Soc. Cogn. Affect. Neurosci.* 8, 123–133. doi: 10.1093/scan/nss119

O'Sullivan, J. A., Power, A. J., Mesgarani, N., Rajaram, S., Foxe, J. J., Shinn-Cunningham, B. G., et al. (2015). Attentional selection in a cocktail party environment can be decoded from single-trial EEG. *Cereb. Cortex* 25, 1697–1706. doi: 10.1093/cercor/bht355

Patterson, K., Nestor, P. J., and Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nat. Rev. Neurosci.* 8, 976–987. doi: 10.1038/nrn2277

Puschmann, S., Steinkamp, S., Gillich, I., Mirkovic, B., Debener, S., and Thiel, C. M. (2017). The right temporoparietal junction supports speech tracking during selective listening: evidence from concurrent EEG-fMRI. *J. Neurosci.* 37, 11505–11516. doi: 10.1523/JNEUROSCI.1007-17.2017

Rimmele, J. M., Zion Golumbic, E., Schroger, E., and Poeppel, D. (2015). The effects of selective attention and speech acoustics on neural speech-tracking in a multi-talker scene. *Cortex* 68, 144–154. doi: 10.1016/j.cortex.2014.12.014

Rinne, T., Koistinen, S., Salonen, O., and Alho, K. (2009). Task-dependent activations of human auditory cortex during pitch discrimination and pitch memory tasks. *J. Neurosci.* 29, 13338–13343. doi: 10.1523/JNEUROSCI.3012-09.2009

Ross, L. A., Saint-Amour, D., Leavitt, V. M., Javitt, D. C., and Foxe, J. J. (2007). Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments. *Cereb. Cortex* 17, 1147–1153. doi: 10.1093/cercor/bhl024

Salmi, J., Rinne, T., Degerman, A., Salonen, O., and Alho, K. (2007). Orienting and maintenance of spatial attention in audition and vision: multimodal and modality-specific brain activations. *Brain Struc. Funct.* 212, 181–194. doi: 10.1007/s00429-007-0152-2

Scott, S. K., Blank, C. C., Rosen, S., and Wise, R. J. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain* 123, 2400–2406. doi: 10.1093/brain/123.12.2400

Scott, S. K., Rosen, S., Wickham, L., and Wise, R. J. (2004). A positron emission tomography study of the neural basis of informational and energetic masking effects in speech perception. *J. Acoust. Soc. Am.* 115, 813–821. doi: 10.1121/1.1639336

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science* 270, 303–304. doi: 10.1126/science.270.5234.303

Sumby, W. H., and Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *J. Acoust. Soc. Am.* 25, 212–215.

Tzourio, N., Massioui, F. E., Crivello, F., Joliot, M., Renault, B., and Mazoyer, B. (1997). Functional anatomy of human auditory attention studied with PET. *NeuroImage* 5, 63–77. doi: 10.1006/nimg.1996.0252

van Wassenhove, V., Grant, K. W., and Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proc. Nat. Acad. Sci. U.S.A.* 102, 41181–41186. doi: 10.1073/pnas.0408949102

Weiner, K. S., and Zilles, K. (2016). The anatomical and functional specialization of the fusiform gyrus. *Neuropsychologia* 83, 48–62. doi: 10.1016/j.neuropsychologia.2015.06.033

Whitfield-Gabrieli, S., and Nieto-Castanon, A. (2012). Conn: a functional connectivity toolbox for correlated and anticorrelated brain networks. *Brain Connect.* 2, 125–141. doi: 10.1089/brain.2012.0073

Wikman, P., and Rinne, T. (2019). Interaction of the effects associated with auditory-motor integration and attention-engaging listening tasks. *Neuropsychologia* 124, 322–336. doi: 10.1016/j.neuropsychologia.2018.11.006

Wikman, P., Sahari, E., Leminen, A., Leminen, M., Laine, M., and Alho, K. (2018). "Selective auditory attention during naturalistic audio-visual dialogues with shuffled lines," in *Neurobiology of Speech and Language, Proceedings of the 2nd International Workshop*, eds O. Shcherbakova and Y. Shtyrov St. Petersburg, 60.

Wikman, P., Ylinen, A. Y. H., Leminen, A., Leminen, M., and Alho, K. (2019). "Task-related effects during natural audio-visual dialogues in fMRI," in *: Proceedings of the 11th Annual Meeting of the Society for the Neurobiology of Language*, Helsinki, 20–22.

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

Willems, R. M., Ozyurek, A., and Hagoort, P. (2009). Differential roles for left inferior frontal and superior temporal cortex in multimodal integration of action and language. *NeuroImage* 47, 1992–2004. doi: 10.1016/j.neuroimage.2009.05.066

Zatorre, R. J., Mondor, T. A., and Evans, A. C. (1999). Auditory attention to space and frequency activates similar cerebral systems. *NeuroImage* 10, 544–554. doi: 10.1006/nimg.1999.0491

Zion Golumbic, E. M., Ding, N., Bickel, S., Lakatos, P., Schevon, C. A., McKhann, G. M., et al. (2013). Mechanisms underlying selective neuronal tracking of attended speech at a "cocktail party". *Neuron* 77, 980–991. doi: 10.1016/j.neuron.2012.12.037

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: 16 June 2020
doi: 10.3389/fnins.2020.00603

Decoding the Attended Speaker From EEG Using Adaptive Evaluation Intervals Captures Fluctuations in Attentional Listening

Manuela Jaeger[1,2]*, Bojana Mirkovic[1,3], Martin G. Bleichner[1,4] and Stefan Debener[1,3,5]

[1] Neuropsychology Lab, Department of Psychology, University of Oldenburg, Oldenburg, Germany, [2] Fraunhofer Institute for Digital Media Technology IDMT, Division Hearing, Speech and Audio Technology, Oldenburg, Germany, [3] Cluster of Excellence Hearing4all, University of Oldenburg, Oldenburg, Germany, [4] Neurophysiology of Everyday Life Lab, Department of Psychology, University of Oldenburg, Oldenburg, Germany, [5] Research Center for Neurosensory Science, University of Oldenburg, Oldenburg, Germany

Listeners differ in their ability to attend to a speech stream in the presence of a competing sound. Differences in speech intelligibility in noise cannot be fully explained by the hearing ability which suggests the involvement of additional cognitive factors. A better understanding of the temporal fluctuations in the ability to pay selective auditory attention to a desired speech stream may help in explaining these variabilities. In order to better understand the temporal dynamics of selective auditory attention, we developed an online auditory attention decoding (AAD) processing pipeline based on speech envelope tracking in the electroencephalogram (EEG). Participants had to attend to one audiobook story while a second one had to be ignored. Online AAD was applied to track the attention toward the target speech signal. Individual temporal attention profiles were computed by combining an established AAD method with an adaptive staircase procedure. The individual decoding performance over time was analyzed and linked to behavioral performance as well as subjective ratings of listening effort, motivation, and fatigue. The grand average attended speaker decoding profile derived in the online experiment indicated performance above chance level. Parameters describing the individual AAD performance in each testing block indicated significant differences in decoding performance over time to be closely related to the behavioral performance in the selective listening task. Further, an exploratory analysis indicated that subjects with poor decoding performance reported higher listening effort and fatigue compared to good performers. Taken together our results show that online EEG based AAD in a complex listening situation is feasible. Adaptive attended speaker decoding profiles over time could be used as an objective measure of behavioral performance and listening effort. The developed online processing pipeline could also serve as a basis for future EEG based near real-time auditory neurofeedback systems.

Keywords: EEG, AAD, speech envelope tracking, online attended speaker decoding, listening effort, selective auditory attention, attentional fluctuations

Edited by:
Jerker Rönnberg,
Linköping University, Sweden

Reviewed by:
Virginia Best,
Boston University, United States
Malte Wöstmann,
University of Lübeck, Germany
Hartmut Meister,
University of Cologne, Germany

***Correspondence:**
Manuela Jaeger
manuela.jaeger@uni-oldenburg.de

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

Received: *06 November 2019*
Accepted: *15 May 2020*
Published: *16 June 2020*

Citation:
Jaeger M, Mirkovic B, Bleichner MG and Debener S (2020) Decoding the Attended Speaker From EEG Using Adaptive Evaluation Intervals Captures Fluctuations in Attentional Listening. Front. Neurosci. 14:603. doi: 10.3389/fnins.2020.00603

INTRODUCTION

The human auditory system enables us to follow a speaker of interest among concurrent other speakers, even in noisy environments (Cherry, 1953). Speech comprehension in a noisy listening situation relies on a listeners' ability to segregate an auditory scene into separate auditory objects, and on the ability to attend to a relevant sound stream while suppressing irrelevant information. Paying attention to a specific sound object facilitates auditory processing and resolves competition between multiple sources (Shinn-Cunningham and Best, 2008; Bizley and Cohen, 2013). Several electroencephalographic (EEG) studies revealed robust modulations of event-related potentials by selective attention, which may act as a sensory gain-control-mechanism enhancing the responses to the attended auditory stimulus and/or downregulating the processing of the to-be-ignored stimulus (Hillyard et al., 1973; Woldorff et al., 1993; Choi et al., 2013; Jaeger et al., 2018).

Hearing impaired and normal hearing listeners differ in their performance when they have to attend to a specific speech stream presented simultaneously with competing sounds (Bronkhorst, 2000; Kidd et al., 2007; Shinn-Cunningham and Best, 2008; Ruggles and Shinn-Cunningham, 2011). These performance differences in speech intelligibility in noise cannot be easily explained by the degree of hearing loss (Peissig and Kollmeier, 1997; Gallun et al., 2013; Glyde et al., 2013) and suggest the involvement of additional cognitive factors. Listening to degraded speech may require the allocation of attentional resources to achieve successful speech comprehension (for a review see: Peelle, 2018). The resources are allocated based on task demands and the allocation is controlled by continuous performance monitoring operations to optimize speech intelligibility (Kuchinsky et al., 2016; Vaden et al., 2016). As a consequence, hearing impaired individuals following a conversation in a complex listening situation may experience higher levels of effort to achieve optimal speech comprehension and may fatigue earlier compared to normal hearing controls (Kramer et al., 2006; Holman et al., 2019; Puschmann et al., 2019).

Given the adaptive nature of attention allocation, it is likely that selective attention does not operate in a stable manner but rather fluctuates over time. This idea is supported by recent research showing that momentary attentional lapses or fluctuations in the level of attention are common (Weissman et al., 2006) and can result in erroneous behavior (Eichele et al., 2008). A time-resolved description of auditory selective attention may provide new insights into auditory processing deficits and may help to explain behavioral variabilities in complex listening situations in hearing impaired as well as normal hearing individuals. Our long-term goal is to provide this information as an auditory neurofeedback signal in near real-time, as this may serve as a basis for future auditory training applications.

Natural speech contains information on different time scales (Poeppel, 2003) and envelope modulations between 4 and 8 Hz seem to be critical for speech intelligibility (Drullman et al., 1994a,b; Ghitza, 2012). It has been found that the speech envelope of single speech streams is represented in ongoing auditory cortex activity (Luo and Poeppel, 2007; Aiken and Picton, 2008; Nourski et al., 2009; Kubanek et al., 2013) and the strength of this representation appears to be correlated with intelligibility (Ahissar et al., 2001; Doelling et al., 2014). A two competing speaker paradigm in which two spatially separated speech streams are presented simultaneously has been established by Broadbent (1952) to examine selective attention effects in a challenging listening situation with ecologically valid stimuli. Selective attention to one of two speech streams results in a stronger cortical phase-locking to the attended compared to the ignored speech envelope (Kerlin et al., 2010; Ding and Simon, 2012; Mesgarani and Chang, 2012; Horton et al., 2013; Zion Golumbic et al., 2013; Kong et al., 2014; Fiedler et al., 2019). Moreover, hearing impaired individuals show a reduced attentional modulation in cortical speech envelope tracking, which may reflect deficits in the inhibition of to be ignored signals (Petersen et al., 2017). Accordingly, monitoring the neural tracking of the to-be-attended and to-be-ignored speech stream may capture individual differences in how selective attention abilities unfold over time.

Identifying the degree and direction of attention in near real-time requires that this information can be extracted from short time intervals. Several studies have shown that attention can be reliably decoded from single-trial EEG data in the two competing speaker paradigm (Horton et al., 2014; Mirkovic et al., 2015, 2016; O'Sullivan et al., 2015; Biesmans et al., 2017; Fiedler et al., 2017; Fuglsang et al., 2017, 2020; Haghighi et al., 2017) using various auditory attention decoding (AAD) methods (for a review see: Alickovic et al., 2019). In these studies, AAD procedures demonstrated above chance-level accuracy for evaluation periods of time ranging from 2 to 60 s. In a neurofeedback application, features should be obtained as quickly as possible. This requires implementation of an online artifact attenuation procedure, as ongoing EEG data are typically contaminated by artifacts. On the other hand, this requires minimizing the evaluation interval. Current AAD procedures do not focus on adapting evaluation intervals online, which would allow the tracking of attentional fluctuations. Most studies ignore the possibility of attentional fluctuations and use a fixed evaluation interval. Yet, it is likely that attentional fluctuations influence the individual AAD accuracy and thereby contribute to performance differences which are not reflected in behavioral performance (Horton et al., 2014; Mirkovic et al., 2015, 2016; O'Sullivan et al., 2015; Puschmann et al., 2019).

Our aim was to develop a simple online AAD processing. Therefore, we implemented a single-trial decoding approach that included a fully automated online EEG artifact attenuation procedure. Individual attended speaker decoding profiles were estimated by combining the previously established AAD method with an adaptive staircase procedure, which modulated the length of the next evaluation interval based on the previous decoding outcome. The staircase served to optimize the trade-off between the duration of an evaluation interval and a participant's individual AAD accuracy. A two competing speaker paradigm was carried out to initially validate the performance of the developed online AAD processing pipeline in a group of normal hearing listeners. By using a well-established paradigm, we expected to examine reliable effects of selective auditory attention in the normal hearing population and the derived AAD decoding performance could be compared to other AAD methods. In a first analysis the adaptive staircase procedure was evaluated offline to demonstrate that decoding performance was better than chance level. Second, the efficiency of the online EEG artifact attenuation procedure was explored by comparing the online AAD performance with AAD performance based on uncorrected EEG data. Third, parameters reflecting the attended speaker decoding performance in each testing block were analyzed and related to behavioral performance, in order to identify a possible link to the selective attention ability over time. Finally, a possible relationship between attended speaker decoding performance and subjective ratings of listening effort, motivation and fatigue was explored based on a group analysis. Listening effort is related to the speech intelligibility (determined by the speech-to-noise ratio) and typically reveals large inter-individual differences (Krueger et al., 2017).

MATERIALS AND METHODS

Participants

Twenty one native German speaking participants between the age of 19 and 30 (mean age = 22.3; *SD* 2.7; 16 female) took part in the study. All reported no present neurological or psychiatric conditions. Audiometric thresholds of 20 dB HL or better in both ears were confirmed by pure tone audiometry at octave frequencies from 250 Hz to 8 kHz. The study was approved by the local ethics committee (University of Oldenburg, Germany) and conforms with the World Medical Association Declaration of Helsinki. All participants signed informed consent prior to the experiment and received monetary reimbursement afterward. One individual had to be excluded from the analysis due to technical problems (data loss) during the experiment, leaving a sample size of $N = 20$ for the EEG analysis.

Task and Stimuli

To investigate if AAD based on envelope tracking is feasible in an online experiment we implemented a paradigm with two competing speakers similar to previously reported studies (Mirkovic et al., 2015; O'Sullivan et al., 2015). Participants were instructed to attend to one of two simultaneously presented speech streams throughout the entire experiment (approximately 60 min). One speech stream was presented from the right and the other from the left side to achieve a natural listening situation in which participants were able to use additional spatial cues to direct selective auditory attention. The to-be-attended speech stream and its side of presentation was not changed during the experimental session but was randomized across participants. The stimulus presentation consisted of six blocks lasting 10 min each and separated by short breaks of approximately 5 min. Before each stimulus presentation block an arrow, presented on a screen, pointed in the direction of the to-be-attended speech stream to remind participants about the attended story and its side of presentation. In the stimulus presentation blocks participants were instructed to keep their eyes open and to focus their gaze on a white fixation cross on a light gray background. During the break, subjects were asked to rate their "subjective listening effort," "subjective motivation level," and "subjective fatigue level." For "subjective listening effort" participants were asked "How much effort does it require for you to follow the speaker?" ("Wie anstrengend ist es für Sie dem Sprecher zu folgen?" in German) using a categorical rating scale with seven labeled categories and six intermediate steps from "no effort" ("mühelos" in German) to "extreme effort" ("extrem anstrengend") according to Krueger et al. (2017). "Subjective motivation level" and "subjective fatigue level" was evaluated by asking "How motivated are you now?" ("Wie motiviert sind Sie jetzt?") and "How tired are you now?" ("Wie müde fühlen Sie sich jetzt?"). Subjective ratings of motivation and fatigue were done on the same categorical scale used for rating listening effort to achieve similar scaling between the items. After rating their subjective listening effort, motivation and fatigue level, participants were asked to fill out a multiple-choice questionnaire containing 10 questions related to the content of each speech stream in the previous block. Participants were instructed to answer as many questions as possible but were discouraged from guessing the answers to any question by choosing to leave a question unanswered if they did not know the answer. Even the questionnaire contained questions related to both speech streams, participants were further encouraged to continue attending only to the indicated speech stream and to ignore the other one.

The two speech streams consisted of fairy tales narrated in German by two professional male speakers. For each speech stream silent gaps longer than 500 ms were reduced to this length. The amplitude of both speech streams was adjusted to achieve equal loudness. A detailed description of the speech material and loudness adjustment is available in Mirkovic et al. (2016). Both speech streams were sampled at a rate of 48 kHz and presented to the participant using Psychophysics Toolbox for Matlab (Brainard, 1997), a HDSP 9632 sound card (RME, Haimhausen, Germany), a ADI 8 DS MK III DA converter (RME, Haimhausen, Germany), PA5 attenuator (Tucker-Davis Technologies, Alachua, United States, a C245BEE amplifier (NAD, Pickering, Canada) and two Sirocco S30 loudspeakers (Cambridge Audio, London, United Kingdom). The loudspeakers were located in front of the participant 45 degree to the right and to the left at ear height. The distance between loudspeaker and ear was 1.1 m. Simultaneous presentation of the two sound streams via loudspeakers resulted in a comfortable sound pressure level of 70 dB SPL, measured at the place of the participants head.

EEG Recordings

During the experiment, participants were seated in a comfortable chair in a sound-attenuated and dimly lit booth. EEG data were collected simultaneously from two different electrode layouts, a high-density EEG cap and two cEEGrids (Debener et al., 2015) placed around each ear of the participant. The cEEGrid data will be presented elsewhere.

The high-density EEG cap consisted of 94 Ag/AgCl electrodes arranged in a customized, infracerebral electrode cap with an equidistant electrode layout (Easycap, Herrsching, Germany). Two additional electrodes were placed below the eyes to record

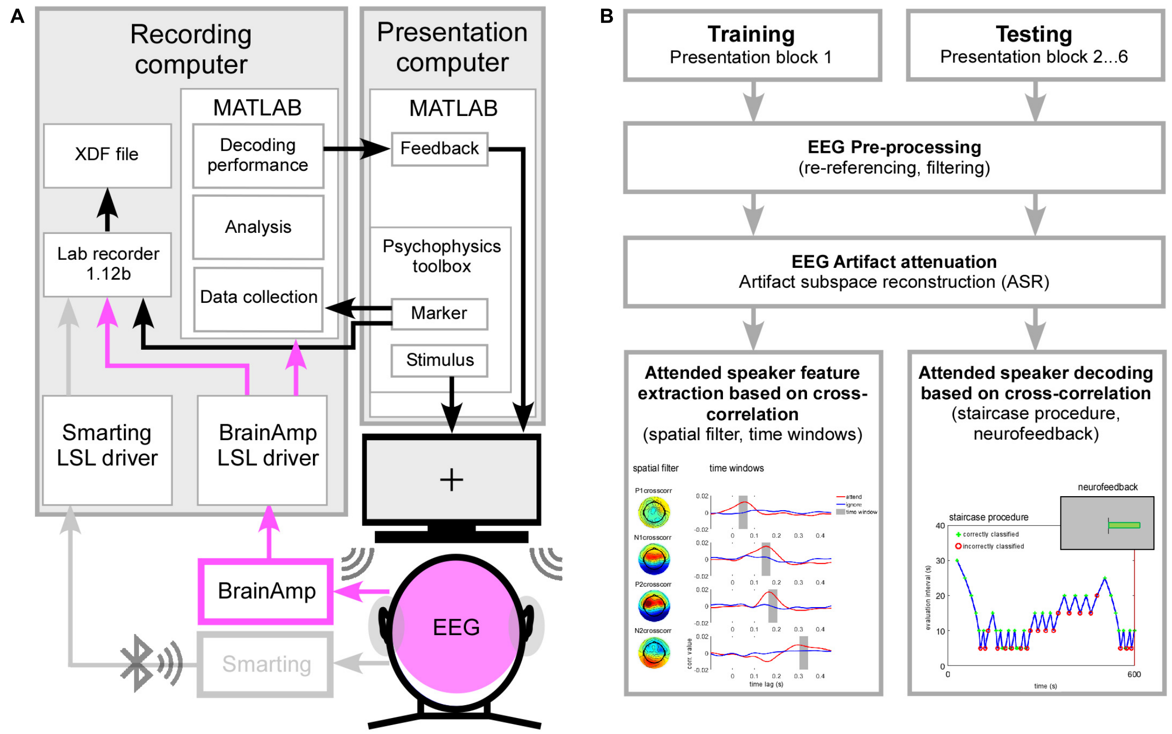


FIGURE 1 | (A) Psychophysics toolbox was used for sound presentation and sending event markers to the Lab recorder and into Matlab on the recording computer. EEG was recorded at 96-channel high density cap (pink) and one cEEGrid attached around each ear (gray, data not shown here). High density EEG was recorded using the BrainAmp amplifiers physically connected to the recording computer. EEG data and event markers were integrated into one XDF file (extensible data format) by using Lab recorder software. High density EEG data and event marker information were collected and analyzed in Matlab. The derived attended speaker decoding performance was condensed into a feedback value, transmitted to the presentation computer and visually presented to the participant. **(B)** Schematic illustration of the training and testing procedure. EEG data derived in the training and testing phase underwent similar pre-processing and artifact attenuation. In the training procedure individual features of selective auditory attention were extracted from the cross-correlation function and related to the main positive and negative deflections (P1crosscorr, N1crosscorr, P2crosscorr, and N2crosscorr). For each deflection a spatial filter was determined and applied to the EEG data. The virtual channel time course was again cross-correlated with the attended and ignored speech envelope in order to determine a time window containing the largest attention effect. During the testing procedure spatial filters and corresponding time windows were applied to extract the attention effect based on cross-correlation values. Averaged across the main deflections a positive value indicated a correctly classified trial while a negative value indicated an incorrectly classified trial. Based on the classification outcome in the previous evaluation interval the length of the next evaluation interval was modulated in steps of 5 s to derive individual attended speaker decoding profiles. Condensed neurofeedback was presented as a visual bar at the end of each testing block.

electro-oculograms (EOG). BrainAmp amplifiers (Brainproducts GmbH, Gilching, Germany) recorded all channels against a nose-tip reference with a sampling rate of 500 Hz and band-pass filtered the data from 0.0159 to 250 Hz. Electrode impedances were kept below 20 kΩ.

Experimental Setup

The experimental setup is shown in **Figure 1A** and consisted of two personal computers connected with ethernet cable to a switch and building a small network. A presentation computer was responsible for auditory stimulus presentation, sound onset marker delivery and presentation of visual instructions and feedback information on a screen located in the booth. High density EEG cap and cEEGrids EEG signals as well as sound presentation onset markers were streamed into the network and integrated using the Lab Recorder software from the Lab Streaming Layer (LSL)[1] package running on the recording computer. LSL enables the collection of time series from different recording modalities by handling the networking, time-synchronization and (near) real-time access to the data (Swartz Center for Computational Neuroscience and Kothe, 2015). On the same recording computer, high-density EEG data and sound presentation onset markers were additionally collected in Matlab (Version 7.14, Mathworks Inc., Natick, MA, United States) to perform the online attended speaker decoding. Using Matlab (as described in section "Online AAD Processing Pipeline") the derived attended speaker decoding performance was condensed into a single feedback value and presented to the participant as a horizontal bar on the screen.

Online AAD Processing Pipeline

During the experiment AAD was performed online on the high-density EEG cap data and by using customized Matlab scripts and the EEGLAB toolbox (Version 13.6.5b; Delorme and Makeig, 2004). Since selective auditory attention modulates

---

[1] https://github.com/sccn/labstreaminglayer

the strength of the attended speech representation in the EEG (Kong et al., 2014; O'Sullivan et al., 2015), speech envelope tracking was realized by analyzing the EEG impulse responses to the temporal envelopes of the presented speech streams. EEG impulse responses were estimated by applying a cross-correlation analysis between EEG signals and corresponding speech envelope information.

The temporal envelopes of the clean speech were extracted following Petersen et al. (2017). After computing the absolute values of the Hilbert transform of the two speech streams the transformed signals were low-pass filtered at 15 Hz. The first order derivative was calculated to highlight prominent changes in the speech signal time course related to sound onsets of words and syllables. After half-wave rectification the resulting speech envelopes were resampled to 250 Hz. The speech envelopes of the presented speech streams were extracted offline and stored for the online EEG data analysis.

The online speech envelope tracking procedure consisted of two parts: (1) training the model needed for making attended speaker prediction was performed after the 1st presentation block on EEG data collected in that block, (2) testing the model in subsequent presentation blocks (2–6) and condensed feedback presentation at the end of each block. Note that the online, adaptive processing pipeline for AAD was fully automated and did not require any action from participants or experimenter. A schematic illustration of the training and testing procedure is shown in **Figure 1B**.

Training Procedure

After finishing the data collection of the 1st presentation block, the EEG raw data were pre-processed. This included re-referencing to common average, low pass filtering at 40 Hz (FIR filter, filter order: 100, window type: Hann), downsampling to 250 Hz and high pass filtering at 1 Hz (FIR filter, filter order: 500, window type: Hann) to remove drifts from the data. The pre-processed EEG data were submitted to a processing pipeline performing EEG artifact attenuation and deriving individual parameters for EEG based attended speaker decoding.

For online EEG artifact reduction Artifact Subspace Reconstruction (ASR) as introduced by Mullen et al. (2013) and available as EEGLAB plugin *clean_rawdata* (version 0.32) was used. ASR is based on a sliding-window Principal Component Analysis and attenuates high-variance signal components in the EEG data (for instance, eye blinks, eye movements, and motion artifacts) relative to some artifact-free calibration data reasonably well (Blum et al., 2019). To derive the required artifact-free calibration data, time windows containing abnormally high-power artifacts were automatically removed from the pre-processed EEG data by running the *clean_window* function. The function is included in the *clean_rawdata* plugin and was called based on default parameters except of the *MaxBadChannels* parameter: aiming for a very clean output we used a value of 0.075. EEG channels containing abnormal data or higher amount of line noise were identified based on inter-channel correlations by submitting the pre-processed EEG data to the *clean_channels* function (included in the *clean_rawdata* plugin). As *CorrelationThreshold* parameter a value of 0.95 was chosen meaning that EEG channels with a lower correlation value relative to the other channels were marked as abnormal. The identified bad channels were excluded from the EEG data analysis during the training and testing procedure. The obtained artifact-free calibration data were submitted to the ASR calibration method (function *asr_calibrate*) to derive a state structure containing the statistical properties of the calibration data. This state structure was submitted together with the original pre-processed EEG data to the ASR processing method (function *asr_process*). During the processing step the ASR method detects artifacts based on their deviant statistical properties and linearly reconstructs the EEG data from the retained signal subspace based on the statistical properties of the calibration data. Since ASR is processing the EEG data in chunks of 500 ms, it makes it suitable for automatic EEG artifact attenuation in online applications.

The artifact attenuated EEG data from the first presentation block were used as a training data set to derive individual parameters for the EEG based AAD. After low pass filtering at 15 Hz (FIR filter, filter order: 100, window type: Hann) the 10 min continuous training data were segmented in time periods of 30 s resulting in 20 consecutive trials. EEG impulse responses to the attended speaker stream were calculated for each channel and trial by running a cross-correlation between EEG signals and corresponding speech envelope information on time lags of −200 to 600 ms. The cross-correlation measures the similarity between EEG and speech envelope as a function of temporal displacement of one relative to the other. The derived cross-correlation coefficients range between −1 and +1. Values closer to 0 indicate no similarity, while values closer to ±1 indicate a strong linear relationship between the two signals.

The obtained EEG impulse responses at each channel were averaged across trials and revealed positive and negative deflections which resemble in their peak latencies and topographies components from the auditory evoked potential literature (Picton, 2013). From the averaged EEG impulse response we extracted the scalp distribution of the cross-correlation coefficients corresponding to the maxima and minima of the main deflections denoted P1crosscorr, N1crosscorr, P2crosscorr, and N2crosscorr based on their polarity and predefined time windows (P1crosscorr: 28–68 ms; N1crosscorr: 76–156 ms; P2crosscorr: 156–396 ms; N2crosscorr: 276–456 ms). Each of the extracted scalp distributions were used as a spatial filter in which the corresponding cross-correlation coefficients were interpreted as filter weights. Multiplying the spatial filter weights with the multi-channel EEG time course derived one virtual channel time course for each deflection (P1crosscorr, N1crosscorr, P2crosscorr, and N2crosscorr). Thereby, EEG channels with higher cross-correlation values were given more weight than those with lower values. Furthermore, EEG channels with negative cross-correlation coefficients, indicating a negative linear relationship between EEG and speech envelope time course, were reversed in phase.

Separately for each deflection, the virtual channel time course was segmented into trials of 30 s length and cross-correlated with the speech envelopes of to-be-attended and to-be-ignored speech stream. The trial averaged EEG impulse response to the to-be-ignored speaker stream was subtracted from the trial averaged EEG impulse response to the to-be-attended speaker stream to quantify the effect of selective attention on the neural tracking of speech. In the derived difference EEG impulse response the time point was determined showing the maximum positive deviation, indicating the largest attention effect. To further optimize performance of the attended speaker decoding algorithm against trial to trial variations in the EEG impulse response, an analysis time window of ±20 ms was centered on each time point. Averaging cross-correlation values across the analysis time window should increase robustness of the discriminative algorithm against random outliers.

At the end of the training procedure, the individually extracted spatial filter weights and corresponding analysis time windows at each deflection (P1crosscorr, N1crosscorr, P2crosscorr, and N2crosscorr) were stored for the attended speaker decoding performed by the testing procedure together with bad channels information and the ASR state structure necessary to run the automatic EEG artifact attenuation.

Testing Procedure and Condensed Feedback Presentation

EEG based AAD including automatic EEG artifact attenuation was performed by running a testing procedure at the end of each presentation block (2–6). The following processing steps were done on a single-trial level to evaluate the feasibility of online data processing. EEG raw data were pre-processed identically to the training procedure. Bad channels identified during training procedure were excluded from the data analysis and pre-processed EEG data were submitted together with the ASR state structure to the ASR processing method (function *asr_process*) to run automatic EEG artifact attenuation. The ASR state structure was updated every time it was called to account for gradual changes in the statistical properties of the EEG data over time. After performing artifact attenuation, the pre-processed EEG data were low pass filtered at 15 Hz (FIR filter, filter order: 100, window type: Hann) and submitted to the attended speaker decoding algorithm.

During AAD the extracted spatial filter weights were applied to the pre-processed multi-channel EEG time course to derive one virtual channel for each deflection (P1crosscorr, N1crosscorr, P2crosscorr, and N2crosscorr). These virtual channel time courses underwent the same processing steps as in the training procedure resulting again in EEG impulse response difference values that were then averaged across all four deflections (P1crosscorr, N1crosscorr, P2crosscorr, and N2crosscorr) and used as a decision criterion to quantify the effect of selective attention. A positive difference value indicated that the to-be-attended speech envelope was more strongly represented in the EEG compared to the to-be-ignored speech envelope and the single trial was marked as correctly classified. A negative difference value indicated a stronger representation of the to-be-ignored speech envelope in the EEG and the single trial was marked as incorrectly classified.

We used a 1-up, 1-down staircase procedure to adapt the evaluation interval (trial length) of the single trial analysis to the individual attended speaker decoding performance. Starting with an evaluation interval of 30 s for the first trial in the 2nd presentation block (i.e., 1st testing block) the evaluation interval of the following trials was varied in steps of ±5 s based on the outcome of the attended speaker decoding in the previous trial. If the current trial was classified correctly, the evaluation interval for the next trial was shortened by 5 s. An incorrect classification resulted in an extension of the next trial evaluation interval by 5 s. The lower edge of the staircase procedure was defined as an evaluation interval of 5 s, while the upper edge was not restricted. During the testing phase (2nd–6th presentation block), the staircase procedure was automatically stopped at the end of each presentation block and the value used for the first trial of the subsequent block, in order to derive a continuous attended speaker decoding profile over time.

To test the feasibility of an auditory neurofeedback application, the AAD performance reflected by the blockwise outcome of the staircase procedure was condensed into a single feedback value. The visual feedback was presented as a bar to the participant, after completing the content related questionnaires and subjective ratings of the previous presentation block. Given the infrequent presentation of the feedback value, we did not expect a benefit of the feedback on subsequent block performance and therefore did not analyze the feedback further.

Offline Validation of the Online AAD Processing Pipeline

Grand Average EEG Impulse Response

In order to explore whether attention influenced the neural tracking, corresponding EEG impulse responses were extracted from the EEG data collected during the testing phase (5 testing blocks, 3000 s in total). The EEG data were pre-processed identically to the online procedure and automatic EEG artifact attenuation was applied. Impulse responses to the to-be-attended and to-be-ignored speech were calculated at each EEG channel using 30 s evaluation intervals (100 trials in total) and averaged across trials for each participant. The grand average EEG impulse response to the attended and ignored speech envelope was used to identify the main positive and negative deflections (P1crosscorr, N1crosscorr, P2crosscorr, and N2crosscorr) and corresponding topographies reflecting which electrode sites contributed most to the neural tracking of the speech envelope.

Grand Average Attended Speaker Decoding Profile

To derive a grand average attended speaker decoding profile the individual profiles determined by the 1-up, 1-down online staircase procedure were interpolated over the complete time course of the testing phase (5 testing blocks, 3000 s in total) in steps of 5 s and averaged across participants. A chance level attended speaker decoding profile was calculated in the offline analysis to identify at which time points the grand average profiles significantly differ from chance performance. To derive a chance level for AAD combined with the 1-up, 1-down staircase procedure we used a permutation approach. For this, individual spatial filter weights and corresponding analysis time windows extracted during the training procedure in the online phase were kept identical, while the attended speaker decoding profile was calculated offline by using the to-be-attended and to-be-ignored speech envelope from a randomly assigned part of the speech material. We repeated this procedure 10 times for each participant to derive a valid chance level decoding profile. Attended speaker decoding profiles were tested with a running Wilcoxon signed rank test across participants. The resulting p values were corrected for multiple-comparisons using the False Discovery Rate (FDR) method (Benjamini and Hochberg, 1995).

AAD With Mean Evaluation Intervals and Fixed Trial Lengths

For further validation of the attended speaker decoding performance, we compared the outcome of the 1-up, 1-down staircase procedure reflected in the mean evaluation interval to a traditional classification method with fixed evaluation segments. Mean evaluation intervals in each testing block were transformed to normal distribution by using inverse transformation. The transformed evaluation intervals were averaged across blocks to derive a single evaluation interval for each participant reflecting the mean performance over time. Furthermore, for each participant we calculated the accuracy of correctly classified trials offline while keeping the evaluation segment at a fixed length of 30 s. Fixed trial length decoding accuracy was correlated with the transformed mean evaluation interval by using a Pearson correlation. Across participants we hypothesized that high AAD performance on fixed 30 s intervals would be related to shorter mean evaluation intervals derived from the staircase procedure and expected a negative relationship between the variables.

Influence of Online EEG Artifact Attenuation on AAD Performance

A possible benefit of applying automated online EEG artifact attenuation (ASR) on AAD performance was explored by comparing the attended speaker decoding profiles against decoding profiles derived from the ASR – uncorrected EEG data. The AAD training and testing procedure was performed identically to the online processing. We expected that the implemented online EEG artifact attenuation procedure (ASR) would result in a better decoding performance, which should be reflected in shorter evaluation intervals.

Evaluation of Behavioral Performance and Decoding Performance Parameters

Behavioral Performance Across Testing Blocks

After completing each testing block participants were asked to fill out a multiple-choice questionnaire containing 10 questions related to the content of each speech stream in the previous block. For each participant and testing block a sensitivity index (d’) was calculated considering the z-transformed proportion of correctly answered questions to the attended story (hits) minus the z-transformed proportion of correctly answered questions to the ignored story (false alarms). We hypothesized that differences in selective attention ability over the time course of the experiment are reflected in the sensitivity index (d’). Effects of time on behavioral performance were tested by conducting a 1 × 5 repeated measures ANOVA on the sensitivity index (d’). The factor “time” (5 levels: testing blocks 1, 2, 3, 4, 5) was defined as within-subject factor and the repeated measures ANOVA was conducted by using a general linear model. The significance level was set at $p < 0.05$. Paired-sample t-tests were performed as *post hoc* analyses and the False Discovery Rate method (Benjamini and Hochberg, 1995) was applied to correct for multiple comparisons.

Decoding Performance Across Testing Blocks

Parameters describing the individual attended speaker decoding performance in each testing block were extracted from the attended speaker decoding profile derived in the online experiment. We analyzed the mean evaluation interval and the standard deviation as descriptive parameters for the mean decoding performance and decoding fluctuation over time, respectively. Both parameters were transformed to normal distribution by using inverse transformation. We hypothesized fluctuations in the extracted parameters over the time course of the experiment. Effects of time on attended speaker decoding performance were tested by conducting 1 × 5 repeated measures ANOVAs (5 levels: testing blocks 1, 2, 3, 4, 5). Again, paired-sample t-tests were used to follow up effects and corrections for multiple comparisons were applied where necessary (FDR).

Decoding Performance and Subjective Ratings of Listening Effort

A median split based on attended speaker decoding in the EEG was used to explore a possible relationship between individual attended speaker decoding performance and subjective ratings of listening effort, motivation and fatigue. In challenging listening situations these subjective ratings may indicate differences in selective attention ability across participants even when speech comprehension is still high and does not indicate significant differences in behavioral performance. Since no explicit hypotheses could be tested this analysis was exploratory and may help to tailor future studies. Participants were divided into two groups based on their transformed mean evaluation intervals averaged across all testing blocks. The 10 participants showing the best attended speaker decoding performance in the EEG were assigned to a group of “good performers,” while the remaining 10 participants formed a group of “poor performers.” Since a non-parametric Friedman test did not indicate significant differences in the subjective ratings of listening effort [$\chi^2(4) = 2.84, p = 0.58, n = 20$], motivation [$\chi^2(4) = 3.54, p = 0.47, n = 20$] and fatigue [$\chi^2(4) = 1.99, p = 0.74, n = 20$] over testing blocks these values were averaged across the testing blocks and compared between performance groups by using a non-parametric Mann-Whitney U-test.

RESULTS

Offline Validation of the Online AAD Processing Pipeline

Grand Average EEG Impulse Response

In order to verify that attention manipulation influenced the neural tracking of the attended and ignored speech stream, corresponding EEG impulse responses were extracted from the

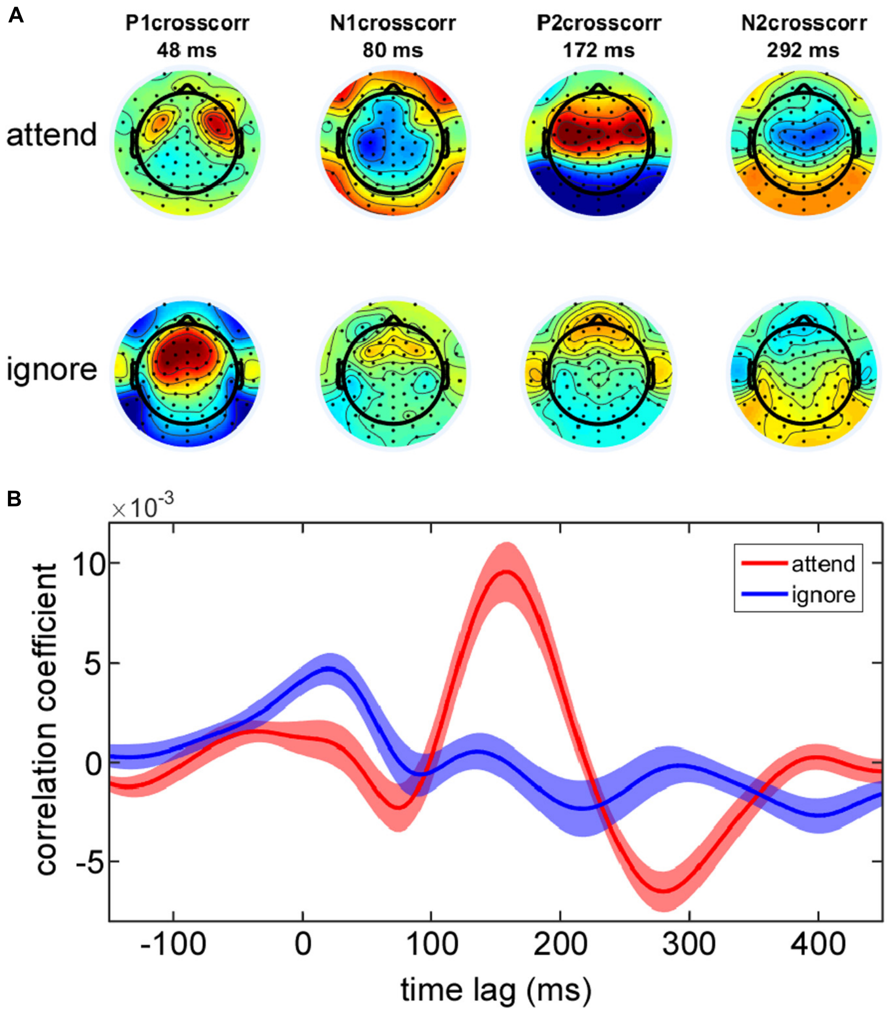


FIGURE 2 | Group average EEG impulse response and corresponding topographies. **(A)** Topographies of the main positive and negative deflections (P1crosscorr, N1crosscorr, P2crosscorr, and N2crosscorr) of the group average EEG impulse response reflect which electrode sites contribute most to the neural tracking of the attended and ignored speech envelope. **(B)** EEG impulse response with ±1 SEM (shaded area) is plotted as a function of time lag separately for the attended and ignored speech envelope at EEG channel Cz.

EEG data collected during the testing phase based on 30 s intervals. In **Figure 2** the grand average EEG impulse response is shown for the attended and ignored speech envelope. Based on cross-correlation we found robust responses to the attended speech envelope with peaks in correlation values at time lag 48, 80, 172, and 292 ms corresponding to the P1crosscorr, N1crosscorr, P2crosscorr and N2crosscorr components from recent speech envelope tracking literature (Horton et al., 2013; Kong et al., 2014; Petersen et al., 2017; Mirkovic et al., 2019). The scalp distributions at the peak latency correlations to the attended speech envelope showed bilateral foci over temporal and frontal electrode sites (**Figure 2A**). Inspection of the EEG impulse response (**Figure 2B**) to the ignored speech envelope suggested that selective attention had a major impact on phase-locking to the ignored speaker stream. While the EEG impulse response to the ignored speech envelope showed a clear positive peak at 48 ms corresponding to the P1crosscorr, all other subsequent deflections were strongly reduced in amplitude, possibly due to a suppression of the to-be-ignored speaker. In accordance with Kong et al. (2014) we found a stronger P1crosscorr amplitude in the EEG impulse response to the ignored speech envelope compared to the attended envelope, which was even reflected in time lags before 0 ms.

Grand Average Attended Speaker Decoding Profile

The group averaged online attended speaker decoding profile, which was derived by an online AAD processing pipeline combined with a staircase procedure is shown in **Figure 3** (red line) over the entire testing phase of 3000 s. Descriptively, during the first testing block a decrease in evaluation interval was visible while from the second testing block on, a modest increase in the group-mean evaluation interval was observable. The estimated chance level decoding performance is represented

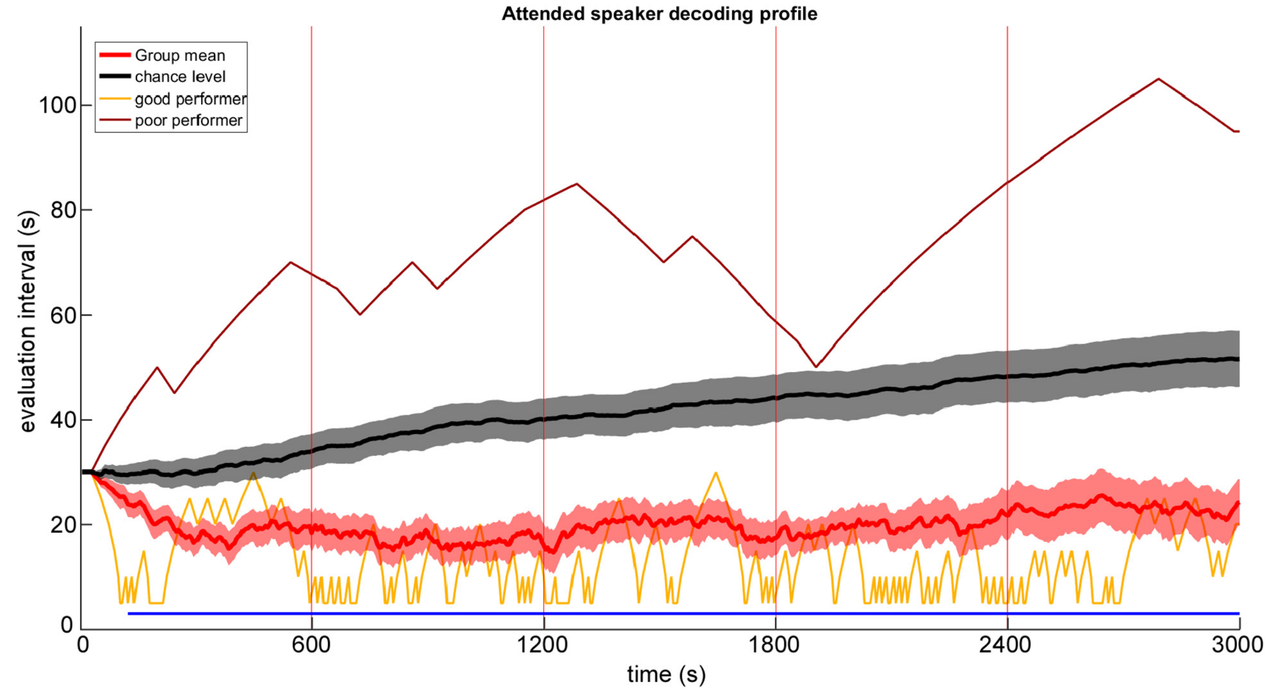


FIGURE 3 | Group mean attended speaker decoding profile (red line) with ±1 SEM (red shaded area) plotted as a function of time. Single subject results are given for the worst and best performing participant. Black line and gray shaded area indicates the chance level with ±1 SEM determined by the permutation approach. The horizontal blue line marks a significant difference between attended speaker decoding profiles and chance level (running Wilcoxon signed rank test, $p < 0.05$, False Discovery Rate corrected). Red vertical lines mark pause intervals between testing blocks.

as a black line, which showed a gradual increase in the length of the evaluation interval across testing blocks over time as well. A running Wilcoxon signed rank test revealed a significant difference between chance level and attended speaker decoding performance from 125 s on and persisted throughout the remaining testing phase. Notably, attended speaker decoding performance differed strongly across participants, as illustrated by showing two individual profiles representing the best (yellow line) and worst (brown line) performance.

AAD With Mean Evaluation Intervals and Fixed Trial Lengths

To prove the validity of the developed AAD processing pipeline further, we compared the outcome of the adaptive staircase procedure reflected in the mean evaluation interval to a traditional classification method with fixed evaluation segments. It was expected that both methods should provide comparable decoding performance on an individual level. Averaged across participants the mean length of the evaluation interval determined by the staircase procedure was 12.2 s (range 8.3–67.6 s). The offline analysis using a fixed trial length of 30 s resulted in a group mean decoding accuracy of 67% (range 44–83%). In **Figure 4**, individual mean evaluation intervals are plotted as a function of individual fixed trial length decoding accuracy. A Pearson correlation revealed a strong negative linear relationship between the variables [$r(18) = -0.93$, $p < 0.001$]. In other words, participants with high fixed trial length decoding accuracy reached smaller mean evaluation intervals determined by the staircase procedure.

Influence of Online EEG Artifact Attenuation on AAD Performance

A possible benefit of applying automated online EEG artifact attenuation (ASR) on AAD performance was explored by comparing the attended speaker decoding profiles against decoding profiles derived from the ASR – uncorrected EEG data. In **Figure 5** the grand average attended speaker decoding profile is shown for corrected (with ASR – red) and uncorrected (without ASR – black) EEG data. As expected, we found that the implemented online EEG artifact attenuation procedure resulted in a better decoding performance. On a descriptive level this performance benefit is reflected in 5–10 s shorter evaluation intervals derived by the adaptive staircase procedure. Especially, in the last testing block the EEG artifact attenuation seemed to outperform the uncorrected processing. A possible explanation may be that participants fatigued earlier toward the end of the experimental duration which is often accompanied by an increase in EEG artifact (i.e., eye blinks and movements) occurrence.

Evaluation of Behavioral Performance and Decoding Performance Parameters

Behavioral Performance Across Testing Blocks

On the behavioral level the analysis of the questionnaire data revealed that all participants followed the instructions

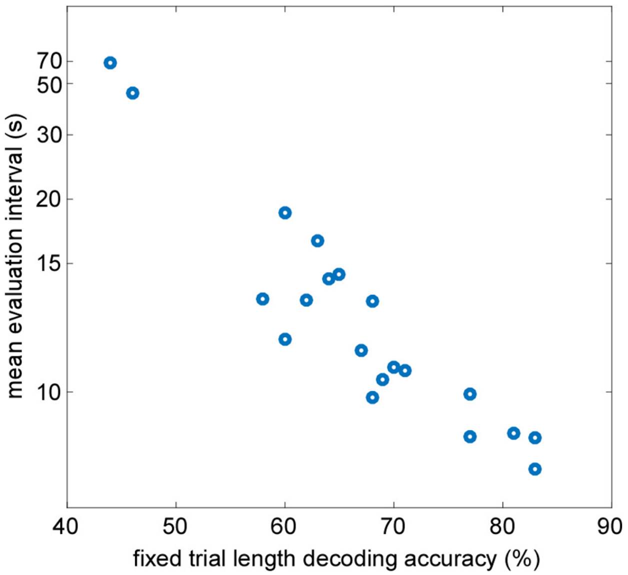


FIGURE 4 | Correlation between individual mean evaluation intervals determined by the staircase procedure and attended speaker decoding accuracy calculated at fixed evaluation intervals of 30 s (r_Pearson = -0.93, $p < 0.001$).

by attending to the indicated speech stream. On average, participants correctly answered 81.25% of the questions on content presented in the to-be-attended speech stream and only 1% of questions on the to-be-ignored speech stream. For each participant and testing block a sensitivity index (d') based on the content related questionnaire was calculated to identify differences in behavioral performance across testing blocks (**Figure 6A**). On a descriptive level, the mean behavioral performance increased from testing block 1–2, as reflected in a steep increase in the sensitivity index (d'), and gradually decreased over time from testing block 2–5. Effects of time on behavioral performance were tested by conducting a 1 × 5 repeated measures ANOVA on the sensitivity index (d') across testing blocks. The 1 × 5 repeated measures ANOVA revealed a significant main effect of time [$F(4, 19) = 7.06$, $p< 0.0001$]. *Post hoc* comparisons using two-tailed paired-sample t-tests identified a significant difference in the sensitivity index (d') between testing blocks 1 and 2 [$t(19) = -5.49$, $p < 0.001$], testing blocks 1 and 3 [$t(19) = -4.59$, $p < 0.01$] as well as between testing blocks 1 and 4 [$t(19) = -2.72$, $p < 0.05$] while a significant decrease in behavioral performance from testing blocks 2–5 was evident [$t(19) = 3.96$, $p < 0.01$].

Decoding Performance Across Testing Blocks

Based on the online attended speaker decoding profiles we extracted the mean evaluation interval and the standard deviation evaluation interval in each testing block as descriptive parameters reflecting the individual decoding performance and its variation over time. We hypothesized those differences in selective attention ability over the time course of the experiment to be reflected in the extracted descriptive parameters. Separate repeated measures ANOVA's were conducted on the mean evaluation intervals (**Figure 6B**) and its standard deviation (**Figure 6C**). Descriptively, both parameters showed a similar behavior over testing blocks. The best attended speaker decoding performance was achieved in testing block 2 reflected in the smallest mean evaluation interval and standard deviation. From testing block 2–5 a gradual increase in mean evaluation interval and standard deviation was apparent. The 1 × 5 repeated measures ANOVA on transformed mean evaluation intervals with the factor 'time' revealed a significant main effect of time [$F(4, 19) = 4.62$, $p< 0.01$]. *Post hoc* comparisons using two-tailed paired-sample t tests identified a significant decrease in mean evaluation interval from testing blocks 1–2 [$t(19) = 3.37$, $p < 0.05$] and a significant increase in mean evaluation interval from testing blocks 2–5 [$t(19) = -3.74$, $p < 0.05$]. A similar behavior in the time course was visible for the standard deviation of the evaluation intervals across testing blocks. Here, the 1 × 5 repeated measures ANOVA revealed a significant main effect of time [$F(4, 19) = 6.39$, $p< 0.001$] too. *Post hoc* comparisons using two-tailed paired-sample t-tests identified a significant decrease in the standard deviation from testing blocks 1–2 [$t(19) = 4.42$, $p < 0.01$] as well as from testing blocks 1–3 [$t(19) = 3.86$, $p < 0.01$] while a significant increase in standard deviation from testing blocks 2–5 was evident [$t(19) = -3.14$, $p < 0.05$].

Attended Speaker Decoding Performance and Subjective Ratings of Listening Effort

A possible relationship between individual attended speaker decoding performance and subjective ratings of listening effort, motivation and fatigue was explored by using a group median split based on the mean evaluation intervals. **Figure 7** shows the good and poor performers separately for the group averaged attended speaker decoding profile (**Figure 7A**), behavioral performance reflected in the sensitivity index (**Figure 7B**) and subjective ratings of listening effort, motivation and fatigue (**Figure 7C**). On a descriptive level, the groups differed strongly in the group averaged attended speaker decoding profile over time. While good performers showed more stable performance in their mean evaluation interval over testing blocks the poor performers' averaged profile indicated more pronounced fluctuations over time. Note that the group differences in attended speaker decoding performance were not reflected by differences in behavioral performance ($Z = 0.95$, $p = 0.34$). Interestingly, good and poor performers differed significantly in their subjective ratings of listening effort and fatigue. The group of good performers reported on average less listening effort ($Z = -2.04$, $p < 0.05$) and less fatigue ($Z = -2.2$, $p < 0.05$) compared to the group of poor performers. Subjective ratings of motivation were descriptively higher in the group of good performers but failed to reach significance ($Z = 1.71$, $p = 0.09$).

DISCUSSION

In this study, we developed an online processing pipeline performing AAD on short segments of EEG data to detect the

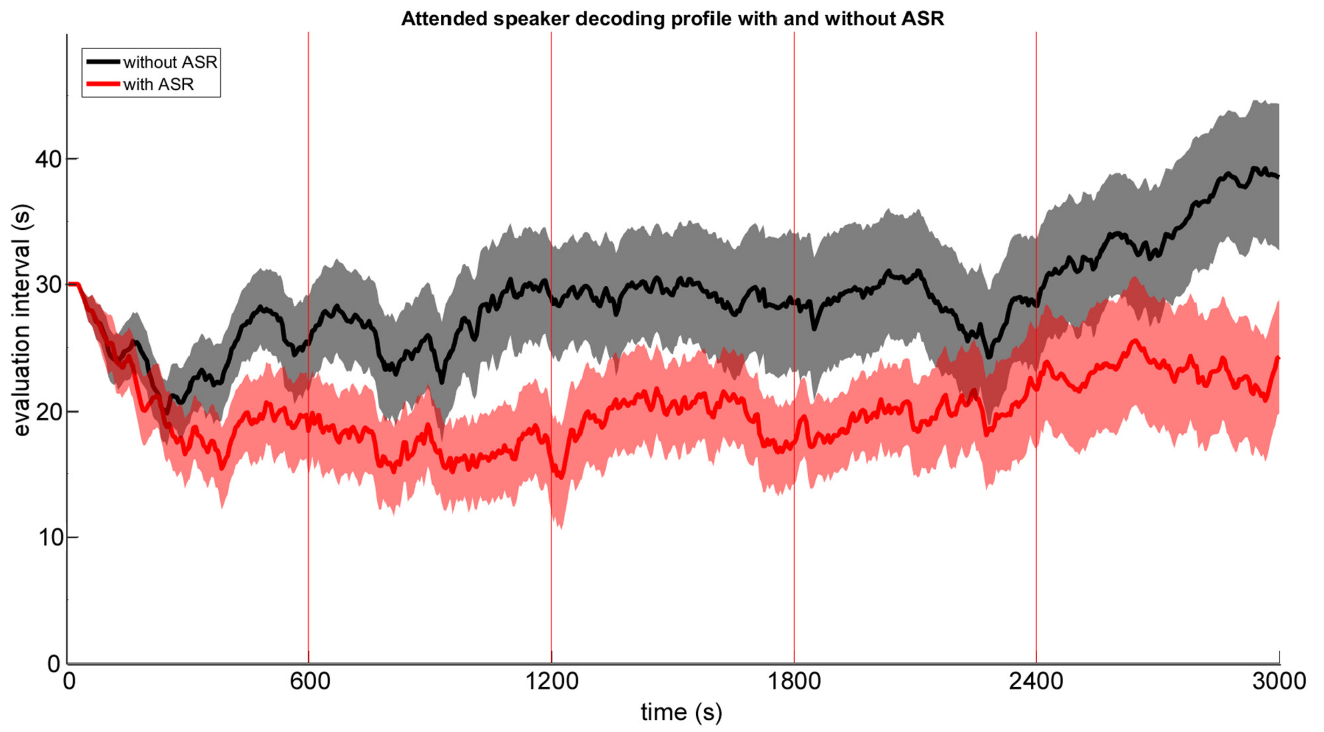


FIGURE 5 | Group mean attended speaker decoding profile with ±1 SEM (shaded area) plotted as a function of time derived from ASR artifact attenuated EEG data (red) and ASR uncorrected EEG data (black).

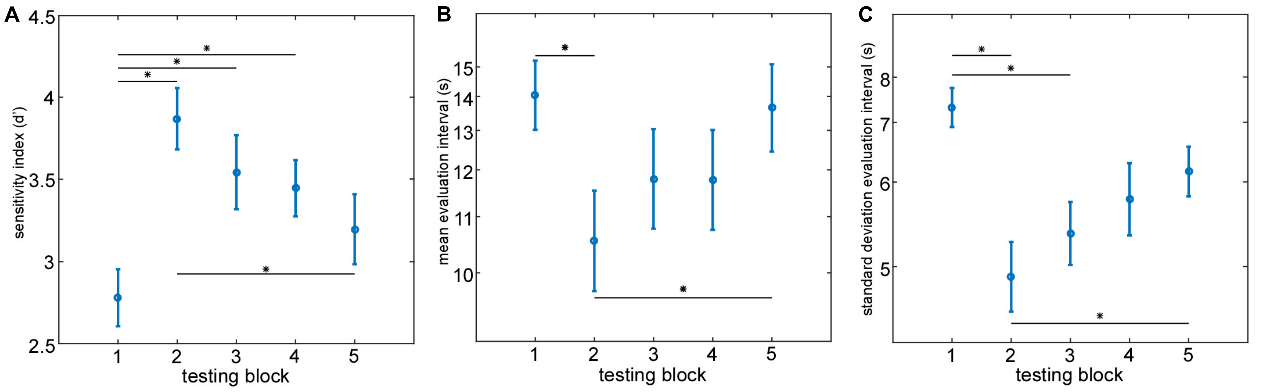


FIGURE 6 | Evaluation of individual behavioral performance and attended speaker decoding performance parameters. Group mean behavioral sensitivity index (d') **(A)**, group mean evaluation intervals **(B)** and group standard deviation evaluation interval **(C)** are plotted as a function of time (testing block). Asterisks mark a significant difference between testing blocks (paired-sample *t*-test, $p < 0.05$, False Discovery Rate corrected). Error bars represent ±1 SEM.

direction and level of attention in a two competing speaker paradigm. The implemented AAD method was combined with an adaptive 1-up, 1-down staircase procedure in order to optimize the trade-off between the duration of evaluation interval and the individual decoding performance. The developed AAD processing pipeline was applied in an online experiment to capture individual attended speaker decoding profiles over time. We hypothesized that exploring these profiles may provide new insights into fluctuations of selective attention and its relation to behavioral performance.

The offline analysis confirmed that the implemented AAD method, which was based on EEG impulse responses to the speech envelope, captured selective auditory attention effects, as reported previously (Horton et al., 2013; Kong et al., 2014; Petersen et al., 2017; Mirkovic et al., 2019). Offline validation of the adaptive procedure revealed a robust relationship between the mean evaluation intervals derived by the staircase method and attended speaker decoding accuracy determined by a classical fixed trial length decoding approach. The implemented online EEG artifact attenuation procedure (ASR) had a beneficial effect

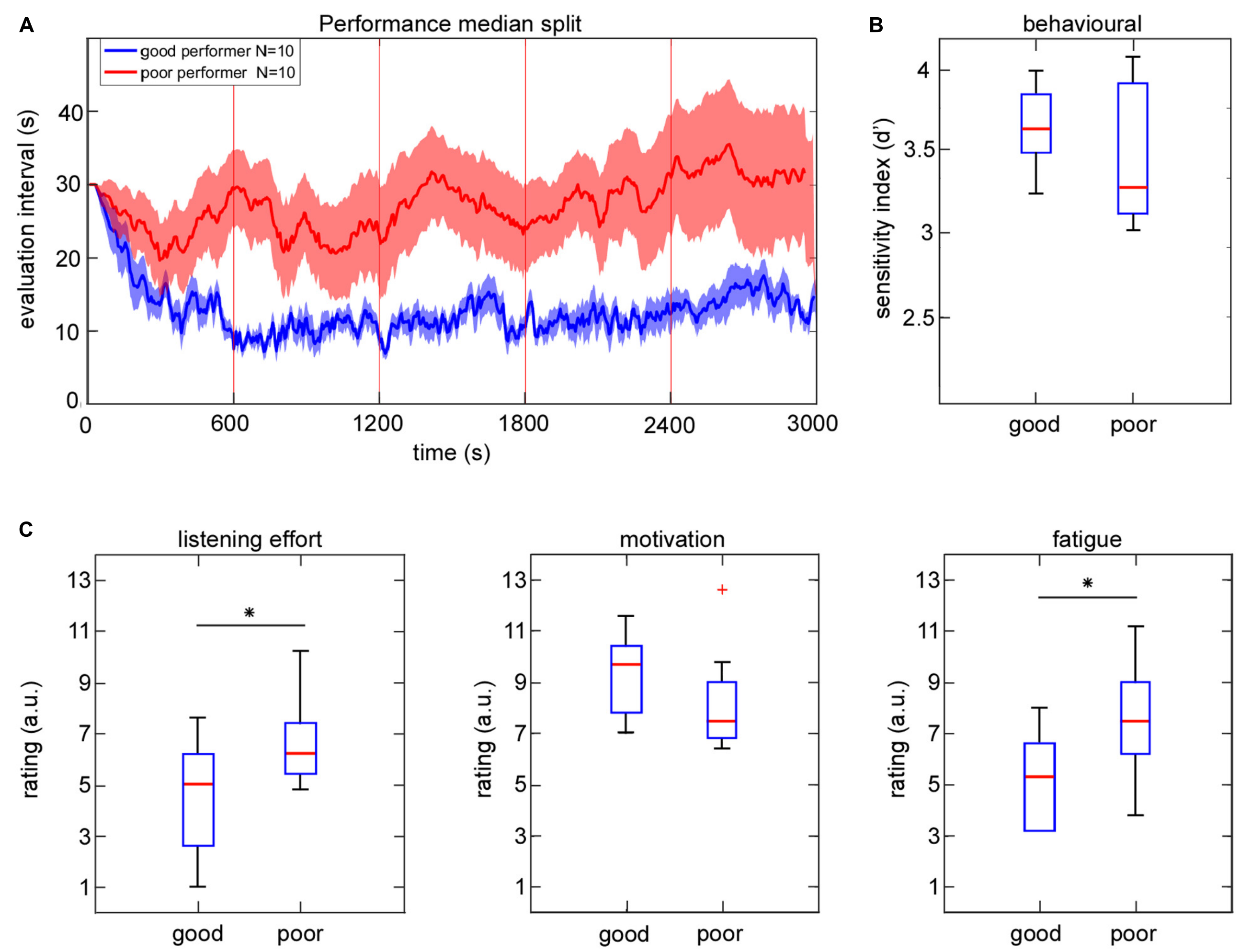


FIGURE 7 | Group median split based on attended speaker decoding performance averaged across testing blocks. **(A)** Group means in performance are plotted as a function of time. The shaded areas represent ±1 SEM. **(B)** Box plot contains the behavioral sensitivity index for good and poor performer. **(C)** Box plots contain subjective ratings of listening effort, motivation and fatigue averaged across all testing blocks for the good and poor performer. Asterisks mark a significant difference in mean subjective ratings between groups (Mann–Whitney *U*-test, $p < 0.05$).

on the attended speaker decoding performance resulting in 5–10 s shorter evaluation intervals compared to the artifact uncorrected EEG data, suggesting that the online processing pipeline was functioning reasonably well. This interpretation is also supported by our analysis of the attended speaker decoding profiles over time. For each participant and testing block we extracted the mean evaluation interval and its standard deviation from the individual attended speaker decoding profiles. On the group level the mean evaluation interval and the standard deviation across testing blocks was closely related to the behavioral performance (d'). Here, shorter mean evaluation intervals and fewer fluctuations in the profile were related to better behavioral performance. Additionally, an exploratory analysis between groups indicated that individuals with poorer attended speaker decoding performance experienced higher listening effort and fatigue over the time course of the experiment. In the following, the benefits and limitations of these procedures will be discussed.

Adaptive staircase procedures are frequently used to determine performance levels (Levitt and Rabiner, 1967). An adaptive 1-up, 1-down procedure is well suited to target a performance level of 50% correct responses, i.e., reveals a detection threshold. This procedure requires careful consideration, since 50% decoding accuracy in a two-class classification problem reflects chance-performance. To exclude the possibility that the captured attended speaker decoding reflected random decision profiles, we estimated the chance-level based on permutation tests (Ojala and Garriga, 2010; Combrisson and Jerbi, 2015). These analyses revealed an increase in the mean chance-level evaluation interval over time. This increase can be explained by the definition of the staircase procedure which was restricted to a minimum length of 5 s for the evaluation interval, but the maximum evaluation interval was not explicitly limited. Hence, for individuals performing at chance-level the evaluation interval would at best fluctuate around the initial evaluation interval of 30 s, or even increase over time, as observed in the permutation tests. Most individuals, however, improved at least initially from the first to the second testing block, suggesting that the attended speaker decoding profiles were not driven by stimulus properties and rather reflected individual profiles of attentive listening.

Looking at the overall decoding performance of the AAD processing pipeline a fixed trial length interval of 30 s revealed in the offline analysis a group average decoding accuracy of 67%. This accuracy level is considerably lower compared to other studies using linear spatio-temporal decoders, which reach decoding accuracies of around 90% (Mirkovic et al., 2015; O'Sullivan et al., 2015; Das et al., 2016). Two reasons may explain this performance difference. First, our processing pipeline followed a strictly chronological approach, that is, we used only the first 10 min of EEG data as a training set to derive individual features of selective attention. More common cross-fold validation strategies may outperform a chronological approach, as they use much more data for classifier training (commonly 90% of all data) and, due to cross-fold sampling, thereby compensate for non-stationarities. However, for online applications these options do not apply, and therefore a chronological processing strategy provides a more realistic result.

A second aspect explaining differences in decoding performance might be the implemented decoding procedure. State-of-the art attended speaker decoding methods focus on optimizing multivariate linear regression models to estimate the speech envelope of the attended speech stream from the EEG data (Mirkovic et al., 2015; O'Sullivan et al., 2015; Das et al., 2016, 2018; Biesmans et al., 2017; Fuglsang et al., 2017), or use deep learning procedures (de Taillez et al., 2017; Ciccarelli et al., 2019). In these studies, the process of model estimation and optimization can be computationally heavy and is therefore not well suited for the near real-time application. Recently, a framework aiming for real-time AAD based on sparse adaptive filtering was proposed by Miran et al. (2018) showing promising results. However, in most online procedures there is a trade-off between algorithm complexity and computation time. As our focus was on a near real-time application, we have opted for a straightforward and low complexity procedure. The chosen cross-correlation approach fitted to these requirements but indicated some temporal smearing of the EEG impulse response, which was visible in the P1crosscorr to the ignored speech envelope. This temporal smearing is caused by the low frequency characteristic of the speech envelope which maps to the EEG signal at overlapping time lags (Crosse et al., 2016). In our study, individual spatial and temporal information of selective attention effects were accounted for by extracting spatial filters and analysis time windows at specific deflections of the EEG impulse response to the speech envelope. The chosen methods condensed features of selective attention across EEG channels and relevant time windows and thereby allowed for a computationally inexpensive attended speaker decoding.

Offline evaluation included a correlation analysis to explore whether the mean online evaluation intervals extracted from the temporal attention profiles are linked to classical decoding accuracy calculated based on fixed trial length intervals of 30 s. A strong correlation revealed that the mean evaluation interval derived by the staircase procedure captured individual decoding performance in the same way as the classical decoding accuracy, while not relying on fixed evaluation intervals. Hence, while the adaptive online procedure does not miss stable individual differences as typically revealed by offline analysis (Choi et al., 2014; Bharadwaj et al., 2015; Puschmann et al., 2019) it optimizes the time interval necessary for attended speaker decoding and thereby reveals attentive listening profiles.

Individual differences in EEG based AAD performance in competing speaker scenarios have been observed before (Horton et al., 2014; Mirkovic et al., 2015, 2016; O'Sullivan et al., 2015). We hypothesized that these decoding performance differences might be related to the individual selective attention ability over time. To reveal more robust interpretations, differences in the attended speaker decoding profiles over time were related to the behavioral performance. For each participant and testing block we extracted the mean evaluation interval and its standard deviation from the attended speaker decoding profile as well as the behavioral sensitivity index (d'). Our statistical analysis revealed that attended speaker decoding performance changed significantly over the time course of the experiment. Best decoding performance was evident in the second testing block followed by a gradual decrease in decoding performance until the fifth testing block. This decrease in decoding performance was characterized by longer mean evaluation intervals necessary for AAD and a higher standard deviation indicating larger fluctuations in the staircase performance. A possible explanation might be that attentional lapses or fluctuations in the level of attention occurred more often toward the end of the experiment as a result of the monotonic and demanding selective listening task. Our behavioral data analysis supports this idea by revealing a similar performance pattern over testing blocks closely following the attended speaker decoding performance. A possible link between decoding performance and behavioral performance is further supported by recent research showing a clear relationship between the speech envelope tracking in the neural data and intelligibility scores to the attended speech signal (Ding and Simon, 2013; Doelling et al., 2014). Taken together, our method seems to be able to evaluate fluctuations in the level of attention, as it is based on an objective measure of attended speaker processing.

As an additional finding, our exploratory analysis indicated that poor attended speaker decoding individuals reported significantly higher listening effort and fatigue when compared to good performing individuals. Note that a group difference in behavioral performance was not evident. This result is in line with research proposing that an increase in cognitive resources (such as working memory) may help to compensate for individual selective attention modulation deficits (Shinn-Cunningham and Best, 2008). Over short periods of time, investment of additional cognitive resources may help to enhance intelligibility. Unfortunately, we did not find significant differences in subjective ratings of listening effort, motivation and fatigue over the time course of the experiment in line with the found behavioral effects. In contrast, the study of Krueger et al. (2017) revealed a clear link between the speech-to-noise ratio determining the speech intelligibility and subjective ratings of listening effort. These differences in outcome may be explained by our study design in which the speech-to-noise ratio was not explicitly modulated but was rather stable over the time course of the experiment. Indeed, good and poor performers differed in their subjective ratings of listening effort and fatigue already from the first testing block on, which was supported by the group difference in their attended speaker decoding profile. We propose that evaluating the selective attention effects between the attended and ignored speech envelope over time in the EEG could potentially serve as an objective measure of listening effort. While theoretical underpinnings and the clinical meaning of listening effort seem poorly developed, an objective measure of listening effort and listening related fatigue would be valuable and complement the variety of methodologies including self-report, behavioral and physiological measures (McGarrigle et al., 2014). Especially, in hearing impaired individuals our procedure may help to capture intra- and inter-individual differences and could be useful in evaluating assistive listening devices.

In summary, our results are consistent with other studies showing that normal-hearing listeners vary widely in their selective attention abilities (Ruggles and Shinn-Cunningham, 2011; Choi et al., 2014). Individual differences in selective auditory attention may be directly related to efficacy of executive cortical processes (Choi et al., 2014) and could be related to differences in subcortical encoding of relevant temporal features necessary for auditory object formation (Ruggles et al., 2011; Bharadwaj et al., 2014). While our analysis does not allow conclusions about the factors contributing to individual differences in temporal auditory attention profiles, we offer an efficient EEG processing pipeline that can help to capture how selective auditory attention fluctuates in complex listening scenarios over time. In a future application, our online AAD processing pipeline could serve as a basis for the development of an auditory neurofeedback training system. Providing information about selective attention fluctuations in near real-time to the participant may help to improve individual listening performance.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Kommission für Forschungsfolgenabschätzung und Ethik, University of Oldenburg, Oldenburg, Germany. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

MJ performed data acquisition. MJ analyzed the data and wrote the manuscript to which BM, MB, and SD contributed with critical revisions. All authors designed the experiment, approved the final version, and agreed to be accountable for this work.

FUNDING

This research was funded by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG Cluster of Excellence 1077 "Hearing4all"), the German Federal Ministry of Education and Research (BMBF, 16SV7784), and the University of Oldenburg Ph.D. program "Signals and Cognition" (Niedersächsisches Ministerium für Wissenschaft und Kultur, Hanover, Germany).

ACKNOWLEDGMENTS

We would like to thank Björn Holtze for help in data collection and Reiner Emkes for technical assistance.

REFERENCES

Ahissar, E., Nagarajan, S., Ahissar, M., Protopapas, A., Mahncke, H., and Merzenich, M. M. (2001). Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proc. Natl. Acad. Sci. U.S.A.* 98, 13367–13372. doi: 10.1073/pnas.201400998

Aiken, S. J., and Picton, T. W. (2008). Human cortical responses to the speech envelope. *Ear Hear.* 29, 139–157.

Alickovic, E., Lunner, T., Gustafsson, F., and Ljung, L. (2019). A tutorial on auditory attention identification methods. *Front. Neurosci.* 13:153. doi: 10.3389/fnins.2019.00153

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Series B Stat. Methodol.* 57, 289–300.

Bharadwaj, H. M., Masud, S., Mehraei, G., Verhulst, S., and Shinn-Cunningham, B. G. (2015). Individual differences reveal correlates of hidden hearing deficits. *J. Neurosci.* 35, 2161–2172. doi: 10.1523/JNEUROSCI.3915-14.2015

Bharadwaj, H. M., Verhulst, S., Shaheen, L., Liberman, M. C., and Shinn-Cunningham, B. G. (2014). Cochlear neuropathy and the coding of supra-threshold sound. *Front. Syst. Neurosci.* 8:26. doi: 10.3389/fnsys.2014.00026

Biesmans, W., Das, N., Francart, T., and Bertrand, A. (2017). Auditory-inspired speech envelope extraction methods for improved EEG-based auditory attention detection in a cocktail party scenario. *IEEE Trans. Neural. Syst. Rehabil. Eng.* 25, 402–412. doi: 10.1109/TNSRE.2016.2571900

Bizley, J. K., and Cohen, Y. E. (2013). The what, where and how of auditory-object perception. *Nat. Rev. Neurosci.* 14, 693–707. doi: 10.1038/nrn3565

Blum, S., Jacobsen, N. S. J., Bleichner, M. G., and Debener, S. (2019). A riemannian modification of artifact subspace reconstruction for EEG artifact handling. *Front. Hum. Neurosci.* 13:141. doi: 10.3389/fnhum.2019.00141

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vis.* 10, 433–436.

Broadbent, D. E. (1952). Listening to one of two synchronous messages. *J. Exp. Psychol.* 44, 51–55. doi: 10.1037/h0056491

Bronkhorst, A. W. (2000). The cocktail party phenomenon: a review of research on speech intelligibility in multiple-talker conditions. *Acustica* 86, 117–128.

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

Choi, I., Rajaram, S., Varghese, L. A., and Shinn-Cunningham, B. G. (2013). Quantifying attentional modulation of auditory-evoked cortical responses from single-trial electroencephalography. *Front. Hum. Neurosci.* 7:115. doi: 10.3389/fnhum.2013.00115

Choi, I., Wang, L., Bharadwaj, H., and Shinn-Cunningham, B. (2014). Individual differences in attentional modulation of cortical responses correlate with selective attention performance. *Hear. Res.* 314, 10–19. doi: 10.1016/j.heares.2014.04.008

Ciccarelli, G., Nolan, M., Perricone, J., Calamia, P. T., Haro, S., O'Sullivan, J., et al. (2019). Comparison of two-talker attention decoding from EEG with nonlinear neural networks and linear methods. *Sci. Rep.* 9:11538. doi: 10.1038/s41598-019-47795-0

Combrisson, E., and Jerbi, K. (2015). Exceeding chance level by chance: the caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy. *J. Neurosci. Methods* 250, 126–136. doi: 10.1016/j.jneumeth.2015.01.010

Crosse, M. J., Di Liberto, G. M., Bednar, A., and Lalor, E. C. (2016). The multivariate temporal response function (mTRF) toolbox: a MATLAB toolbox for relating neural signals to continuous stimuli. *Front. Hum. Neurosci.* 10:604. doi: 10.3389/fnhum.2016.00604

Das, N., Bertrand, A., and Francart, T. (2018). EEG-based auditory attention detection: boundary conditions for background noise and speaker positions. *J. Neural Eng.* 15:066017. doi: 10.1088/1741-2552/aae0a6

Das, N., Biesmans, W., Bertrand, A., and Francart, T. (2016). The effect of head-related filtering and ear-specific decoding bias on auditory attention detection. *J. Neural Eng.* 13:056014. doi: 10.1088/1741-2560/13/5/056014

de Taillez, T., Kollmeier, B., and Meyer, B. T. (2017). Machine learning for decoding listeners' attention from electroencephalography evoked by continuous speech. *Eur. J. Neurosci.* 51, 1234–1241. doi: 10.1111/ejn.13790

Debener, S., Emkes, R., De Vos, M., and Bleichner, M. (2015). Unobtrusive ambulatory EEG using a smartphone and flexible printed electrodes around the ear. *Sci. Rep.* 5:16743. doi: 10.1038/srep16743

Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–12.

Ding, N., and Simon, J. Z. (2012). Emergence of neural encoding of auditory objects while listening to competing speakers. *Proc. Natl. Acad. Sci. U.S.A.* 109, 11854–11859. doi: 10.1073/pnas.1205381109

Ding, N., and Simon, J. Z. (2013). Adaptive temporal encoding leads to a background-insensitive cortical representation of speech. *J. Neurosci.* 33, 5728–5735. doi: 10.1523/JNEUROSCI.5297-12.2013

Doelling, K. B., Arnal, L. H., Ghitza, O., and Poeppel, D. (2014). Acoustic landmarks drive delta-theta oscillations to enable speech comprehension by facilitating perceptual parsing. *Neuroimage* 85(Pt 2), 761–768. doi: 10.1016/j.neuroimage.2013.06.035

Drullman, R., Festen, J. M., and Plomp, R. (1994a). Effect of temporal envelope smearing on speech reception. *J. Acoust. Soc. Am.* 95, 1053–1064.

Drullman, R., Festen, J. M., and Plomp, R. (1994b). Effect of reducing slow temporal modulations on speech reception. *J. Acoust. Soc. Am.* 95, 2670–2680.

Eichele, T., Debener, S., Calhoun, V. D., Specht, K., Engel, A. K., Hugdahl, K., et al. (2008). Prediction of human errors by maladaptive changes in event-related brain networks. *Proc. Natl. Acad. Sci. U.S.A.* 105, 6173–6178. doi: 10.1073/pnas.0708965105

Fiedler, L., Wostmann, M., Graversen, C., Brandmeyer, A., Lunner, T., and Obleser, J. (2017). Single-channel in-ear-EEG detects the focus of auditory attention to concurrent tone streams and mixed speech. *J. Neural Eng.* 14:036020. doi: 10.1088/1741-2552/aa66dd

Fiedler, L., Wostmann, M., Herbst, S. K., and Obleser, J. (2019). Late cortical tracking of ignored speech facilitates neural selectivity in acoustically challenging conditions. *Neuroimage* 186, 33–42. doi: 10.1016/j.neuroimage.2018.10.057

Fuglsang, S. A., Dau, T., and Hjortkjaer, J. (2017). Noise-robust cortical tracking of attended speech in real-world acoustic scenes. *Neuroimage* 156, 435–444. doi: 10.1016/j.neuroimage.2017.04.026

Fuglsang, S. A., Marcher-Rorsted, J., Dau, T., and Hjortkjaer, J. (2020). Effects of sensorineural hearing loss on cortical synchronization to competing speech during selective attention. *J. Neurosci.* 40, 2562–2572. doi: 10.1523/JNEUROSCI.1936-19.2020

Gallun, F. J., Diedesch, A. C., Kampel, S. D., and Jakien, K. M. (2013). Independent impacts of age and hearing loss on spatial release in a complex auditory environment. *Front. Neurosci.* 7:252. doi: 10.3389/fnins.2013.00252

Ghitza, O. (2012). On the role of theta-driven syllabic parsing in decoding speech: intelligibility of speech with a manipulated modulation spectrum. *Front. Psychol.* 3:238. doi: 10.3389/fpsyg.2012.00238

Glyde, H., Cameron, S., Dillon, H., Hickson, L., and Seeto, M. (2013). The Effect of hearing impairment and aging on spatial processing. *Ear Hear.* 34, 15–28. doi: 10.1097/AUD.0b013e3182617f94

Haghighi, M., Moghadamfalahi, M., Akcakaya, M., Shinn-Cunningham, B., and Erdogmus, D. (2017). A graphical model of online auditory scene modulation using EEG evidence for attention. *IEEE Trans. Neural. Syst. Rehabil. Eng.* 25, 1970–1977. doi: 10.1109/TNSRE.2017.2712419

Hillyard, S. A., Hink, R. F., Schwent, V. L., and Picton, T. W. (1973). Electrical signs of selective attention in the human brain. *Science* 182, 177–180.

Holman, J. A., Drummond, A., Hughes, S. E., and Naylor, G. (2019). Hearing impairment and daily-life fatigue: a qualitative study. *Int. J. Audiol.* 58, 408–416. doi: 10.1080/14992027.2019.1597284

Horton, C., D'Zmura, M., and Srinivasan, R. (2013). Suppression of competing speech through entrainment of cortical oscillations. *J. Neurophysiol.* 109, 3082–3093. doi: 10.1152/jn.01026.2012

Horton, C., Srinivasan, R., and D'Zmura, M. (2014). Envelope responses in single-trial EEG indicate attended speaker in a 'cocktail party'. *J. Neural Eng.* 11:046015. doi: 10.1088/1741-2560/11/4/046015

Jaeger, M., Bleichner, M. G., Bauer, A. R., Mirkovic, B., and Debener, S. (2018). Did you listen to the beat? Auditory steady-state responses in the human electroencephalogram at 4 and 7 Hz modulation rates reflect selective attention. *Brain Topogr.* 31, 811–826. doi: 10.1007/s10548-018-0637-8

Kerlin, J. R., Shahin, A. J., and Miller, L. M. (2010). Attentional gain control of ongoing cortical speech representations in a "cocktail party". *J. Neurosci.* 30, 620–628. doi: 10.1523/JNEUROSCI.3631-09.2010

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

Kong, Y. Y., Mullangi, A., and Ding, N. (2014). Differential modulation of auditory responses to attended and unattended speech in different listening conditions. *Hear. Res.* 316, 73–81. doi: 10.1016/j.heares.2014.07.009

Kramer, S. E., Kapteyn, T. S., and Houtgast, T. (2006). Occupational performance: comparing normally-hearing and hearing-impaired employees using the Amsterdam checklist for hearing and work. *Int. J. Audiol.* 45, 503–512. doi: 10.1080/14992020600754583

Krueger, M., Schulte, M., Brand, T., and Holube, I. (2017). Development of an adaptive scaling method for subjective listening effort. *J. Acoust. Soc. Am.* 141:4680. doi: 10.1121/1.4986938

Kubanek, J., Brunner, P., Gunduz, A., Poeppel, D., and Schalk, G. (2013). The tracking of speech envelope in the human cortex. *PLoS One* 8:e53398. doi: 10.1371/journal.pone.0053398

Kuchinsky, S. E., Vaden, K. I. Jr., Ahlstrom, J. B., Cute, S. L., and Humes, L. E. (2016). Task-related vigilance during word recognition in noise for older adults with hearing loss. *Exp. Aging Res.* 42, 50–66. doi: 10.1080/0361073X.2016.1108712

Levitt, H., and Rabiner, L. R. (1967). Use of a sequential strategy in intelligibility testing. *J. Acoust. Soc. Am.* 42, 609–612. doi: 10.1121/1.1910630

Luo, H., and Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron* 54, 1001–1010. doi: 10.1016/j.neuron.2007.06.004

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

Mesgarani, N., and Chang, E. F. (2012). Selective cortical representation of attended speaker in multi-talker speech perception. *Nature* 485, 233–236. doi: 10.1038/nature11020

Miran, S., Akram, S., Sheikhattar, A., Simon, J. Z., Zhang, T., and Babadi, B. (2018). Real-time tracking of selective auditory attention from M/EEG: a bayesian filtering approach. *Front. Neurosci.* 12:262. doi: 10.3389/fnins.2018.00262

Mirkovic, B., Bleichner, M. G., De Vos, M., and Debener, S. (2016). Target speaker detection with concealed EEG around the ear. *Front. Neurosci.* 10:349. doi: 10.3389/fnins.2016.00349

Mirkovic, B., Debener, S., Jaeger, M., and De Vos, M. (2015). Decoding the attended speech stream with multi-channel EEG: implications for online, daily-life applications. *J. Neural Eng.* 12:046007. doi: 10.1088/1741-2560/12/4/046007

Mirkovic, B., Debener, S., Schmidt, J., Jaeger, M., and Neher, T. (2019). Effects of directional sound processing and listener's motivation on EEG responses to continuous noisy speech: Do normal-hearing and aided hearing-impaired listeners differ? *Hear. Res.* 377, 260–270. doi: 10.1016/j.heares.2019.04.005

Mullen, T., Kothe, C., Chi, Y. M., Ojeda, A., Kerth, T., Makeig, S., et al. (2013). Real-time modeling and 3D visualization of source dynamics and connectivity using wearable EEG. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 2013, 2184–2187. doi: 10.1109/EMBC.2013.6609968

Nourski, K. V., Reale, R. A., Oya, H., Kawasaki, H., Kovach, C. K., Chen, H., et al. (2009). Temporal envelope of time-compressed speech represented in the human auditory cortex. *J. Neurosci.* 29, 15564–15574. doi: 10.1523/JNEUROSCI.3065-09.2009

Ojala, M., and Garriga, G. C. (2010). Permutation tests for studying classifier performance. *J. Mach. Learn. Res.* 11, 1833–1863.

O'Sullivan, J. A., Power, A. J., Mesgarani, N., Rajaram, S., Foxe, J. J., Shinn-Cunningham, B. G., et al. (2015). Attentional selection in a cocktail party environment can be decoded from single-trial EEG. *Cereb. Cortex* 25, 1697–1706. doi: 10.1093/cercor/bht355

Peelle, J. E. (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

Peissig, J., and Kollmeier, B. (1997). Directivity of binaural noise reduction in spatial multiple noise-source arrangements for normal and impaired listeners. *J. Acoust. Soc. Am.* 101, 1660–1670.

Petersen, E. B., Wostmann, M., Obleser, J., and Lunner, T. (2017). Neural tracking of attended versus ignored speech is differentially affected by hearing loss. *J. Neurophysiol.* 117, 18–27. doi: 10.1152/jn.00527.2016

Picton, T. W. (2013). Hearing in time: evoked potential studies of temporal processing. *Ear Hear.* 34, 385–401.

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

Puschmann, S., Daeglau, M., Stropahl, M., Mirkovic, B., Rosemann, S., Thiel, C. M., et al. (2019). Hearing-impaired listeners show increased audiovisual benefit when listening to speech in noise. *Neuroimage* 196, 261–268. doi: 10.1016/j.neuroimage.2019.04.017

Ruggles, D., Bharadwaj, H., and Shinn-Cunningham, B. G. (2011). Normal hearing is not enough to guarantee robust encoding of suprathreshold features important in everyday communication. *Proc. Natl. Acad. Sci. U.S.A.* 108, 15516–15521. doi: 10.1073/pnas.1108912108

Ruggles, D., and Shinn-Cunningham, B. (2011). Spatial selective auditory attention in the presence of reverberant energy: individual differences in normal-hearing listeners. *J. Assoc. Res. Otolaryngol.* 12, 395–405. doi: 10.1007/s10162-010-0254-z

Shinn-Cunningham, B. G., and Best, V. (2008). Selective attention in normal and impaired hearing. *Trends Amplif.* 12, 283–299. doi: 10.1177/1084713808325306

Swartz Center for Computational Neuroscience, and Kothe, C. (2015). *Lab Streaming Layer (LSL) [Online]*. San Diego, CA: Swartz Center for Computational Neuroscience.

Vaden, K. I. Jr., Kuchinsky, S. E., Ahlstrom, J. B., Teubner-Rhodes, S. E., and Dubno, J. R. (2016). Cingulo-opercular function during word recognition in noise for older adults with hearing loss. *Exp. Aging Res.* 42, 67–82. doi: 10.1080/0361073X.2016.1108784

Weissman, D. H., Roberts, K. C., Visscher, K. M., and Woldorff, M. G. (2006). The neural bases of momentary lapses in attention. *Nat. Neurosci.* 9, 971–978. doi: 10.1038/nn1727

Woldorff, M. G., Gallen, C. C., Hampson, S. A., Hillyard, S. A., Pantev, C., Sobel, D., et al. (1993). Modulation of early sensory processing in human auditory cortex during auditory selective attention. *Proc. Natl. Acad. Sci. U.S.A.* 90, 8722–8726.

Zion Golumbic, E. M., Ding, N., Bickel, S., Lakatos, P., Schevon, C. A., McKhann, G. M., et al. (2013). Mechanisms underlying selective neuronal tracking of attended speech at a "cocktail party". *Neuron* 77, 980–991. doi: 10.1016/j.neuron.2012.12.037

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: 01 July 2021
doi: 10.3389/fnins.2021.645702

Short-Term Audiovisual Spatial Training Enhances Electrophysiological Correlates of Auditory Selective Spatial Attention

Christina Hanenberg[1], Michael-Christian Schlüter[1], Stephan Getzmann[2] and Jörg Lewald[1]*

[1] Faculty of Psychology, Ruhr University Bochum, Bochum, Germany, [2] Leibniz Research Centre for Working Environment and Human Factors, Dortmund, Germany

OPEN ACCESS

Edited by:
Isabelle Peretz,
Université de Montréal, Canada

Reviewed by:
Claude Alain,
Rotman Research Institute (RRI),
Canada
Benjamin Rich Zendel,
Memorial University of Newfoundland,
Canada

***Correspondence:**
Jörg Lewald
Joerg.Lewald@rub.de

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

Received: 23 December 2020
Accepted: 09 June 2021
Published: 01 July 2021

Citation:
Hanenberg C, Schlüter M-C,
Getzmann S and Lewald J (2021)
Short-Term Audiovisual Spatial
Training Enhances
Electrophysiological Correlates
of Auditory Selective Spatial Attention.
Front. Neurosci. 15:645702.
doi: 10.3389/fnins.2021.645702

Audiovisual cross-modal training has been proposed as a tool to improve human spatial hearing. Here, we investigated training-induced modulations of event-related potential (ERP) components that have been associated with processes of auditory selective spatial attention when a speaker of interest has to be localized in a multiple speaker ("cocktail-party") scenario. Forty-five healthy participants were tested, including younger (19–29 years; $n = 21$) and older (66–76 years; $n = 24$) age groups. Three conditions of short-term training (duration 15 min) were compared, requiring localization of non-speech targets under "cocktail-party" conditions with either (1) synchronous presentation of co-localized auditory-target and visual stimuli (audiovisual-congruency training) or (2) immediate visual feedback on correct or incorrect localization responses (visual-feedback training), or (3) presentation of spatially incongruent auditory-target and visual stimuli presented at random positions with synchronous onset (control condition). Prior to and after training, participants were tested in an auditory spatial attention task (15 min), requiring localization of a predefined spoken word out of three distractor words, which were presented with synchronous stimulus onset from different positions. Peaks of ERP components were analyzed with a specific focus on the N2, which is known to be a correlate of auditory selective spatial attention. N2 amplitudes were significantly larger after audiovisual-congruency training compared with the remaining training conditions for younger, but not older, participants. Also, at the time of the N2, distributed source analysis revealed an enhancement of neural activity induced by audiovisual-congruency training in dorsolateral prefrontal cortex (Brodmann area 9) for the younger group. These findings suggest that cross-modal processes induced by audiovisual-congruency training under "cocktail-party" conditions at a short time scale resulted in an enhancement of correlates of auditory selective spatial attention.

Keywords: cognitive training, auditory selective spatial attention, cocktail-party effect, dorsolateral prefrontal cortex, sound localization

INTRODUCTION

Numerous lines of human and animal research have provided clear evidence that the representation of sound sources in space can be modulated by vision. In particular, in the so-called ventriloquism effect the perceived sound location is shifted toward a spatially disparate, temporally coincident visual event (e.g., Klemm, 1910; Thomas, 1941; Jackson, 1953; Howard and Templeton, 1966; Jack and Thurlow, 1973; Lewald et al., 2001; Lewald and Guski, 2003; for review, see Radeau, 1994). Moreover, exposure to a consistent audiovisual spatial disparity over a certain period of time can induce a systematic shift in sound localization such that the representation of the auditory space is shifted to that of the visual space (Helmholtz, 1867; Stratton, 1896, 1897; Held, 1955; Kalil and Freedman, 1967; Canon, 1970, 1971; Radeau and Bertelson, 1977, 1978; Recanzone, 1998; Lewald, 2002b). These cross-modal adaptive changes, which can emerge over short time scales from seconds to minutes (cf. Bosen et al., 2018), have been termed ventriloquism after-effect. Inspired by animal studies, which demonstrated similar (though more long-term) plasticity of auditory and visual neural representations (Knudsen and Knudsen, 1985, 1989; Brainard and Knudsen, 1993; Knudsen, 1999; Hyde and Knudsen, 2000, 2002; Zheng and Knudsen, 2001), experiments on this phenomenon led to the conception that vision calibrates human auditory spatial perception (Recanzone, 1998; Lewald, 2002b). Results obtained in blind and blindfolded sighted humans as well as in patients with visual-field loss demonstrating specific alterations of sound localization were in accordance with this view (e.g., Zwiers et al., 2001a,b; Lewald, 2002a,c, 2013; Lewald et al., 2009a,b, 2013; Feierabend et al., 2019).

In terms of beneficial effects of multisensory learning (for review, see Shams and Seitz, 2008), approaches of sensory training have been developed, in which auditory stimuli were presented in spatio-temporal alignment with visual stimuli. For example, in patients with pure hemianopia, who suffer from a loss of one half of the visual field due to brain damage while having sufficient audiospatial performance (Lewald et al., 2009b, 2013), neuro-rehabilitative audiovisual training or even auditory unimodal stimulation have been shown to induce long-lasting improvements of visual functions in the anopic hemifield (Bolognini et al., 2005; Passamonti et al., 2009; Lewald et al., 2012). Conversely, cross-modal approaches of sensory training have also demonstrated improvements of auditory functions, suggesting that persons with hearing impairments could benefit from them (Zahorik et al., 2006; Strelnikov et al., 2011; Majdak et al., 2013; Kuk et al., 2014; Grasso et al., 2016; Cai et al., 2018). In the audiospatial domain, audiovisual training has been shown to significantly increase the accuracy of localization of single sound sources under monaural (Strelnikov et al., 2011) and binaural conditions in healthy adults (Cai et al., 2018). In particular, auditory-visual training was found to induce a stronger improvement in sound localization compared to auditory-only training and a significant reduction of front-back confusion for both, trained and untrained sound positions (Cai et al., 2018).

While previous approaches to improve audiospatial performance by sensory training used single sound sources presented in isolation (e.g., Cai et al., 2018), spatial hearing in everyday life requires more complex functions of selective spatial attention, since auditory objects of interest have to be detected and localized among several distractor sound sources. Listening in such a "cocktail-party" situation (Cherry, 1953) is a remarkable ability of the human auditory system, which allows to orient the focus of attention to a sound source of interest in noisy, multi-speaker scenarios (for review, see Bregman, 1990; Bronkhorst, 2015). Such conditions of listening can be challenging already for normal-hearing people, but become substantially more difficult in hearing-impaired persons and at older age, resulting in serious restrictions of communication and social interaction in everyday life (Lewald and Hausmann, 2013; Getzmann et al., 2014, 2015c; see also Pichora-Fuller et al., 2017). This leads to the question of whether hearing performance under these conditions could be improved by training interventions. On the basis of the previous sensory training approaches mentioned above, it seems reasonable to assume that audiovisual stimulation could be an effective tool in this respect. Thus, in the present study a bimodal spatial training was developed in order to enhance brain functions associated with selective spatial attention under multiple-speaker conditions. Two types of sensory training were employed and compared with a control condition: (1) an audiovisual-congruency training, in which auditory targets were presented in spatiotemporal alignment with light stimuli, and (2) a visual-feedback training, in which correct responses on target location were indicated by light flashes. We hypothesized that audiovisual-congruency training may result in more effective learning due to the specific enhancement of multimodal brain circuits by audiovisual spatiotemporal alignment (Shams and Seitz, 2008). Due to findings on auditory short-term or so-called rapid learning (e.g., Alain et al., 2007, 2010), we expected relatively short training periods of about 15 min to enhance earlier evoked responses. For example, Alain et al. (2010) observed rapid improvements in performance within the first few of ten training blocks of trials (<1 h) in a sound identification task of two spectrally identical consonant vowel syllables, differing in voice onset time. The improvements in performance were accompanied by rapid physiological changes in the human auditory system (N1, P2, N2, and LPC).

Here we focused on training-induced modulations of event-related potential (ERP) components that have been associated with processes of auditory selective spatial attention in "cocktail-party" situations, using a multiple-speaker sound-localization task (Lewald et al., 2016; Hanenberg et al., 2019). In particular, we expected a training-induced increase of the N2 component. Localization of predefined auditory target stimuli in multiple-distracter environments has been shown to result in a substantially stronger N2 component of the ERP compared with single-source localization (Lewald and Getzmann, 2015). More generally, the N2 component has been regarded as a neural correlate of processes of cognitive control and orienting attention (Potts, 2004; Folstein and Van Petten, 2008), and has been related to conflict processing and suppression of irrelevant information in the auditory domain (Falkenstein et al., 2002;

Bertoli et al., 2005; Getzmann et al., 2015c; Rueda et al., 2015). Even in patients suffering from anxiety, attention training resulted in larger N2 amplitudes and a better ability of avoiding "threats," which the authors interpret as higher demand of orienting attention away from the "threat" (Dennis and Chen, 2007; Eldar and Bar-Haim, 2010). Regarding age, the N2 is usually reduced in older, compared with younger, adults (e.g., Anderer et al., 1996; Wascher and Getzmann, 2014). This has been interpreted to reflect an age-related decline in inhibitory control over concurrent speech information (Getzmann et al., 2015b,c), in line with the more general inhibitory deficit hypothesis (Hasher and Zacks, 1988; for review, see Gazzaley and D'Esposito, 2007). It has been shown that explicit training of divided attention led to increased N2 amplitudes while also improving older adults' ability to allocate their attention (Zendel et al., 2016). Therefore, as we were interested in potential age-related differences in training effects on ERPs (as also demonstrated for the P300 or Late positive component, LPC; e.g., Yang et al., 2018), groups of younger (19–29 years) and older participants (66–76 years) were compared.

MATERIALS AND METHODS

Participants

Forty-five adult participants took part in this study, assigned to either a younger (n = 21; 12 women; mean age 25.0 years, SE 0.7 years, age range 19–29 years) or older group (n = 24; 12 women; mean age 71.0 years, SE 0.7 years, age range 66–76 years). Data from eight further participants (three younger and five older) were excluded from the analyses since the participants responded in less than 50% of all trials. All participants spoke German fluently and wrote with their right hand. Audiometric thresholds of younger participants were normal (mean across 11 pure-tone frequencies $\leq$25 dB hearing level), while the older group showed mild impairments ($\leq$40 dB hearing level; 0.125–8 kHz; Oscilla USB100, Inmedico, Lystrup, Denmark), within the normal range of age-related hearing loss. This study conformed to the Code of Ethics of the World Medical Association (Declaration of Helsinki), printed in the British Medical Journal (July 18, 1964) and was approved by the Ethical Committee of the Leibniz Research Centre for Working Environment and Human Factors, Dortmund. All participants gave their written informed consent for participation.

Experimental Setup

As in previous studies, the experiments were carried out in a dimly illuminated sound-proof and echo-reduced room (5.0 × 3.3 × 2.4 m^3; Getzmann et al., 2014). Participants were seated on a comfortable chair, positioned with equal distances to the front- and side-walls of the room. At a distance of 1.5 m from the participant's head, a semicircular array with four active broadband loudspeakers (SC5.9; Visaton, Haan, Germany; housing volume 340 cm^3) was arranged in the horizontal plane at −60° and −20° to the left, and at 20° and 60° to the right of the participant's median plane (**Figure 1**). The participant's head position was stabilized by a chin rest. In the median plane of the participant's head, a red light-emitting diode (LED) (diameter 3 mm, luminous intensity 0.025 mcd), mounted at the central position of the semicircular array, served as fixation target during

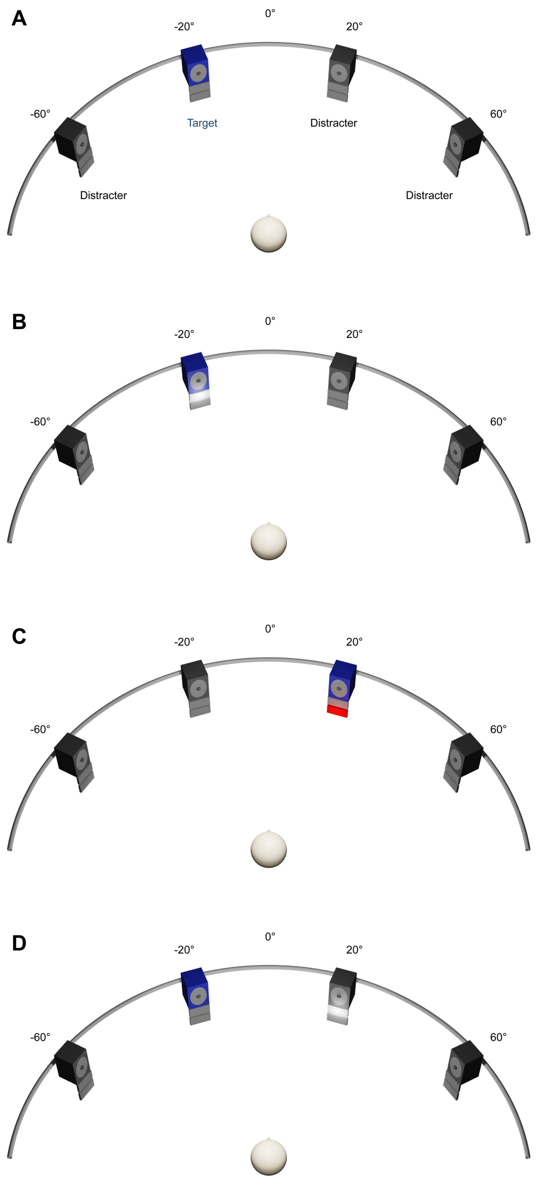


FIGURE 1 | Experimental set-up. **(A)** Auditory spatial attention task. Four numerals were presented simultaneously from four loudspeakers located at –60° and –20° to the left and 20° and 60° to the right. Participants had to localize a predefined target numeral by using a four-button response box. **(B–D)** Training Interventions. Four animal vocalizations were presented simultaneously from four loudspeakers located at –60°, −20°, 20°, and 60°. Participants had to localize a predefined target vocalization by using the same response box as used in spatial attention task. In the audiovisual-congruency training **(B)**, auditory target and visual stimuli were presented in spatiotemporal alignment. In the visual-feedback training **(C)**, visual feedback on correct or incorrect localization was given immediately after each response. In the control condition **(D)**, spatially incongruent auditory target and visual stimuli were presented at random positions with synchronous onset.

testing, while no fixation was required during training. For audiovisual-congruency and visual-feedback training (see section "General Procedure"), an array consisting of a red (4×7 cm^2; 70 cd/m^2) and a white LED screen (4×7 cm^2; 200 cd/m^2) was mounted below each loudspeaker. The two screens were mounted below each other, with the white screen in the upper position, immediately adjacent to the lower edge of the loudspeaker housing (see **Figure 1**).

Auditory Stimuli

The auditory stimuli used for testing were as in previous studies (for further details, see Lewald et al., 2016). Four different German one syllable numerals, spoken by four different speakers (duration about 600 ms) were presented with synchronous stimulus onset from the four loudspeaker positions. The numerals used were "eins" (1), "vier" (4), "acht" (8), and "zehn" (10), each of them spoken by two male and two female native German speakers. Each numeral was presented from each of the four loudspeakers with equal frequency of occurrence at an overall sound pressure level of 66 dB(A), as was measured at the position of the participant's head, using a sound level meter with a 1/2″ free-field measuring microphone (Types 2226 and 4175, Brüel & Kjaer, Naerum, Denmark).

Stimuli used for training have already been used previously (for details, see Lewald, 2016). Sounds consisted of animal vocalizations taken from a sound library (Marcell et al., 2000). Four different vocalizations (denoted "birds chirping;" "dog barking;" "frog;" "sheep" in Marcell et al., 2000) were presented with synchronous stimulus onset from the four loudspeakers. Stimulus durations were adjusted to about 600 ms by cutting the original sound files. Stimuli were digitized at 48 kHz sampling rate and 16-bit resolution and converted to analog form via a PC-controlled soundcard (Terrasoniq TS88 PCI, TerraTec Electronic, Nettetal, Germany). Each animal vocalization was presented from each of the four loudspeakers with equal frequency of occurrence. Stimuli were presented at an overall sound pressure level of 70 dB(A), measured as for the stimuli used for testing (see above).

General Procedure

Following a within-participant repeated-measures crossover design, each participant was tested in three sessions conducted on different days, with intervals of at least 1 week and maximally 3 weeks in between. Each session comprised four blocks, one training block (288–316 trials, duration 15.0–16.5 min), and three identical blocks of testing (prior to, immediately after, and 1 h after training) with an auditory spatial-attention task (see below), each consisting of 288 trials (duration 15 min). The first (pre-training) test block was used as baseline measurement, immediately followed by the training block. Training blocks differed in the type of training conducted at each of the three testing days: a task of localizing a pre-defined target sound (animal vocalization) among three distracters was combined with (1) synchronous presentation of co-localized auditory and visual stimuli (*audiovisual-congruency*), (2) immediate visual feedback on correct or incorrect localization responses (*visual-feedback*), or (3) presentation of spatially incongruent auditory-target and visual stimuli presented at random positions with synchronous onset (*control condition*). The sequence of conditions was counterbalanced across participants in each of the two groups (one-sample Kolmogorov–Smirnov test; $D \leq 0.24$, $p \geq 0.16$).

The second test block (post-training) was started immediately following the training, and the third test block (1-h post-training) began 60 min after the end of the training block. Between pre-training and training-blocks, the instructions given prior to experiments related to the specific training condition (see section "Training Conditions") were briefly repeated. Between post-training and 1-h post-training blocks, participants were allowed to rest, remaining seated in the experimental chair. Prior to the experiment, all participants were informed that they would receive a type of audiovisual training in each session.

Auditory Spatial Attention Task

The auditory spatial attention task used for testing was similar to previous studies (Lewald et al., 2016; Hanenberg et al., 2019). In each trial, participants had to localize a predefined target numeral out of three distracter numerals by pressing one out of four response buttons (**Figure 1A**). The buttons were semi-circularly arranged on a response box, representing the four possible target positions (i.e., far left, mid left, mid right, and far right). Participants were instructed to use the right index finger for responding. For each participant, one numeral (1, 4, 8, or 10) was defined as target, which was kept constant for all measurements. Targets were counterbalanced across participants. During testing, participants had to keep the eyes open and to fixate on the central LED to reduce artifacts due to eye movements and alpha activity in the EEG. Participants were instructed to respond as accurately as possible within about 2 s after stimulus offset. Each trial lasted 3.125 s. Target position, distracter positions, and speakers changed between trials following a pseudo-random order (for details, see Lewald et al., 2016). The timing of the stimuli and the recording of the participants' responses were controlled by custom-written software. Before the beginning of the experiment, participants completed about ten practice trials. They did not receive any feedback on their performance in the auditory spatial attention task during the experiment.

Training Conditions

In all sessions, participants were instructed to indicate target locations using the response box, as described for the auditory spatial attention task (see above). Target sounds used with training were assigned to target numerals during test blocks, that is, when the target numeral was "1" during testing, the target animal vocalization during training was always "birds chirping," "4" was combined with "dog barking," "8" with "frog," and "10" with "sheep." In the *audiovisual-congruency training* condition (**Figure 1B**), the target sound appeared simultaneously with illumination of the white LED screen (duration 600 ms) in 288 of a total of 316 trials (trial duration 3.125 s). The LED screen was mounted immediately below the target loudspeaker. To keep constant the participants' spatial attention and preventing them from relying only on visual cues, we included 28 catch trials, in which target and LED screen appeared at incongruent positions. Accordingly, participants were informed that white

LED screens did not reliably predict the position of the target and were instructed to rely on audition, not vision. Incorrect responses were indicated by flashing of a second, red LED screen (duration 600 ms) at the actual target position immediately after button pressing. The red LED screen was mounted immediately below the white LED screen. Participants were briefed about this procedure prior to training.

In the *visual-feedback training* condition (**Figure 1C**), the sound-localization task and the presentation of *auditory* stimuli were as in the audiovisual-congruency training condition, but without presentation of spatio-temporally congruent visual stimuli. Thus, there was no use of catch trials, thus resulting in a total of 288 trials. For visual feedback, each response was immediately followed by flashing of one of the two LED screens mounted below the loudspeaker that emitted the target sound (duration 600 ms). Correct responses were indicated by white light, incorrect responses by red light at the actual target position. Prior to training, the participant was informed about the visual feedback.

The *control* condition (**Figure 1D**) was similar to the audiovisual-congruency training condition (288 trials). However, the white LED screen appeared at random positions, always diverging from the auditory target (duration 600 ms). Participants were instructed to localize the predefined target sound while ignoring the light flashes. In this condition, no feedback was provided after pressing a button.

Analysis of Behavioral Data

In order to investigate potential effects of the different training paradigms on performance, absolute error was taken as the main measure of localization accuracy, in addition to the percentage of correct responses. The rationale for using this measure and the computation of absolute error has been described previously in detail (Lewald, 2016). In short, the participants' responses were assigned to the azimuth indicated by the position of the response button (−60°; −20°; 20°; 60°), and the unsigned deviation of the response from the actual target azimuth was taken as absolute error. Responses to targets presented at ±60° were excluded from analyses since these data did not provide information on errors to more eccentric positions (Lewald, 2016). Absolute errors were normalized with reference to the pre-training block in the same way as described below for ERP data.

EEG Recording and ERP Analysis

The continuous EEG was sampled at 1 kHz using a QuickAmp-72 amplifier (Brain Products, Gilching, Germany) and 58 Ag/AgCl electrodes, with electrode positions based on the International 10–10 system. Horizontal and vertical electro-oculograms were recorded from four additional electrodes positioned around the left and right eyes. The ground electrode was placed on the center of the forehead, just above the nasion. Two additional electrodes were placed on the left and right mastoids. Electrode impedance was kept below 5 kΩ. The raw data were band pass filtered off-line (cut-off frequencies 0.5 and 25 Hz; slope 48 dB/octave), re-referenced to the average of 58 channels (56 EEG and 2 mastoid electrodes), and segmented into 2000-ms stimulus-locked epochs covering the period from −200 to 1800 ms relative to sound onset. Data were corrected for ocular artifacts using the procedure proposed by Gratton et al. (1983). Individual epochs exceeding a maximum-minimum difference of 200 μV were excluded from further analysis, using the automatic artifact rejection implemented in the BrainVision Analyzer software (Version 2.0; Brain Products, Gilching, Germany). The remaining epochs were baseline corrected to a 200-ms pre-stimulus window and averaged for each participant, separately for each training condition (audiovisual-congruency; visual feedback; control) and each test block (pre-training; post-training; 1-h post-training).

Peaks of five primary ERP components (P1, N1, P2, N2, and LPC) were defined as the maximum positivity or negativity within particular latency windows of specific waveforms after sound onset (P1: 10–110 ms at FCz; N1: 60–160 ms at Cz; P2: 155–255 ms at FCz; N2: 240–340 ms at Cz; LPC: 400–700 ms at Pz). The choice of electrode positions was based on previous knowledge of the ERPs topographical scalp distribution (e.g., Anderer et al., 1996; Folstein and Van Petten, 2008; Martin et al., 2008) and confirmed by visual inspection of the grand average waveforms. As we were interested in effects of training on successful sound localization, only trials with correct responses were included in ERP analyses. To take account of placebo and learning effects, ERP data (peak amplitudes and latencies) were normalized by subtraction of pre-training results (baseline) using the formulae:

$$\text{baseline-normalized post-training value} = \text{post-training value} - \text{pre-training value}$$

and

$$\text{baseline-normalized 1-h post-training value} = \text{1-h post-training value} - \text{pre-training value}.$$

The baseline-normalized data were submitted to 3 × 2 × 2 analyses of variance (ANOVAs) with training condition (audiovisual-congruency; visual feedback; control) and block (post-training; 1-h post-training) as within-participant factors and group (younger; older) as between-participants factor. In addition, to investigate effects of training on ERP amplitudes and latencies, baseline-normalized data were analyzed using one-sample *t*-tests against zero.

Cortical Source Localization

The cortical sources of the effect of training on ERP amplitudes were localized using standardized low-resolution brain electromagnetic tomography (sLORETA) (Pascual-Marqui, 2002), which is part of the LORETA-KEY software package (v20171101) of the KEY Institute for Brain-Mind Research, Zurich, Switzerland[1]. Data were baseline corrected to a 200-ms pre-stimulus window for each participant, separately for each training and each test block. Then, data obtained in the

---

[1] www.uzh.ch/keyinst/loreta

pre-training block were subtracted from data obtained in post-training blocks, and the resulting baseline-normalized data for training conditions were contrasted against the related baseline-normalized data for the control condition (paired groups, test [A-A2] = [B-B2], with A = post/training, A2 = pre/training, B = post/control, B2 = pre/control). We employed sLORETA within 5-ms time windows around the individual ERP peak-amplitude values of P1, N1, P2, N2, and LPC (using all electrodes) for each participant, with the individual latencies taken from the ERP analyses described above.

RESULTS

Behavioral Data

Prior to the main analyses, performance levels were tested against chance-level in order to ensure that the task has been adequately solved. Also, performance in pre-training blocks was compared between younger and older groups to assess effects of age. Both groups showed high levels of percentages of correct responses, differing significantly from chance-level (25%; younger group: $t[20] = 28.55$, $p < 0.001$; older group: $t[23] = 13.09$, $p < 0.001$). For pre-training trials, an ANOVA including the within-participant factor condition (audiovisual-congruency; visual feedback; control) and the between-participants factor group (younger; older) revealed a significant main effect of group ($F[1,43] = 10.30$, $p = 0.003$, $\eta_p{}^2 = 0.19$), indicating a higher percentage of correct responses in younger ($M = 84.3\%$, SE 2.4%), than older ($M = 69.9\%$, SE 3.6%), participants. Also, prior to training, absolute errors were significantly smaller in the younger ($M = 11.7°$, SE 3.1°), than in the older ($M = 26.5°$, SE 2.9°, $F[1,43] = 11.94$, $p = 0.001$, $\eta_p{}^2 = 0.22$), group. There were no significant differences between training conditions within groups in the pre-training blocks, neither in the percentages of correct responses (all $F \leq 0.60$, $p \geq 0.55$), nor in absolute errors (all $F \leq 0.70$, $p \geq 0.50$).

In order to investigate potential effects of the different training paradigms on performance, absolute error was taken as the main measure of localization accuracy. A 3 × 2 × 2 ANOVA on baseline-normalized absolute errors with training condition (audiovisual-congruency; visual feedback; control) and block (post-training; 1-h post-training) as within-participant factors and group (younger; older) as between-participants factor did not indicate main effects or interactions (all $F \leq 3.13$, $p \geq 0.084$). However, across conditions, blocks, and groups, baseline-normalized absolute errors were significantly below zero ($t[44] = -3.98$, $p = 0.0003$), thus indicating general improvement in accuracy that was independent of the type of training condition.

ERPs

In both groups, sound onset elicited a prominent response at vertex position Cz, mainly consisting of a positive deflection (P1), a negative deflection (N1), a second positive deflection (P2), a second negative deflection (N2), and a third positive deflection (LPC; **Figure 2**). The P2 and N2 waves were less prominent in amplitude in older, than younger, participants, which is in line with previous results (e.g., Getzmann et al., 2015b). Mean latencies (with reference to sound onset) were 116 ms (*SE* 3 ms) for N1, 213 ms (*SE* 3 ms) for P2, 296 ms (*SE* 5 ms) for N2 waves, and 570 ms (*SE* 5 ms) for LPC (averaged across groups and conditions). Baseline-normalized data (see section "EEG Recording and ERP Analysis") were submitted to 3 × 2 × 2 ANOVAs, with training condition (audiovisual-congruency; visual feedback; and control) and block (post-training; 1-h post-training) as within-participant factors and group (younger; older) as between-participants factor, to detect potential impacts of the different types of training on amplitude and latency of ERP components, indicating short-term processes of auditory spatial attention.

P1 Wave

For baseline-normalized P1 peak amplitudes at electrode position Cz, the ANOVA indicated a significant interaction of condition and block ($F[1,43] = 3.19$, $p = 0.046$, $\eta_p{}^2 = 0.07$), but no further main effects or interactions (all $F \leq 3.00$, $p \geq 0.99$). However, *t*-tests against zero, conducted separately for training conditions and blocks, did not reveal significant results (all $t[44] \leq 1.33$, $p \geq 0.19$; one-sample *t*-test, two-tailed; Bonferroni-adjusted $\alpha = 0.008$). Also, paired *post hoc t*-tests, comparing baseline-normalized peaks obtained after audiovisual-congruency training or visual-feedback training with the control condition did not reveal any significant differences (all $t[44] \leq -1.61$, $p \geq 0.12$; Bonferroni-adjusted $\alpha = 0.007$). For baseline-normalized P1 peak latencies, there were no significant interactions or main effects (all $F \leq 1.17$, $p \geq 0.29$).

N1 Wave

For baseline-normalized N1 peak amplitudes at electrode position Cz, the ANOVA indicated a main effect of block ($F[1,43] = 9.90$, $p = 0.003$, $\eta_p{}^2 = 0.19$), but no further main effects or interactions (all $F \leq 1.73$, $p \geq 0.18$). Baseline-normalized amplitudes in post-training ($M = 0.49\ \mu V$, $SE = 0.17$; $t[44] = 2.84$, $p = 0.007$), but not in 1-h post-training, blocks ($M = 0.08\ \mu V$, $SE = 0.20$; $t[44] = 0.39$, $p = 0.70$) were significantly in the positive range (one-sample *t*-tests, two-tailed; Bonferroni-adjusted $\alpha = 0.025$), thus indicating reduction in N1 amplitude immediately after training. Baseline-normalized peak amplitudes were more positive in post-training blocks than in 1-h post-training blocks ($t[44] = 3.21$, $p = 0.002$; paired *t*-test, two-tailed).

For baseline-normalized N1 peak latencies, the ANOVA indicated a significant block × group interaction ($F[1,43] = 6.99$, $p = 0.011$, $\eta_p{}^2 = 0.14$), as well as a main effect of group ($F[1,43] = 7.45$, $p = 0.009$, $\eta_p{}^2 = 0.15$). No further main effects or interactions were obtained (all $F \leq 1.10$, $p \geq 0.34$). N1 peak latencies significantly differed between groups in the post-training (younger group: $M = 3.49$ ms, $SE = 1.56$ ms; older group: $M = -5.26$ ms, $SE = 1.97$; $t[43] = -3.42$, $p = 0.001$), but not in the 1-h post-training, block (younger group: $M = 0.00$, $SE = 1.42$; older group: $M = -2.36$ ms, $SE = 1.60$ ms; $t[43] = -1.09$, $p = 0.28$; Bonferroni-adjusted $\alpha = 0.025$). However, *t*-tests against zero for baseline-normalized N1 latencies (conducted separately for blocks and groups) did not reveal significant results (younger

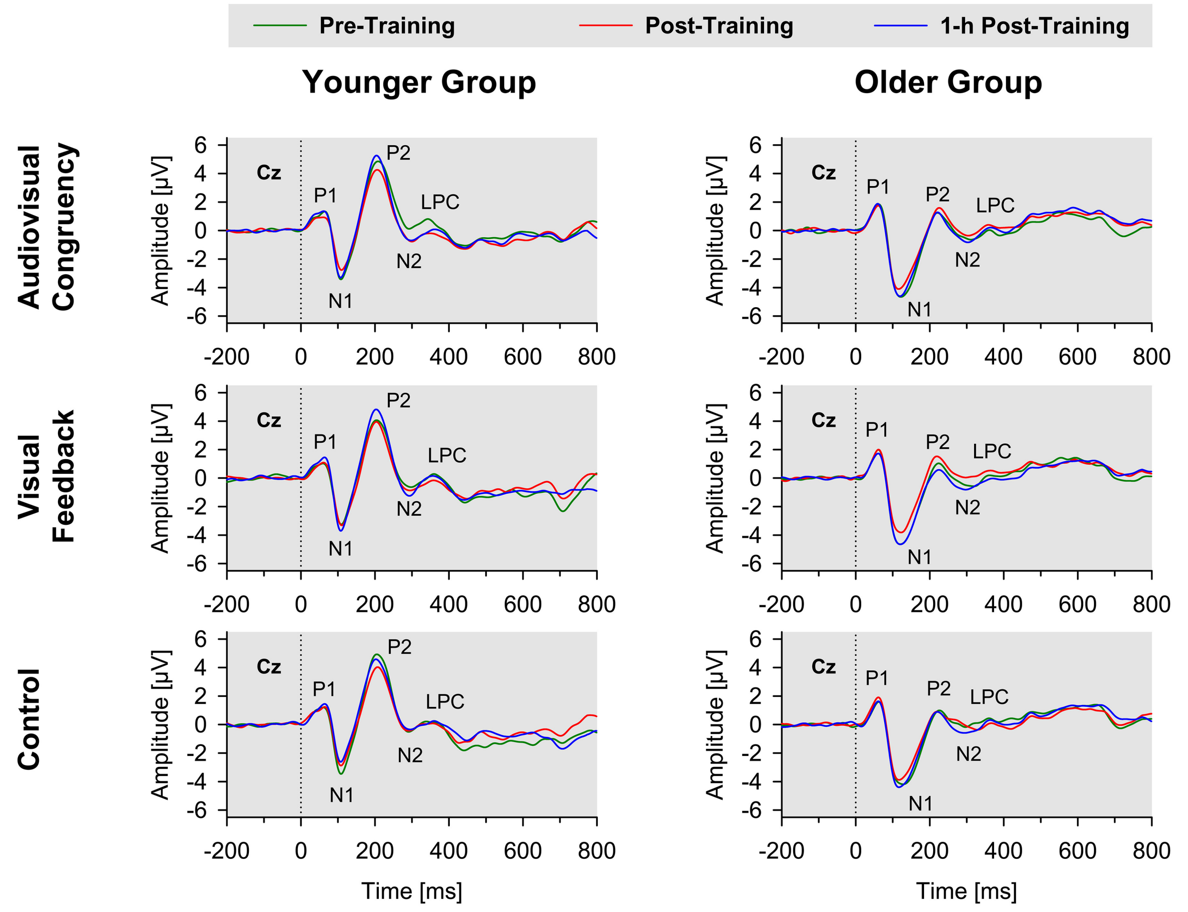


FIGURE 2 | Grand-average ERPs to stimulus onset for younger (left panels) and older participants (right panels) at Cz electrode position. Data obtained before training (pre-training), immediately after training (post-training), and 1 h after training (1-h post-training) are shown separately for audiovisual-congruency training, visual-feedback training, and control condition. P1, N1, P2, and N2 components and LPC are marked in each diagram.

group: all $t[20] \leq 2.24$, $p \geq 0.04$; older group: all $t[23] \leq -2.67$, $p \geq 0.014$; Bonferroni-adjusted $\alpha = 0.0125$).

P2 Wave

For baseline-normalized P2 peak amplitudes at electrode position FCz, there was a significant interaction of block × group ($F[1,43] = 9.24$, $p = 0.004$, $\eta_p^2 = 0.18$), but no further main effects or interactions (all $F \leq 2.54$, $p \geq 0.12$). As revealed by *post hoc* testing, amplitudes in the younger, but not in the older, group were significantly lower in post-training, than in 1-h post-training, blocks (younger group: post-training $M = -0.73$ µV, $SE = 0.41$ µV, 1-h post-training $M = 0.24$ µV, $SE = 0.36$ µV, $t[20] = -3.27$, $p = 0.004$; older group: post-training $M = 0.34$ µV, $SE = 0.29$ µV, 1-h post-training $M = 0.04$ µV, $SE = 0.28$ µV, $t[23] \leq 1.03$, $p > 0.31$; two paired t-tests, one for each group, two-tailed; Bonferroni-adjusted $\alpha = 0.0125$). However, t-tests against zero for baseline-normalized P2 peak amplitudes showed neither significant results when conducted separately for each block (all $t[44] \leq -0.62$, $p \geq 0.54$; one-sample t-test, two-tailed; Bonferroni-adjusted $\alpha = 0.025$), nor when conducted separately for each group (younger group: all $t[20] \leq -1.77$, $p > 0.09$; older group: all $t[23] \leq 1.17$, $p > 0.26$; one-sample t-tests, two-tailed; Bonferroni-adjusted $\alpha = 0.0125$).

For baseline-normalized P2 peak latencies, the ANOVA indicated a significant interaction of training condition × group ($F[1,43] = 5.70$, $p = 0.005$, $\eta_p^2 = 0.12$), as well as a main effect of block ($F[1,43] = 7.44$, $p = 0.009$, $\eta_p^2 = 0.15$), with no further main effects or interactions (all $F \leq 1.24$, $p \geq 0.27$). *Post hoc* testing revealed that baseline-normalized peak latencies were significantly shorter than zero in the 1-h post-training block ($M = -4.38$ ms, $SE = 1.11$ ms; $t[44] = -3.94$, $p < 0.0001$), but not in the post-training block ($M = -0.52$ ms, $SE = 1.44$ ms; $t[44] = -0.36$, $p = 0.72$; one-sample t-test, two-tailed; Bonferroni-adjusted $\alpha = 0.025$). Baseline-normalized latencies, analyzed separately for training conditions and groups, were significantly shorter than zero in the younger, but not in the older, group for the audiovisual-congruency condition in the1-h post-training block, but not for the other conditions and the post-training block (younger group: $M = -10.19$ ms, SE = 3.08 ms, $t[20] = -3.31$, $p = 0.003$; all other $t[20] \leq -2.67$, $p > 0.015$; older group: all $t[23] \leq -3.11$, $p \geq 0.005$; one-sample t-tests, two-tailed; Bonferroni-adjusted $\alpha = 0.004$).

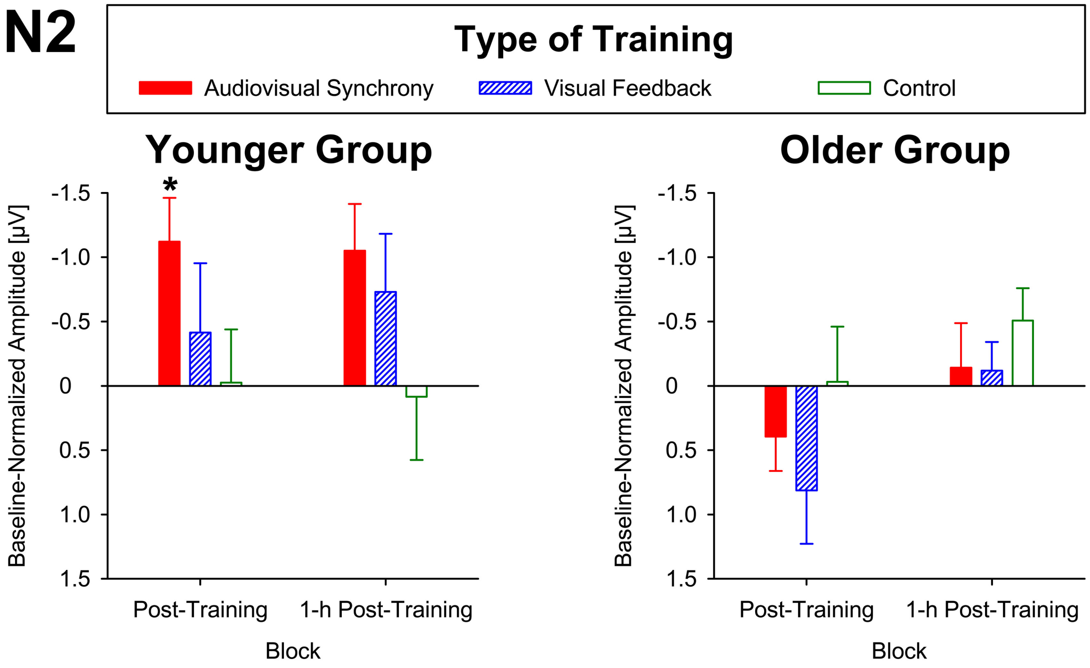


FIGURE 3 | Baseline-normalized N2 peak amplitudes at electrode position Cz in younger and older participants. Post-training and 1-h post-training data are shown separately for audiovisual-congruency training, visual-feedback training, and control condition. A significant negative deviation from zero was found exclusively in the post-training block after audiovisual-congruency training in the younger group, *$p < 0.05$.

N2 Wave

For baseline-normalized N2 peak amplitudes at electrode position Cz, the ANOVA indicated a training condition × group interaction ($F[2,86] = 3.34$, $p = 0.04$, $\eta_p^2 = 0.07$; **Figure 3**), but merely non-significant numerical trends for the factors block ($F[1,43] = 3.70$, $p = 0.06$, $\eta_p^2 = 0.08$) and group ($F[1,43] = 3.79$, $p = 0.06$, $\eta_p^2 = 0.08$). No further main effects or interactions were found (all $F \leq 2.78$, $p \geq 0.1$). Averaged across post-training and 1-h post-training blocks, baseline-normalized peak amplitudes were significantly larger than zero for the younger participants after audiovisual-congruency training ($t[20] = -3.39$, $p = 0.003$; two one-sample t-tests, two-tailed; Bonferroni-adjusted $\alpha = 0.008$), but not for the other conditions (all $|t[20]| \leq -1.24$, $p \geq 0.23$). Neither, there were any significant results for the older group with all conditions (all $t[23] \leq 1.27$, $p \geq 0.22$; one-sample t-test, two-tailed; Bonferroni-corrected $\alpha = 0.025$). In line with that, baseline-normalized N2 peak amplitudes in the audiovisual-congruency condition were significantly larger than in the visual-feedback training and control conditions exclusively for the younger group and for the post-training block, ($M = -1.12$ µV, $SE = 0.34$ µV; $t[20] = -3.29$, $p = 0.004$; all other $|t[20]| \leq -0.77$, $p \geq 0.45$; one-sample t-test, two-tailed; Bonferroni-adjusted $\alpha = 0.008$), indicating a specific increase in N2 amplitude after audiovisual-congruency training. In the 1-h post-training block, baseline-normalized N2 peak amplitudes in the audiovisual-congruency condition were numerically larger than in the visual-feedback training and control conditions, again, exclusively for younger participants ($M = -1.05$ µV, $SE = 0.36$ µV; $t[20] = -2.90$, $p = 0.009$; all other $|t[20]| \leq -1.62$, $p \geq 0.12$; one-sample t-test, two-tailed; Bonferroni-adjusted $\alpha = 0.008$). *Post hoc* testing did not reveal any significant differences in the older group (all $|t[23]| \leq -2.00$, $p \geq 0.06$; one-sample t-test, two-tailed; Bonferroni-adjusted $\alpha = 0.008$). In line with these findings on N2 peak amplitudes, baseline-normalized topographies showed a fronto-central/parietal negativity after audiovisual-congruency training exclusively in the younger group, with its maximum in the left hemisphere (**Figure 4**).

For baseline-normalized peak latencies, the ANOVA revealed a significant main effect of block ($F[1,43] = 7.10$, $p = 0.011$, $\eta_p^2 = 0.14$). No further main effects or interactions were obtained (all $F \leq 2.09$, $p \geq 0.13$). Averaged across training conditions and groups, N2 latencies were significantly longer in post-training ($M = 6.28$ ms, $SE = 3.20$ ms), than 1-h post-training, blocks ($M = -0.06$ ms, $SE = 2.27$ ms; $t[44] = 6.34$, $p = 0.009$). However, there were no significant differences from zero (all $t[44] \leq 1.96$, $p \geq 0.056$; one-sample t-test, two-tailed; Bonferroni-corrected $\alpha = 0.025$).

To estimate potential effects of the sequence of the training conditions on these results, a four-factor ANOVA on the dependent variable baseline-normalized N2 amplitude was performed with the within-participant factors day of testing (first, second, or third) and block, and the between-participant factors sequence (all six permutations of sequences of audiovisual-congruency training, visual-feedback training, and control condition) and group. Neither in the younger, nor in the older group, the sequence of conditions had an impact on the N2 amplitude. The ANOVA indicated a significant interaction of day

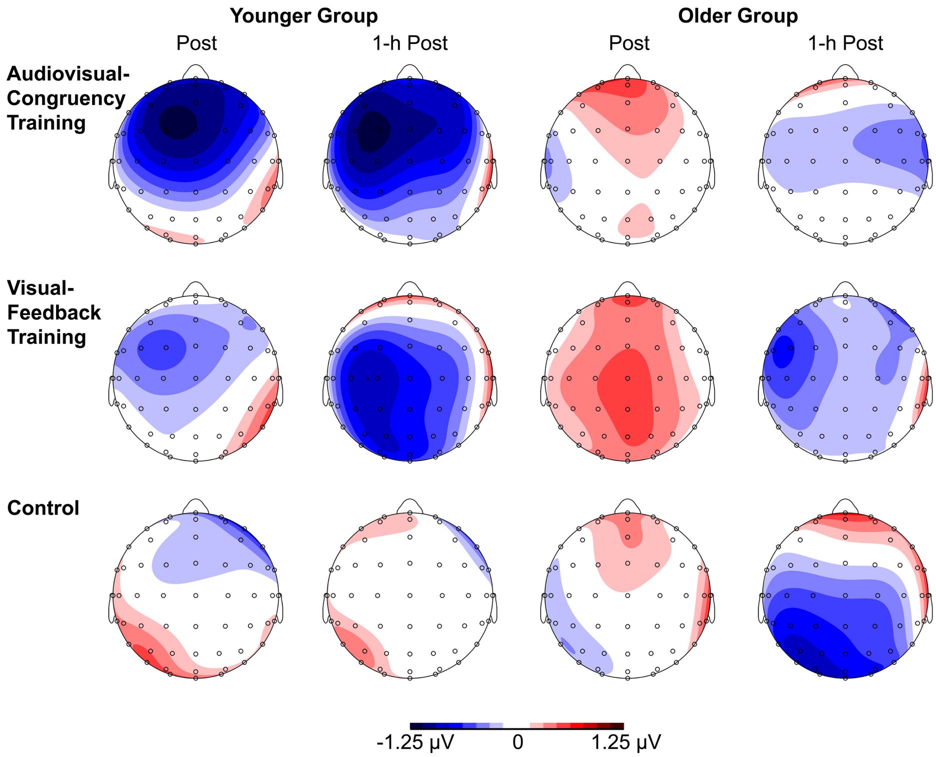


FIGURE 4 | Topographies of N2 components. Baseline-normalized topographies, assessed immediately after (post) and 1 h after training (1-h post) in younger and older participants are shown separately for audiovisual-congruency training, visual-feedback training, and control condition.

of testing and group ($F[2,66] = 4.55$, $p = 0.014$, $\eta_p{}^2 = 0.12$), but no effects involving the factor sequence (all $F \leq 1.48$, $p \geq 0.17$). A *post hoc* inspection of the interaction of day of testing and group using two one-factor ANOVAs separately for each group suggested a non-significant trend of an increase of baseline-normalized N2 amplitude between the first and the last day of testing in the younger group ($F[2,40] = 3.22$, $p = 0.050$, $\eta_p{}^2 = 0.14$), but not in older participants ($F[2,46] = 0.23$, $p = 0.80$, Bonferroni-corrected $\alpha = 0.025$).

LPC

For baseline-normalized LPC peak amplitudes at electrode position Pz, the ANOVA did not reveal any main effects or interactions (all $F \leq 1.71$, $p \geq 0.19$). For baseline-normalized LPC peak latencies no main effects or interactions were obtained (all $F \leq 2.71$, $p \geq 0.11$).

Cortical Source of Electrical Activity at the Time of the N2

The cortical source of the enhancing effect of training on the N2 amplitude found for the younger group in the post-training block of the audiovisual-congruency condition was localized using sLORETA. Data obtained in the pre-block were subtracted from data obtained in the post-block and the resulting difference values were contrasted against the related difference values for the control condition. The analysis revealed an enhancement of electrical activity induced by audiovisual-congruency training at a focal peak location at MNI coordinates $X = 25$ mm, $Y = 50$ mm, $Z = 40$ mm ($t = 6.45$, $p = 0.005$, two-tailed), in right superior frontal gyrus (Brodmann area, BA 9) in the anterior region of superior frontal sulcus (SFS) (**Figure 5**). Results for the five brain structures with the highest *t*-scores revealed by this contrast are summed up in **Table 1**.

Post hoc, related analyses using sLORETA were conducted for all ERP components, conditions, blocks, and groups, in order to control for the specificity of this result obtained at the time of the N2 for the younger group in the post-training block of the audiovisual-congruency condition. As shown in **Table 2**, all results except that described above failed to reach significance level (all $|t| \leq 4.92$, $p \geq 0.06$, two-tailed). Imaging data for all analyses are shown in **Supplementary Figure 1**.

DISCUSSION

These results demonstrated an age-specific effect of audiovisual spatial training on neurophysiological correlates of auditory selective spatial attention in a simulated "cocktail-party" scenario. After audiovisual-congruency training, but not after visual-feedback training and the control condition, the N2 peak amplitude was increased. This effect was observed in younger, but not older, participants with a non-significant decline within 1 h after training. At the time of the N2, electrical

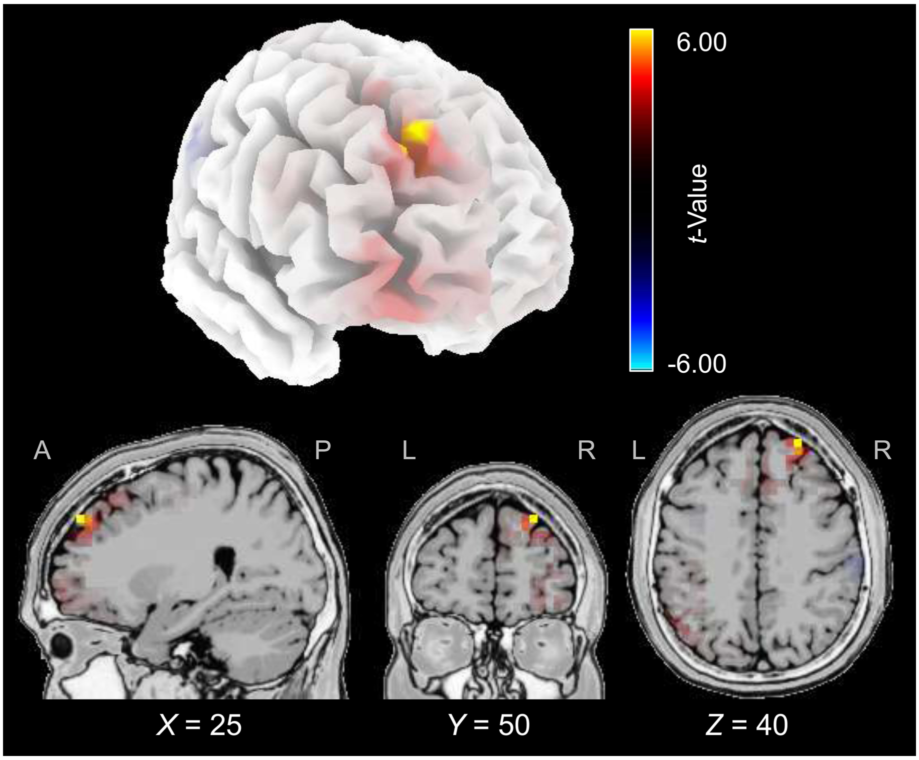


FIGURE 5 | Electrical brain activity at the time of the N2 after audiovisual-congruency training in the younger group. Electrical activity as revealed by sLORETA in the post-block was normalized with reference to pre-training activity and the resulting difference values were contrasted against the related difference values for the control condition. Data were mapped onto a standard 3-D brain template (top) or onto sagittal, coronal, and horizontal slices (T2 MNI-template "Colin 27" of sLORETA) positioned at X, Z, and Y coordinates (MNI) as given in the figure (bottom). Color coding shows *t*-values, with warm colors indicating increase of electrical activity and cold colors indicating decrease of electrical activity after training intervention (peak activity: $t = 6.45$, $p = 0.005$).

TABLE 1 | Brain structures revealed by the contrast of baseline-normalized activities obtained with audiovisual-congruency training vs. control condition at the time of the N2.

t-value	*p*-value	Structure	*X*	*Y*	*Z*
6.45	**0.005****	Right Superior Frontal Gyrus	25	50	40
4.15	0.245	Left Postcentral Gyrus	−40	−30	55
3.77	0.411	Right Middle Frontal Gyrus	30	45	40
3.39	0.638	Left Angular Gyrus	−50	−75	30
3.31	0.688	Left Middle Temporal Gyrus	−45	−75	10

*Peak activities immediately after training in the younger group are shown for the five structures with highest t-scores. Coordinates (X, Y, and Z) are in MNI space and refer to the voxel with the maximum t-score in each structure. Significant effect in bold, **p < 0.01.*

imaging revealed an increase in activity after audiovisual-congruency training that was located in right dorsolateral prefrontal cortex (DLPFC) (BA 9), in the anterior region of SFS. These findings provided, on the one hand, clear evidence that specific training characteristics, namely temporal synchrony and spatial congruency of auditory and visual target stimuli, can enhance ERP correlates of auditory spatial attention. On the other hand, we failed to find a related specificity for behavioral performance, which was generally improved after all training and control conditions.

Enhanced N2 Amplitude After Audiovisual-Congruency Training

In younger participants, audiovisual-congruency training induced a specific post-intervention enhancement of the N2, rather than P1, N1, and P2 components or LPC, for which no consistent changes were observed. This finding is in alignment with previous studies suggesting the N2 to be a decisive correlate of auditory selective spatial attention in "cocktail-party" scenarios (Gamble and Luck, 2011; Getzmann et al., 2015b; Lewald and Getzmann, 2015; Lewald et al., 2016; Hanenberg et al., 2019). An enhanced N2 was present not only immediately following training, but also in the 1-h-post-block, with a non-significant tendency to decrease as a function of time after training. Thus, it seems as if there were longer-lasting modulations in cortical processing after audiovisual-congruency training, which might be due to processes of neural plasticity induced by performing multisensory target localization.

An effect of cognitive training on the N2 has not been reported thus far in the context of auditory spatial attention. Previous studies focusing on improvements of working memory reported increases in N2 amplitude following *n*-back training in healthy participants (Covey et al., 2019) as well as in Multiple Sclerosis patients (Covey et al., 2018). Thus, on the basis of the limited data available at present, it remains open

TABLE 2 | Contrast of training vs. control conditions, as revealed by sLORETA.

			Younger Group		Older Group	
Training Condition	**Block**	**ERP Component**	t_{ext}	p_{ext}	t_{ext}	p_{ext}
Audiovisual Congruency						
	Post-Training	P1	−3.21	0.862	2.85	0.962
		N1	3.70	0.473	−3.78	0.437
		P2	3.32	0.705	−3.25	0.711
		N2	**6.45**	**0.005****	2.24	1.000
		LPC	−3.85	0.395	−3.08	0.919
	1-h Post-Training	P1	4.22	0.237	−2.65	0.993
		N1	3.64	0.355	3.74	0.416
		P2	4.92	0.060	−4.60	0.098
		N2	4.15	0.234	−3.59	0.578
		LPC	3.59	0.563	3.59	0.536
Visual Feedback						
	Post-Training	P1	3.14	0.902	−4.15	0.258
		N1	3.23	0.701	3.32	0.704
		P2	3.23	0.734	−3.33	0.729
		N2	3.73	0.482	−4.04	0.290
		LPC	3.37	0.703	−3.46	0.624
	1-h Post-Training	P1	3.85	0.465	−3.72	0.444
		N1	3.57	0.405	−3.61	0.498
		P2	3.63	0.478	−3.78	0.407
		N2	3.52	0.570	−3.58	0.596
		LPC	4.17	0.256	−3.62	0.554

Extreme t-scores (t_{ext}) and related p-values (p_{ext}) resulting from comparisons of baseline-normalized electrical brain activities are shown for training conditions (audiovisual-congruency training, visual-feedback training), blocks (post-training, 1-h post-training), and ERP components (P1, N1, P2, N2, and LPC). Significant effect in bold, $^{**}p < 0.01$.

whether audiovisual-congruency training had an effect on N2-related processes specifically involved in auditory attentional functions or rather enhanced more general cognitive processes associated with the N2.

The N2 enhancement following training was restricted to the younger group, which was an unexpected finding. The absence of an effect of training in the older group could possibly be due to the less flexible cognitive system of older adults, which is generally more occupied by task difficulty than that of younger persons (Yordanova et al., 2004; Salthouse, 2009; Gajewski and Falkenstein, 2012; Getzmann et al., 2015a,b,c; Olfers and Band, 2018). Our finding may be compatible with ERP research on age differences in speech perception. While older adults often showed reduced N2 amplitudes (Daffner et al., 2015), probably reflecting a less successful reorienting toward the stimulus of interest, younger adults showed more pronounced N2 amplitudes, reflecting an enhanced inhibitory control (Getzmann et al., 2014, 2015c). Thus, that the N2 enhancement was observed here in younger, but not older, adults could indicate a specific impact of the audiovisual-congruency training on processes of cognitive control. The question of whether related ERP enhancements, as found here for the N2 in younger participants, can also be induced in elderly people by more intense and repeated daily training over longer periods has to be addressed by future studies.

Cortical Sources of Training-Induced N2 Enhancement

At the time of the N2, an increase of electrical activity after audiovisual-congruency training was found to be located in right superior frontal gyrus, in the anterior region of SFS (cf. **Figure 5**). Previous studies, focusing on the cortical correlates of selective auditory spatial attention using various methods, have revealed several cortical regions, composing a complex network. This network comprises auditory cortex, posterior superior temporal gyrus (pSTG) and planum temporale (PT), inferior parietal lobule (IPL), superior parietal lobule and precuneus, inferior frontal gyrus, frontal eye field (FEF), as well as regions of BA 9 and SFS, which were nearby the location of training-induced activity change found here (Pugh et al., 1996; Nakai et al., 2005; Zündorf et al., 2013, 2014; Kong et al., 2014; Lee et al., 2014; Lewald and Getzmann, 2015; Braga et al., 2016, 2017; Lewald, 2016, 2019; Lewald et al., 2016, 2018). In particular, the area including SFS and FEF has been related to the so-called N2ac (Lewald et al., 2016), an anterior contralateral N2 subcomponent, which has been regarded as a correlate of auditory selective spatial attention (Gamble and Luck, 2011). Generally, the SFS is well-known as auditory spatial region of dorsofrontal cortex. This region has been demonstrated in many studies to be involved in sound localization (e.g., Weeks et al., 2000; Alain et al., 2001; Zatorre et al., 2002; Lewald et al., 2008; Zündorf et al., 2016) and has been assigned to the auditory posterodorsal (spatial) pathway (Arnott et al., 2004). However, it has to be noted that the activity increase revealed here was located at a more anterior position of SFS, compared with areas described in the studies cited above, which reported positions in caudal SFS. Because of the low spatial resolution of the electrical imaging method used here, any clear-cut conclusions on the location of activity in specific subareas of DLPFC might be difficult to draw, and further studies using imaging techniques with higher spatial resolution, such as functional magnetic resonance imaging (fMRI), may have to clarify this issue.

The present results may be related to recent findings by transcranial direct current stimulation (tDCS). In a preceding study using the same "cocktail-party" task as used here, a significant enhancement of the N2 was observed after monopolar anodal tDCS of right pSTG, including PT and auditory cortex (Hanenberg et al., 2019). Also, bilateral monopolar anodal tDCS over this area has been shown to induce clear offline improvements in behavioral performance with this task (Lewald, 2019). These findings have been related to the crucial role of PT in "cocktail-party" sound localization, as had been revealed by fMRI in healthy persons (Zündorf et al., 2013) and voxel-based lesion-behavior mapping analyses in stroke patients (Zündorf et al., 2014). The PT may represent an initial stage of auditory spatial processing within the hierarchically organized posterodorsal cortical pathway, channeling information to frontoparietal areas for further analyses (Griffiths and Warren, 2002; Krumbholz et al., 2005), including those relevant for selective spatial attention (Zündorf et al., 2014). In general alignment with this view, Hanenberg et al. (2019) found a reduction of activity in right IPL at the time of the enhanced N2 after anodal tDCS over right pSTG. The IPL is connected with ipsilateral DLPFC via dorsal components of the superior longitudinal fasciculus (SLF), a white-matter bundle that is crucially involved in functions of spatial orienting and awareness, as well as attentional control (Doricchi and Tomaiuolo, 2003; Makris et al., 2005; Thiebaut de Schotten et al., 2005, 2008; Doricchi et al., 2008; Bernal and Altman, 2010; Suchan et al., 2014; Nakajima et al., 2020). Zündorf et al. (2014) showed that left-sided lesions of the SLF were associated with deficits in the "cocktail-party" task, suggesting an important role of this structure and its frontal target areas in auditory selective spatial attention. Thus, the DLPFC might be part of a temporo-parieto-frontal network concerned with auditory functions subserving "cocktail-party" listening.

The specific post-intervention increase in activity found after audiovisual-congruency training suggests that the related processes in DLPFC were specifically strengthened by bimodal stimulation. This result may be compatible with previous findings indicating multi- or supramodal properties of the dorsofrontal networks that have been usually associated with selective spatial attention in the visual modality (e.g., Slotnick and Moo, 2006; Macaluso, 2010; Lee et al., 2012; Bharadwaj et al., 2014; Lewald et al., 2018). For the monkey DLPFC, it has been suggested that neuronal processes exist for visual and auditory location information and spatial working memory (Fuster et al., 2000; Kikuchi-Yorioka and Sawaguchi, 2000; Artchakov et al., 2007; Hwang and Romanski, 2015; for review, see Plakke and Romanski, 2016), and the human DLPFC has been shown to be involved in transforming auditory and visual inputs into multimodal spatial representations that can be used to guide saccades (Tark and Curtis, 2013). The monkey DLPFC receives projections from posterior auditory cortex areas known to be involved in spatial processing and from the posterior parietal cortex (Chavis and Pandya, 1976; Rauschecker et al., 1995; Romanski et al., 1999a,b). The latter area, which also has reciprocal connections with posterior auditory cortex, has been shown to be critically concerned with auditory and visual spatial processing in human and non-human primates (e.g., Mishkin et al., 1983; Mazzoni et al., 1996; Bushara et al., 1999; Romanski et al., 1999b; Lewald et al., 2002, 2004, 2016). Thus, in conclusion, we assume that the present finding of activity enhancement in BA 9 induced by repetitive processing of spatially and temporally congruent audiovisual stimuli during training may be related to the auditory-visual bimodal properties of the dorsal attention network composed of the DLPFC region and its connections with posterior parietal areas via SLF.

Audiovisual-Congruency as a Key Factor for Training Effects

An effect of training on the N2 was found exclusively for the audiovisual-congruency condition. This result may be in alignment with the multitude of studies on audiovisual integration, which have demonstrated bimodal enhancement by spatiotemporal alignment using several behavioral (Recanzone, 1998; Lewald et al., 2001; Lewald, 2002b; Lewald and Guski, 2003; Lovelace et al., 2003; Alais and Burr, 2004; Pages and Groh, 2013) and neurophysiological approaches (Besle et al., 2004; Stekelenburg and Vroomen, 2007, 2012; Talsma et al., 2007; Santangelo et al., 2008; Stein and Stanford, 2008; for review, see Stein and Meredith, 1993). Also, few studies in animals and humans have already demonstrated at the behavioral level that spatiotemporally congruent audiovisual stimulation can be used to improve accuracy of localization of single sound sources (Strelnikov et al., 2011; Isaiah et al., 2014; Berger et al., 2018; Cai et al., 2018; Kumpik et al., 2019). Even though we failed to find specific effects of audiovisual-congruency training at the behavioral level, the present study extended these previous approaches by showing that electrophysiological correlates of audiospatial attention in the presence of multiple distractor sources were enhanced by this type of training, while no effect was observed for visual-feedback training.

We assume that this result is related to the experience of phenomenal causality of auditory and visual events (i.e., the impression of a common cause) during audiovisual-congruency training, as typically occurs in the ventriloquism effect (Lewald and Guski, 2003). It is important to note that such binding phenomena do not require complex stimuli, with a highly compelling, meaningful association of auditory and visual information. Rather, simple light spots and tone bursts have been shown to be sufficient to induce audiovisual binding if presented in close spatiotemporal proximity (e.g., Thomas, 1941; Bertelson and Radeau, 1981; Radeau and Bertelson, 1987; Lewald et al., 2001; Slutsky and Recanzone, 2001; Lewald and Guski, 2003). In the audiovisual-congruency training used here, light flashes and target animal vocalizations were presented with identical stimulus onset and duration from roughly the same location. Participants may have experienced binding of auditory and visual events, if distractors were successfully suppressed by the occurrence of the "cocktail-party" effect. That is, this type of training may have induced neural processes resulting in more effective distractor suppression and, thus, more accurate representation of auditory targets in the presence of distractor sources. These processes can be described in terms of short-term neural plasticity based on the ventriloquisms effect, as has been discussed in the context of the normally occurring continuous calibration of auditory space by visual experience or its counterpart, the ventriloquism after-effect, in which repetitive or trial-wise presentation of synchronized auditory and visual stimuli with consistent spatial disparity shifts the representation of auditory space relative to the visual space (Recanzone, 1998; Lewald, 2002b). In an EEG source-imaging study focused on the neural basis of the ventriloquism aftereffect, Park and Kayser (2020) recently reported that prolonged exposure to consistent auditory-visual discrepancies recruits, in addition to sensory (occipital and temporal) cortices and multisensory parietal areas, prefrontal regions, including inferior frontal, middle frontal, and superior frontal gyri about 240 ms after stimulus onset. This finding could, potentially, be related to the enhancement of DLPFC activity found here after audiovisual-congruency training at the time of the N2. Taken together, spatiotemporal congruency of auditory and visual stimuli during training appears to be a key feature enhancing neural processes of auditory selective spatial attention. Our results suggested that this training-induced short-term plasticity occurs particularly in the DLPFC region at the time of the N2 component of the ERP.

Training-Induced Effects on N1 Amplitude and P2 Latency

Only minor and rather non-specific post-intervention changes were observed for N1 and P2 components, and no consistent effects at all for the P1 component and the LPC. The N1 amplitude was generally reduced in the post-training-blocks compared with 1-h-post-blocks, independently of the training condition. In terms of learning, changes in N1 amplitude are often referred to as early automatic stimulus processing (Näätänen and Picton, 1987; Lange, 2013) depending on attentional phenomena (Hillyard et al., 1973; Eimer, 2014; for review, see Luck et al., 2000; McEvoy et al., 2001). Increased N1 amplitudes have been associated with attention catching properties of auditory stimuli and task difficulty (Näätänen et al., 2011). On the other hand, decreased N1 amplitudes could indicate less attentional effort after training due to improved early processing of the stimuli, task familiarity, or the participants' impression that the task was less demanding after already having performed it (Tallus et al., 2015). Furthermore, the N1 typically decreases with repeated stimulation (Näätänen and Picton, 1987), but also recovers with time. This might be reflected by the result that N1 amplitudes were comparably strong in the pre-training and 1-h-post-blocks, but reduced in the post-block. Neural refractoriness as well as adaptation has been suggested as underlying mechanisms of this amplitude attenuation in N1 (Budd et al., 1998; Rosburg et al., 2010). These factors could also be relevant for our findings, given that the participants performed the test blocks three times per session.

P2 latencies were specifically shortened in the younger group 1 h after audiovisual-congruency training, but not for the other conditions, the older group, and the post-training block. On the one hand, this result suggests an accelerated occurrence of P2-related processes after this type of training. On the other hand, it seems difficult to interpret since one may generally expect stronger effects of training on ERPs immediately after the interventions, rather than delayed by 1 h. As the analyses also revealed a general shortening of P2 latencies in the 1-h-post-blocks, compared with post-blocks, it seems likely that an unspecific effect occurred that was due to the repeated execution of the task, independently of the type of intervention. To which extent such unspecific effects have superimposed potential effects specific to the audiovisual-congruency training remains unclear.

If one compares the N1 and P2 waves in the two age groups (**Figure 2**), it seems as if older adults have overall larger N1 amplitude and reduced P2 amplitude. This could be due to a longer processing negativity overlapping the N1 and P2 wave, which might explain why the N1 wave appears broader in older adults and why older adults show an overall more negative amplitude between 80 and 220 ms.

Limitations

Unlike the clear-cut electrophysiological result, task performance was found to be unspecifically improved after all training interventions and independently of age. This outcome was probably due to the multiple repetition of the task, as is often found in research on learning and memory (Ebbinghaus, 1885). That we failed to find specific training effects could, possibly, be a result of ceiling effects (Brungart and Simpson, 2007) after repeated training, as were found previously in a similar task (Hanenberg et al., 2019). However, given mean rates of correct responses of about 84% in the younger and 70% in the older group, ceiling effects should not play a major role here. Alternatively, the effect of the audiovisual-congruency training observed for the N2 amplitude could be confined to specific sub-processes required to solve a “cocktail-party” speech localization task. Assuming that the N2 especially reflects cognitive control processes mainly related to the inhibition of task-relevant information (as argued above), the audiovisual-congruency training might have enhanced specific cognitive control processes (reflected by the increase in N2 amplitude), rather than speech-in-noise localization in general.

This study left open the question of whether more extended and more long-lasting improvements in task performance in “cocktail-party” scenarios could be achieved by training in both, younger and older adults. Motivation has been shown to have a significant impact on task performance (Green and Seitz, 2015), and future studies offering more attractive training paradigms, such as game-based training, should be considered here. Recent work on this topic has shown that various age groups profit from action video game training, showing enhanced performances in task switching abilities after playing for 3 weeks (Basak et al., 2008; Strobach et al., 2012; Wang et al., 2016). Recently, Schuchert and Lewald (2020), using a similar “cocktail-party” task as used here, demonstrated that both audio action game training and video non-action game training improved auditory selective spatial attention in younger adults. The present results thus suggest that a bimodal (audio-visual synchronous) game training may also be promising in this respect. It has, however, to be noted that other training approaches, such as music training have also been shown to have beneficial effects on auditory processing in acoustically complex situations, especially in older adults. For example, Zendel et al. (2019) recently found for a group of older adults that learning to play the piano improved understanding words presented in loud background noise after 6 months. Also, a 10-week participation in a choir was shown to improve speech-in-noise perception, pitch discrimination ability, and the strength of the neural representation of speech fundamental frequency, suggesting that short-term choir participation may be considered as an effective intervention for mitigating age-related hearing losses (Dubinsky et al., 2019). Future studies will show which of these quite different training approaches is best suited to improve hearing performance of older people in everyday situations requiring auditory selective spatial attention.

CONCLUSION

In summary, the present study showed that short-term audiovisual-congruency training, but not visual-feedback training and a control condition, enhanced the N2 component in a multiple-speaker target localization task. The increase in N2 was associated with an increase of electrical activity in DLPFC and may indicate enhancement of neural processes of auditory selective spatial attention. Both effects were observed in younger, but not older, participants. Further experiments are necessary in order to examine whether more intensive, longer lasting and more realistic audio-visual training settings are suitable to obtain improvements also in behavioral 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 Ethical Committee of the Leibniz Research Centre for Working Environment and Human Factors, Dortmund. The participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

CH, SG, and JL contributed to conception and design of the study. CH collected the data and wrote the first draft of the manuscript. CH and JL wrote sections of the manuscript. JL supervised the project and takes final responsibility for this article. All authors contributed to the analyses, manuscript revision, read, and approved the submitted version.

FUNDING

This work was supported by the German Federal Ministry of Education and Research in the framework of the TRAIN-STIM project (Grant number 01GQ1424E).

ACKNOWLEDGMENTS

The authors wish to thank Jonas Heyermann and Stefan Weber for help in running the experiments, and Tobias Blanke and Peter Dillmann for preparing software and parts of the electronic equipment.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2021.645702/full#supplementary-material

REFERENCES

Alain, C., Arnott, S. R., Hevenor, S., Graham, S., and Grady, C. L. (2001). 'What' and 'where' in the human auditory system. *Proc. Natl. Acad. Sci. U.S.A.* 98, 12301–12306.

Alain, C., Campeanu, S., and Tremblay, K. (2010). Changes in sensory evoked responses coincide with rapid improvement in speech identification performance. *J. Cogn. Neurosci.* 22, 392–403. doi: 10.1162/jocn.2009.21279

Alain, C., Snyder, J. S., He, Y., and Reinke, K. S. (2007). Changes in auditory cortex parallel rapid perceptual learning. *Cereb. Cortex* 17, 1074–1084. doi: 10.1093/cercor/bhl018

Alais, D., and Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Curr. Biol.* 14, 257–262. doi: 10.1016/j.cub.2004.01.029

Anderer, P., Semlitsch, H. V., and Saletu, B. (1996). Multichannel auditory event-related brain potentials: effects of normal aging on the scalp distribution of N1. P2, N2 and P300 latencies and amplitudes. *Electroencephalogr. Clin. Neurophysiol.* 99, 458–472. doi: 10.1016/s0013-4694(96)96518-9

Arnott, S. R., Binns, M. A., Grady, C. L., and Alain, C. (2004). Assessing the auditory dual-pathway model in humans. *NeuroImage* 22, 401–408. doi: 10.1016/j.neuroimage.2004.01.014

Artchakov, D., Tikhonravov, D., Vuontela, V., Linnankoski, I., Korvenoja, A., and Carlson, S. (2007). Processing of auditory and visual location information in the monkey prefrontal cortex. *Exp. Brain Res.* 180, 469–479. doi: 10.1007/s00221-007-0873-8

Basak, C., Boot, W. R., Voss, M. W., and Kramer, A. F. (2008). Can training in a real-time strategy video game attenuate cognitive decline in older adults? *Psychol. Aging* 23, 765–777. doi: 10.1037/a0013494

Berger, C. C., Gonzalez-Franco, M., Tajadura-Jiménez, A., Florencio, D., and Zhang, Z. (2018). Generic HRTFs may be good enough in virtual reality. Improving source localization through cross-modal plasticity. *Front. Neurosci.* 12:21. doi: 10.3389/fnins.2018.00021

Bernal, B., and Altman, N. (2010). The connectivity of the superior longitudinal fasciculus: a tractography DTI study. *Magn. Reson. Imaging* 28, 217–225. doi: 10.1016/j.mri.2009.07.008

Bertelson, P., and Radeau, M. (1981). Cross-modal bias and perceptual fusion with auditory-visual spatial discordance. *Percept. Psychophysics* 29, 578–584. doi: 10.3758/bf03207374

Bertoli, S., Smurzynski, J., and Probst, R. (2005). Effects of age, age-related hearing loss, and contralateral cafeteria noise on the discrimination of small frequency changes: psychoacoustic and electrophysiological measures. *J. Assoc. Res. Otolaryngol.* 6, 207–222. doi: 10.1007/s10162-005-5029-6

Besle, J., Fort, A., Delpuech, C., and Giard, M.-H. (2004). Bimodal speech: early suppressive visual effects in human auditory cortex. *Eur. J. Neurosci.* 20, 2225–2234. doi: 10.1111/j.1460-9568.2004.03670.x

Bharadwaj, H. M., Lee, A. K. C., and Shinn-Cunningham, B. G. (2014). Measuring auditory selective attention using frequency tagging. *Front. Integr. Neurosci.* 8:6. doi: 10.3389/fnint.2014.00006

Bolognini, N., Rasi, F., Coccia, M., and Làdavas, E. (2005). Visual search improvement in hemianopic patients after audio-visual stimulation. *Brain* 128, 2830–2842. doi: 10.1093/brain/awh656

Bosen, A. K., Fleming, J. T., Allen, P. D., O'Neill, W. E., and Paige, G. D. (2018). Multiple time scales of the ventriloquism aftereffect. *PLoS One* 13:e0200930. doi: 10.1371/journal.pone.0200930

Braga, R. M., Fu, R. Z., Seemungal, B. M., Wise, R. J. S., and Leech, R. (2016). Eye movements during auditory attention predict individual differences in dorsal attention network activity. *Front. Hum. Neurosci.* 10:164. doi: 10.3389/fnhum.2016.00164

Braga, R. M., Hellyer, P. J., Wise, R. J. S., and Leech, R. (2017). Auditory and visual connectivity gradients in frontoparietal cortex: frontoparietal audiovisual gradients. *Hum. Brain Mapp.* 38, 255–270. doi: 10.1002/hbm.23358

Brainard, M. S., and Knudsen, E. I. (1993). Experience-dependent plasticity in the inferior colliculus: a site for visual calibration of the neural representation of auditory space in the barn owl. *J. Neurosci.* 13, 4589–4608. doi: 10.1523/jneurosci.13-11-04589.1993

Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound*. Cambridge, MA: MIT Press.

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

Brungart, D. S., and Simpson, B. D. (2007). Cocktail party listening in a dynamic multitalker environment. *Percept. Psychophys.* 69, 79–91. doi: 10.3758/bf03194455

Budd, T. W., Barry, R. J., Gordon, E., Rennie, C., and Michie, P. T. (1998). Decrement of the N1 auditory event-related potential with stimulus repetition: habituation vs. refractoriness. *Int. J. Psychophysiol.* 31, 51–68. doi: 10.1016/s0167-8760(98)00040-3

Bushara, K. O., Weeks, R. A., Ishii, K., Catalan, M. J., Tian, B., Rauschecker, J. P., et al. (1999). Modality-specific frontal and parietal areas for auditory and visual spatial localization in humans. *Nat. Neurosci.* 2, 759–766. doi: 10.1038/11239

Cai, Y., Chen, G., Zhong, X., Yu, G., Mo, H., Jiang, J., et al. (2018). Influence of audiovisual training on horizontal sound localization and its related ERP response. *Front. Hum. Neurosci.* 12:423. doi: 10.3389/fnhum.2018.00423

Canon, L. K. (1970). Intermodality inconsistency of input and directed attention as determinants of the nature of adaptation. *J. Exp. Psychol.* 84, 141–147. doi: 10.1037/h0028925

Canon, L. K. (1971). Directed attention and maladaptive 'adaptation' to displacement of the visual field. *J. Exp. Psychol.* 88, 403–408. doi: 10.1037/h0030878

Chavis, D. A., and Pandya, D. N. (1976). Further observations on corticofrontal connections in the rhesus monkey. *Brain Res.* 117, 369–386. doi: 10.1016/0006-8993(76)90089-5

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

Covey, T. J., Shucard, J. L., Benedict, R. H., Weinstock-Guttman, B., and Shucard, D. W. (2018). Improved cognitive performance and event-related potential changes following working memory training in patients with multiple sclerosis. *Mult. Scler. J. Exp. Transl. Clin.* 4:2055217317747626.

Covey, T. J., Shucard, J. L., and Shucard, D. W. (2019). Working memory training and perceptual discrimination training impact overlapping and distinct neurocognitive processes: evidence from event-related potentials and transfer of training gains. *Cognition* 182, 50–72. doi: 10.1016/j.cognition.2018.08.012

Daffner, K. R., Alperin, B. R., Mott, K. K., Tusch, E. S., and Holcomb, P. J. (2015). Age-related differences in early novelty processing: using PCA to parse the overlapping anterior P2 and N2 components. *Biol. Psychol.* 105, 83–94. doi: 10.1016/j.biopsycho.2015.01.002

Dennis, T. A., and Chen, C. C. (2007). Neurophysiological mechanisms in the emotional modulation of attention: the interplay between threat sensitivity and attentional control. *Biolo. Psychol.* 76, 1–10. doi: 10.1016/j.biopsycho.2007.05.001

Doricchi, F., Thiebaut de Schotten, M., Tomaiuolo, F., and Bartolomeo, P. (2008). White matter (dis)connections and gray matter (dys)functions in visual neglect: gaining insights into the brain networks of spatial awareness. *Cortex* 44, 983–995. doi: 10.1016/j.cortex.2008.03.006

Doricchi, F., and Tomaiuolo, F. (2003). The anatomy of neglect without hemianopia: a key role for parietal–frontal disconnection? *Neuroreport* 14, 2239–2243. doi: 10.1097/00001756-200312020-00021

Dubinsky, E., Wood, E. A., Nespoli, G., and Russo, F. A. (2019). Short-term choir singing supports speech-in-noise perception and neural pitch strength in older adults with age-related hearing loss. *Front. Neurosci.* 13:1153. doi: 10.3389/fnins.2019.01153

Ebbinghaus, H. (1885). *Über das Gedächtnis: Untersuchungen zur Experimentellen Psychologie*. Berlin: Duncker & Humblot.

Eimer, M. (2014). The neural basis of attentional control in visual search. *Trends Cogn. Scie.* 18, 526–535. doi: 10.1016/j.tics.2014.05.005

Eldar, S., and Bar-Haim, Y. (2010). Neural plasticity in response to attention training in anxiety. *Psychol. Med.* 40, 667–677. doi: 10.1017/S0033291709990766

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

Feierabend, M., Karnath, H.-O., and Lewald, J. (2019). Auditory space perception in the blind: horizontal sound localization in acoustically simple and complex situations. *Perception* 48, 1039–1057. doi: 10.1177/0301006619872062

Folstein, J. R., and Van Petten, C. (2008). Influence of cognitive control and mismatch on the N2 component of the ERP: a review. *Psychophysiology* 45, 152–170.

Fuster, J. M., Bodner, M., and Kroger, J. K. (2000). Cross-modal and cross-temporal association in neurons of frontal cortex. *Nature* 405, 347–351. doi: 10.1038/35012613

Gajewski, P. D., and Falkenstein, M. (2012). Training-induced improvement of response selection and error detection in aging assessed by task switching: effects of cognitive, physical, and relaxation training. *Front. Hum. Neurosci.* 6:130. doi: 10.3389/fnhum.2012.00130

Gamble, M. L., and Luck, S. J. (2011). N2ac: an ERP component associated with the focusing of attention within an auditory scene: the N2ac component. *Psychophysiology* 48, 1057–1068. doi: 10.1111/j.1469-8986.2010.01172.x

Gazzaley, A., and D'Esposito, M. (2007). Top-down modulation and normal aging. *Ann. N. Y. Acad. Sci.* 1097, 67–83. doi: 10.1196/annals.1379.010

Getzmann, S., Falkenstein, M., and Wascher, E. (2015a). ERP correlates of auditory goal-directed behavior of younger and older adults in a dynamic speech perception task. *Behav. Brain Res.* 278, 435–445. doi: 10.1016/j.bbr.2014.10.026

Getzmann, S., Hanenberg, C., Lewald, J., Falkenstein, M., and Wascher, E. (2015b). Effects of age on electrophysiological correlates of speech processing in a dynamic 'cocktail-party' situation. *Front. Neurosci.* 9:341. doi: 10.3389/fnins.2015.00341

Getzmann, S., Wascher, E., and Falkenstein, M. (2015c). What does successful speech-in-noise perception in aging depend on? Electrophysiological correlates of high and low performance in older adults. *Neuropsychologia* 70, 43–57. doi: 10.1016/j.neuropsychologia.2015.02.009

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

Grasso, P. A., Làdavas, E., and Bertini, C. (2016). Compensatory recovery after multisensory stimulation in hemianopic patients: behavioral and neurophysiological components. *Front. Syst. Neurosci.* 10:45. doi: 10.3389/fnsys.2016.00045

Gratton, G., Coles, M. G., and Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalogr. Clin. Neurophysiol.* 55, 468–484. doi: 10.1016/0013-4694(83)90135-9

Green, C. S., and Seitz, A. R. (2015). The impacts of video games on cognition (and how the government can guide the industry). *Policy Insights Behav. Brain Sci.* 2, 101–110. doi: 10.1177/2372732215601121

Griffiths, T. D., and Warren, J. D. (2002). The planum temporale as a computational hub. *Trends Neurosci.* 25, 348–353. doi: 10.1016/s0166-2236(02)02191-4

Hanenberg, C., Getzmann, S., and Lewald, J. (2019). Transcranial direct current stimulation of posterior temporal cortex modulates electrophysiological correlates of auditory selective spatial attention in posterior parietal cortex. *Neuropsychologia* 131, 160–170. doi: 10.1016/j.neuropsychologia.2019.05.023

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*, Vol. 22, ed. G. H. Bower (Cambridge, MA: Academic Press), 193–225. doi: 10.1016/S0079-7421(08)60041-9

Held, R. (1955). Shifts in binaural localization after prolonged exposures to atypical combinations of stimuli. *Am. J. Psychol.* 68, 526–548. doi: 10.2307/1418782

Helmholtz, H. V. (1867). *Handbuch der Physiologischen Optik*. Bonn: Voss.

Hillyard, S. A., Hink, R. F., Schwent, V. L., and Picton, T. W. (1973). Electrical signs of selective attention in the human brain. *Science* 182, 177–180. doi: 10.1126/science.182.4108.177

Howard, I. P., and Templeton, W. B. (1966). *Human Spatial Orientation*. Hoboken, NJ: Wiley.

Hwang, J., and Romanski, L. M. (2015). Prefrontal neuronal responses during audiovisual mnemonic processing. *J. Neurosci.* 35, 960–971. doi: 10.1523/jneurosci.1328-14.2015

Hyde, P. S., and Knudsen, E. I. (2000). Topographic projection from the optic tectum to the auditory space map in the inferior colliculus of the barn owl. *J. Comp. Neurol.* 421, 146–160. doi: 10.1002/(sici)1096-9861(20000529)421:2<146::aid-cne2>3.0.co;2-5

Hyde, P. S., and Knudsen, E. I. (2002). The optic tectum controls visually guided adaptive plasticity in the owl's auditory space map. *Nature* 415, 73–76. doi: 10.1038/415073a

Isaiah, A., Vongpaisal, T., King, A. J., and Hartley, D. E. H. (2014). Multisensory training improves auditory spatial processing following bilateral cochlear implantation. *J. Neurosci.* 34, 11119–11130. doi: 10.1523/jneurosci.4767-13.2014

Jack, C. E., and Thurlow, W. R. (1973). Effects of degree of visual association and angle of displacement on the 'ventriloquism' effect. *Percept. Mot. Skills* 37, 967–979. doi: 10.2466/pms.1973.37.3.967

Jackson, C. V. (1953). Visual factors in auditory localization. *Q. J. Exp. Psychol.* 5, 52–65. doi: 10.1080/17470215308416626

Kalil, R., and Freedman, S. J. (1967). Compensation for auditory re-arrangement in the absence of observer movement. *Percept. Mot. Skills* 24, 475–478. doi: 10.2466/pms.1967.24.2.475

Kikuchi-Yorioka, Y., and Sawaguchi, T. (2000). Parallel visuospatial and audiospatial working memory processes in the monkey dorsolateral prefrontal cortex. *Nat. Neurosci.* 3, 1075–1076. doi: 10.1038/80581

Klemm, O. (1910). Lokalisation von Sinneseindrücken bei disparaten Nebenreizen. *Psychol. Stud.* 5, 73–162.

Knudsen, E. I. (1999). Mechanisms of experience-dependent plasticity in the auditory localization pathway of the barn owl. *J. Comp. Physiol. A* 185, 305–321. doi: 10.1007/s003590050391

Knudsen, E. I., and Knudsen, P. F. (1985). Vision guides the adjustment of auditory localization in young barn owls. *Science* 230, 545–548. doi: 10.1126/science.4048948

Knudsen, E. I., and Knudsen, P. F. (1989). Vision calibrates sound localization in developing barn owls. *J. Neurosci.* 9, 3306–3313. doi: 10.1523/jneurosci.09-09-03306.1989

Kong, L., Michalka, S. W., Rosen, M. L., Sheremata, S. L., Swisher, J. D., Shinn-Cunningham, B. G., et al. (2014). Auditory spatial attention representations in the human cerebral cortex. *Cereb. Cortex* 24, 773–784. doi: 10.1093/cercor/bhs359

Krumbholz, K., Schönwiesner, M., Rübsamen, R., Zilles, K., Fink, G. R., and von Cramon, D. Y. (2005). Hierarchical processing of sound location and motion in the human brainstem and planum temporale. *Eur. J. Neurosci.* 21, 230–238. doi: 10.1111/j.1460-9568.2004.03836.x

Kuk, F., Keenan, D. M., Lau, C., Crose, B., and Schumacher, J. (2014). Evaluation of a localization training program for hearing impaired listeners. *Ear Hear.* 35, 652–666. doi: 10.1097/aud.0000000000000067

Kumpik, D. P., Campbell, C., Schnupp, J. W. H., and King, A. J. (2019). Re-weighting of sound localization cues by audiovisual training. *Front. Neurosci.* 13:1164. doi: 10.3389/fnins.2019.01164

Lange, K. (2013). The ups and downs of temporal orienting: a review of auditory temporal orienting studies and a model associating the heterogeneous findings on the auditory N1 with opposite effects of attention and prediction. *Front. Hum. Neurosci.* 7:263. doi: 10.3389/fnhum.2013.00263

Lee, A. K. C., Larson, E., and Maddox, R. K. (2012). Mapping cortical dynamics using simultaneous MEG/EEG and anatomically-constrained minimum-norm estimates: an auditory attention example. *J. Vis. Exp.* 68:e4262.

Lee, A. K. C., Larson, E., Maddox, R. K., and Shinn-Cunningham, B. G. (2014). Using neuroimaging to understand the cortical mechanisms of auditory selective attention. *Hear. Res.* 307, 111–120. doi: 10.1016/j.heares.2013.06.010

Lewald, J. (2002a). Opposing effects of head position on sound localization in blind and sighted human subjects. *Eur. J. Neurosci.* 15, 1219–1224. doi: 10.1046/j.1460-9568.2002.01949.x

Lewald, J. (2002b). Rapid adaptation to auditory-visual spatial disparity. *Learn. Mem.* 9, 268–278. doi: 10.1101/lm.51402

Lewald, J. (2002c). Vertical sound localization in blind humans. *Neuropsychologia* 40, 1868–1872. doi: 10.1016/s0028-3932(02)00071-4

Lewald, J. (2013). Exceptional ability of blind humans to hear sound motion: implications for the emergence of auditory space. *Neuropsychologia* 51, 181–186. doi: 10.1016/j.neuropsychologia.2012.11.017

Lewald, J. (2016). Modulation of human auditory spatial scene analysis by transcranial direct current stimulation. *Neuropsychologia* 84, 282–293. doi: 10.1016/j.neuropsychologia.2016.01.030

Lewald, J. (2019). Bihemispheric anodal transcranial direct-current stimulation over temporal cortex enhances auditory selective spatial attention. *Exp. Brain Res.* 237, 1539–1549. doi: 10.1007/s00221-019-05525-y

Lewald, J., Ehrenstein, W. H., and Guski, R. (2001). Spatio-temporal constraints for auditory–visual integration. *Behav. Brain Res.* 121, 69–79. doi: 10.1016/s0166-4328(00)00386-7

Lewald, J., Foltys, H., and Töpper, R. (2002). Role of the posterior parietal cortex in spatial hearing. *J. Neurosci.* 22:RC207.

Lewald, J., and Getzmann, S. (2015). Electrophysiological correlates of cocktail-party listening. *Behav. Brain Res.* 292, 157–166. doi: 10.1016/j.bbr.2015.06.025

Lewald, J., and Guski, R. (2003). Cross-modal perceptual integration of spatially and temporally disparate auditory and visual stimuli. *Brain Res. Cogn. Brain Res.* 16, 468–478. doi: 10.1016/s0926-6410(03)00074-0

Lewald, J., Hanenberg, C., and Getzmann, S. (2016). Brain correlates of the orientation of auditory spatial attention onto speaker location in a 'cocktail-party' situation. *Psychophysiology* 53, 1484–1495. doi: 10.1111/psyp.12692

Lewald, J., and Hausmann, M. (2013). Effects of sex and age on auditory spatial scene analysis. *Hear. Res.* 299, 46–52. doi: 10.1016/j.heares.2013.02.005

Lewald, J., Kentridge, R. W., Peters, S., Tegenthoff, M., Heywood, C. A., and Hausmann, M. (2013). Auditory-visual localization in hemianopia. *Neuropsychology* 27, 573–582. doi: 10.1037/a0033451

Lewald, J., Peters, S., Tegenthoff, M., and Hausmann, M. (2009a). Dissociation of auditory and visual straight ahead in hemianopia. *Brain Res.* 1287, 111–117. doi: 10.1016/j.brainres.2009.06.085

Lewald, J., Peters, S., Tegenthoff, M., and Hausmann, M. (2009b). Distortion of auditory space in hemianopia. *Eur. J. Neurosci.* 30, 1401–1411. doi: 10.1111/j.1460-9568.2009.06905.x

Lewald, J., Riederer, K. A. J., Lentz, T., and Meister, I. G. (2008). Processing of sound location in human cortex. *Eur. J. Neurosci.* 27, 1261–1270. doi: 10.1111/j.1460-9568.2008.06094.x

Lewald, J., Schlüter, M.-C., and Getzmann, S. (2018). Cortical processing of location changes in a "cocktail-party" situation: spatial oddball effects on electrophysiological correlates of auditory selective attention. *Hear. Res.* 365, 49–61. doi: 10.1016/j.heares.2018.04.009

Lewald, J., Tegenthoff, M., Peters, S., and Hausmann, M. (2012). Passive auditory stimulation improves vision in hemianopia. *PLoS One* 7:e31603. doi: 10.1371/journal.pone.0031603

Lewald, J., Wienemann, M., and Boroojerdi, B. (2004). Shift in sound localization induced by rTMS of the posterior parietal lobe. *Neuropsychologia* 42, 1598–1607. doi: 10.1016/j.neuropsychologia.2004.04.012

Lovelace, C. T., Stein, B. E., and Wallace, M. T. (2003). An irrelevant light enhances auditory detection in humans: a psychophysical analysis of multisensory integration in stimulus detection. *Brain Re. Cogn. Brain Res.* 17, 447–453. doi: 10.1016/s0926-6410(03)00160-5

Luck, S. J., Woodman, G. F., and Vogel, E. K. (2000). Event-related potential studies of attention. *Trends Cog. Sci.* 4, 432–440.

Macaluso, E. (2010). Orienting of spatial attention and the interplay between the senses. *Cortex* 46, 282–297. doi: 10.1016/j.cortex.2009.05.010

Majdak, P., Walder, T., and Laback, B. (2013). Effect of long-term training on sound localization performance with spectrally warped and band-limited head-related transfer functions. *J. Acoust. Soc. Am.* 134, 2148–2159. doi: 10.1121/1.4816543

Makris, N., Kennedy, D. N., McInerney, S., Sorensen, A. G., Wang, R., Caviness, V. S. Jr., et al. (2005). Segmentation of subcomponents within the superior longitudinal fascicle in humans: a quantitative, in vivo, DT-MRI study. *Cereb. Cortex* 15, 854–869. doi: 10.1093/cercor/bhh186

Marcell, M. M., Borella, D., Greene, M., Kerr, E., and Rogers, S. (2000). Confrontation naming of environmental sounds. *J. Clin. Exp. Neuropsychol.* 22, 830–864. doi: 10.1076/jcen.22.6.830.949

Martin, B. A., Tremblay, K. L., and Korczak, P. (2008). Speech evoked potentials: from the laboratory to the clinic. *Ear Hear.* 29, 285–313. doi: 10.1097/aud.0b013e3181662c0e

Mazzoni, P., Bracewell, R. M., Barash, S., and Andersen, R. A. (1996). Spatially tuned auditory responses in area LIP of macaques performing delayed memory saccades to acoustic targets. *J. Neurophysiol.* 75, 1233–1241. doi: 10.1152/jn.1996.75.3.1233

McEvoy, L. K., Pellouchoud, E., Smith, M. E., and Gevins, A. (2001). Neurophysiological signals of working memory in normal aging. *Brain Res. Cogn. Brain Res.* 11, 363–376. doi: 10.1016/s0926-6410(01)00009-x

Mishkin, M., Ungerleider, L. G., and Macko, K. A. (1983). Object vision and spatial vision: two cortical pathways. *Trends Neurosci.* 6, 414–417. doi: 10.1016/0166-2236(83)90190-x

Näätänen, R., Kujala, T., and Winkler, I. (2011). Auditory processing that leads to conscious perception: a unique window to central auditory processing opened by the mismatch negativity and related responses. *Psychophysiology* 48, 4–22. doi: 10.1111/j.1469-8986.2010.01114.x

Näätänen, R., and Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. *Psychophysiology* 24, 375–425. doi: 10.1111/j.1469-8986.1987.tb00311.x

Nakai, T., Kato, C., and Matsuo, K. (2005). An FMRI study to investigate auditory attention: a model of the cocktail party phenomenon. *Magn. Reson. Med. Sci.* 4, 75–82. doi: 10.2463/mrms.4.75

Nakajima, R., Kinoshita, M., Shinohara, H., and Nakada, M. (2020). The superior longitudinal fascicle: reconsidering the fronto-parietal neural network based on anatomy and function. *Brain Imaging Behav.* 14, 2817–2830. doi: 10.1007/s11682-019-00187-4

Olfers, K. J. F., and Band, G. P. H. (2018). Game-based training of flexibility and attention improves task-switch performance: near and far transfer of cognitive training in an EEG study. *Psychol. Res.* 82, 186–202. doi: 10.1007/s00426-017-0933-z

Pages, D. S., and Groh, J. M. (2013). Looking at the ventriloquist: visual outcome of eye movements calibrates sound localization. *PLoS One* 8:e72562. doi: 10.1371/journal.pone.0072562

Park, H., and Kayser, C. (2020). The neurophysiological basis of short- and long-term ventriloquism aftereffects. *bioRxiv* [Preprint]. doi: 10.1101/2020.06.16.154161

Pascual-Marqui, R. D. (2002). Standardized low-resolution brain electromagnetic tomography (sLORETA): technical details. *Methods Find. Exp. Clin. Pharmacol.* 24(Suppl. D), 5–12.

Passamonti, C., Bertini, C., and Làdavas, E. (2009). Audio-visual stimulation improves oculomotor patterns in patients with hemianopia. *Neuropsychologia* 47, 546–555. doi: 10.1016/j.neuropsychologia.2008.10.008

Pichora-Fuller, M. K., Alain, C., and Schneider, B. A. (2017). "Older adults at the cocktail party," in *The Auditory System at the Cocktail Party*. Springer Handbook of Auditory Research, Vol. 60, eds J. C. Middlebrooks, J. Z. Simon, A. N. Popper, and R. R. Fay (Berlin: Springer), 227–259. doi: 10.1007/978-3-319-51662-2_9

Plakke, B., and Romanski, L. M. (2016). Neural circuits in auditory and audiovisual memory. *Brain Res.* 1640, 278–288. doi: 10.1016/j.brainres.2015.11.042

Potts, G. F. (2004). An ERP index of task relevance evaluation of visual stimuli. *Brain Cogn.* 56, 5–13. doi: 10.1016/j.bandc.2004.03.006

Pugh, K. R., Offywitz, B. A., Shaywitz, S. E., Fulbright, R. K., Byrd, D., Skudlarski, P., et al. (1996). Auditory selective attention: an fMRI investigation. *NeuroImage* 4(3 Pt 1), 159–173.

Radeau, M. (1994). Auditory-visual spatial interaction and modularity. *Curr. Psychol. Cogn.* 13, 3–51.

Radeau, M., and Bertelson, P. (1977). Adaptation to auditory-visual discordance and ventriloquism in semirealistic situations. *Percept. Psychophys.* 22, 137–146. doi: 10.3758/bf03198746

Radeau, M., and Bertelson, P. (1978). Cognitive factors and adaptation to auditory-visual discordance. *Percept. Psychophys.* 23, 341–343. doi: 10.3758/bf03199719

Radeau, M., and Bertelson, P. (1987). Auditory-visual interaction and the timing of inputs: thomas (1941) revisited. *Psychol. Res.* 49, 17–22. doi: 10.1007/bf00309198

Rauschecker, J. P., Tian, B., and Hauser, M. (1995). Processing of complex sounds in the macaque nonprimary auditory cortex. *Science* 268, 111–114. doi: 10.1126/science.7701330

Recanzone, G. H. (1998). Rapidly induced auditory plasticity: the ventriloquism aftereffect. *Proc. Natl. Acad. Sci. U.S.A.* 95, 869–875. doi: 10.1073/pnas.95.3.869

Romanski, L. M., Bates, J. F., and Goldman-Rakic, P. S. (1999a). Auditory belt and parabelt projections to the prefrontal cortex in the rhesus monkey. *J. Comp. Neurol.* 403, 141–157. doi: 10.1002/(sici)1096-9861(19990111)403:2<141::aid-cne1>3.0.co;2-v

Romanski, L. M., Tian, B., Fritz, J., Mishkin, M., Goldman-Rakic, P. S., and Rauschecker, J. P. (1999b). Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nat. Neurosci.* 2, 1131–1136. doi: 10.1038/16056

Rosburg, T., Zimmerer, K., and Huonker, R. (2010). Short-term habituation of auditory evoked potential and neuromagnetic field components in dependence of the interstimulus interval. *Exp. Brain Res.* 205, 559–570. doi: 10.1007/s00221-010-2391-3

Rueda, M., Pozuelos, J., and Cómbita, L. (2015). Cognitive neuroscience of attention. From brain mechanisms to individual differences in efficiency. *AIMS Neurosci.* 2, 183–202. doi: 10.3934/Neuroscience.2015.4.183

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

Santangelo, V., Van der Lubbe, R. H. J., Olivetti Belardinelli, M., and Postma, A. (2008). Multisensory integration affects ERP components elicited by exogenous cues. *Exp. Brain Res.* 185, 269–277. doi: 10.1007/s00221-007-1151-5

Schuchert, J. B., and Lewald, J. (2020). Training with audio and video games improves audiospatial performance in a "cocktail-party" task: a controlled intervention study in young adults. *bioRxiv* [Preprint]. doi: 10.1101/2020.11.17.386300

Shams, L., and Seitz, A. R. (2008). Benefits of multisensory learning. *Trends Cogn. Sci.* 12, 411–417. doi: 10.1016/j.tics.2008.07.006

Slotnick, S. D., and Moo, L. R. (2006). Prefrontal cortex hemispheric specialization for categorical and coordinate visual spatial memory. *Neuropsychologia* 44, 1560–1568. doi: 10.1016/j.neuropsychologia.2006.01.018

Slutsky, D. A., and Recanzone, G. H. (2001). Temporal and spatial dependency of the ventriloquism effect. *Neuroreport* 12, 7–10. doi: 10.1097/00001756-200101220-00009

Stein, B. E., and Meredith, M. A. (1993). *The Merging of the Senses*. Cambridge, MA: MIT Press.

Stein, B. E., and Stanford, T. R. (2008). Multisensory integration: current issues from the perspective of the single neuron. *Nat. Rev. Neurosci.* 9, 255–266. doi: 10.1038/nrn2331

Stekelenburg, J. J., and Vroomen, J. (2007). Neural correlates of multisensory integration of ecologically valid audiovisual events. *J. Cogn. Neurosci.* 19, 1964–1973. doi: 10.1162/jocn.2007.19.12.1964

Stekelenburg, J. J., and Vroomen, J. (2012). Electrophysiological correlates of predictive coding of auditory location in the perception of natural audiovisual events. *Front. Integr. Neurosci.* 6:26. doi: 10.3389/fnint.2012.00026

Stratton, G. M. (1896). Some preliminary experiments on vision without inversion of the retinal image. *Psychol. Rev.* 3, 611–617. 10.1037/h0072918

Stratton, G. M. (1897). Vision without inversion of the retinal image. *Psychol. Rev.* 4, 341–360. doi: 10.1037/h0075482

Strelnikov, K., Rosito, M., and Barone, P. (2011). Effect of audiovisual training on monaural spatial hearing in horizontal plane. *PLoS One* 6:e18344. doi: 10.1371/journal.pone.0018344

Strobach, T., Liepelt, R., Schubert, T., and Kiesel, A. (2012). Task switching: effects of practice on switch and mixing costs. *Psychol. Res.* 76, 74–83. doi: 10.1007/s00426-011-0323-x

Suchan, J., Umarova, R., Schnell, S., Himmelbach, M., Weiller, C., Karnath, H.-O., et al. (2014). Fiber pathways connecting cortical areas relevant for spatial orienting and exploration: fiber pathways relevant for spatial orienting. *Hum. Brain Mapp.* 35, 1031–1043. doi: 10.1002/hbm.22232

Tallus, J., Soveri, A., Hämäläinen, H., Tuomainen, J., and Laine, M. (2015). Effects of auditory attention training with the dichotic listening task: behavioural and neurophysiological evidence. *PLoS One* 10:e0139318. doi: 10.1371/journal.pone.0139318

Talsma, D., Doty, T. J., and Woldorff, M. G. (2007). Selective attention and audiovisual integration: is attending to both modalities a prerequisite for early integration? *Cereb. Cortex* 17, 679–690. doi: 10.1093/cercor/bhk016

Tark, K.-J., and Curtis, C. E. (2013). Deciding where to look based on visual, auditory, and semantic information. *Brain Res.* 1525, 26–38. doi: 10.1016/j.brainres.2013.06.002

Thiebaut de Schotten, M., Kinkingnéhun, S., Delmaire, C., Lehéricy, S., Duffau, H., Thivard, L., et al. (2008). Visualization of disconnection syndromes in humans. *Cortex* 44, 1097–1103. doi: 10.1016/j.cortex.2008.02.003

Thiebaut de Schotten, M., Urbanski, M., Duffau, H., Volle, E., Lévy, R., Dubois, B., et al. (2005). Direct evidence for a parietal-frontal pathway subserving spatial awareness in humans. *Science* 309, 2226–2228. doi: 10.1126/science.1116251

Thomas, G. J. (1941). Experimental study of the influence of vision on sound localization. *J. Exp. Psychol.* 28, 163–177. doi: 10.1037/h0055183

Wang, P., Liu, H.-H., Zhu, X.-T., Meng, T., Li, H.-J., and Zuo, X.-N. (2016). Action video game training for healthy adults: a meta-analytic study. *Front. Psychol.* 7:907. doi: 10.3389/fpsyg.2016.00907

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

Weeks, R., Horwitz, B., Aziz-Sultan, A., Tian, B., Wessinger, C. M., Cohen, L. G., et al. (2000). A positron emission tomographic study of auditory localization in the congenitally blind. *J. Neurosci.* 20, 2664–2672. doi: 10.1523/jneurosci.20-07-02664.2000

Yang, W., Guo, A., Li, Y., Qiu, J., Li, S., Yin, S., et al. (2018). Audio-visual spatiotemporal perceptual training enhances the P300 component in healthy older adults. *Front. Psychol.* 9:2537. doi: 10.3389/fpsyg.2018.02537

Yordanova, J., Falkenstein, M., Hohnsbein, J., and Kolev, V. (2004). Parallel systems of error processing in the brain. *NeuroImage* 22, 590–602. doi: 10.1016/j.neuroimage.2004.01.040

Zahorik, P., Bangayan, P., Sundareswaran, V., Wang, K., and Tam, C. (2006). Perceptual recalibration in human sound localization: learning to remediate front-back reversals. *J. Acoust. Soc. Am.* 120, 343–359. doi: 10.1121/1.2208429

Zatorre, R. J., Bouffard, M., Ahad, P., and Belin, P. (2002). Where is 'where' in the human auditory cortex? *Nat. Neurosci.* 5, 905–909. doi: 10.1038/nn904

Zendel, B. R., de Boysson, C., Mellah, S., Démonet, J. F., and Belleville, S. (2016). The impact of attentional training on event-related potentials in older adults. *Neurobiol. Aging* 47, 10–22. doi: 10.1016/j.neurobiolaging.2016.06.023

Zendel, B. R., West, G. L., Belleville, S., and Peretz, I. (2019). Musical training improves the ability to understand speech-in-noise in older adults. *Neurobiol. Aging* 81, 102–115. doi: 10.1016/j.neurobiolaging.2019.05.015

Zheng, W., and Knudsen, E. I. (2001). Gabaergic inhibition antagonizes adaptive adjustment of the owl's auditory space map during the initial phase of plasticity. *J. Neurosci.* 21, 4356–4365. 10.1523/jneurosci.21-12-04356.2001

Zündorf, I. C., Karnath, H.-O., and Lewald, J. (2014). The effect of brain lesions on sound localization in complex acoustic environments. *Brain* 137, 1410–1418. doi: 10.1093/brain/awu044

Zündorf, I. C., Lewald, J., and Karnath, H.-O. (2013). Neural correlates of sound localization in complex acoustic environments. *PLoS One* 8:e64259. doi: 10.1371/journal.pone.0064259

Zündorf, I. C., Lewald, J., and Karnath, H.-O. (2016). Testing the dual-pathway model for auditory processing in human cortex. *NeuroImage* 124, 672–681. doi: 10.1016/j.neuroimage.2015.09.026

Zwiers, M. P., Van Opstal, A. J., and Cruysberg, J. R. (2001a). A spatial hearing deficit in early-blind humans. *J. Neurosci.* 21, RC142: 1–5.

Zwiers, M. P., Van Opstal, A. J., and Cruysberg, J. R. (2001b). Two-dimensional sound-localization behavior of early-blind humans. *Exp. Brain Res.* 140, 206–222. doi: 10.1007/s002210100800

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 Human Neuroscience

ORIGINAL RESEARCH
published: 30 November 2021
doi: 10.3389/fnhum.2021.771711

The Effects of Working Memory Load on Auditory Distraction in Adults With Attention Deficit Hyperactivity Disorder

Rina Blomberg[1,2]*, *Andrea Johansson Capusan*[3,4,5], *Carine Signoret*[1,2], *Henrik Danielsson*[1,2] and *Jerker Rönnberg*[1,2,6]

[1] *Department of Behavioural Sciences and Learning, Linköping University, Linköping, Sweden,* [2] *Swedish Institute for Disability Research, Linköping University, Linköping, Sweden,* [3] *Department of Psychiatry, Linköping University, Linköping, Sweden,* [4] *Department of Clinical and Experimental Medicine, Linköping University, Linköping, Sweden,* [5] *Center for Social and Affective Neuroscience, Linköping University, Linköping, Sweden,* [6] *Center for Medical Image Science and Visualization, Linköping University, Linköping, Sweden*

Edited by:
Jessica A. Turner,
Georgia State University,
United States

Reviewed by:
Joaquim Radua,
Institut de Recerca Biomèdica August Pi i Sunyer (IDIBAPS), Spain
Timothy Michael Ellmore,
The City College of New York (CUNY),
United States

***Correspondence:**
Rina Blomberg
rina.blomberg@liu.se

Specialty section:
This article was submitted to Brain Health and Clinical Neuroscience,
a section of the journal
Frontiers in Human Neuroscience

Received: *07 September 2021*
Accepted: *08 November 2021*
Published: *30 November 2021*

Citation:
Blomberg R, Johansson Capusan A, Signoret C, Danielsson H and Rönnberg J (2021) The Effects of Working Memory Load on Auditory Distraction in Adults With Attention Deficit Hyperactivity Disorder. Front. Hum. Neurosci. 15:771711. doi: 10.3389/fnhum.2021.771711

Cognitive control provides us with the ability to *inter alia*, regulate the locus of attention and ignore environmental distractions in accordance with our goals. Auditory distraction is a frequently cited symptom in adults with attention deficit hyperactivity disorder (aADHD)–yet few task-based fMRI studies have explored whether deficits in cognitive control (associated with the disorder) impedes on the ability to suppress/compensate for exogenously evoked cortical responses to noise in this population. In the current study, we explored the effects of auditory distraction as function of working memory (WM) load. Participants completed two tasks: an auditory target detection (ATD) task in which the goal was to actively detect salient oddball tones amidst a stream of standard tones in noise, and a visual *n*-back task consisting of 0-, 1-, and 2-back WM conditions whilst concurrently ignoring the same tonal signal from the ATD task. Results indicated that our sample of young aADHD (n = 17), compared to typically developed controls (n = 17), had difficulty attenuating auditory cortical responses to the task-irrelevant sound when WM demands were high (2-back). Heightened auditory activity to task-irrelevant sound was associated with both poorer WM performance and symptomatic inattentiveness. In the ATD task, we observed a significant increase in functional communications between auditory and salience networks in aADHD. Because performance outcomes were on par with controls for this task, we suggest that this increased functional connectivity in aADHD was likely an adaptive mechanism for suboptimal listening conditions. Taken together, our results indicate that aADHD are more susceptible to noise interference when they are engaged in a primary task. The ability to cope with auditory distraction appears to be related to the WM demands of the task and thus the capacity to deploy cognitive control.

Keywords: attention deficit hyperactivity disorder, adults, attention, cognitive control, auditory distraction, salience network (SN), working memory, task-based fMRI

INTRODUCTION

Cognitive control refers to a set of complex cognitive mechanisms that collectively coordinate flexible and goal-directed behavior and include working memory (WM), attention, conflict monitoring, contextual anticipation and inference, inhibition and action selection (Egner, 2017). Importantly, these cognitive mechanisms provide us with the ability to not only selectively prioritize goals/requirements but to also suppress intrusive thoughts, inhibit inappropriate actions and ignore environmental distractions in accordance with those goals/requirements (Badre, 2020). Attention deficit hyperactivity disorder (ADHD) is a neuropsychiatric disorder in which symptoms encompass developmental deficits in cognitive control (Barkley, 1997; Castellanos and Tannock, 2002; van Lieshout et al., 2017; Pievsky and McGrath, 2018). Although once considered a childhood disorder, deficits in cognitive control have been shown to persist into adulthood (Faraone et al., 2015, 2021). In adults, ADHD is heterogenous but some symptoms are more representative of adult-ADHD (aADHD) than others. Symptoms of hyperactivity for instance, have been shown to decrease (or at least manifest in different ways compared to children) whereas attention and WM related difficulties tend to persist into adulthood (Mostert et al., 2015; Uchida et al., 2018). Indeed, deficient WM capacity is one of the most robust associations of impairments in daily functioning in aADHD (Willcutt et al., 2005; Alderson et al., 2013).

Cognitive neuroscience has known for some time that attention modulates sensory processing. Directing attention to a particular sensory modality for instance, can increase cortical activity in primary and secondary processing regions whilst directing attention away from the sensory source can reduce neural activity in said cortical regions. The effect of the latter is considered a mechanism for inhibiting sensory distraction. WM capacity–as an index for more general cognitive control ability–has long been theorized to play a prominent role in constraining distraction. In the auditory domain, empirical studies have shown that individuals with low WM capacity have difficulties hampering the disruptive effects of involuntary orienting to task-irrelevant acoustic stimulation (e.g., Conway et al., 2001; Dige et al., 2010; Sörqvist, 2010a; Hughes et al., 2013; Yurgil and Golob, 2013; Pelletier et al., 2016), although results are not always consistent (e.g., Beaman, 2004; Körner et al., 2017; Nagaraj et al., 2020).

One theory (Sörqvist, 2010b; Sörqvist and Rönnberg, 2014; Sörqvist and Marsh, 2015; Marsh and Campbell, 2016) suggests that WM capacity shields against auditory distraction in two main ways. First, high WM capacity is associated with better cognitive control, thus individuals with high WM capacity are able to maintain a more steadfast locus of attention in the face of challenging demands than their low WM capacity counterparts. Secondly, this uptake in attentional engagement deploys inhibitory mechanisms that suppress neural responses to task irrelevant sounds in accordance with these demands. The effect is reciprocal, in that challenging requirements both increase attentional engagement and decrease susceptibility for distraction within the limits of the individual's capacity for cognitive control. In support of this theoretical perspective, several studies in typically developed adults (TDa) have shown that increasing WM load in the visual modality results in greater neural attenuation of task-irrelevant auditory stimulation in the brainstem, and auditory cortical processing regions (e.g., Gisselgård et al., 2003; Regenbogen et al., 2012; Sörqvist et al., 2012, 2016).

Heightened auditory distraction is a commonly reported symptom in adults with ADHD (Schulze et al., 2020) and clinically, is associated with more general impairments in attention. Some behavioral studies in adults with ADHD indicate that their susceptibility to auditory distraction is related to a deficient WM capacity (e.g., Dige et al., 2010; Pelletier et al., 2016). However, neuroimaging studies investigating the effects of attentional engagement upon auditory distraction in aADHD are scarce. The main purpose of the current study, therefore, is to investigate whether our sample of aADHD, compared to healthy controls, demonstrate heightened cortical responses to task-irrelevant acoustic stimulation whilst engaged in a visual WM task; and to explore whether cortical attenuation of task-irrelevant acoustic stimulation is associated with individual differences in WM capacity.

To this end, we adopted a similar task paradigm from previous neuroimaging studies in TDa (Regenbogen et al., 2012; Sörqvist et al., 2016) that reported decreases in task-irrelevant auditory processing as a function of visual WM load. In all experimental condition's participants were exposed to a monotonous tonal signal which included an occasional deviant pitch whilst viewing a sequence of letters in the center of their visual field. In three of four experimental conditions, participants were instructed to ignore the auditory stream and perform one of three *n*-back conditions: 0-, 1-, and 2-back, on the visual letter sequence. In the remaining condition, participants were asked to explicitly attend to the auditory signal and to detect the deviant pitch, i.e., oddball target, whilst ignoring the visual sequence of letters.

fMRI imaging studies in TDa have shown that exogenous responses to task-irrelevant changes in pitch activate auditory core and belt regions as well as the posterior insula (e.g., Mayer et al., 2006; Sabri et al., 2011; Alho et al., 2014, 2015; Huang et al., 2015). These regions are thought to be activated pre-attentively and may generate the early N1 and MMN (mismatch negativity) auditory components in EEG recordings (Edwards et al., 2005; Garrido et al., 2009; Menon and Uddin, 2010; El Karoui et al., 2015; Blenkmann et al., 2018; Citherlet et al., 2020). From previous studies (Gisselgård et al., 2003; Regenbogen et al., 2012; Sörqvist et al., 2012, 2016), we expected that neural responses to the streaming acoustic signal, in these core and belt auditory regions, would attenuate when participants focused their attention on the visual WM task. And we hypothesized that the magnitude of attenuation in the control group, would increase as WM load transitioned from low to high. In contrast, we hypothesized that the ADHD group would show heightened auditory activity under high visual WM load relative to controls and that this heightened activity would negatively correlate with performance on the high-load task. Additionally, based upon findings from Sörqvist et al. (2012) we also expected WM capacity to correlate with auditory attenuation under the high-load condition.

The auditory oddball task was a simple target detection task that was not expected to place demands on WM capacity. The salience network (SN), a network critically involved in cognitive control operations, has been shown to be consistently activated in target detection tasks (Crottaz-Herbette and Menon, 2006). The network is generally thought to be triggered exogenously via communications from sensory processing regions and plays an active role in the vigilant anticipation, detection, and response-mediation of behaviorally salient stimuli (Corbetta et al., 2008; Menon and Uddin, 2010; Cabeza et al., 2012; Vossel et al., 2014). We therefore expected to see increased SN activation in the auditory target detection (ATD) task relative to a resting baseline in both participant groups. Although some resting-state fMRI studies have observed aberrant SN connectivity in ADHD (see Castellanos and Aoki, 2016 for a review), task-based fMRI investigations of the SN in aADHD are rare, particularly in the auditory domain. We therefore additionally tested for between-group differences in auditory–SN connectivity during the ATD task.

MATERIALS AND METHODS

Participants

We recruited two groups of participants: clinically stable adults with ADHD and healthy controls. Inclusion criteria were assessed via a two-step procedure. First, all applicants were required to fill in a digital questionnaire regarding age, health, handedness, alcohol/substance use, diagnosis/es and medications. Applicants were excluded at this stage if they were older than 50 years; were dominantly left-handed; reported having medical or psychiatric conditions or disabilities that could affect the quality of the data (e.g., severe acute psychiatric disorders such as but not exclusively: psychotic disorder, bipolar disorder, current severe MDD; ASD or hearing loss) or reported frequent use of alcohol/substances. Participants medicated with medication that could affect attention or wakefulness such as neuroleptics, sedatives, and/or opioids were excluded. Because a vast majority of aADHD (up to 75%) also have comorbid anxiety and/or depressive symptoms (Kessler et al., 2005; Kooij et al., 2012), ADHD-applicants on stable doses for at least 2 months of common antidepressants were not excluded. Stable medication with SSRI or SNRI indicates stability in the comorbid condition, while not unnecessarily excluding a representative group of participants. In addition, applicants with ADHD were included only if medicating with central stimulants (methylphenidate, dexamphetamine, lisdexamfetamine, etc.) and prepared to undergo a 48 h washout period prior to testing or were currently unmedicated for their ADHD.

Second, inclusion criteria were further assessed on the day of participation. Participants were screened for normal hearing thresholds (<20 dB HL) with pure-tone audiometry at six frequencies ranging from 250 to 8000 Hz (as described in: American Speech-Language-Hearing Association, 2005). Clinical assessments of attention, ADHD-symptom severity, and the presence of comorbid disorders and problems with substance abuse were investigated in more detail via the d2-R Test of Attention (Brickenkamp et al., 2010), the 18-item adult ADHD self-report scale (ASRS) v.1.1 (Kessler et al., 2005; Rodriguez et al., 2007) and the Mini-International Neuropsychiatric Interview (MINI) 7.0.2 DSM-5 for ADHD studies (Sheehan et al., 1998) respectively. Individual scores for the ASRS were calculated by summing scores from items associated with self-reported attentional difficulties (Part A: 1–4; Part B: 1–5), hyperactivity/impulsivity problems (Part A: 5–6; Part B: 6–12), as well as the sum of all 18-items as a general index of symptom severity. Standard scoring procedures were used for each index in the d2-R and MINI.

The data for this study was obtained from 34 participants who met the inclusion and exclusion criteria, forming two equal groups. The ADHD-group consisted of eleven females (age: M = 27 years, SD = 7.0) and six males (age: M = 29 years, SD = 7.0). 15 out of the 17 ADHD participants were prescribed stimulant medication (eight lisdexamfetamine, six methylphenidate and one dexamphetamine), and abstained from their medication for 48 h prior to testing. The remaining two participants were unmedicated for ADHD. In the ADHD group five participants had SSRI medication and one had lamotrigine in stable doses during at least 2 months. The control group consisted of 13 females (age: M = 25 years, SD = 4.9) and 4 males (age: M = 26 years, SD = 6.2).

Materials and Experimental Protocol

The scanning protocol utilized a block design and was programmed in Presentation Software (21.1, build September 05, 2019).[1] Participants used their right index finger to trigger the response button and response times (RTs), performance accuracy and false alarms (i.e., responding when no target was present) were recorded via a response box (LUMINA, Cedrus Corporation, San Pedro, CA, United States) interfaced with the stimulus presentation. In all experimental blocks, participants were exposed to both auditory and visual stimuli. Auditory stimuli were presented to participants via OptoActive^{TM} active noise canceling headphones (OptoAcoustics Ltd., Tel Aviv, Israel). The noise canceling headphones both passively and actively attenuated the background echo planar imaging (EPI) gradient noise to $\sim$ 58 dB SPL. And the headphones were kept in place via inflatable positioning pads (Pearltec MRI/CT Multipad Plus, MagMedix, Fitchburg, MA, United States) which also served to minimize head movements within the head cage. Visual stimuli were presented in the center of the visual field via MRI-goggles (Resonance Technology Company, Inc., Los Angeles, CA, United States).

Prior to the scan, subjects underwent a training session to ensure that they clearly understood all the task requirements. For the visual n-back task, participants were told to perform one of three n-back conditions whilst ignoring a streaming acoustic signal. N-back stimuli consisted of a sequence of 15 letters drawn pseudo-randomly from the set: K, M, Q, R, S, T, and W (white text on black background, font size: 18 points). Letters were individually presented for a duration of 500 ms followed by an interstimulus interval indicated by a fixation cross of 1022 ms,

---

[1] www.neurobs.com

and for a total block duration of 22.8 s (equivalent to 30 repetition times; TR = 761 ms). The 0-back condition required participants to press the response button to the target letter, K. In the 1-back condition, the response button was to be triggered when participants detected two consecutive, identical letters. The 2-back condition required participants to respond when they saw a letter identical to a letter presented two trials prior. Each *n*-back block contained four target letters in total.

In the ATD task, participants were told to ignore the visual *n*-back stimuli and instead shift their attention to the streaming auditory stimulus. The task required participants to press the response button every time they heard a deviant tone (1000 Hz) amidst a stream of standard tones (500 Hz). Tones of 150 ms (rise and fall = 22 ms) were presented every 104 ms with a total of four deviant tones within random distances of circa 2–6 s per 22.8 s experimental block. The perceived loudness of the standard and deviant tones was set at −16 LUFS and presented to participants at a sound pressure level of ∼ 75 dB. The presentation order of the experimental conditions was partially counterbalanced over the entire experiment and divided into three runs of eight task blocks, wherein each experimental condition was presented twice per run (see **Figure 1**). Every experimental block was proceeded by a 15.2 s resting baseline condition (=20 TR) and initiated by a 5.3 s (=7 TR) task instruction.

Outside of the scanner, and in a quiet room, participants completed two WM span tests. The Reading span task (Rönnberg et al., 1989) to measure WM maintenance, and the Size-comparison span test (Sörqvist et al., 2010) as a measure of both WM gating and maintenance; the procedures for which are published in Blomberg et al. (2019).

Image Acquisition

Whole-head fMRI scans were performed on a Siemens Prisma 3T scanner with a 64-channel head coil at the Centre for Medical Imaging and Visualization (CMIV), Linköping University Hospital, Sweden. A 3D, T1-weighted MPRAGE (magnetization prepared rapid gradient echo) anatomical scan was acquired with the following parameters: repetition time (TR) = 2300 ms; echo time (TE) = 2.36 ms; flip angle (FA) = 8°; field of view (FOV) = 250 × 250 × 225 mm; acquisition matrix = 288 × 288 × 208; slice orientation = sagittal; slice thickness = 0.9 mm; number of slices = 208; voxel size = 0.87 × 0.68 × 0.9 mm. Whole brain, blood oxygen level dependent (BOLD) fMRI was conducted using EPI with the following parameters: TR = 761 ms; TE = 24 ms; FA = 53°; FOV = 204 × 204 mm; acquisition matrix = 68 × 68; number of slices = 45; slice thickness = 3 mm; voxel size = 3 × 3 × 3 mm. Field map imaging was performed with a double-echo spoiled gradient echo sequence, generating one magnitude and two phase images [TR = 520 ms; TE = 4.92/7.38 ms; FA = 60°; total EPI readout time = 16.415 ms; blip direction = 1]. Participants underwent a 12 min resting state scan (this data is reported elsewhere) immediately prior to the current study's task-based scan. Before commencing the task-based session, participants were removed from the scanner so that the ear plugs they wore during the resting state scan could be removed and so that the active noise-canceling headphones could be refitted and recalibrated to ensure the quality of the signal-to-noise ratio.

Pre-processing

The CONN toolbox v.20.b (Whitfield-Gabrieli and Nieto-Castanon, 2012, 2017; RRID:SCR_009550)[2], which is powered by SPM12 (Statistical Parametric Mapping v. 12, Wellcome Department of Imaging Neuroscience, University College London, United Kingdom), was used to pre-process the data in MATLAB R2020B software. We used the default pre-processing pipeline for volume-based analyses but with indirect normalization to standard stereotactic (MNI) space as we had obtained gradient field maps during image acquisition (Nieto-Castanon, 2020). The pipeline consisted of the following five steps and parameters:

(1) Functional realignment and unwarp with the use of fieldmaps for susceptibility distortion correction.
(2) Slice-timing correction wherein the predefined "Siemens interleaved" acquisition sequence was selected from the CONN toolbox user interface.
(3) Outlier identification in which framewise displacements greater than 0.9 mm or global BOLD signal changes above five SD were flagged as potential outliers.
(4) Indirect segmentation and normalization in which the functional data was first co-registered to the anatomical data using SPM12 inter-modality co-registration procedure with a normalized mutual cost function (Collignon et al., 1995; Studholme et al., 1998). Second, the anatomical data was normalized into standard MNI-space and segmented into gray matter, cerebrospinal fluid, and white matter tissue classes using SPM12 unified segmentation and normalization procedure (Ashburner and Friston, 2005) with the T1-weighted volume as a reference image. Third, the same estimates of the deformation field from the unified segmentation and normalization procedure on the anatomical data was applied to the functional data. CONN toolbox's default probability tissue maps were selected: 180 × 216 × 180 mm bounding box with 1 and 2 mm isotropic voxels for the anatomical and functional data respectively.
(5) The functional data was smoothed using spatial convolution with CONN toolbox's default Gaussian kernel recommendation of 8 mm FWHM (full width half maximum).

Regions of Interest Definitions

Auditory Regions of Interests

We selected four anatomical regions of interests (ROIs) within each hemisphere that have been consistently associated with the early detection of changes in frequency/pitch; these included Heschl's gyrus (HG) of the auditory core; the planum polare (PP) and planum temporale (PT) of the auditory belt; and the posterior insula (Bamiou et al., 2003; Moerel et al., 2014). The Harvard-Oxford structural atlas implemented in the CONN toolbox was

---

[2] www.nitrc.org/projects/conn

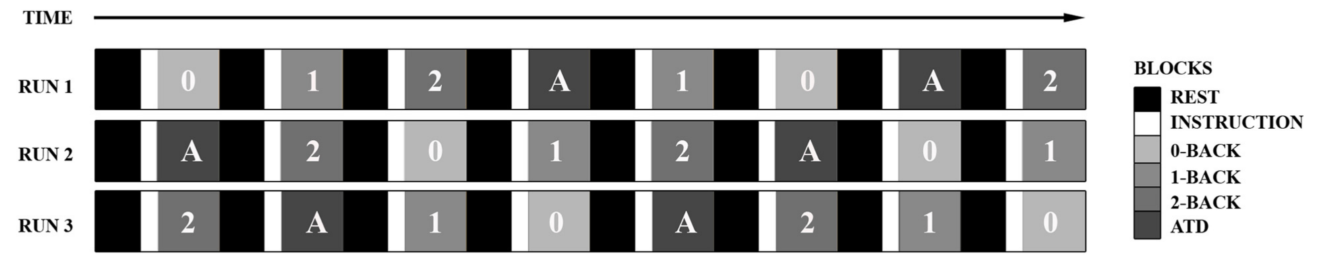


FIGURE 1 | The scanning protocol utilized a block design. Each experimental block (30 TR) was proceeded by a resting baseline condition (20 TR) and every experimental block was initiated by task instruction (7 TR). The figure depicts the presentation order of the experimental conditions which were partially counterbalanced over the entire experiment and divided into three runs of eight task blocks (3 × 456 TR). Each experimental condition was presented twice per run.

used to define the HG, PP, and PT ROIs. Because the insula is a relatively large structure, we used the SPM anatomy toolbox 3.0 (Eickhoff et al., 2005, 2006, 2007) to generate two posterior histological subdivisions: a dysgranular ROI (dpI) just posterior to the central sulcus comprising of Id2 and Id5; and a granular posterior ROI (gpI) comprising of Ig1, Ig2, and Ig3. In total, we investigated 10 auditory ROIs (five in each hemisphere).

Salience Network Regions of Interests

The CONN-toolbox's network atlas was used to select SN-hub ROIs with predefined MNI centroids (x, y, z) consisting of the: dorsal anterior cingulate (dACC; 0, 22, 35); left (−44, 13, 1) and right (47, 14, 0) anterior insula (aI); left (−32, 45, 27) and right (32, 46, 27) rostral prefrontal cortex (rPFC); and the left (−60, −39, 31) and right (62, −35, 32) supramarginal gyrus (SMG).

Statistical Analysis

Group Descriptives and Behavioral Analysis

One-way ANOVAs were used to test differences in means between groups, except when the assumption of equal variances was violated in which a Welch-test was applied. A response on one of the ASRS items was missing for one of the control participants on the hyperactivity/impulsivity scale, so this participant was excluded from the F-test of group differences in hyperactivity/impulsivity. Also, one of the ADHD-participants' behavioral results from the scanner task was lost due to a technical error. Statistical comparisons of fMRI task performance (accuracy, RT, and false-alarms) were therefore conducted on 16 of the 17 ADHD participants. The RTs we report are based on accurate trials only.

Univariate Activation Analysis

The univariate activation analysis was conducted in the SPM12 analysis software. The pre-processed functional data for each participant was entered into a general linear model (GLM) that included for each run; five regressors representing the four experimental conditions and the resting baseline condition. Six motion parameters (obtained from the realignment procedure during pre-processing) were included as covariates of no interest. The model additionally included three regressors representing the mean signal across the three runs. SPM12's default high-pass filter of 128 s was applied prior to parameter estimation to control for low-frequency signal confounds. Contrast estimates for the effects of ATD (ATD–rest), and the change in auditory activity from attending to the auditory modality and attending to the visual modality for each WM load condition (0-back–ATD, 1-back–ATD, and 2-back–ATD) were extracted for each participant by way of the SPM-compatible REX (ROI Extraction) tool[3]; and then used for group-level ROI analyses in IBM SPSS v.27 statistical analysis software. Because we did not have specific hypotheses of interactions between hemispheres, auditory ROI analyses were performed per hemisphere, which also allowed us to preserve statistical power. Differences between hemispheres are thus only qualitatively assessed in the results.

Independent t-tests were used to explore group-level SN, and auditory ROI BOLD activation for the ATD task and resulting p-values were Bonferroni corrected for multiple comparisons. Two-tailed Spearman's correlation analysis was also performed to determine if performance on the ATD task correlated with activity in any of the auditory/SN ROIs. For the visual WM task, a mixed repeated measures ANOVA (one per hemisphere) was used primarily to analyze the main effect of WM load and the WM load × Group interaction; and also, to determine whether there was a linear relationship between auditory ROI attenuation and visual WM load. Greenhouse-Geisser correction of degrees of freedom was applied when the sphericity assumption was violated, and p-values from relevant *post hoc* tests were Bonferroni corrected. Two-tailed Spearman's correlation analysis was used to test the hypothesis that WM capacity would positively correlate with auditory attenuation under high cognitive load. We additionally explored whether the strength of auditory attenuation in the high-load condition was correlated with improved task performance and individual differences in self-rated inattentiveness (derived from the ASRS, see **Table 1**).

Although not associated with the main hypotheses of this study, for completeness we additionally conducted a voxel-wise whole brain analysis. Group level, one-sample (within-groups) and two-sample (between-groups) t-tests were performed on the four contrasts of interest of which both ADHD > Controls and Controls > ADHD comparisons were explored. Resulting statistical maps were family wise error (FWE) corrected using $p < 0.05$ with a minimum extent threshold of $k = 10$ voxels.

---

[3] https://www.nitrc.org/projects/rex

TABLE 1 | Group descriptives for education, hearing acuity and neuropsychological assessments of ADHD-related symptomology.

	Controls (N = 17)	ADHD (N = 17)
Current/highest education level	***N***	***N***
Upper secondary	1	4
Undergraduate	14	13
Post-graduate	2	–
Hearing acuity (dB)	***M* (SD)**	***M* (SD)**
Pure-tone average	−2.5 (4.1)	−0.3 (7.4)
d2-Test Attention (standard score)	***M* (SD)**	***M* (SD)**
Concentration	103 (7.1)	100 (9.0)
Processing speed	103 (15.6)	106 (15.2)
Precision	100 (9.9)	97 (11.6)
ASRS v.1.1 (aggregate score)	***M* (SD)**	***M* (SD)**
Inattention	12.8 (4.7)	27.5 (3.9)
Hyperactivity/Impulsivity	11.9[†] (5.7)	23.0 (8.1)
Total ASRS	24.6[†] (9.2)	50.5 (11.5)
MINI 7.0.1 for ADHD studies	***N***	***N***
ADHD: Inattentive	1	6
ADHD: Impulsive/Hyperactive	–	–
ADHD: Combined	–	11

[†] *Mean derived from 16 out of the 17 Controls as item 8 (Part B) was missing for one of the participants.*

Functional Connectivity Analysis

An additional denoising procedure was performed on the pre-processed data) using the CONN-toolbox's default denoising pipeline (Whitfield-Gabrieli and Nieto-Castanon, 2012; Nieto-Castanon, 2020). The pipeline consisted of the following two steps:

(1) Nuisance covariates derived from CONN-toolbox's implementation of anatomical component-based correction (aCompCor) were entered into an ordinary least squares regression in order to remove confounding effects on the estimated BOLD signal in each voxel per subject and run. The covariates included five noise components from cerebral white matter; five noise components from cerebrospinal areas; 12 subject motion components (three translation, three rotation and their first-order temporal derivatives), outlier scans identified in the pre-processing procedure (see section "Pre-processing," step 3) and components representing the effect of each task-condition convolved with the canonical hemodynamic response function in order to reduce the influence of slow trends, initial magnetization transients as well as constant task-related effects.

(2) Temporal band pass filtering (high pass: 0.008 Hz, low pass: 0.09 Hz) on the BOLD signal was applied in order to minimize the influence of physiological head motion and other noise sources.

First, and second level functional connectivity analysis was further conducted in the CONN-toolbox. Each participant's denoised, voxel-wise BOLD time series data (concatenated over runs) was averaged within each auditory, and SN ROI per experimental condition. Then, for each task condition, first-level HRF-weighted ROI–ROI connectivity analyses were performed in which the correlation coefficient of each ROI to all other ROIs was calculated. Resulting correlation coefficients were Fisher z-transformed. Each participant's ROI-ROI connectivity matrices were then entered into a second level GLM to obtain group-level estimates for the ATD condition. Univariate parametric statistics were used to perform connection-based, between-group inferences (ADHD > Controls) across all pairs of ROIs. The connection-level significance threshold of $p < 0.05$ was conservatively corrected by way of FDR-correction (false discovery rate).

RESULTS

Group Descriptives

Table 1 presents group descriptives for education, hearing acuity, MINI, ASRS and d2-R measures. The majority of participants (both ADHD and Controls) were studying (or had completed studies) at an undergraduate level of education. There were no differences between groups in age $F(1, 33) = 0.35$, $p = 0.557$; or hearing acuity, *Welch's* $F(1, 25.3) = 1.4$, $p = 0.252$. The majority of individuals in the control group reported having relatively infrequent difficulties with inattention and hyperactivity/impulsivity as measured by the ASRS and MINI. Although one control participant was categorized as inattentive on the MINI, this participant only scored 17 for inattentiveness on the ASRS which is well under the cut-off score (24) for diagnostic evaluation. However, excluding this participant from the analysis did not change the pattern of results. The reason for this discrepancy between scales may be because the MINI requires a forced choice, yes/no answer for each item, whereas the ASRS is more nuanced and allows the participant to reflect and grade in more detail the frequency of symptoms. The majority of ADHD participants (11) had the combined subtype, the remaining six had mainly inattentive subtype according to MINI. In addition, our sample of aADHD had, as expected, significantly higher ASRS scores on both inattentive $F(1, 32) = 97.1$, $p < 0.000$, $\eta^2 = 0.75$; and hyperactivity/impulsivity $F(1, 31) = 20.6$, $p < 0.000$, $\eta^2 = 0.40$; subscales compared to Controls. The ADHD group did not however demonstrate significantly poorer performance than Controls on the d2-Test measures of concentration *Welch's* $F(1, 30.4) = 1.3$, $p = 0.262$; processing speed $F(1, 32) = 0.31$, $p = 0.581$; or precision $F(1, 32) = 0.63$, $p = 0.432$.

Behavioral Results

Statistical details of group differences in WM capacity and in-scanner performance measures are reported in **Table 2**. In measures of WM capacity, our sample of aADHD performed on par with Controls. In the n-back task, no differences in mean RTs,

TABLE 2 | Descriptives and one-way ANOVA results (Welch test was applied when the assumption of homogeneity was violated) for group comparisons of working memory (WM) capacity and in-scanner task performance on the *n*-back and auditory target detection (ATD) conditions. RT: response times in milliseconds and correspond to accurate trials only. Due to a technical error, fMRI-task performance data was missing for one ADHD participant.

		Controls *M* (SD)	**ADHD *M* (SD)**			
WM capacity		***N* = 17**	***N* = 17**	***F***	***df***	***p***
Reading span		0.51 (0.12)	0.46 (0.16)	1.29	1, 32	0.265
Size comparison span		0.71 (0.12)	0.63 (0.23)	1.35	1, 24.8	0.256
fMRI-task performance		***N* = 17**	***N* = 16**	***F***	***df***	***p***
RT	0-back	375 (38)	390 (57)	0.71	1, 31	0.405
	1-back	462 (83)	503 (71)	2.28	1, 31	0.141
	2-back	525 (119)	534 (105)	0.06	1, 31	0.814
	ATD	283 (31)	296 (44)	0.87	1, 31	0.358
Accuracy	0-back	0.99 (0.02)	0.95 (0.13)	1.57	1, 31	0.219
	1-back	0.93 (0.10)	0.86 (0.17)	2.33	1, 31	0.137
	2-back	0.89 (0.06)	0.74 (0.16)	13.45	1, 31	0.001
	ATD	0.87 (0.07)	0.84 (0.08)	0.91	1, 31	0.349
False alarms	0-back	0.2 (0.6)	0.6 (0.8)	2.62	1, 31	0.116
	1-back	1.2 (1.8)	2.7 (2.8)	3.36	1, 31	0.077
	2-back	3.2 (2.0)	4.3 (3.3)	1.32	1, 31	0.259
	ATD	3.1 (1.8)	3.6 (2.0)	0.91	1, 31	0.349

or the number of false alarms were observed between groups. Group means in accuracy performance were slightly lower in the ADHD group compared to Controls, but the difference in means was only significant in the highest load (2-back) condition ($\eta^2 = 0.303$). We were surprised to see relatively low accuracy scores on the ATD task from both groups. Closer inspection of the data revealed that the majority of errors were associated with false alarms and occurring milliseconds–seconds prior to the onset of the oddball. Thus, the low accuracy scores in the ATD task was likely an outcome of participants responding impulsively to the anticipation of an oddball (see section "Auditory Target Detection in Noise" in the Discussion).

Whole Brain Analysis

Two-sample *t*-tests of whole brain voxel-wise activity did not detect significant differences between groups in any contrasted condition after controlling for multiple comparisons. Within-group analyses however did reveal significant cluster-level activations in all contrast conditions. Cluster-level statistics and slice-by-slice maps of significant effects for within-groups analyses are reported in **Supplementary Table 1** and **Supplementary Figures 1–4**. For the ATD–rest contrast, activity was observed in auditory, superior temporal and supplementary motor cortices within both groups. For the *n*-back–ATD contrasts, the number of active regions increased with increasing WM load and was observed mainly in higher order visual processing regions, the exterior cerebellum as well as superior partial and middle frontal gyri in both groups.

Auditory Target Detection

Figure 2 presents the detailed results of the univariate ROI analysis for the ATD task (see also **Supplementary Table 2**). Significant increases relative to the resting baseline for the ATD condition were observed in all SN-hubs in Controls, and all except the dACC and left rPFC in ADHD when correcting for multiple comparisons (i.e., all seven hub ROIs) but uncorrected *p*-values were significant. Between-group comparisons revealed no significant differences for the SN-rest contrast suggesting that all seven SN-hubs were on average, more active relative to baseline in both groups. Initial analysis of the auditory ROIS indicated that only the core, and belt ROIs were significantly more activated than baseline during target detection. Hence, contrary to expectations from previous literature, our posterior insula ROIs were not actively involved in the ATD task; and this was evident in both groups. However, we considered that this may be a result of averaging over relatively large regions of the posterior insula and that a within ROI cluster analysis may prove more informative. To investigate, we used the SPM wfu_pickatlas tool to perform cluster analyses. Results identified significantly active clusters of 26 voxels (*p* FWE-corr = 0.024) in the left dpI (peak coordinates: −46, −8, 6) and eight voxels (*p* FWE-corr = 0.035) in the gpI (peak coordinates: −34, −30, 8) in the ADHD group. In the Control group, a significant cluster of 38 voxels (*p* FWE-corr = 0.011) was identified in the left dpI (peak coordinates: −40, −2, 2). Thus, in line with previous studies implicating the posterior insula in early pitch discrimination tasks, our data does indicate that at least the left posterior insula was involved in our ATD task. Statistical between-group comparisons did not reveal any differences in auditory cortical activation, indicating that all auditory ROIs were similarly activated in both groups throughout the ATD task.

Functional connectivity analysis revealed group differences in ROI–ROI connectivity for the ATD task. ADHD participants demonstrated stronger auditory-SN connectivity than Controls of which the majority of connections involved left lateralized auditory communications with dACC, aI and SMG hubs of the SN (see **Figure 3** for details). We additionally investigated if individual differences in behavioral performance on the ATD

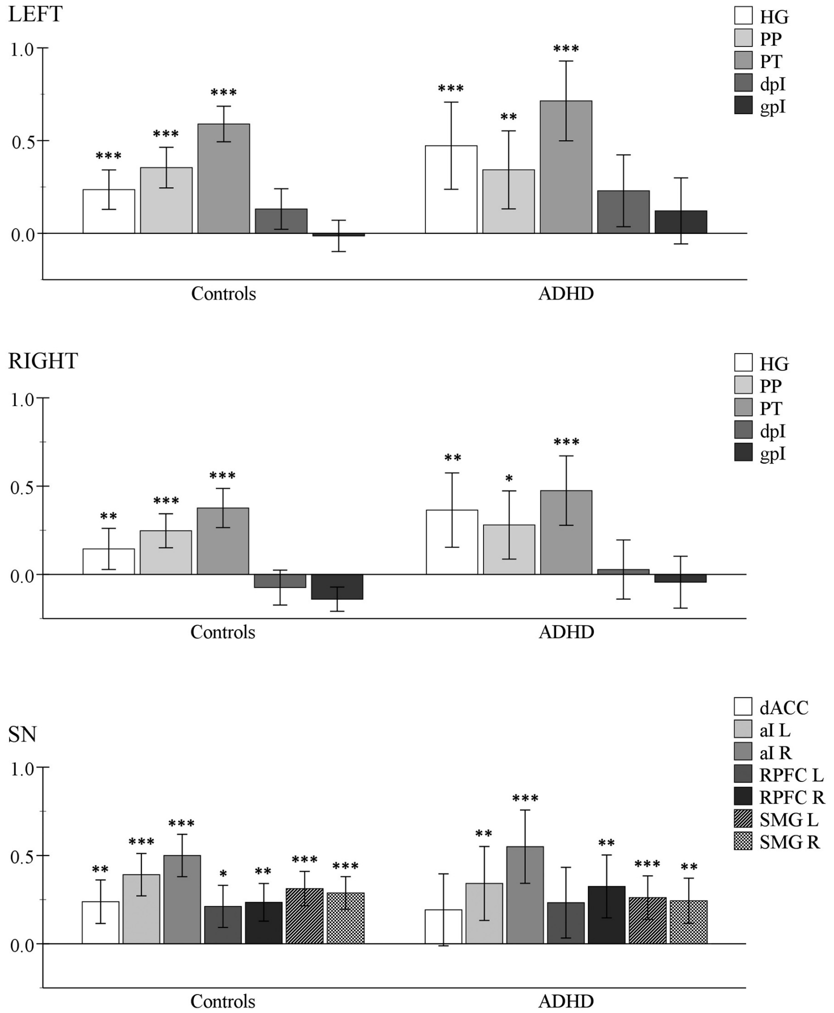


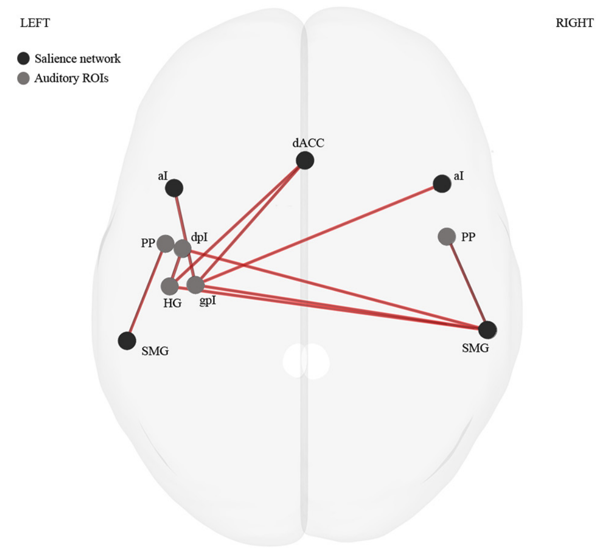
FIGURE 2 | Within-group analysis results for the auditory target detection contrast (ATD > rest). One-sample *t*-tests were used to determine if the mean BOLD activity within each ROI (per group) was greater than zero. *Y*-axis corresponds to mean beta image values. Asterisks indicate Bonferroni-corrected levels of significance: *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$; error bars = 95% CI. Top panel: left auditory ROIs. Middle panel: right auditory ROIs. Bottom panel: salient network (SN) hub-regions. HG, Heschl's gyrus; PP, planum polare, PT, planum temporale, dpI, dysgranular posterior insula; gpI, granular posterior insula; dACC, dorsal anterior cingulate cortex; aI, anterior insula; RPFC, rostral prefrontal cortex; SMG, supramarginal gyrus; L, left; R, right.

task would correlate with activity levels in any of the auditory or SN ROIs. Here we observed a negative correlation between RTs associated with accurate trials and increases in BOLD activity the right SMG, $r = -0.39$, $p = 0.027$. Hence, faster RTs were associated with heightened activity of the right SMG. No other significant brain–performance correlations were observed.

Effect of Working Memory Load on Auditory Attenuation

Results of the mixed repeated-measures ANOVAs are displayed in **Figure 4**. Both hemispheres presented evidence for a main effect of auditory attenuation as a function of WM load: F_{left} (2, 64) = 34.6, p = 0.000; F_{right} (2, 64) = 29.3, p = 0.000); and within-participant contrasts confirmed that the effect was linear: F_{left} (1, 32) = 30.9, p = 0.000; F_{right} (1, 32) = 45.6, p = 0.000). The interaction term, WM load × Group, was also significant in both hemispheres: F_{left} (2, 64) = 8.8, p = 0.000; F_{right} (2, 64) = 6.2, p = 0.003, and *post hoc* inspection of the interaction indicated that the Control group's auditory responses to the task-irrelevant sound source decreased significantly as WM load increased, and was more suppressed compared to ADHD participants in both hemispheres during the high-load condition (see **Figure 4A** and **Supplementary Table 3** for *post hoc* results). Auditory attenuation increased marginally in the ADHD group across load conditions and was shown to be statistically significant only in the right hemisphere between the 0- and 2-back conditions.

A significant, main effect of ROI was also observed in both hemispheres: F_{left} (2.5, 78.9) = 13.2, p = 0.000; F_{right} (2.1, 67.7) = 14.5, p = 0.000).; and the interaction term ROIs × Group was consequently significant: F_{left} (4.128) = 5.6, p = 0.000; F_{right} (4.128) = 3.7, p = 0.007. The interactions:



ADHD > Controls	t (32)	p FDR-corr
Auditory – SN		
dACC – HG$_{\text{left}}$	3.5	.028
dACC – gpI$_{\text{left}}$	3.7	.028
aI$_{\text{left}}$ – gpI$_{\text{left}}$	3.4	.028
aI$_{\text{right}}$ – gpI$_{\text{left}}$	3.5	.028
SMG$_{\text{left}}$ – PP$_{\text{left}}$	3.8	.028
SMG$_{\text{right}}$ – PP$_{\text{right}}$	3.7	.028
SMG$_{\text{right}}$ – HG$_{\text{left}}$	3.4	.028
SMG$_{\text{right}}$ – dpI$_{\text{left}}$	3.7	.028
SMG$_{\text{right}}$ – gpI$_{\text{left}}$	3.3	.030
Auditory – Auditory		
HG$_{\text{left}}$ – dpI$_{\text{left}}$	3.5	.028

FIGURE 3 | Superior glass-brain perspective and tabulated statistics of ROI–ROI connections that were significantly stronger ($p_{FDR-corr} < 0.05$) for the ADHD group in the ATD task. HG, Heschl's gyrus; PP, planum polare, dpI, dysgranular posterior insula; gpI, granular posterior insula; dACC, dorsal anterior cingulate cortex; aI, anterior insula; SMG, supramarginal gyrus.

WM load × ROI, and WM load × ROI × Group were not significant in either hemisphere, nonetheless *post hoc* analysis of the WM load × ROI × Group interaction was conducted as we consider it necessary to report the specific ROIs where group differences in attenuation were observed. In both hemispheres, attenuation was significantly less evident in the PP and the dpI for the ADHD group compared to Controls under the high-load condition (see **Figure 4B** and **Supplementary Table 3** for detailed results). Trends in the remaining ROIs were observable but the *post hoc* test did not reach the conservatively corrected threshold for significance.

We correlated the mean change in auditory attenuation under the high-load condition within each ROI with our WM capacity measures: Reading span and SIC-span. Contrary to expectations, we did not observe a significant relationship with WM capacity in any auditory ROIs. However, the magnitude of auditory attenuation in the high-load condition did correlate significantly with individual differences in self-rated inattentiveness across the majority of auditory ROIs (see **Figure 5A** for details) which suggests that participants who have difficulties suppressing exogenously triggered responses to task-irrelevant auditory stimulation are also highly susceptible to distraction. In addition, accuracy performance on the 2-back condition also correlated with attenuation levels in the majority of auditory ROIs (see **Figure 5B**). No other performance related correlations were observed. Taken together, these findings support the hypothesis that heightened cortical responses to task-irrelevant auditory stimulation can both negatively impact task performance and contribute to subjective experiences of auditory distraction.

DISCUSSION

The main purpose of the current study was to investigate how attentional engagement impacts exogenous cortical responses to acoustic stimulation, and whether adults with ADHD (aADHD) have difficulties hampering auditory distraction. As far as we can determine, we are the first fMRI study to demonstrate that aADHD have difficulty hampering early auditory cortical responses to task-irrelevant sound when required to focus on a cognitively demanding task. When participants' attention was focused on the auditory modality, auditory cortical activity was enhanced relative to a resting baseline; and when attention was directed away from the auditory modality and toward a visual WM task, auditory processing was attenuated. The degree of attenuation in auditory regions was relative to cognitive load demands, and by extension, endogenous attentional engagement toward the visual task. Importantly, the relationship between attentional engagement and auditory attenuation proved less efficient in aADHD than our matched sample (age and gender) of TDa. In particular, aADHD were had heightened cortical responses to task-irrelevant auditory stimulation and poorer performance capacity in the most challenging WM condition. In addition, although aADHD performed on par with TDa in the ATD task, functional communications between the SN and auditory ROIs were stronger in aADHD. Taken together, the results indicate that aADHD are more susceptible to noise interference when they are engaged in a primary task. How well they cope with noise interference appears to be related to the WM capacity demands of the task. We discuss these results in more detail in the proceeding subsections.

Working Memory Capacity and Auditory Distraction

We were not expecting aADHD to perform on par with Controls in measures of WM capacity. In a previous study performed by our lab group (Blomberg et al., 2019), adolescents (<18 years) with ADHD demonstrated significantly reduced capacity than matched (age and gender) healthy controls on both the Reading span and the Size-comparison span tasks. This finding, in combination with general associations of ADHD with deficient WM capacity (Willcutt et al., 2005; Alderson et al., 2013) was a major premise for using these measures again in our adult sample. Although ADHD is known to persist into adulthood in over half of patients, numerous researchers have noted that many neurocognitive deficits normalize after the developmental transition into adulthood and suggest that ADHD may mainly be attributable to a developmental delay

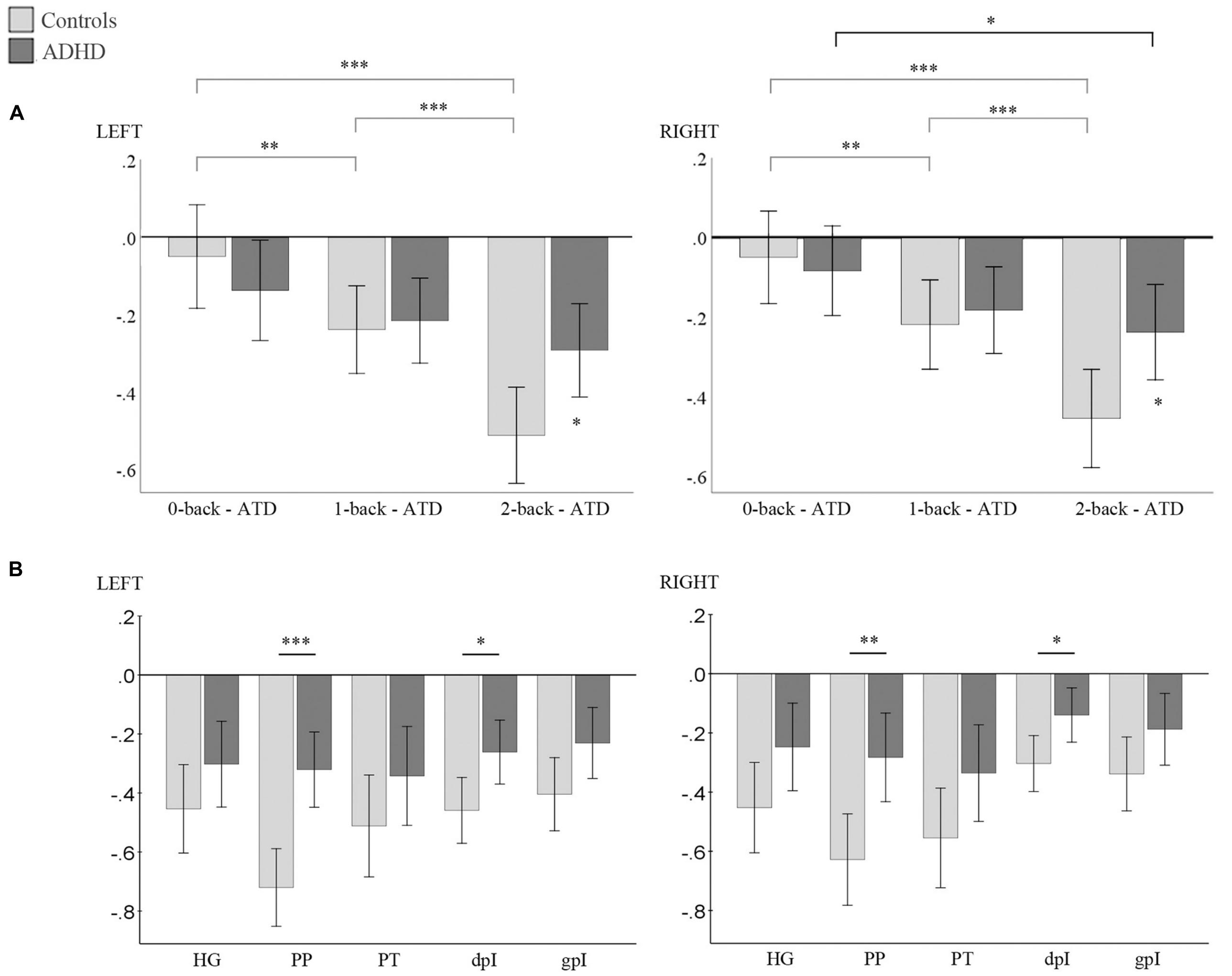


FIGURE 4 | (A) Results of *post hoc* analysis for the WM load × Group interaction. Within group differences indicated by horizontal lines above graph. Between group differences indicated by asterisk above error bars. **(B)** Results of *post hoc* analysis for the ROI × Group × High load interaction. Between-group differences indicated by horizontal lines above graph. Left panel: left hemisphere; right panel: right hemisphere. Bonferroni correct *p*-values: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$; error bars = 95% CI. ATD = auditory target detection. *y*-axis = estimated marginal means. HG, Heschl's gyrus; PP, planum polare, PT, planum temporale, dpI, dysgranular posterior insula; gpI, granular posterior insula.

(Shaw et al., 2007; de Zeeuw and Durston, 2017). A related, yet different perspective, associates ADHD with life-time subcortical (basal ganglia and cerebellum) dysfunction, and that age-related reductions in neurocognitive deficits in some cases are attributable to a healthy maturation of frontal lobes and the improved ability to issue cognitive control over subcortical systems (Halperin and Schulz, 2006). Presumably, such developmental differences between groups were exemplified through our WM capacity measures in our previous study with adolescents (Blomberg et al., 2019). We similarly note that concentration scores from the d2-Test of attention were also differential between groups in our adolescent study, but not in the current study. Because our ADHD participants consisted mostly of young adults undergoing higher education and performed on par with Controls in WM and the d2-Test, it is tempting to consider our sample as relatively "high functioning"–a notion that fits well with the aforementioned developmental models of ADHD. However, we must take into consideration that our ADHD group had poorer accuracy performance in the demanding, in-scanner WM condition (2-back). Hence, there is some evidence to suggest that some WM related difficulties reside in our sample of young adults.

Possibly, the neurocognitive profile of ADHD in adults may be more readily observable under challenging experimental contexts. Our sample of aADHD may well have been able to maintain a commensurate number of items in WM to that of TDa in a quiet, isolated room when undergoing the Reading span and Size-comparison span tests; but during the scanning session, the addition of irrelevant sound stimuli and background noise whilst performing the *n*-back task likely rendered the ADHD group vulnerable to cognitive interference. The latter conclusion

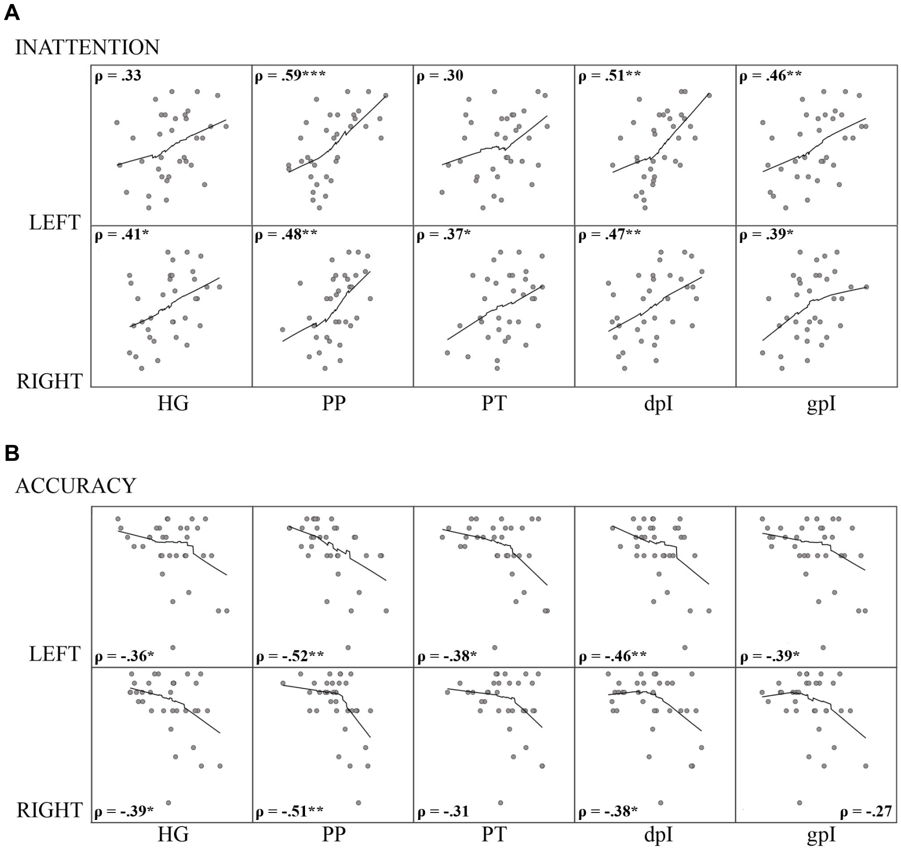


FIGURE 5 | Results of auditory ROI Spearman's rho (ρ) correlation analysis between the mean change in attenuation under high visual working memory load (2-back–ATD) and **(A)** self-rated inattention; and **(B)** 2-back task accuracy performance. Upper and lower rows of scatter plots correspond to left and right hemispheres, respectively. Asterisks indicate the correlation was significant (two-tailed) at thresholds: $^*p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$. Loess trend lines were approximated from a data span of 0.75 with a gaussian kernel. HG, Heschl's gyrus; PP, planum polare, PT, planum temporale, dpI, dysgranular posterior insula; gpI, granular posterior insula.

is consistent with two other studies in aADHD. Pelletier et al. (2016) reported that visual serial recall performance in their sample of aADHD was more strongly affected by the presence of irrelevant sound; and Dige et al. (2010) observed greater noise interference effects upon verbal memory in a dichotic memory task relative to controls.

We had hypothesized that the additional demands placed by the acoustic environment in the scanner would impact cognitive processing, particularly under the high-load condition, given that control resources needed to be shared between both regulating attention and WM toward one modality whilst suppressing exogenously evoked distractions in another. On this basis, we also expected that participants with high WM capacity would demonstrate more efficient neural attenuation of the acoustic environment and perform overall better on the 2-back condition, a hypothesis supported in previous work by Sörqvist et al. (2012) from our work group (see Sörqvist and Rönnberg, 2014 for a review). This was not supported in our results. We note that in Sörqvist et al. (2012), the authors derived a composite measure of WM capacity from several complex span tasks and the composite score only correlated with auditory brainstem attenuation to task-irrelevant sound in their 3-back condition (i.e., not 2-back). It is thus possible that the WM capacity measures we used in the current study were not sufficiently sensitive indexes of the type of control mechanisms involved in the successful regulation of attention in our 2-back condition.

We did however observe a relatively robust negative correlation between individual differences in auditory attenuation of task-irrelevant sound and 2-back task accuracy. This result provides us with an indication that better cognitive performance on the high-load WM task was intricately related to participants' capacity to regulate resources between modalities and inhibit distraction. In additional support of this conclusion, participants' subjective experiences of inattentiveness in daily life (as determined by the ASRS) also robustly correlated with the degree of auditory cortical attenuation under the high-load condition. Our participants with ADHD, that consisted of inattentive or combined (i.e., both inattentive and hyperactive/impulsive) subtypes, were thus more susceptible to auditory distraction and demonstrated poorer performance capacity under the most challenging WM condition.

Auditory Target Detection in Noise

The ATD task, in which an infrequent "oddball" tone was to be detected amongst a rapid stream of standard tones, places relatively little demand upon cognitive control systems when performed in quiet and should result in high accuracy scores–especially in TDa with normal hearing thresholds. Despite our use of an active noise canceling system which enabled us to present the acoustic stimulus at an audible signal-to-noise ratio of ∼17 dB SPL, behavioral responses from both groups contained a relatively high number of false alarms that occurred just milliseconds-to-seconds prior to the onset of a target. This result suggests that participants were likely responding impulsively to the anticipation of an oddball. The presentation times of oddball-targets were randomly distributed over the 22 s duration of each task-block (four targets per block), so anticipating the exact timing of a target was not possible; however, the audibility of background scanner noise may have been interfering with expectations of an oddball occurring, resulting in impulsive false-alarm responding.

We expected the SN to be heavily involved in the ATD task due to its involvement in vigilant anticipation, detection, and response-mediation of behaviorally salient stimuli; and we explored this network on the grounds that a number of studies have reported aberrant SN functional connectivity in ADHD (see Castellanos and Aoki, 2016 for a review). Significant hyper auditory–SN connectivity was observed in aADHD. Although our test protocol makes it difficult to discern the exact reason for these stronger SN–auditory interactions in the ADHD group, the explanation we consider aligns well with our more general thesis that noise places increased demands on control systems. Given their reported difficulties with inattentiveness, the suboptimal listening situation may have impacted ADHD participants such that the informational exchange between auditory ROIs and SN-hubs was enhanced to facilitate successful oddball detection. As we observed in the 2-back task, heightened auditory responses to the acoustic environment were associated with poorer accuracy performance and symptomatic inattentiveness. And we argued that the challenging 2-back condition taxed cognitive resources and impeded on ADHD participants' ability to suppress exogenous responses to the acoustic environment. Even though we observed differences in functional connectivity between groups, there were no differences in performance outcomes on the ATD task. Because we do not consider the ATD task cognitively challenging, it is likely our sample of aADHD had the resources available to enhance oddball detection and reduce erroneous responding under suboptimal listening conditions; hence compensating for their presumed symptomatic susceptibility to noise interference.

Increased activity in the right SMG of the SN was also negatively correlated with individual differences in RTs on the ATD task, of which the RTs corresponded only to accurate trials. Hence, faster, accurate RTs were associated with increased activity in the right SMG. The SMG is a core hub of the SN, and the right lateralized region in particular has been implicated in the mediation of exogenous attention to behaviorally salient events (Corbetta and Shulman, 2002; Corbetta et al., 2008; Cabeza et al., 2012; Vossel et al., 2014). Given this proposed functional role of the right SMG in sensory target detection, our combined results suggests that rapid and successful detection of the oddball was associated with effective assignment of saliency to the oddball tone mediated by the right SMG.

Clinical Implications

With respect to our results, it appears that "high-functioning" aADHD perform relatively well during cognitive tasks but at a greater cost. This notion is in line with earlier literature indicating that a college student with ADHD may need to work twice as hard as their non-ADHD counterparts in order to achieve satisfactory grades (Faraone et al., 2015). Hence, in order to perform in the complex reality of a college or work environment, aADHD may need to deploy more cognitive effort in order to inhibit distraction. The implication of this increased effort is potentially a contributing factor to the reportedly higher levels of perceived stress in college students with ADHD symptoms and the prevalence of anxiety and stress-related pathologies in aADHD (Salla et al., 2019; Gbessemehlan et al., 2020; Öster et al., 2020). Given that some participants were unmedicated during this study, we suggest that more research is needed to understand how medication can compensate for these deficiencies. To prevent stress related problems later in life, clinical interventions should also address adapting work/study environments in order to minimize disturbance and utilize psychoeducation to better improve stress and time-management.

Limitations

This study has several limitations. First, comparisons of ROIs and networks are difficult across literature due to differences in nomenclature and methods of definition. Hence, our decision for using predefined ROIs and networks through freely available analysis applications, together with as many default analysis settings as possible, was in the hope of facilitating future researchers' ability to confer/replicate our results. Second, scanner noise was audible throughout the entire experiment. Even during the resting baseline condition. Although we were able to present the auditory stimulus at a SNR of ∼ 17 dB SPL, the active noise canceling system was not 100% stable in that the dampened background could fluctuate; and we speculate that the effect of background noise interfered with performance in both participant groups on the ATD task. Nonetheless, the addition of background noise was not a problematic influence with respect to our most important finding–that our sample of aADHD were more susceptible to auditory distraction when task demands taxed cognitive resources. Third, care must be taken when generalizing results. The manifestation of ADHD in adults is highly heterogenous and often confounded by comorbidities. Around half of our sample were on stable SSRI medication which is indicative of earlier problems with anxiety and depression however, both anxiety and depression reside at the lower spectrum of expected of psychiatric comorbidities in adults with ADHD (Katzman et al., 2017). Our sample was also imbalanced in male:female ratio, of which there were more females than males with ADHD. Although childhood ADHD is more common in boys, differences in prevalence between sexes diminish almost completely in adulthood (Faraone et al., 2015; Matte et al., 2015) and symptomatic differences in hyperactivity and inattention between sexes also wane with older age (Ramtekkar et al., 2010); so we should not expect the gender imbalance in our sample to dramatically affect more general conclusions of our results. That notwithstanding, our sample was small and relatively high-functioning and may not be entirely representative for the more severe spectrum of the disorder. If, however subjects with ADHD in general are expected to have more severe symptoms and functional impairment, group differences detected between our sample and controls are likely to underestimate the overall differences rendering our results conservative. With respect to these limitations, we have cautiously opted to confine discussion to our small sample rather than boldly extrapolate our findings to the population as a whole. That being said, our general findings contribute important evidence to extant theories of cognitive control, auditory distraction, and the pathophysiology of ADHD in adults.

CONCLUSION

To conclude, our sample of relatively high functioning young adults with ADHD were able to modulate auditory and SN systems in response to noise interference and perform at a commensurate level to controls when WM task demands were low. But when WM task demands were high, ADHD participants had difficulties attenuating task-irrelevant auditory cortical processing. Heightened auditory activity to task-irrelevant sound was associated with both poorer task performance and symptomatic inattentiveness. Our findings contribute to developmental models of persistent ADHD and more generally, WM capacity models of distraction; and demonstrate that an individual's ability to regulate attentional engagement and impede auditory distraction is intricately related to their capacity for cognitive control. The study also has important clinical implications for aADHD underscoring the need for early interventions to adapt study/work environments, develop effective coping strategies, and minimize risk for chronic stress and anxiety in this vulnerable group.

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 Regional Ethical Review Board in Linköping, Sweden (Dnr 2019-06158). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

JR, RB, AJC, and HD contributed to the conception and design of the study. AJC and RB were responsible for participant recruitment and data collection. RB was responsibility for analysis and manuscript drafting under the supervision of CS, HD, and JR. All authors scrutinized the statistical analysis, contributed to the manuscript's revision and approved submitted version.

FUNDING

The study was funded by the Swedish Research Council (2015-01917).

ACKNOWLEDGMENTS

We would like to extend a special thank you to Paul Hamilton and Robin Kämpe for the MRI expertise necessary for this study. We also thank Per Gustafsson for his collaborative recommendations and theoretical advice.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnhum.2021.771711/full#supplementary-material

REFERENCES

Alderson, R. M., Kasper, L. J., Hudec, K. L., and Patros, C. H. G. (2013). Attention-deficit/hyperactivity disorder (ADHD) and working memory in adults: a meta-analytic review. *Neuropsychology* 27, 287–302. doi: 10.1037/a0032371

Alho, K., Rinne, T., Herron, T. J., and Woods, D. L. (2014). Stimulus-dependent activations and attention-related modulations in the auditory cortex: a meta-analysis of fMRI studies. *Hear. Res.* 307, 29–41. doi: 10.1016/j.heares.2013.08.001

Alho, K., Salmi, J., Koistinen, S., Salonen, O., and Rinne, T. (2015). Top-down controlled and bottom-up triggered orienting of auditory attention to pitch activate overlapping brain networks. *Brain Res.* 1626, 136–145. doi: 10.1016/j.brainres.2014.12.050

American Speech-Language-Hearing Association (2005). *Guidelines for Manual Pure-Tone Threshold Audiometry [Guidelines]*. Rockville, MD: American Speech-Language-Hearing Association, doi: 10.1044/policy.GL2005-00014

Ashburner, J., and Friston, K. J. (2005). Unified segmentation. *NeuroImage* 26, 839–851. doi: 10.1016/j.neuroimage.2005.02.018

Badre, D. (2020). "Brain networks for cognitive control: four unresolved questions," in *Intrusive Thinking across Neuropsychiatric Disorders: From Molecules to Free Will.* Strûngmann forum reports, eds P. W. Kalivas and M. P. Paulus (Cambridge, MA: MIT press), 203–228.

Bamiou, D.-E., Musiek, F. E., and Luxon, L. M. (2003). The insula (Island of Reil) and its role in auditory processing: literature review. *Brain Res. Rev.* 42, 143–154. doi: 10.1016/S0165-0173(03)00172-3

Barkley, R. A. (1997). Behavioral inhibition, sustained attention, and executive functions: constructing a unifying theory of ADHD. *Psychol. Bull.* 121, 65–94.

Beaman, C. P. (2004). The irrelevant sound phenomenon revisited: what role for working memory capacity? *J. Exp. Psychol. Learn. Mem. Cogn.* 30, 1106–1118. doi: 10.1037/0278-7393.30.5.1106

Blenkmann, A. O., Collavini, S., Lubell, J., Llorens, A., Funderud, I., Ivanovic, J., et al. (2018). Auditory deviance detection in the human insula: an intracranial EEG study. *bioRxiv* [Preprint] doi: 10.1101/487306

Blomberg, R., Danielsson, H., Rudner, M., Söderlund, G. B. W., and Rönnberg, J. (2019). Speech processing difficulties in attention deficit hyperactivity disorder. *Front. Psychol.* 10:1536. doi: 10.3389/fpsyg.2019.01536

Brickenkamp, R., Schmidt-Atzert, L., and Liepmann, D. (2010). *d2-R*. Göttingen: Hogrefe.

Cabeza, R., Ciaramelli, E., and Moscovitch, M. (2012). Cognitive contributions of the ventral parietal cortex: an integrative theoretical account. *Trends Cogn. Sci.* 16, 338–352. doi: 10.1016/j.tics.2012.04.008

Castellanos, F. X., and Aoki, Y. (2016). Intrinsic functional connectivity in attention-deficit/hyperactivity disorder: a science in development. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* 1, 253–261. doi: 10.1016/j.bpsc.2016.03.004

Castellanos, F. X., and Tannock, R. (2002). Neuroscience of attention-deficit/hyperactivity disorder: the search for endophenotypes. *Nat. Rev. Neurosci.* 3, 617–628. doi: 10.1038/nrn896

Citherlet, D., Boucher, O., Tremblay, J., Robert, M., Gallagher, A., Bouthillier, A., et al. (2020). Spatiotemporal dynamics of auditory information processing in the insular cortex: an intracranial EEG study using an oddball paradigm. *Brain Struct. Funct.* 225, 1537–1559. doi: 10.1007/s00429-020-02072-z

Collignon, A., Maes, F., Delaere, D., Vandermeulen, D., Suetens, P., and Marchal, G. (1995). "Automated multi-modality image registration based on information theory," in *Proceedings of the 14th International Conference on Information Processing In Medical Imaging-IPMI95. Computational Imaging and Vision*, eds Y. Bizais, C. Barillot, and R. Di Paola (Ile de Berder: Kluwer Academic Publishers), 263–274.

Conway, A. R. A., Cowan, N., and Bunting, M. F. (2001). The cocktail party phenomenon revisited: the importance of working memory capacity. *Psychon. Bull. Rev.* 8, 331–335. doi: 10.3758/BF03196169

Corbetta, M., Patel, G., and Shulman, G. L. (2008). The reorienting system of the human brain: from environment to theory of mind. *Neuron* 58, 306–324. doi: 10.1016/j.neuron.2008.04.017

Corbetta, M., and Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nat. Rev. Neurosci.* 3, 201–215. doi: 10.1038/nrn755

Crottaz-Herbette, S., and Menon, V. (2006). Where and when the anterior cingulate cortex modulates attentional response: combined fMRI and ERP evidence. *J. Cogn. Neurosci.* 18, 766–780. doi: 10.1162/jocn.2006.18.5.766

de Zeeuw, P., and Durston, S. (2017). "Cognitive control in attention deficit hyperactivity disorder," in *The Wiley Handbook of Cognitive Control*, ed. T. Egner (Hoboken, NJ: John Wiley & Sons, Ltd), 602–618. doi: 10.1002/9781118920497.ch33

Dige, N., Maahr, E., and Backenroth-Ohsako, G. (2010). Reduced capacity in a dichotic memory test for adult patients with ADHD. *J. Atten. Disord.* 13, 677–683. doi: 10.1177/1087054709347245

Edwards, E., Soltani, M., Deouell, L. Y., Berger, M. S., and Knight, R. T. (2005). High gamma activity in response to deviant auditory stimuli recorded directly from human cortex. *J. Neurophysiol.* 94, 4269–4280. doi: 10.1152/jn.00324.2005

Egner, T. (ed.) (2017). *The Wiley Handbook of Cognitive Control | Wiley*. Hoboken, NJ: Wiley-Blackwell.

Eickhoff, S. B., Heim, S., Zilles, K., and Amunts, K. (2006). Testing anatomically specified hypotheses in functional imaging using cytoarchitectonic maps. *NeuroImage* 32, 570–582. doi: 10.1016/j.neuroimage.2006.04.204

Eickhoff, S. B., Paus, T., Caspers, S., Grosbras, M.-H., Evans, A. C., Zilles, K., et al. (2007). Assignment of functional activations to probabilistic cytoarchitectonic areas revisited. *NeuroImage* 36, 511–521. doi: 10.1016/j.neuroimage.2007.03.060

Eickhoff, S. B., Stephan, K. E., Mohlberg, H., Grefkes, C., Fink, G. R., Amunts, K., et al. (2005). A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *NeuroImage* 25, 1325–1335. doi: 10.1016/j.neuroimage.2004.12.034

El Karoui, I., King, J.-R., Sitt, J., Meyniel, F., Van Gaal, S., Hasboun, D., et al. (2015). Event-related potential, time-frequency, and functional connectivity facets of local and global auditory novelty processing: an intracranial study in humans. *Cereb. Cortex* 25, 4203–4212. doi: 10.1093/cercor/bhu143

Faraone, S. V., Asherson, P., Banaschewski, T., Biederman, J., Buitelaar, J. K., Ramos-Quiroga, J. A., et al. (2015). Attention-deficit/hyperactivity disorder. *Nat. Rev. Dis. Primer* 1:15020. doi: 10.1038/nrdp.2015.20

Faraone, S. V., Banaschewski, T., Coghill, D., Zheng, Y., Biederman, J., Bellgrove, M. A., et al. (2021). The world federation of ADHD international consensus statement: 208 evidence-based conclusions about the disorder. *Neurosci. Biobehav. Rev.* 128, 789–818. doi: 10.1016/j.neubiorev.2021.01.022

Garrido, M. I., Kilner, J. M., Stephan, K. E., and Friston, K. J. (2009). The mismatch negativity: a review of underlying mechanisms. *Clin. Neurophysiol.* 120, 453–463. doi: 10.1016/j.clinph.2008.11.029

Gbessemehlan, A., Arsandaux, J., Orri, M., Montagni, I., Macalli, M., Tournier, M., et al. (2020). Perceived stress partially accounts for the association between attention deficit hyperactivity disorder (ADHD) symptoms and suicidal ideation among students. *Psychiatry Res.* 291:113284. doi: 10.1016/j.psychres.2020.113284

Gisselgård, J., Petersson, K. M., Baddeley, A., and Ingvar, M. (2003). The irrelevant speech effect: a PET study. *Neuropsychologia* 41, 1899–1911. doi: 10.1016/S0028-3932(03)00122-2

Halperin, J. M., and Schulz, K. P. (2006). Revisiting the role of the prefrontal cortex in the pathophysiology of attention-deficit/hyperactivity disorder. *Psychol. Bull.* 132, 560–581. doi: 10.1037/0033-2909.132.4.560

Huang, S., Li, Y., Zhang, W., Zhang, B., Liu, X., Mo, L., et al. (2015). Multisensory competition is modulated by sensory pathway interactions with Fronto-sensorimotor and default-mode network regions. *J. Neurosci.* 35, 9064–9077. doi: 10.1523/JNEUROSCI.3760-14.2015

Hughes, R. W., Hurlstone, M. J., Marsh, J. E., Vachon, F., and Jones, D. M. (2013). Cognitive control of auditory distraction: Impact of task difficulty, foreknowledge, and working memory capacity supports duplex-mechanism account. *J. Exp. Psychol. Hum. Percept. Perform.* 39, 539–553. doi: 10.1037/a0029064

Katzman, M. A., Bilkey, T. S., Chokka, P. R., Fallu, A., and Klassen, L. J. (2017). Adult ADHD and comorbid disorders: clinical implications of a dimensional approach. *BMC Psychiatry* 17:302. doi: 10.1186/s12888-017-1463-3

Kessler, R. C., Adler, L., Ames, M., Demler, O., Faraone, S., Hiripi, E., et al. (2005). The World Health Organization adult ADHD self-report scale (ASRS): a short screening scale for use in the general population. *Psychol. Med.* 35, 245–256. doi: 10.1017/S0033291704002892

Kooij, J. J. S., Huss, M., Asherson, P., Akehurst, R., Beusterien, K., French, A., et al. (2012). Distinguishing comorbidity and successful management of adult ADHD. *J. Atten. Disord.* 16, 3S–19S. doi: 10.1177/1087054711435361

Körner, U., Röer, J. P., Buchner, A., and Bell, R. (2017). Working memory capacity is equally unrelated to auditory distraction by changing-state and deviant sounds. *J. Mem. Lang.* 96, 122–137. doi: 10.1016/j.jml.2017.05.005

Marsh, J. E., and Campbell, T. A. (2016). Processing complex sounds passing through the rostral brainstem: the new early filter model. *Front. Neurosci.* 10:136. doi: 10.3389/fnins.2016.00136

Matte, B., Anselmi, L., Salum, G. A., Kieling, C., Gonçalves, H., Menezes, A., et al. (2015). ADHD in DSM-5: a field trial in a large, representative sample of 18- to 19-year-old adults. *Psychol. Med.* 45, 361–373. doi: 10.1017/S0033291714001470

Mayer, A. R., Harrington, D., Adair, J. C., and Lee, R. (2006). The neural networks underlying endogenous auditory covert orienting and reorienting. *NeuroImage* 30, 938–949. doi: 10.1016/j.neuroimage.2005.10.050

Menon, V., and Uddin, L. Q. (2010). Saliency, switching, attention and control: a network model of insula function. *Brain Struct. Funct.* 214, 655–667. doi: 10.1007/s00429-010-0262-0

Moerel, M., De Martino, F., and Formisano, E. (2014). An anatomical and functional topography of human auditory cortical areas. *Front. Neurosci.* 8:225. doi: 10.3389/fnins.2014.00225

Mostert, J. C., Onnink, A. M. H., Klein, M., Dammers, J., Harneit, A., Schulten, T., et al. (2015). Cognitive heterogeneity in adult attention deficit/hyperactivity disorder: a systematic analysis of neuropsychological measurements. *Eur. Neuropsychopharmacol.* 25, 2062–2074. doi: 10.1016/j.euroneuro.2015.08.010

Nagaraj, N. K., Magimairaj, B. M., and Schwartz, S. (2020). Auditory distraction in school-age children relative to individual differences in working memory capacity. *Atten. Percept. Psychophys.* 82, 3581–3593. doi: 10.3758/s13414-020-02056-5

Nieto-Castanon, A. (2020). *Handbook of Functional Connectivity Magnetic Resonance Imaging Methods in CONN.* Boston: Hilbert Press.

Öster, C., Ramklint, M., Meyer, J., and Isaksson, J. (2020). How do adolescents with ADHD perceive and experience stress? An interview study. *Nord. J. Psychiatry* 74, 123–130. doi: 10.1080/08039488.2019.1677771

Pelletier, M.-F., Hodgetts, H. M., Lafleur, M. F., Vincent, A., and Tremblay, S. (2016). Vulnerability to the irrelevant sound effect in adult ADHD. *J. Atten. Disord.* 20, 306–316. doi: 10.1177/1087054713492563

Pievsky, M. A., and McGrath, R. E. (2018). The neurocognitive profile of attention-deficit/hyperactivity disorder: a review of meta-analyses. *Arch. Clin. Neuropsychol.* 33, 143–157. doi: 10.1093/arclin/acx055

Ramtekkar, U. P., Reiersen, A. M., Todorov, A. A., and Todd, R. D. (2010). Sex and age differences in Attention-Deficit/Hyperactivity Disorder symptoms and diagnoses: implications for DSM-V and ICD-11. *J. Am. Acad. Child Adolesc. Psychiatry* 49, 217–28.e-13.

Regenbogen, C., Vos, M. D., Debener, S., Turetsky, B. I., Mößnang, C., Finkelmeyer, A., et al. (2012). Auditory processing under cross-modal visual load investigated with simultaneous EEG-fMRI. *PLoS One* 7:e52267. doi: 10.1371/journal.pone.0052267

Rodriguez, A., Ginsberg, Y., Fernholm, A., and Nyberg, L. (2007). ADHD difficult to diagnose in adults. ASRS v1.1 self-report scales valuable help–now translated to Swedish. *Lakartidningen* 104, 1398–1400.

Rönnberg, J., Arlinger, S., Lyxell, B., and Kinnefors, C. (1989). Visual evoked potentials: relation to adult speech reading and cognitive function. *J. Speech Hear. Res.* 32, 725–735.

Sabri, M., Humphries, C., Binder, J. R., and Liebenthal, E. (2011). Neural events leading to and associated with detection of sounds under high processing load. *Hum. Brain Mapp.* 34, 587–597. doi: 10.1002/hbm.21457

Salla, J., Galéra, C., Guichard, E., Tzourio, C., and Michel, G. (2019). ADHD symptomatology and perceived stress among french college students. *J. Atten. Disord.* 23, 1711–1718. doi: 10.1177/1087054716685841

Schulze, M., Lux, S., and Philipsen, A. (2020). Sensory processing in adult ADHD – a systematic review. *Research Squre* [Preprint] doi: 10.21203/rs.3.rs-71514/v1

Shaw, P., Eckstrand, K., Sharp, W., Blumenthal, J., Lerch, J. P., Greenstein, D., et al. (2007). Attention-deficit/hyperactivity disorder is characterized by a delay in cortical maturation. *Proc. Natl. Acad. Sci. U.S.A.* 104, 19649–19654. doi: 10.1073/pnas.0707741104

Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiller, E., et al. (1998). The mini-international neuropsychiatric interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J. Clin. Psychiatry* 59 Suppl. 20, 22–33; quiz 34–57.

Sörqvist, P. (2010a). High working memory capacity attenuates the deviation effect but not the changing-state effect: further support for the duplex-mechanism account of auditory distraction. *Mem. Cognit.* 38, 651–658. doi: 10.3758/MC.38.5.651

Sörqvist, P. (2010b). The role of working memory capacity in auditory distraction: a review. *Noise Health* 12:217. doi: 10.4103/1463-1741.70500

Sörqvist, P., and Marsh, J. E. (2015). How concentration shields against distraction. *Curr. Dir. Psychol. Sci.* 24, 267–272. doi: 10.1177/0963721415577356

Sörqvist, P., and Rönnberg, J. (2014). Individual differences in distractibility: an update and a model. *PsyCh J.* 3, 42–57. doi: 10.1002/pchj.47

Sörqvist, P., Dahlström, Ö, Karlsson, T., and Rönnberg, J. (2016). Concentration: the neural underpinnings of how cognitive load shields against distraction. *Front. Hum. Neurosci.* 10:221. doi: 10.3389/fnhum.2016.00221

Sörqvist, P., Ljungberg, J. K., and Ljung, R. (2010). A sub-process view of working memory capacity: evidence from effects of speech on prose memory. *Memory* 18, 310–326. doi: 10.1080/09658211003601530

Sörqvist, P., Stenfelt, S., and Rönnberg, J. (2012). Working memory capacity and visual–verbal cognitive load modulate auditory–sensory gating in the brainstem: toward a unified view of attention. *J. Cogn. Neurosci.* 24, 2147–2154. doi: 10.1162/jocn_a_00275

Studholme, C., Hawkes, D. J., and Hill, D. L. G. (1998). "Normalized entropy measure for multimodality image alignment," in *Proceedings of the Medical Imaging 1998: Image Processing*, (Bellingham, WAS: International Society for Optics and Photonics), 132–143. doi: 10.1117/12.310835

Uchida, M., Spencer, T. J., Faraone, S. V., and Biederman, J. (2018). Adult outcome of ADHD: an overview of results from the MGH longitudinal family studies of pediatrically and psychiatrically referred youth with and without ADHD of both sexes. *J. Atten. Disord.* 22, 523–534. doi: 10.1177/1087054715604360

van Lieshout, M., Luman, M., Twisk, J. W. R., Faraone, S. V., Heslenfeld, D. J., Hartman, C. A., et al. (2017). Neurocognitive predictors of ADHD outcome: a 6-year follow-up study. *J. Abnorm. Child Psychol.* 45, 261–272. doi: 10.1007/s10802-016-0175-3

Vossel, S., Geng, J. J., and Fink, G. R. (2014). Dorsal and ventral attention systems. *The Neuroscientist* 20, 150–159. doi: 10.1177/1073858413494269

Whitfield-Gabrieli, S., and Nieto-Castanon, A. (2012). Conn: a functional connectivity toolbox for correlated and anticorrelated brain networks. *Brain Connect.* 2, 125–141. doi: 10.1089/brain.2012.0073

Whitfield-Gabrieli, S., and Nieto-Castanon, A. (2017). *Functional Connectivity SPM Toolbox 2017.* Available online at: http://www.nitrc.org/projects/conn (Accessed January 27, 2020).

Willcutt, E. G., Doyle, A. E., Nigg, J. T., Faraone, S. V., and Pennington, B. F. (2005). Validity of the executive function theory of attention-deficit/hyperactivity disorder: a meta-analytic review. *Biol. Psychiatry* 57, 1336–1346. doi: 10.1016/j.biopsych.2005.02.006

Yurgil, K. A., and Golob, E. J. (2013). Cortical potentials in an auditory oddball task reflect individual differences in working memory capacity. *Psychophysiology* 50, 1263–1274. doi: 10.1111/psyp.12140

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

ORIGINAL RESEARCH
published: 23 June 2022
doi: 10.3389/fnins.2022.930932

Acoustic Stimuli Can Improve and Impair Somatosensory Perception

Matthias Nuernberger[1,2*†], **Denise Schaller**[1†], **Carsten Klingner**[1,2], **Otto Witte**[1] and **Stefan Brodoehl**[1,2]

[1] Department of Neurology, Jena University Hospital, Jena, Germany, [2] Biomagnetic Center, Jena University Hospital, Jena, Germany

The integration of stimuli from different sensory modalities forms the basis for human perception. While the relevant impact of visual stimuli on the perception of other sensory modalities is recognized, much less is known about the impact of auditory stimuli on general sensory processing. This study aims to investigate the effect of acoustic stimuli on the processing of somatosensory stimuli using real noise (i.e., unpleasant everyday noise, RN) and neutral white noise (*WN*). To this purpose, we studied 20 healthy human subjects between 20 and 29 years of age (mean: 24, SD: ±1.9 years sex ratio 1:1). Somatosensory perception was evaluated using mechanical detection threshold (MDT) of the skin on the back of the dominant hand. To investigate the underlying mechanisms in the brain, fMRI was performed while applying acoustic stimulation (RN and WN) and tactile stimulation of the dominant hand. Here we show that acoustic stimulation with noise alters the perception of touch on the skin. We found that the effect of RN and *WN* differed. *RN* leads to an improved tactile perception, whereas *WN* impaired tactile perception. These changes go along with significant differences in brain activity and connectivity. *WN* is associated with a significant increase in brain activity in multiple brain areas such as the auditory and somatosensory cortex, parietal association cortex, and the thalamus compared to *RN*. With tactile stimulation of the skin, the flow of information in these brain areas is altered. While with *RN* the information flow from the thalamus to the somatosensory cortex is prominent, the network activity pattern changes under *WN* revealing an increase in interaction between multiple networks. Unpleasant noise inhibits the multisensory integration and enables a more efficient unimodal perception in the somatosensory system, improving perception. Whether this is to be interpreted as a temporary increase in phasic alertness or by a stronger filter function of the thalamus with a preference for unimodal stimuli is still open for debate.

Keywords: somatosensory perception, MDT, fMRI, acoustic noise, connectivity, crossmodal interaction, white noise, sensory integration

OPEN ACCESS

Edited by:
Anu Sharma,
University of Colorado Boulder,
United States

Reviewed by:
Giulia Cartocci,
Sapienza University of Rome, Italy
Terry Blumenthal,
Wake Forest University, United States

***Correspondence:**
Matthias Nuernberger
matthias.nuernberger@med.uni-jena.de

†*These authors have contributed equally to this work*

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

Received: *28 April 2022*
Accepted: *30 May 2022*
Published: *23 June 2022*

Citation:
Nuernberger M, Schaller D, Klingner C, Witte O and Brodoehl S (2022) Acoustic Stimuli Can Improve and Impair Somatosensory Perception. Front. Neurosci. 16:930932. doi: 10.3389/fnins.2022.930932

INTRODUCTION

The concept of crossmodal interactions extends the classical doctrine of unimodal processing in primary sensory brain areas (Macaluso and Driver, 2005; Stein, 2012). Well-known examples of multisensory perceptual illusions caused by crossmodal interactions include the McGurk effect (McGurk and MacDonald, 1976) and the ventriloquist effect (Vroomen et al., 2001). These illusions have in common that spatial distance of otherwise associated visual and auditory stimuli lead to mislocalization and misinterpretation of sensory input. In case of contradictory information, the visual stimulus is often evaluated as more valid or a compromise between the sensory information is generated to bridge the prediction error. For example, if the spatial source of what is heard and what is seen do not match, the spatial information of what is seen can often overwrite that of what is heard. This tendency is discussed as visual dominance (Colavita, 1974). However, in addition to visual information, auditory stimuli can also influence the perception of another sensory modality (Ma et al., 2009; Kayser and Shams, 2015). The perception of high-frequency sounds leads to the perception of a surface as smoother and drier in tactile perception. Low-frequency sounds lead to a rougher and moister perception of the same surface. This effect is called parchment skin illusion (Jousmäki and Hari, 1998). Acoustic stimulation can also distract attention and impair motor learning processes (Barutchu et al., 2010). Besides specific acoustic stimuli (i.e., sounds like a car horn) and noise (i.e., a random cluster of familiar and unfamiliar sounds with pleasant and unpleasant features), the most relevant auditory research instrument is white noise (*WN*). *WN* can be described as a hissing sound likely to/h/in constant aspiration. It carries an audio signal in form of a flat spectrum across all audible frequencies. Its relevance for cognitive information processing is subject to debate, but there are also indications of beneficial aspects to learning processes (Rausch et al., 2014). Somatosensory perception can also be influenced by other sensory modalities like visual input or deprivation, as we were able to show in previous work (Brodoehl et al., 2015a,b). Closing the eyes leads to an enhanced perception of subtle touch at the expense of spatial integration of this information. The brain switches between a mode with enhanced thalamo-somatosensory coupling and a mode with enhanced multimodal integration. By extending our findings to the auditory and somatosensory systems' interactions, we hypothesize that modulation of activity in the auditory system alters the perception of somatosensory stimuli. We assume that in addition to top-down modulation by complex stimuli, bottom-up modulation *via* crossmodal interactions also plays a relevant role. We aimed to investigate how the perception of a simple tactile touch on the skin is altered by different acoustic stimulation. The acoustic stimuli we used were a sound generally perceived as unpleasant real noise (*RN*) and a sound perceived as *WN*. To study the underlying neural mechanisms, we used functional magnetic resonance imaging to evaluate brain network activity and quantify causal information flow.

MATERIALS AND METHODS

Participants

We recruited 20 healthy human volunteers (range: 20–29 years, mean: 23.95, SD: ±1.91 years, 10 male) without neurological or otological afflictions. All participants identified as either male or female. All participants were right-handed. Typical exclusion criteria were considered (Brodoehl et al., 2016). All subjects were informed about the procedure of the trial in written form as well as personally and gave their written consent according to the Declaration of Helsinki. All 20 recruited participants participated in all experiments described in this analysis. The trial was approved by the Ethics Committee of the Medical Faculty of the Friedrich Schiller University Jena, Germany (registration number: 4301-01/15). We performed pilot trials with five subjects to evaluate our experimental design. Data of pilot trials is not part of this analysis. The subjects of the pilot trials were matched to the planned participants of our experiments.

Acoustic Stimuli

We used two different sounds for acoustic stimulation: *RN* and *WN*. *RN* was intended to represent very unpleasant acoustic information while *WN* was assessed as neutral to slightly unpleasant. From a variety of sounds presumed as unpleasant noise (e.g., the sound of a jackhammer, traffic noise, or people yelling), a sound sample with intense instrumental heavy metal music was selected as *RN*. It offers a wide range of frequencies and extensive temporal modulations. In a rating of 0 (not unpleasant) to 10 (very unpleasant), this sound sample was rated highest (mean: 8.3) in our pilot trials, so it was used in the final study. In the same manner *WN* was rated slightly unpleasant in our pilot trial (mean: 2.8). *WN* can best be described as a static monotonous sound without specific characteristics. Using the software program mp3Gain (developed by Glen Sawyer, Version 1.3.4), the samples was normalized to 90 dB (0.633 Pa) and a length of 10 min. As *WN*, a freely available sample which is composed of all frequencies in the human hearing range was used. It was normalized to 90 dB (0.633 Pa) and a length of 10 min as well. The volume of the acoustic stimuli was adjusted to 75 dB before application so that it was rated as loud but still tolerable by the subjects. Sound pain threshold was not reached or surpassed. As baseline apart from the forementioned two acoustic stimuli we used silence without any auditory input (*rest*).

Somatosensory Perception Testing by Mechanical Detection Threshold

Mechanical Detection Threshold (MDT) was measured using a quantitative sensory testing (QST)-compliant set of Von-Frey-Hairs® (Optihair2, Marstock Nervtest, Germany). These plastic filaments apply pressure between 0.125 and 64 mN (grating-factor: 2) when they touch the skin of the dominant backhand on a hairless spot with a diameter of 0.5 mm for 1 s while being bended to S-shape. The described procedure is standardized in our lab (Brodoehl et al., 2013). All subjects were blindfolded and received the stimulation on the same area on the back of the right hand (diameter: 1 cm, shaved area marked by colored pencil). Threshold determinations (10 stimuli each) were acquired by alternately descending until the subject failed to notice the stimulus and ascending until re-noticing occurred ("method of limits"). Means and standard errors of the results of all blocks were calculated and analyzed as surrogates for the actual threshold.

Somatosensory Perception Experiment

The examination took place in a darkened, anechoic room. The subjects were instructed to keep their eyes closed during the examination. Furthermore, they received a blindfold and standard noise protection headphones to prevent them from being influenced by any ambient noise. The acoustic stimuli *RN* and *WN* were presented using standard in-ear headphones. To measure the perception of a simple touch on the skin, we determined the tactile mechanical detection threshold (MDT). The actual examination was performed in a total of 12 blocks of 4 min each. During each block, either *RN*, *WN*, or no sound (*rest*) was played. The order of the blocks was arranged pseudorandomized. Within each block and starting at minute 2, MDT was determined five times. For this purpose – starting from 16 mN – the filaments were presented in descending order of strength until the subject no longer perceived any touch and then again in ascending order of strength until touch was again perceptible. The experiment is devised as a repeated measures design. A schematic representation of the procedure is shown in **Figure 1** (left).

fMRI Experiments

fMRI Data Acquisition Parameters

All experiments were performed using a 3.0-T MR scanner (Trio, Siemens, Erlangen, Germany) to obtain echo-planar T2*-weighted image volumes (EPI) and transaxial T1-weighted structural images. The high-resolution T1-weighted structural images had a voxel size of 1 mm × 1 mm × 1 mm to enable precise anatomical localization. EPI images were acquired using the following parameters: voxel size = 3 mm × 3 mm × 3 mm, TR = 2.52 s, TE = 35 ms, and 40 transaxial slices (including the entire cerebrum and cerebellum).

Experimental Setup

The fMRI examination was performed the day after the assessment of MDT. Subjects were positioned in the scanner, given standard headphones with acoustic shielding, and instructed to keep their eyes closed during the examination. After each scan, the subjects' level of vigilance was inquired.

fMRI Experiment 1: Acoustic Stimulation

Experiment 1 lasted a total of 36 min. In a block design, either *RN* or *WN* was played through the headphones. The exposure to this noise in combination with the headphones makes the scanner hardly audible for the subjects. A total of 12 blocks, each lasting 3 min, were executed in a pseudorandomized order; each block consequently lasted approximately 72 EPI images. The task was passive, so no active participation of the subjects was necessary. The experiment is devised as a repeated measures design.

fMRI Experiment 2: Acoustic Stimulation Combined With Tactile Stimulation of the Right Hand

Duration and block design were identical to Experiment 1. Additionally, tactile stimulation with a balloon-controlled air-driven device (*air-puff*) was applied to the fingers of the right hand during the blocks. Von Frey filaments could not be used in the scanner room, because additional personal was not allowed in the scanner room during image acquisition. During 1 block, 10 stimulations occurred (interstimulus interval 15–30 s, duration 1.5 s, onset 10th second after block start). The task was passive, so no active participation of the subjects was necessary. **Figure 1** (right) shows the sequence of the fMRI experiments. The experiment is devised as a repeated measures design.

fMRI Data Preprocessing

For each subject, all images were realigned to the first volume using six-parameter rigid-body transformations to correct for motion artifacts. The images were co-registered with the corresponding anatomical (T1-weighted) images of the subject, re-sliced to correct for acquisition delays (referenced to the 10th slice only in the event-related design), normalized to the Montreal Neurological Institute (MNI) standard brain to report MNI coordinates and smoothed using a 6-mm full-width-at-half-maximum Gaussian kernel.

Statistical Analysis

Statistical analysis was performed using IBM SPSS27® (Version: 27.0.0.0). Data was tested for normal distribution by Shapiro–Wilk and Kolmogorov–Smirnov test. Data was tested for homogeneity of variance by Levene's test. Data was tested for equal sphericity by Mauchly's test. Where normal distribution was attained, we used independent samples t-test. Where normal distribution was not attained, the Mann–Whitney U test was used. Where equal sphericity was not attained, Greenhouse–Geisser correction was applied. Findings were considered significant at $p < 0.05$ (two-sided). Results were corrected for family wise error (FWE) induced by multiple comparisons using Bonferroni-correction ($p < 0.05$). Standard confidence interval (95% CI: mean ± 2 SD) was used. For the evaluation of the mechanical detection threshold (MDT) in the somatosensory perception experiment we followed standardized protocol (Rolke et al., 2006): the geometric mean was calculated for each block of each stimulus (*rest*, *RN*, *WN*), and subsequently the mean was calculated for each of the stimuli. To perform repeated measures analysis of variance (repeated-measures ANOVA) logarithmic transformation of the data was performed using the natural logarithm. Data analysis for the fMRI experiments was performed on a PC using MATLAB (Version 2019a, MathWorks, Natick, MA, United States) and SPM12 software (Wellcome Department of Cognitive Neurology, London, United Kingdom[1]) (Friston, 2007). fMRI Experiment 1: multiple regression analysis using a general linear model was performed to obtain statistical parametric maps calculated for the somatosensory stimulation. The fMRI signal time courses were high-pass filtered (128 s) and modeled as an experimental-stimulus onset function convolved by the canonical hemodynamic response function (low-pass filter). Two contrasts of interest were examined, resulting in two t-statistical (paired t-test) maps ($RN > WN$) for the first fMRI experiment. Individual results were projected onto their respective co-registered high-resolution T1-weighted 3-D data set. The anatomical localization of the activated areas was analyzed regarding the standard stereotaxic atlas and was mapped using the anatomical toolbox of SPM12 (Eickhoff et al., 2005). Threshold free cluster enhancement (TFCE) was applied, and results were corrected for FWE induced by multiple comparisons using Bonferroni-correction ($p < 0.05$). fMRI Experiment 2: several regions of interest (ROIs) were defined based on our hypotheses. The analyzed ROIs are shown in **Supplementary Table 1**. The time-series data from these identified regions were extracted, and cluster-specific time series were then estimated

---

[1] http://www.fil.ion.ucl.ac.uk/spm

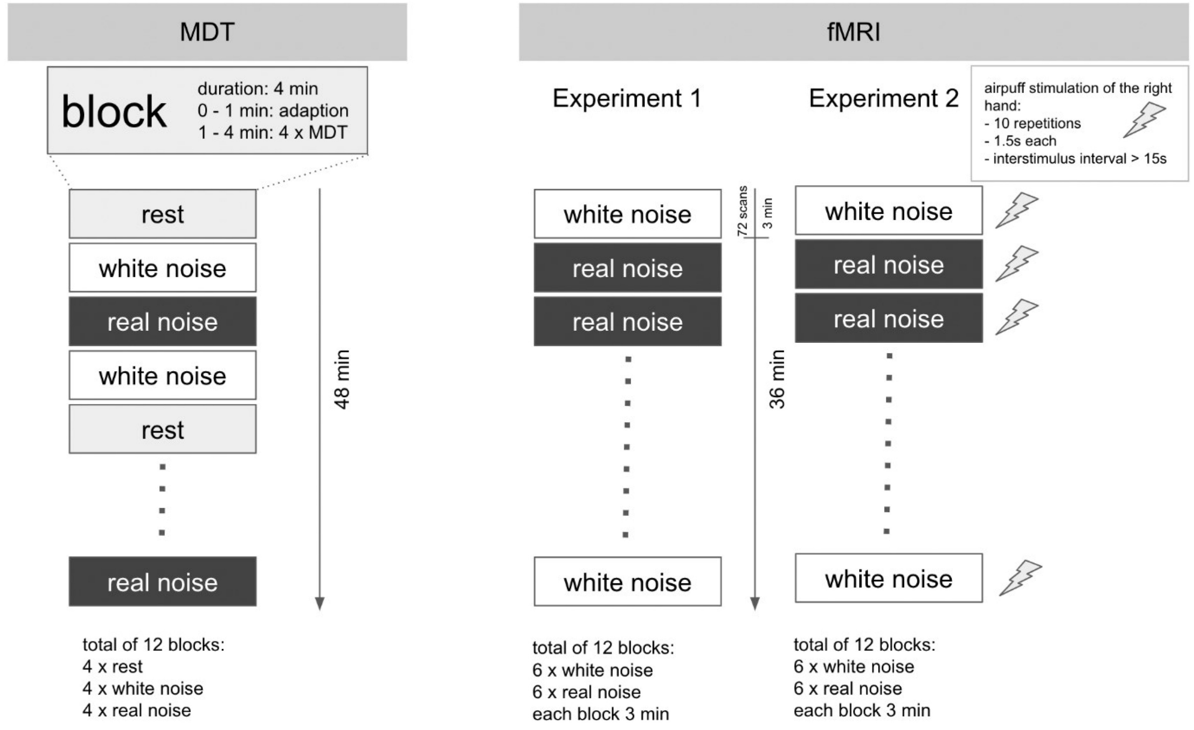


FIGURE 1 | Experimental design for MDT and fMRI experiments. Schematic illustration of the experimental design. Left: MDT: a total of 12 blocks with a duration of 4 min each. The determination of the MDT was performed in each block from the second minute onward (4× per block). In sum, the study lasted 48 min per subject. Right: fMRI: the experimental setup for Experiments 1 and 2 was in major parts identical. Each of the 12 blocks with acoustic stimulus lasted 3 min. The overall duration per experiment was 36 min. In Experiment 2, however, the fingers of the right hand were additionally stimulated tactilely (*air-puff*) within each block: 10 repetitions, duration 1.5 s, interstimulus interval 15–30 s. In contrast to the MDT determination, no block with *rest* condition (silence) was performed in the fMRI experiments.

by averaging the time series of all voxels within a cluster. Several sources of variance were removed from the data using linear regression: (1) six parameters obtained by rigid body correction of head motion, (2) a signal from a ventricular region of interest, and (3) a signal from a region centered in the white matter. All signal intensity time courses were band-pass filtered ($0.01 < f < 0.1$ Hz) to reduce the effects of low-frequency drift and high-frequency noise. Conditional Granger causality analysis (GCA) was applied to explore the dynamic causal relationship between the time series. This approach has been widely used in previous fMRI studies. In our study, GCA was performed using the toolbox implemented by Seth (2010). The detailed theory behind Granger causality has been previously described (Granger, 1969). TFCE was applied, and results were corrected for multiple comparisons using Bonferroni-correction ($p < 0.05$).

RESULTS

Effects of Acoustic Stimulation Upon the Somatosensory Perception of Touch

In 20 healthy subjects (24 ± 1.9 years, 10 females) the mechanical detection threshold (MDT) was assessed under different acoustic stimuli (*rest*, *WN*, *RN*). Starting from a baseline at *rest*, 75% of the subjects showed a deterioration in perceptual performance in *WN* and 90% of the subjects showed an improvement in *RN*. In direct comparison of the two noise variants, the participants demonstrated a statistically significantly better perceptual performance with *RN* compared to *WN*, $U = 284.000$, $p = 0.024$. The mean values of the group analysis were 1.40 (±0.83 mN) for *rest*, 1.74 (±1.18 mN) for *WN*, and 1.13 (±0.55 mN) for *RN* condition. In group analysis (repeated-measures ANOVA), all pairwise comparisons were statistically significant at a significance level of $p < 0.05$ (*rest* vs. *WN*: $p = 0.049$, *rest* vs. *RN*: $p = 0.027$, and *WN* vs. *RN*: $p < 0.001$). There was no statistically significant influence age or gender. To control for fluctuations of alertness during the experiment, we tested the second block of every condition against the last block of the same condition. There was no statistically significant difference between these blocks, $p > 0.05$. Intraindividual results are shown in **Table 1**.

Changes in Brain Activity Caused by Sound Stimulation

Experiment 1

fMRI scans of all 20 subjects were analyzed to investigate the effect of the 2 conditions *WN* and *RN* upon brain activity

TABLE 1 | Results of somatosensory perception testing with the mechanical detection threshold (MDT).

Subjects		**Rest**		**WN**		**RN**	
#	**Age/sex**	**mN**	**SD**	**mN**	**SD**	**mN**	**SD**
1	21/f	1.04	±0.09	1.03	±0.07	0.75	±0.05
2	22/f	1.05	±0.33	1.05	±0.31	0.78	±0.19
3	22/f	1.43	±0.23	1.98	±0.4	1.22	±0.21
4	22/f	0.71	±0.12	0.72	±0.17	0.6	±0.08
5	22/m	3.12	±0.53	2.93	±0.12	2.6	±0.27
6	22/m	1.49	±0.97	2.44	±0.93	1.23	±0.47
7	23/m	1.14	±0.73	1.75	±0.66	1.18	±0.94
8	24/f	1.8	±1.12	1.35	±0.35	0.9	±0.17
9	24/f	0.29	±0.07	0.41	±0.15	0.27	±0.08
10	24/f	1.37	±0.12	1.38	±0.41	1.35	±0.32
11	24/m	1.53	±0.26	2	±0.59	1.28	±0.09
12	24/m	1.59	±0.92	1.25	±0.08	0.83	±0.12
13	24/m	2.15	±1.35	2.54	±1.19	1.37	±1.02
14	25/f	1.02	±0.49	1.56	±0.81	0.58	±0.14
15	25/f	0.41	±0.27	0.75	±0.45	0.33	±0.18
16	25/f	0.58	±0.13	0.67	±0.27	0.33	±0.14
17	25/m	1.77	±0.65	2.26	±1.14	1.04	±0.15
18	25/m	0.64	±0.17	0.66	±0.19	0.39	±0.09
19	27/m	3.61	±0.83	5.7	±3.56	4.69	±2.35
20	29/m	1.33	±1.04	2.28	±1.85	0.91	±0.84
Mean	24	1.40	±0.83	1.74	±1.18	1.13	±0.55

The mean value of the mechanical detection threshold (MDT) in mN is shown for each subject and each block (rest, white noise, real noise). WN, white noise; RN, real noise; SD, standard deviation.

(Experiment 1). The results of the second-level analysis with SPM are shown in **Figure 2**. Statistically significant activation (significance level at $p < 0.05$, FWE-corrected) was found only for the contrast of *WN* > *RN*. Acoustic stimulation with *WN* resulted in significantly increased brain activity patterns, especially in the temporal, parietal, and occipital cortex. Discrete activations were also detectable in the primary somatosensory cortex Clusters with increased contrast are listed in **Table 2**. Furthermore, increased brain activity compared to the *RN* condition was found in the hippocampus, parahippocampal gyrus, cingulate gyrus, and primary motor cortex. In summary, an increase in brain activity during *WN* was found especially in the auditory cortex and in the parietal association cortex (see **Figure 2**).

Experiment 2

First, the activation patterns while stimulating the right hand were analyzed. Two subjects did not show activation in the primary somatosensory cortex (S1, hand knob), so they were excluded from further analysis regarding Experiment 2. The activation maps of the tactile stimulation of the second-level analysis were defined as ROIs (primary and secondary somatosensory cortex) in the extraction of the time series (S1 and S2 in **Supplementary Table 1**). A summarized illustration of the GCA is shown in **Figure 3**. A more detailed illustration of the matrix representation of the pairwise causality results is included in **Supplementary Figure 1**.The results of the causality analysis revealed two basic patterns of information flow in the investigated regions. For *RN* compared to *WN*, a direct flow of information from the thalamus to the somatosensory cortex is prominent. In acoustic stimulation with *WN*, on the other hand, a complex interaction of all involved network partners becomes apparent. In particular, an intense information exchange between the auditory cortex, the integrative association areas in the parietal cortex, and the thalamus emerge. The somatosensory cortex continues to receive input primarily from the thalamus, but now directly exchanges information with the auditory system and the association cortex. Significance level at $p < 0.05$, FWE-corrected.

DISCUSSION

The results of our study showed that acoustic stimuli such as sounds generally perceived as *RN* or *WN* change the perception of a touch on the skin. More importantly, both categories of noise

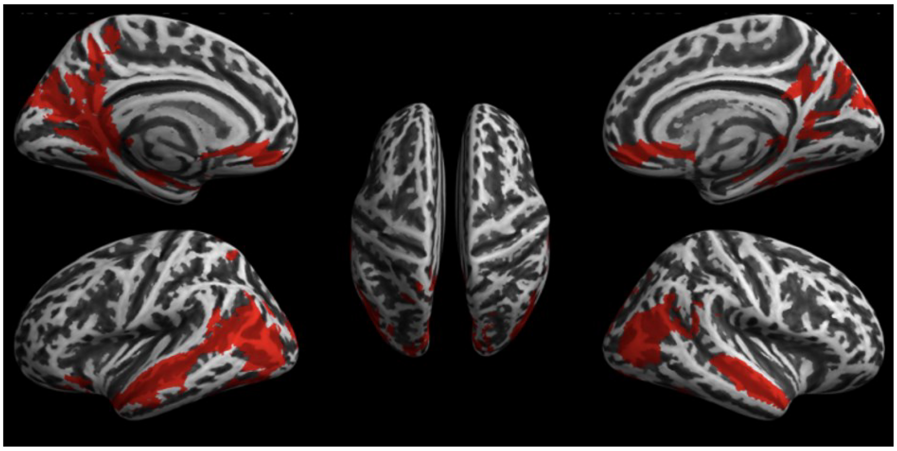


FIGURE 2 | Results of the fMRI Experiment 1: WN > RN.

TABLE 2 | Results of the fMRI Experiment 1: WN > RN.

Cluster	Voxel	t_{max}	MNI			Brain regions
#	*n*		*X*	*Y*	*Z*	
1	273	7.47	−60	−26	−5	Temporal lobe left Amygdala (LB), area tE 3, area Id1 (insula), entorhinal cortex
2	161	5.7	39	4	−29	Temporal lobe right Area tE 3; area Id1 (insula)
3	117	6.08	−48	−68	19	Temporal lobe left Area PGp (IPL); area Pga (IPL), area PFm (IPL)
4	53	4.73	51	−68	19	Temporal lobe right Area PGp (IPL); area PGa (IPL)
5	52	4.44	−6	25	−20	Gyrus rectus bilateral n.d.
6	50	5.17	−15	−44	4	Gyrus fusiformis left, "lingual gyrus" Subiculum; DG (hippocampus); CA1 (hippocampus); temporal thalamus
7	38	5.07	−9	−50	52	Precuneus left Area 5m (SPL), area 5Ci (SPL), area 51 (SPL)
8	38	4.95	−12	7	−14	Olfactory cortex left, insular lobe left; IFG (p. orbitalis) left Amygdala left
9	35	4.73	6	16	−14	Gyrus rectus right, olfactory cortex right n.d.
10	32	6.47	−33	−20	−11	Hippocampus left, parahippocampal gyrus left CA3, CA2, CA1, DG (hippocampus), subiculum
11	21	4.67	−48	−80	1	Occipital lobe (inferior and medius) hOc5 (V5/Mt); area FG1; area FG2; hOc4v [V4(v)]
12	15	4.64	27	−41	4	Right hippocampus, right parahippocampal gyrus DG, CA1, CA2, CA3 (hippocampus); temporal thalamus, subiculum
13	14	4.09	12	−44	55	MCC (right cingulate gyrus) Area 5Ci (SPL); area 5m (SPL), area 3a, area 4p
14	12	4.38	6	−92	28	Cuneus right hOc3d (V3d) bds, hOc2 right (V2), hOc4d (V3A)
15	11	3.89	45	−83	10	Occipital and temporal lobes hOc5 (V5/Mt)
16	11	4.36	30	−20	−14	Hippocampus right, parahippocampal gyrus right CA1, CA2, CA3, DG (hippocampus), subiculum

Activations are shown on an inflated brain model. Results were corrected after TFCE at $p < 0.05$ (FWE-corrected) and are shown in red. The contrast displayed is the pairwise t-test for the comparison: "white noise > real noise." The table shows the number of clusters found with size (in voxels), maximum t-value, MNI coordinates, and anatomical location/description.

have opposite effects on the perception of touch. *RN* improves the perception while *WN* impairs it. Additionally, the brain activity under application of these two types of noise again reveals significant differences. In the case of *WN*, brain activity increases compared to *RN* in brain regions belonging to the auditory, visual, somatosensory, and integrative systems. This is also associated with significantly increased network activity during the processing of a simple touch on the skin. Networks outside the somatosensory system are particularly affected. Communication between the thalamus, integrative brain areas, and the auditory system are amplified. In the presence of *RN*, however, this network activity decreases significantly and is replaced by a more unidirectional flow of information from the thalamus to the somatosensory cortex. The discussion of our findings will mainly focus on two topics: the nature of crossmodal interactions and the effect of noise on sensory perception.

Processing of Touch: Uni- and Multisensory Processing

Simple, non-painful touch is presumably processed *via* two distinct pathways (Dijkerman and de Haan, 2007). Both begin in the thalamus, projecting to the primary somatosensory cortex (S1) (Brodmann areas 3a, 3b, 1, and 2). From there, they either pass through S2 into the posterior insula or terminate in the posterior parietal cortex (PPC). A good overview is presented by Klingner et al. (2011). These two pathways represent the classical hierarchical processing of somatosensory information. Recognition and perception take place in both target areas (posterior insula and PPC). However, there is an increasing departure from this classical view in which primary sensory cortices process only one specific modality of sensory stimulus processing. Many recent studies show that there is interconnectivity between primary sensory cortices of different modalities. This observation raises the question of whether any cortex can be truly unisensory (Macaluso and Driver, 2005; Ghazanfar and Schroeder, 2006). Accordingly, the processing of one sensory modality would automatically have effects on the processing of other modalities. In one of our earlier works, we were able to show that, even in the absence of a visual stimulus, opening and closing the eyes can switch between a more uni- or multisensory oriented processing pathway (Brodoehl et al., 2013). Another general observation is that weak stimuli are in particular receptive to multisensory interactions (Stein et al., 1993; Macaluso and Driver, 2005; Stein, 2012).

Interactions of Hearing and Touch

There are many examples in which conflicting sensory information can lead to misinterpretations. Furthermore, there is concrete evidence that hearing certain sounds (e.g., scratching one's fingernails across a blackboard) (Halpern et al., 1986) or seeing certain scenes (e.g., when a spider walks across a neck) (Keysers et al., 2004) can trigger corresponding activity in the somatosensory system. While numerous studies are describing auditory-visual and visual-tactile interactions (Kennett et al., 2001; Vroomen et al., 2001; Ro et al., 2004) there

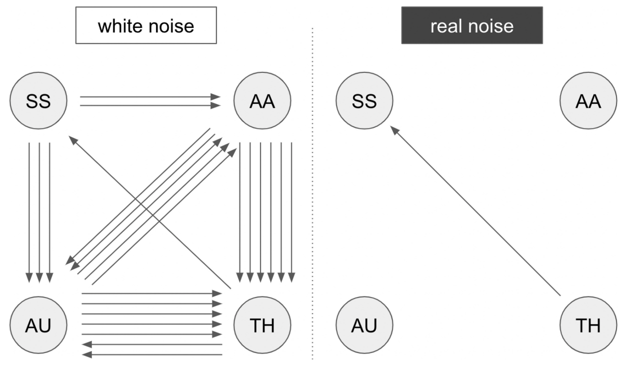


FIGURE 3 | Results of the GCA analysis. Summarized results of the Granger causality analysis of fMRI Experiment 2 with pairwise comparison "white noise > real noise" (white noise) and "real noise > white noise" (real noise). The arrows represent significant connections in the comparison matrix (**Supplementary Figure 1**). An arrow represents a registered causal link (compare **Supplementary Figure 1**). SS, somatosensory areas; AU, auditory areas; AA, postparietal association cortex; TH, thalamus. A more detailed description of the brain regions is provided in **Supplementary Table 1**.

are relatively few that examine interactions between sound and touch. Although there are reports of interactions (Gescheider et al., 1969; Gillmeister and Eimer, 2007; Navarra et al., 2007; Serino et al., 2007), little is known about the physiological principles. Yet some aspects of hearing and touch, such as vibration, seem particularly close. Loud vibrations of a car can be heard and felt. In addition to the apparent proximity of important anatomical structures (S2 and auditory cortex), neuroimaging studies have shown that there are direct interactions between somatosensory and auditory stimuli in the auditory cortex (Foxe et al., 2000). Like between the visual and auditory systems (Falchier et al., 2002), there are direct anatomical connections between the somatosensory cortex and the auditory cortex (Schroeder et al., 2001).

That sound can alter the perception of somatosensory stimuli has already been compellingly demonstrated by Ro et al. (2009). In their experiments, they used a 500 Hz sound and were able to show that the perception threshold for an electrical stimulus on the middle finger of the left hand is improved when sound and somatosensory stimulation occur simultaneously. Besides, they showed that there was a clear spatial effect. The improvement in recognition performance occurred only when the acoustic stimulus was also presented on the corresponding side of the body.

Major differences from our study should be highlighted: the 500 Hz tone used by Ro et al. (2009) was played in synchrony with the somatosensory stimulus. In our study, the acoustic sounds were played as a sustained stimulation. The 500 Hz tone itself can be classified as very uniform and of high frequency. We used a real touch on the skin in our experiments (both in the MDT and in the fMRI experiments). Typically, electrical stimulation is considered artificial (Burke and Gandevia, 1988; Dean et al., 2006). We want to emphasize these differences because partially contradictory results (to those of Ro et al., 2009) emerged in our study. The results of Ro et al. (2009) could be interpreted as a consequence of a temporal and spatial orientation of alertness, analogous to Spence and Driver (1997) and similar to cueing mechanisms. However, this is not the case for our results, where there is no direct temporal and spatial relationship between auditory and tactile stimuli. However, we were able to show that two different types of noise had different effects on the perception of touch and brain activity.

General Effects of Noise on Brain Activity

Certain types of noise can have a calming effect on humans. Othman et al. (2020) exemplified that *WN* can lead to an improvement in auditory working memory. This was associated with significantly increased activity in the auditory system, cingulate, and frontal brain, among others. The main argument here was that *WN* can create an ideal configuration of background noise in the brain (Faisal et al., 2008). Positive effects of *WN* have also been shown for other cognitive functions (Soderlund et al., 2010). But there is only very limited evidence for the benefit of the most widespread use of *WN* in our population: supporting sleep initiation (Hong et al., 2021; Riedy et al., 2021). Not surprisingly, studies show negative effects of long-term everyday noise exposure (e.g., traffic noise), especially for the cognitive development of children (Stansfeld et al., 2005; Szalma and Hancock, 2011; Klatte et al., 2013; Schlittmeier et al., 2015). Additionally, there is no doubt about the adverse effects of long-term noise exposure especially on the cardiovascular system (Maschke, 2011; Munzel et al., 2014, 2018).

How Noise Can Improve Attention and Perception

In a recent study by Schlittmeier et al. (2015), the effects of different noise levels of traffic sounds on attention-based cognition tasks were investigated. Here it was shown that moderate noise (of 50 dB) led to an improvement in a mental arithmetic task, whereas this effect did not occur at 70 dB. At 70 dB there was even a deterioration in the Stroop test, which is a well-established word-color-interference test. The main explanation given here is that traffic noise leads to an increased attentional focus, which has beneficial effects on performance in the arithmetic task and, in contrast, negative effects on performance in the Stroop test (Kahneman, 1973; Broadbent, 1978; Smith and Broadbent, 1981). This increase in phasic attention may have contributed to the improved perceptual performance of touch in the condition with *RN* as described in our present work.

Under normal circumstances, noise is perceived as a disturbance. In this context, an improvement of the signal-to-noise ratio is often considered desirable. However, in certain, often non-linear, systems, noise can help to amplify weak signals. This phenomenon is called stochastic resonance (Wiesenfeld and Moss, 1995). This is particularly relevant for the processing of sensory stimuli since external stimuli are always affected by either thermodynamic or quantum mechanical effects due to their nature (Faisal et al., 2008). It has been shown that the perception of a sensory stimulus can be significantly enhanced by a certain level of noise. Here, however, noise is related to the specific stimulus itself and not an acoustic stimulus. Noise around a subthreshold tactile stimulus can act as a kind of negative marker and increase the chance of perceiving that weak stimulus (Moss et al., 1996). With the auditory system, it has also been shown that certain levels of noise can lead to improved perception and discrimination of acoustic signals (Zeng et al., 2000). These findings were obtained both with purely acoustic stimulation and with direct electrical stimulation.

Noise in one system can enhance perception in a different modality, as has been shown in the context of *WN* and visual perception (Gleiss and Kayser, 2014). This has been investigated for the first time for the interaction of auditory and somatosensory systems in our work.

Results of This Study in the Context of Crossmodal Interactions, Shifting Attentional Focus, and Stochastic Resonance

The analysis of brain activity in our study indicated that cerebral activity differs between the two acoustic stimuli (*WN* and *RN*). It is, however, a problem that no reliable baseline activity could be defined since no rest condition (without noise) could be realized in the scanner room due to the scanning noises (Shellock et al., 1998). We interpret our results in a way that the unpleasant *RN* creates a brain state with an optimized unimodal procession of somatosensory stimuli. This might be favored by focused phasic attention (Schlittmeier et al., 2015). This results in a lower perception threshold as demonstrated by the MDT. The *WN* environment, on the other hand, led to significantly increased activity and connectivity in the auditory and somatosensory cortex, the association cortex, and the thalamus. At first, it may appear contradictory that this results in a decline in the perceptual performance of touch. However, some of our previous work has shown that increased connectivity of sensory and integrative brain areas, while not associated with the improved perceptual performance of simple stimuli, can lead to improved processing of more complex stimuli, that involve higher hierarchies of sensory and integrative processing (Brodoehl et al., 2016).

Shortcomings of Our Trial

To address limitations of our trial, the constricted comparability between passive fMRI experiments and active somatosensory perception experiments must be discussed. Active participation of subjects while in the MRI scanner would involve artifacts through movement and presume additional personal in the scanner room during image acquisition. This limitation could be handled by physical tools which present a somatosensory stimulus on one hand while providing the possibility for active feedback on the other hand. Furthermore, *WN* could trigger a more pronounced cerebral activation because it includes the whole frequency range, and the auditory cortex is organized tonotopically. However, this aspect is still a matter of discussion and not yet clarified. Another possibly relevant aspect is the difference in valence and structure of the applied acoustic stimuli. *WN* might offer a calming effect upon some participants, especially in contrast to the chosen *RN*. An additional acoustic stimulus with positive valence and even structure would help distinguish the detected effects and offer control of this aspect.

CONCLUSION

The current data provide evidence for a behavioral relevant influence of acoustic noise on the cerebral processing of somatosensory information. Depending on the nature of acoustic noise we found contrary effects with increased perceptual sensitivity due to *RN* and decreased sensitivity due to *WN*. Our further analyses of the cerebral information processing provide evidence that interactions of acoustic and somatosensory stimuli occur at multiple levels in a complex and spatial distributed network. Subsequent studies could investigate changes in information processing while experiencing different acoustic stimuli.

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 Committee of the Medical Faculty of the Friedrich Schiller University Jena, Germany. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

DS and SB: conceptualization. SB, DS, and MN: methodology, investigation, and writing – original draft. MN and SB: visualization. OW, SB, and CK: supervision. MN, DS, SB, CK, and OW: writing – review and editing. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by the Interdisciplinary Center of Clinical Research of the Medical Faculty Jena.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2022.930932/full#supplementary-material

REFERENCES

Barutchu, A., Danaher, J., Crewther, S. G., Innes-Brown, H., Shivdasani, M. N., and Paolini, A. G. (2010). Audiovisual integration in noise by children and adults. *J. Exp. Child Psychol.* 105, 38–50. doi: 10.1016/j.jecp.2009.08.005

Broadbent, D. E. (1978). The current state of noise research: reply to poulton. *Psychol. Bull.* 85, 1052–1067. doi: 10.1037/0033-2909.85.5.1052

Brodoehl, S., Klingner, C., Stieglitz, K., and Witte, O. W. (2013). Age-related changes in the somatosensory processing of tactile stimulation–an fMRI study. *Behav. Brain Res.* 238, 259–264. doi: 10.1016/j.bbr.2012.10.038

Brodoehl, S., Klingner, C., Stieglitz, K., and Witte, O. W. (2015a). The impact of eye closure on somatosensory perception in the elderly. *Behav. Brain Res.* 293, 89–95. doi: 10.1016/j.bbr.2015.07.014

Brodoehl, S., Klingner, C. M., and Witte, O. W. (2015b). Eye closure enhances dark night perceptions. *Sci. Rep.* 5:10515. doi: 10.1038/srep10515

Brodoehl, S., Witte, O. W., and Klingner, C. M. (2016). Measuring eye states in functional MRI. *BMC Neurosci.* 17:48. doi: 10.1186/s12868-016-0282-7

Burke, D., and Gandevia, S. C. (1988). Interfering cutaneous stimulation and the muscle afferent contribution to cortical potentials. *Electroencephalogr. Clin. Neurophysiol.* 70, 118–125. doi: 10.1016/0013-4694(88)90112-5

Colavita, F. B. (1974). Human sensory dominance. *Percept. Psychophys.* 16, 409–412. doi: 10.3758/BF03203962

Dean, J., Bowsher, D., and Johnson, M. I. (2006). The effects of unilateral transcutaneous electrical nerve stimulation of the median nerve on bilateral somatosensory thresholds. *Clin. Physiol. Funct. Imaging* 26, 314–318. doi: 10.1111/j.1475-097X.2006.00689.x

Dijkerman, H. C., and de Haan, E. H. (2007). Somatosensory processes subserving perception and action. *Behav. Brain Sci.* 30, 189–201. doi: 10.1017/s0140525x07001392

Eickhoff, S. B., Stephan, K. E., Mohlberg, H., Grefkes, C., Fink, G. R., Amunts, K., et al. (2005). A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *Neuroimage* 25, 1325–1335. doi: 10.1016/j.neuroimage.2004.12.034

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

Falchier, A., Clavagnier, S., Barone, P., and Kennedy, H. (2002). Anatomical evidence of multimodal integration in primate striate cortex. *J. Neurosci.* 22, 5749–5759. doi: 10.1523/jneurosci.22-13-05749.2002

Foxe, J. J., Morocz, I. A., Murray, M. M., Higgins, B. A., Javitt, D. C., and Schroeder, C. E. (2000). Multisensory auditory-somatosensory interactions in early cortical processing revealed by high-density electrical mapping. *Brain Res. Cogn. Brain Res.* 10, 77–83. doi: 10.1016/s0926-6410(00)00024-0

Friston, K. (2007). "Chapter 2 Statistical parametric mapping," in *Statistical Parametric Mapping*, eds K. Friston, J. Ashburner, S. Kiebel, T. Nichols, and W. Penny (London: Academic Press), 10–31.

Gescheider, G. A., Barton, W. G., Bruce, M. R., Goldberg, J. H., and Greenspan, M. J. (1969). Effects of simultaneous auditory stimulation on the detection of tactile stimuli. *J. Exp. Psychol.* 81, 120–125. doi: 10.1037/h0027438

Ghazanfar, A. A., and Schroeder, C. E. (2006). Is neocortex essentially multisensory? *Trends Cogn. Sci.* 10, 278–285. doi: 10.1016/j.tics.2006.04.008

Gillmeister, H., and Eimer, M. (2007). Tactile enhancement of auditory detection and perceived loudness. *Brain Res.* 1160, 58–68. doi: 10.1016/j.brainres.2007.03.041

Gleiss, S., and Kayser, C. (2014). Acoustic noise improves visual perception and modulates occipital oscillatory states. *J. Cogn. Neurosci.* 26, 699–711. doi: 10.1162/jocn_a_00524

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* 37, 424–438. doi: 10.2307/1912791

Halpern, D. L., Blake, R., and Hillenbrand, J. (1986). Psychoacoustics of a chilling sound. *Percept. Psychophys.* 39, 77–80. doi: 10.3758/bf03211488

Hong, S. A., Kuziez, D., Das, N., Harris, D., and Brunworth, J. D. (2021). Hazardous sound outputs of white noise devices intended for infants. *Int. J. Pediatr. Otorhinolaryngol.* 146:110757. doi: 10.1016/j.ijporl.2021.110757

Jousmäki, V., and Hari, R. (1998). Parchment-skin illusion: sound-biased touch. *Curr. Biol.* 8:R190. doi: 10.1016/s0960-9822(98)70120-4

Kahneman, D. (1973). *Attention and Effort*. Englewood Cliffs, NJ: Prentice-Hall.

Kayser, C., and Shams, L. (2015). Multisensory causal inference in the brain. *PLoS Biol.* 13:e1002075. doi: 10.1371/journal.pbio.1002075

Kennett, S., Taylor-Clarke, M., and Haggard, P. (2001). Noninformative vision improves the spatial resolution of touch in humans. *Curr. Biol.* 11, 1188–1191. doi: 10.1016/s0960-9822(01)00327-x

Keysers, C., Wicker, B., Gazzola, V., Anton, J. L., Fogassi, L., and Gallese, V. (2004). A touching sight: SII/PV activation during the observation and experience of touch. *Neuron* 42, 335–346. doi: 10.1016/s0896-6273(04)00156-4

Klatte, M., Bergstrom, K., and Lachmann, T. (2013). Does noise affect learning? A short review on noise effects on cognitive performance in children. *Front. Psychol.* 4:578. doi: 10.3389/fpsyg.2013.00578

Klingner, C. M., Nenadic, I., Hasler, C., Brodoehl, S., and Witte, O. W. (2011). Habituation within the somatosensory processing hierarchy. *Behav. Brain Res.* 225, 432–436. doi: 10.1016/j.bbr.2011.07.053

Ma, W. J., Zhou, X., Ross, L. A., Foxe, J. J., and Parra, L. C. (2009). Lip-reading aids word recognition most in moderate noise: a bayesian explanation using high-dimensional feature space. *PLoS One* 4:e4638. doi: 10.1371/journal.pone.0004638

Macaluso, E., and Driver, J. (2005). Multisensory spatial interactions: a window onto functional integration in the human brain. *Trends Neurosci.* 28, 264–271. doi: 10.1016/j.tins.2005.03.008

Maschke, C. (2011). Cardiovascular effects of environmental noise: research in Germany. *Noise Health* 13, 205–211. doi: 10.4103/1463-1741.80150

McGurk, H., and MacDonald, J. (1976). Hearing lips and seeing voices. *Nature* 264, 746–748. doi: 10.1038/264746a0

Moss, F., Chiou-Tan, F., and Klinke, R. (1996). Will there be noise in their ears? *Nat. Med.* 2, 860–862. doi: 10.1038/nm0896-860

Munzel, T., Gori, T., Babisch, W., and Basner, M. (2014). Cardiovascular effects of environmental noise exposure. *Eur. Heart J.* 35, 829–836. doi: 10.1093/eurheartj/ehu030

Munzel, T., Schmidt, F. P., Steven, S., Herzog, J., Daiber, A., and Sorensen, M. (2018). Environmental noise and the cardiovascular system. *J. Am. Coll. Cardiol.* 71, 688–697. doi: 10.1016/j.jacc.2017.12.015

Navarra, J., Soto-Faraco, S., and Spence, C. (2007). Adaptation to audiotactile asynchrony. *Neurosci. Lett.* 413, 72–76. doi: 10.1016/j.neulet.2006.11.027

Othman, E., Yusoff, A. N., Mohamad, M., Abdul Manan, H., Abd Hamid, A. I., and Giampietro, V. (2020). Effects of white noise on word recall performance and brain activity in healthy adolescents with normal and low auditory working memory. *Exp. Brain Res.* 238, 945–956. doi: 10.1007/s00221-020-05765-3

Rausch, V. H., Bauch, E. M., and Bunzeck, N. (2014). White noise improves learning by modulating activity in dopaminergic midbrain regions and right superior temporal sulcus. *J. Cogn. Neurosci.* 26, 1469–1480. doi: 10.1162/jocn_a_00537

Riedy, S. M., Smith, M. G., Rocha, S., and Basner, M. (2021). Noise as a sleep aid: a systematic review. *Sleep Med. Rev.* 55, 101385. doi: 10.1016/j.smrv.2020.101385

Ro, T., Hsu, J., Yasar, N. E., Elmore, L. C., and Beauchamp, M. S. (2009). Sound enhances touch perception. *Exp. Brain Res.* 195, 135–143. doi: 10.1007/s00221-009-1759-8

Ro, T., Wallace, R., Hagedorn, J., Farne, A., and Pienkos, E. (2004). Visual enhancing of tactile perception in the posterior parietal cortex. *J. Cogn. Neurosci.* 16, 24–30. doi: 10.1162/089892904322755520

Rolke, R., Baron, R., Maier, C., Tölle, T. R., Treede, D. R., Beyer, A., et al. (2006). Quantitative sensory testing in the German Research Network on Neuropathic Pain (DFNS): standardized protocol and reference values. *Pain* 123, 231–243. doi: 10.1016/j.pain.2006.01.041

Schlittmeier, S. J., Feil, A., Liebl, A., and Hellbrück, J. R. (2015). The impact of road traffic noise on cognitive performance in attention-based tasks depends on noise level even within moderate-level ranges. *Noise Health* 17, 148–157. doi: 10.4103/1463-1741.155845

Schroeder, C. E., Lindsley, R. W., Specht, C., Marcovici, A., Smiley, J. F., and Javitt, D. C. (2001). Somatosensory input to auditory association cortex in the macaque monkey. *J. Neurophysiol.* 85, 1322–1327. doi: 10.1152/jn.2001.85.3.1322

Serino, A., Bassolino, M., Farne, A., and Ladavas, E. (2007). Extended multisensory space in blind cane users. *Psychol. Sci.* 18, 642–648. doi: 10.1111/j.1467-9280.2007.01952.x

Seth, A. K. (2010). A MATLAB toolbox for granger causal connectivity analysis. *J. Neurosci. Methods* 186, 262–273. doi: 10.1016/j.jneumeth.2009.11.020

Shellock, F. G., Ziarati, M., Atkinson, D., and Chen, D. Y. (1998). Determination of gradient magnetic field-induced acoustic noise associated with the use of echo planar and three-dimensional, fast spin echo techniques. *J. Magn. Reson. Imaging* 8, 1154–1157. doi: 10.1002/jmri.1880080522

Smith, A. P., and Broadbent, D. E. (1981). Noise and levels of processing. *Acta Psychol.* 47, 129–142. doi: 10.1016/0001-6918(81)90004-4

Soderlund, G. B., Sikstrom, S., Loftesnes, J. M., and Sonuga-Barke, E. J. (2010). The effects of background white noise on memory performance in inattentive school children. *Behav. Brain Funct.* 6:55. doi: 10.1186/1744-9081-6-55

Spence, C., and Driver, J. (1997). Audiovisual links in exogenous covert spatial orienting. *Percept. Psychophys.* 59, 1–22. doi: 10.3758/bf03206843

Stansfeld, S. A., Berglund, B., Clark, C., Lopez-Barrio, I., Fischer, P., Ohrstrom, E., et al. (2005). Aircraft and road traffic noise and children's cognition and health: a cross-national study. *Lancet* 365, 1942–1949. doi: 10.1016/S0140-6736(05)66660-3

Stein, B. E. (2012). *The New Handbook of Multisensory Processing*. Cambridge, MA: MIT Press.

Stein, B. E., Meredith, M. A., and Wallace, M. T. (1993). The visually responsive neuron and beyond: multisensory integration in cat and monkey. *Prog. Brain Res.* 95, 79–90. doi: 10.1016/s0079-6123(08)60359-3

Szalma, J. L., and Hancock, P. A. (2011). Noise effects on human performance: a meta-analytic synthesis. *Psychol. Bull.* 137, 682–707. doi: 10.1037/a0023987

Vroomen, J., Bertelson, P., and de Gelder, B. (2001). The ventriloquist effect does not depend on the direction of automatic visual attention. *Percept. Psychophys.* 63, 651–659. doi: 10.3758/bf03194427

Wiesenfeld, K., and Moss, F. (1995). Stochastic resonance and the benefits of noise: from ice ages to crayfish and SQUIDs. *Nature* 373, 33–36. doi: 10.1038/373033a0

Zeng, F. G., Fu, Q. J., and Morse, R. (2000). Human hearing enhanced by noise. *Brain Res.* 869, 251–255. doi: 10.1016/s0006-8993(00)02475-6

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: 04 August 2022
doi: 10.3389/fpsyg.2022.935475

Age-Related Differences in Early Cortical Representations of Target Speech Masked by Either Steady-State Noise or Competing Speech

Bruce A. Schneider[1]*, Cristina Rabaglia[1], Meital Avivi-Reich[1,2], Dena Krieger[1], Stephen R. Arnott[3] and Claude Alain[3,4]

[1]Department of Psychology, Human Communication Laboratory, University of Toronto Mississauga, Mississauga, ON, Canada, [2]Department of Communication Arts, Sciences, and Disorders, Brooklyn College, City University of New York, Brooklyn, NY, United States, [3]Rotman Research Institute, Baycrest Centre, Toronto, ON, Canada, [4]Department of Psychology, St. George Campus, University of Toronto, Toronto, ON, Canada

Edited by:
Patrik Sörqvist,
University of Gävle,
Sweden

Reviewed by:
Erol J. Ozmeral,
University of South Florida,
United States
Ahmad Nazlim Bin Yusoff,
Universiti Kebangsaan Malaysia,
Malaysia

***Correspondence:**
Bruce A. Schneider
bruce.schneider@utoronto.ca

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

Received: 04 May 2022
Accepted: 13 June 2022
Published: 04 August 2022

Citation:
Schneider BA, Rabaglia C,
Avivi-Reich M, Krieger D,
Arnott SR and Alain C (2022)
Age-Related Differences in Early
Cortical Representations of Target
Speech Masked by Either Steady-
State Noise or Competing Speech.
Front. Psychol. 13:935475.
doi: 10.3389/fpsyg.2022.935475

Word in noise identification is facilitated by acoustic differences between target and competing sounds and temporal separation between the onset of the masker and that of the target. Younger and older adults are able to take advantage of onset delay when the masker is dissimilar (Noise) to the target word, but only younger adults are able to do so when the masker is similar (Babble). We examined the neural underpinning of this age difference using cortical evoked responses to words masked by either Babble or Noise when the masker preceded the target word by 100 or 600 ms in younger and older adults, after adjusting the signal-to-noise ratios (SNRs) to equate behavioural performance across age groups and conditions. For the 100 ms onset delay, the word in noise elicited an acoustic change complex (ACC) response that was comparable in younger and older adults. For the 600 ms onset delay, the ACC was modulated by both masker type and age. In older adults, the ACC to a word in babble was not affected by the increase in onset delay whereas younger adults showed a benefit from longer delays. Hence, the age difference in sensitivity to temporal delay is indexed by early activity in the auditory cortex. These results are consistent with the hypothesis that an increase in onset delay improves stream segregation in younger adults in both noise and babble, but only in noise for older adults and that this change in stream segregation is evident in early cortical processes.

Keywords: aging, release from masking, ERP, word in noise, acoustic change complex

INTRODUCTION

Communication in everyday life often requires listeners to navigate complex auditory scenes, full of competing information arriving at the listeners' ears concurrently with the target message. Further, the challenging task of processing soundscapes becomes increasingly difficult as we age. Importantly, older adults exhibit difficulty comprehending speech when competing sounds are present (e.g., Anderson et al., 2018; Avivi-Reich et al., 2018)—holding a coherent conversation in a crowded restaurant with piped-in music, for example, might pose something akin to a herculean task for older adults, even if they possess clinically normal hearing for their age group (Humes, 2020).

In adverse listening situations, listeners must be able to isolate a reasonably veridical sensory representation of the target message, thus allowing further processing to take place. In order to do so, the auditory scene must be parsed into its auditory components (stream segregation, Bregman, 1990), thereby allowing listeners to focus their attention on the target signal. This can be a demanding task, requiring processing at both peripheral and central levels. Sound sources that temporally and spectrally overlap the target signal create excitation patterns in the cochlea and along the auditory nerve that overlap with those of the target signal. This type of interference often is referred to as energetic masking or peripheral masking (e.g., Durlach et al., 2003; Vander Werff et al., 2021). In addition, when the masker contains speech, it is likely to initiate lexical processing that could potentially allow irrelevant content to interfere with the processing of the target message at more central levels. This type of interference is referred to as informational masking (Freyman et al., 1999; Durlach et al., 2003; Schneider et al., 2007, 2010; Kidd et al., 2008; Jagadeesh and Uppunda, 2021), and is thought to affect higher more central processes than energetic masking (Arbogast et al., 2002; Freyman et al., 2004; Ihlefeld and Shinn-Cunningham, 2008; Szalárdy et al., 2019; Vander Werff et al., 2021).

Listeners can alleviate the interference cause by competing sound sources if they are able to segregate the incoming auditory input into separate auditory streams and correctly identify the target stream. Successful stream segregation, leading to a reduction in the interference caused by the maskers, is referred to as “release from masking” (e.g., Brungart et al., 2001; Durlach et al., 2003; King et al., 2020). The ability to do so depends on the perceptual similarities and dissimilarities between the target signal and the other competing sound sources present in the same auditory scene. Any differences among the sound sources could assist stream segregation, thereby providing a release from masking (Bregman, 1990). Different acoustic cues that could assist stream segregation have been previously investigated (e.g., Brungart et al., 2001; Humes et al., 2006; Vongpaisal and Pichora-Fuller, 2007; King et al., 2020; Rajasingam et al., 2021). These cues include acoustic dissimilarities between the target and masker/s (such as differences in F0 and spectrum) and temporal differences in the onset of successive sounds. Beyond these acoustic factors, knowledge-driven or top-down assisting cues, such as expectations, prior exposure and attention have also been found to affect stream segregation (e.g., Shinn-Cunningham and Best, 2008; Ragert et al., 2014).

Older adults, even those who are considered to have normal hearing for their age, show a reduced ability to use certain cues to enhance speech in noise perception (e.g., Dubno et al., 2002; Helfer and Freyman, 2008; Avivi-Reich et al., 2014; Stevenson et al., 2015; Roque et al., 2019). Importantly, all types of maskers do not have a similar effect on listeners across the lifespan. Maskers that seem particularly detrimental as one ages are those that contain competing speech (Tun and Wingfield, 1999; Helfer and Freyman, 2008; Rajan and Cainer, 2008). The disproportional difficulty older adults experience in multi-talker scenes compared with younger adults could be related to difficulties segregating the target stream from competing speech streams due to the acoustic similarity between them. While segregating a speech stream from noise streams, that contain no semantic information and significantly differ acoustically from the target, seems to be relatively automatic and less demanding (Snyder et al., 2006), segregating a target speech stream from other competing speech streams may require more attention and resources and result in less release from masking (Alain, 2004). In addition, it has been suggested that older adults benefit less from acoustic cues and perceptual opportunities, such as an onset delay between speech maskers and the target speech (Ben-David et al., 2012; Getzmann and Näätänen, 2015) compared to young adults. Considering these age-related findings, it is important to further examine how older adults differ from young adults in the ability to release speech from masking when attempting to identify word in noise from different types of maskers, and with different temporal relationships between maskers and target words.

In the present study, we focus on the degree of acoustic similarity between the target and competing auditory inputs, and the temporal cues derived from differences in sound onset. When the onset of a target sound and the onset of one or more competing auditory streams are separated in time, listeners take advantage of this temporal discrepancy to segregate a target sound within an auditory scene (Zwicker, 1965; McFadden and Wright, 1990; Wright, 1997; Wagener and Brand, 2005; Ben-David et al., 2012). However, there is also evidence that acoustic similarity and temporal coherence may interact. Stream segregation is not achieved instantly, and the time it takes for it to develop depends both on the stimuli used as well as on the listener (Bregman and Campbell, 1971; Bregman, 1990). The segregation of target speech from competing speech appears to take longer than the segregation of target speech from competing noise. Ezzatian et al. (2012) found that segregation of a speech target takes longer to complete when masked by other, competing two-talker speech than when masked by steady-state noise. When younger adult participants were asked to identify a target word presented in a semantically anomalous sentence (aka, for which sentential context could not provide a valid clue to the target word’s identity), there was a relationship between the serial position of the target word in the sentence and recognition accuracy—but only when the masker was competing speech, not when the masker was noise.

Ageing, Streaming, and Word in Noise Identification

The impact of both acoustic similarity and temporal factors on auditory stream segregation may change with ageing. Ben-David et al. (2012) asked younger and older adults to repeat single words that were presented with either 100, 225, 350, 600, or 1,100 ms delay from the onset of a masking sound that consisted of either multi-talker babble or steady-state speech spectrum noise. In general, older adults needed a higher signal-to-noise ratio (SNR) to reach 50% word identification accuracy in both maskers, and, in general, longer delays between the target and masker onset resulted in better performance—thresholds decreased exponentially with increased delay between the target and masker. Younger adults, further, exhibited this same relationship between onset delay and performance regardless of masker type (steady-state noise or babble). In contrast, for older adults, target-masker onset delays were only beneficial for the noise masker, where the effect of onset delay was similar for older and younger adults. With babble maskers, older adults appeared unable to take advantage of the delay in onset between the target and the babble (see **Figure 1**). Hence, while older and younger adults are both able to benefit from onset delay when the masker is noise, only younger adults are able to do so when the masker is babble.

In the present study, we use cortical auditory evoked potentials (CAEPs) to examine if these differences in the ability to use temporal cues to stream segregation are reflected in the very early stages of sensory processing. We focus here on the acoustic change complex (ACC), which is elicited by changes within a continuous stream of sounds. The ACC comprises N1 and P2 deflections, analogous to those elicited by sound onset, and is thought to represent early stages of sensory encoding of the stimulus (e.g., Ostroff et al., 1998; Niemczak and Vander Werff, 2019). If there are age-related differences in the unfolding processes of stream segregation, then we might expect these differences to be reflected in a cortical marker (i.e., the ACC) reflecting sensory registration of the speech stimulus embedded in noise. Prior work suggests that the ACC is responsive to masking of the target; in general, adding a competing auditory source to a speech signal delays the N1 peak onset, and reduces N1 peak amplitude (Billings et al., 2009, 2011). However, there is evidence that attentional factors also affect the N1 response. For example, the amplitude of the N1 wave is often larger when attention is directed to speech sounds than during passive listening (Alain et al., 2004; Billings et al., 2011; Zendel et al., 2015). Modulating a masking noise (vs. a steady-state masker) can result in a CAEP-related release from masking, allowing for a detectable CAEP response for modulated maskers where none exists with a steady-state masker when sounds are presented close to threshold (with a target tone; Androulidakis and Jones, 2006). Introducing interruptions to a masker (vs. a continuous masker) can also affect the CAEP for a speech target relative to a continuous masker (Faucette and Stuart, 2017, 2020). Indeed, there is evidence that the release from masking due to certain characteristics of the stimulus is comparable in magnitude between behavioural and electrophysiological domains. Tanner et al. (2019) examined the CAEPs evoked by a /ba/ presented concurrently with either steady-state maskers of 30 and 60 SPL, or a masker that was modulating between the two levels, and examined the electrophysiological and behavioural threshold for detection under all three maskers. The authors found a release from masking of about 13.5 dB in magnitude in both the behavioural and electrophysiological domains. Hence, there is evidence that release from masking can be reflected in changes to the N1 response, and even that these changes may be comparable in magnitude to behavioural release from masking under certain conditions.

The present study used CAEPs to clarify why older adults do not experience an improvement in word recognition with an increase in onset delay between a speech masker and target speech whereas younger adults do. In designing the experiment, we opted to adjust the ratio of the speech target to the masker to produce equivalent word identification scores in all four combinations of Masker Type (Noise vs. Babble) and Onset Delay (100 vs. 600 ms) in both younger and older adults. There were several reasons for doing so. The first reason was to ensure that we would observe a measurable cortical evoked potential. Previous work has suggested it is difficult to measure stable cortical evoked potentials when speech stimuli are presented at threshold levels (in masking: Whiting et al., 1998; Androulidakis and Jones, 2006).

A second reason is that a number of studies have shown that younger and older adults, when tested under identical stimulus conditions, tend to engage different neural mechanisms when performing the same task (see reviews by Wong et al., 2009;

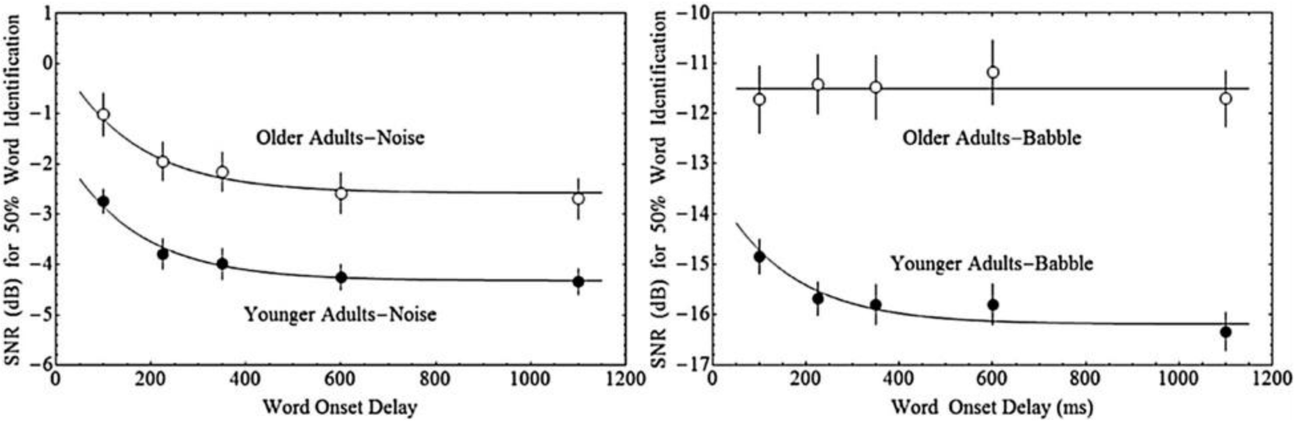


FIGURE 1 | Fifty percent word identification thresholds as a function of the onset delay between the masker and the target word for younger and older adults in either Steady-State Noise, or multi-talker Babble. Adapted with permission from Ben-David et al. (2012).

Velia Cardin, 2016). However, when there are age-related differences in behavioural performance, it becomes difficult to determine the reasons for any associated age differences in neural activity. It could be, for example, that older adults may need to engage different and/or additional neural mechanisms because hearing and/or visual losses make the task more difficult for them. Or, it could be that there are systemic age-related changes in neural functioning that require that different brain mechanisms and/or areas to be engaged to accomplish the task in older adults irrespective of the level of task difficulty. Adjusting, for example, the SNR to produce equivalent levels of behavioural performance (i.e., equivalent task difficulty), can allow us to distinguish between these two different possibilities. In addition, Alain et al. (2004) found that age-related differences in CAEPs were minimized when participants were attending to the auditory stimulus, and performing at equivalent behavioural levels. Hence, the disappearance of age-related differences in neural activity when age-related differences in behavioural performance are eliminated (for instance, by adjusting the SNR) would be consistent with the notion that younger and older adults engage the same neural mechanisms when task difficulty is adjusted to produce equivalent behavioural performance. On the other hand, a finding that age-related neural processing differences persisted after equating younger and older adults with respect to behavioural output, would be consistent with the notion that older adults must engage different neural processes to perform a task. The latter result would suggest that the neural circuitry available to younger adults when required to perform a certain task, such as unmasking an auditory target, is not as available to older adults as it is to younger adults.

A similar argument could be made with respect to the engagement of different brain mechanisms when there is a change in task (e.g., a change in the Onset Delay between masker and speech target). If, after adjusting for behavioural performance across the two delays, we find differences in CAEPs, we can conclude that there are neural processing differences between the two delay conditions that are relatively independent of behavioural performance. Finally, we might find an interaction between masker similarity, onset delay, and age that will be easier to interpret if behavioural performance is equated across all eight combinations of these three factors.

Hence, in the present study, we were searching for neural evidence of the behavioural result that an increase in Onset Delay between a babble masker and a speech target makes listening easier for younger adults but not for older adults. We conducted this search when both age groups had equivalent word identification scores. Any differences in early cortical responses under such circumstances would indicate an age difference in the way speech in babble was processed.

MATERIALS AND METHODS

Participants

Twenty-four younger adults ($M_{\text{age}} = 21.6$; $SD_{\text{age}} = 2.7$; range = 18–27 years) and 24 older adults ($M_{\text{age}} = 72.5$; $SD_{\text{age}} = 5.7$; range = 65–85 years) received a modest stipend in exchange for participation in this study. Younger adult participants were students at the University of Toronto Mississauga; older adults were community-dwelling volunteers. All participants indicated they were native speakers of English who were not fluent in any additional languages, and achieved a minimum score of 9/20 on the Mill Hill vocabulary test (Raven, 1965). Hearing screenings conducted within the year prior to participation showed that all participants had pure-tone air-conduction thresholds within clinically normal limits between 200 and 3,000 Hz (see average hearing thresholds presented in **Figure 2**). In addition, all participants indicated *via* self-report that they were in good health with no history of auditory pathology or neurological trauma. We were unable to obtain readable CAEPs in one younger adult and this participant was excluded from the ERP analysis.

Stimuli and Apparatus

Five hundred and sixteen bi-syllabic recorded words, spoken by a female actor with a southern Ontario accent, taken from Murphy et al. (2000), were used for this experiment. Words were divided into four lists of 128. The word lists were derived from 10 lists featuring the same words, as used in Ben-David et al. (2012) in that the first eight word lists from this study were combined, and the remaining two lists were split in half and added to each list. Thus, the lists were well equated on word frequency, density of lexical neighbourhood, and duration (for further details, see Ben-David et al., 2012). Spoken words were presented to participants on a background of either continuous speech spectrum noise or multi-talker babble taken from the "Revised Speech Perception in Noise" (R-SPIN) test (Bilger et al., 1984). All of the 516 words were equated with respect to root mean square amplitude. Word stimuli were delivered binaurally by converting the digital signal to analogue form (using a 16-bit digital-to-analogue converter TDT DD1), and controlling the analogue output using an Enhanced Real-time Processor (TDT RP2.1) and programmable attenuator (TDT PA5), before delivering the signal to a headphone buffer (TDT HB7) and a Sennheiser HD 265 headphone.

Procedure

Participants were tested individually in a single-walled sound attenuated booth. Each experimental session consisted of four blocks; each block consisted of 128 consecutive single-word trials. For two of the blocks, participants heard the target words masked by multi-talker babble; in the other two blocks, participants heard target words masked by speech-spectrum noise. For each masker type (babble and noise), participants completed one block with a 100 ms delay between the onset of the masker and the subsequent onset of the to-be-repeated word, and one block with a 600 ms delay between the onset of the masker and the subsequent onset of the to-be-repeated word. The order of presentation of the four possible masker/delay conditions (Babble 100 ms; Babble 600 ms; Noise 100 ms; Noise 600 ms) was counterbalanced across participants, with each participant completing a randomly assigned block order set, such that each of the four possible conditions was presented at each of the four possible serial block locations (1st, 2nd,

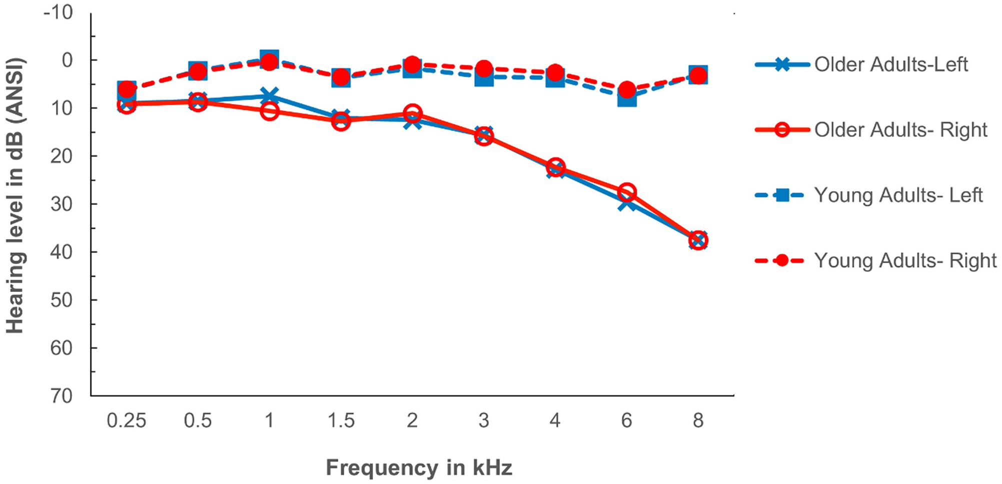


FIGURE 2 | Average audiograms for the two age groups (Young vs. Older adults). Left and right ears are plotted separately.

3rd, or 4th) an equal number of times within each age group, and across the entire sample. The four word lists were always presented in the same order across all participants. In this way, each individual word list was presented in each of the four possible masker/delay combinations an equal number of times.

Participants were told to repeat the word they heard and encouraged to guess if they were somewhat uncertain, but say "pass" if they were very unsure of the word. Participants were not given practice trials or feedback. Optional short breaks were permitted between each of the four blocks. Accuracy was coded during the experimental session by a native English-speaking experimenter who listened to participant responses *via* headphones. After recording the participant response, the experimenter then cued the next word, with a minimum of 250 ms between the end of the participants' utterance and the beginning of the next stimulus in the set.

Words were always presented at 60 dB SPL. The levels of the competing speech or noise were determined according to the following procedure. The psychometric functions reported in Ben-David et al. (2012) were used to determine the SNR that produced a level of 95% correct word recognition for the two groups of participants (young and old) at each of the four conditions in this experiment. These SNRs reflect the average level at which participants in each age group, at each masker-target onset delay, were 95% accurate at determining the target word in each condition, rounded to two decimal places. For younger adults, these levels, in dB, were: Noise$_{100\text{ ms}}$: +6.93; Noise$_{600\text{ ms}}$: +4.95; Babble$_{100\text{ ms}}$: −1.65; Babble$_{600\text{ ms}}$: −2.70. For older adults, these levels, in dB, were: Noise$_{100\text{ ms}}$: +9.41; Noise$_{600\text{ ms}}$: +7.03; Babble$_{100\text{ ms}}$: +3.41; Babble$_{600\text{ ms}}$: +1.46.

Electrophysiology Recording

Neuroelectric brain activity was recorded continuously using a 128-channel HydroCel Geodesic Sensor Net (EGI technology) with a sample rate of 500 Hz. During recording, data were referenced to Cz with a bandpass of DC-100 Hz, and stored for offline analysis. EEG recordings were preprocessed offline using Brain Electrical Source Analysis software (BESA Research version 7.0; MEGIS GmbH, Gräfelfing, Germany).

EEG Preprocessing

All trials, regardless of behavioural accuracy, were included in the EEG analysis. The EEG data were visually inspected to identify segments contaminated by defective electrodes. Noisy electrodes were interpolated using data from the surrounding electrodes, and no more than eight electrodes were interpolated per participant. The EEG was then re-referenced to the average of all electrodes and digitally filtered with a 1 Hz high-pass filter (forward, 6 dB/octave) and 30 Hz low-pass filter (zero phase, 24 dB/octave), which was identical to the filters used by others (Billings et al., 2011). For each participant, a set of ocular movements was identified from the continuous EEG recording and used to generate spatial components to best account for eye movement artefacts. The spatial topographies were then subtracted from the continuous EEG to correct for lateral and vertical eye movements as well as for eye blinks. The data were parsed into 700 ms epochs that were time-locked to either noise onset or word onset, including 100 ms of pre-stimulus activity (which served as the baseline). Epochs with EEG signal exceeding ±60 µV were marked and excluded from further analysis. The processed data consisted of a minimum of 75% of the epochs per experimental condition and participant for the young and older adult group. The epochs were averaged according to the experimental conditions: babble noise, speech-spectrum noise; 100 ms noise preceding word onset; and 600 ms noise preceding word onset. Each average was then baseline-corrected with respect to the 100 ms pre-stimulus baseline interval.

The effect of noise type and noise duration on CAEP amplitude and latency was quantified using 15 electrodes over the midline central and fronto-central scalp area. This cluster of electrodes best capture the dominant (i.e., tangential orientation) source activity for N1 and P2 waves from the auditory cortices located in the superior temporal gyrus. For the 600 ms noise duration, the N1 and P2 was measured related to word onset. The N1 peak latency and amplitude was defined as the maximum negativity between 100 and 250 ms. The P2 peak latency and amplitude was defined as the maximum positivity between 200 and 400 ms.

For the 100 ms delay between masker and word onset, the N1 and P2 waves elicited by noise onset partly overlapped with those elicited by the word because of the short delay between masker and word onset. To isolate the response to the word as much as possible, we subtracted the auditory evoked responses elicited in the 600 ms condition from the 100 ms condition. The masker onset and masker duration were identical in both conditions, with the only difference being the presence of a spoken word starting at 100 ms in the 100 ms condition. This subtraction procedure is based on the assumption that the auditory evoked responses elicited by the noise and word onset sum together linearly. The difference wave is thought to index processing of the masked word in the 100 ms condition, with the response related to masker onset removed. We then measured the N1 and P2 peak latency and amplitude from this difference waveform. The N1 peak latency and amplitude were defined as the maximum negativity between 200 and 350 ms after masker onset. The P2 peak latency and amplitude were defined as the maximum positivity between 300 and 500 ms after masker onset. Because the word was presented 100 ms after masker onset, we subtracted 100 ms from the peaks found in the difference waveform to allow comparison with the 600 ms delay condition.

RESULTS

Behavioural Results

After each experimental session, a native English speaker scored the accuracy of each participant by listening to an audio recording of each session. Average percentage agreement for the online coding by the experimenter and the offline coding by the second scorer was 98.5% for younger adults (min = 93.8%) and 98.7% for older adults (min = 93.0%). For individual word trials where the two accuracy scores disagreed, a third rater listened to the recording and the judgement (correct or error) endorsed by two out of the three scorers was used; these resolved accuracy totals were retained as the accuracy scores for each individual. Average percentage correct for each age group and condition are displayed in **Table 1**.

Participants from both age groups performed within two percentage points of 95% correct for all conditions. Since word identification performance for all groups was centred at an extreme end (i.e., 95% correct) of the percentage scale, a Stevens arcsine transform was used to convert word identification performance into sau units (see Sherbecoe and Studebaker, 2004). A 2 (masker type) by (2 onset delay) by 2 (age group) ANOVA was performed on these transformed values with Onset Delay and Masker Type as within-subject factors, and Age Group as a between-subjects factor. This analysis revealed no main effect of masker, $F(1,46) = 1.041$, $p = 0.313$, indicating that word identification score did not differ overall across the two types of maskers. There was a main effect of delay, $F(1, 46) = 16.64$, $p < 0.001$, reflecting that word identification performance—collapsing across masker type and age group—was statistically significantly higher in the 100 ms delay condition ($M = 87.29$, $SE = 0.478$) than the 600 ms delay condition ($M = 85.93$, $SE = 0.448$); however, the magnitude of this discrepancy equates to a difference in word identification accuracy of less than one word out of the 128 total words per list. This effect of delay, additionally, did not vary between the age groups, $F(1,46) < 1$, nor between masker types, $F(1,46) = 2.089$, $p = 0.155$. There was a significant main effect of age group on overall word identification performance, $F(1,46) = 4.48$, $p = 0.04$, reflecting that older adults word identification performance ($M = 85.69$, $SE = 0.611$), collapsing across maskers and delays, was slightly worse than younger adults ($M = 87.53$, $SE = 0.611$). Again, however, the magnitude of this difference was such that older adults, collapsing across masker condition and delay duration, identified on average only 1.79 fewer words per condition block than younger adults. The effect of age group on word identification performance did not differ across the masker types, $F(1, 46) < 1$, $p = 0.907$, or onset delay (see above). The three-way interaction between masker, delay, and age group was also not significant, $F(1,46) < 1$.

TABLE 1 | Average percentage correct word identification performance by Condition (Noise or Babble) and Delay (100 or 600 ms) for all 24 Older and 24 Younger Participants.

	Noise		**Babble**	
	100 ms	**600 ms**	**100 ms**	**600 ms**
Younger Adults	96.97	95.80	96.84	96.45
Older Adults	95.55	94.53	95.61	94.76

Electrophysiological Results

The impetus for this study was to search for electrophysiological correlates of the effects of release from masking due to an increase in the onset delay between masker and target words (from 100 to 600 ms) on the initial processing of words heard in two kinds of acoustic interference (noise vs. babble) for younger and older adults. A previous behavioural study found a significant release from masking with an increase in onset delay for young adults in Noise, young adults in Babble, and older adults in Noise, but not for older adults in Babble (Ben-David et al., 2012).

In both groups and in all experimental conditions, words presented with a masker generated an ACC that comprised N1 and P2 waves that peaked at central sites. **Figure 3** shows the group mean ACC from the midline fronto-central electrode in young and older adults as a function of the experimental conditions.

When the word is masked by Noise at either Onset Delay, there does not appear to be any age differences with respect to the N1 peak. With respect to the P2 wave, it appears to peak later in older than in younger adults with the amplitude of the peak being greater in older adults at both Onset Delays.

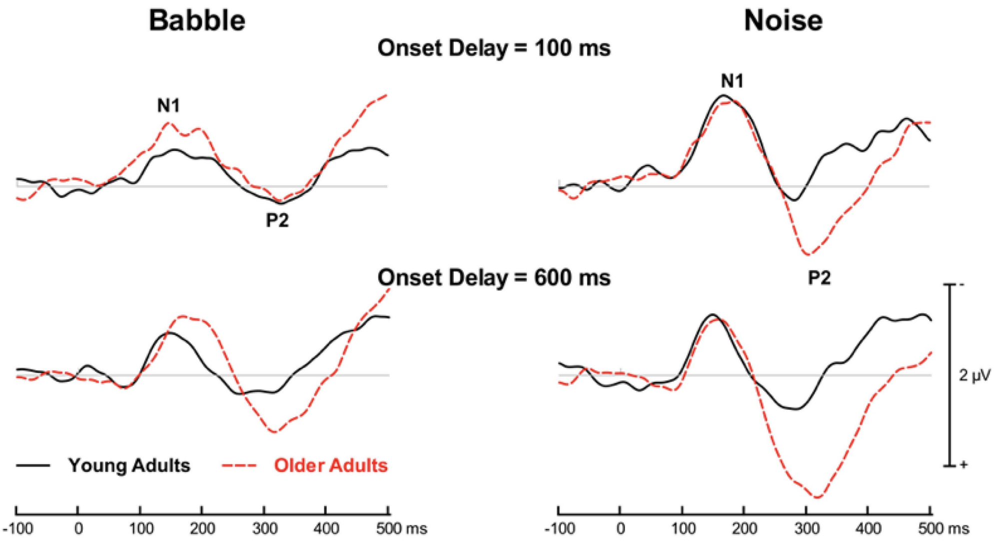


FIGURE 3 | Group mean event-related potentials time-locked on word onset. Traces are shown when the target word was preceded by 100 ms of masker (top row) or 600 ms of masker (bottom row). The electrophysiological responses to the target word embedded in babble are shown on the left whereas those obtained when the word was embedded in speech spectrum noise (i.e., noise) are shown on the right.

When the masker is Babble and the Onset Delay is 100 ms, the N1 peak latency does not appear to differ between younger and older adults, although the magnitude of the N1 peak appears to be larger in older adults. There also does not appear to be any significant differences in the location and amplitude of the P2 in Babble when the delay is 100 ms. However, when the Onset Delay is 600 ms, both the N1 and P2 waveforms appear to peak later in older than in younger adults with slightly higher amplitudes in both cases.

The traces in **Figure 3** indicate that both the amplitude and locations of the N1 and P2 peaks differ with respect to the Age Group to which participants belong, and that the extent of this difference differs with both the Type of Masker and Onset Delay. Hence, the degree of release from masking that presumably occurs with an increase in Onset Delay from 100 to 600 ms in both maskers for young adults, but only in Noise for older adults, can differentially affect both the time between the onset of the word and the peak of each wave (its latency), as well as the amplitude of the electrophysiological response at its peak (its amplitude). Estimates of these two parameters of N1 and P2 were obtained as described in the Methods Section. Because the factors in this experiment can affect the two parameters of the waveforms in different ways, we looked for a way to simultaneously represent both parameters of a waveform together.

Specifically, we computed 95% confidence limits for both N1 and P2 latency and amplitude in each group by experimental condition combination: (young-noise-100, young-noise-600, young-babble-100, young-babble-600, old-noise-100, old-noise-600, old-babble-100, and old-babble-600). For example, to represent the joint effects of latency and amplitude on N1 for young adults in Noise, at a delay = 100 ms, we constructed a rectangle in a two-dimensional plot whose *x*-axis was the latency of the peak of the N1 wave relative to the onset of the target word, and whose *y*-axis was its amplitude. This rectangle is labelled as YN and appears in grey in the left-hand panel of **Figure 4** in the section reserved for the N1 waveform. The 95% CI for latency is specified by the *x* coordinate of the right-hand side of the grey rectangle minus the *x* coordinate of the left-hand side of the rectangle. The 95% CI for peak amplitude is specified by the difference between the *y* coordinate of the upper boundary of the grey rectangle minus its lower boundary. The probability that both the population mean latency and population mean peak for this group fall within this rectangle is 0.95*0.95 = 0.9025.

Now if two of the groups had the same population mean latency and the same population mean peak amplitude, we would expect considerable overlap between the two rectangles. Conversely, if the two groups had significantly different population mean latencies, and/or significantly different population mean peak amplitudes, we would expect to find very little overlap between the rectangles for these two groups.

In the rest of the left panel of **Figure 4**, we plot the eight rectangles representing the joint CIs for latency and amplitude when the target word was presented with a Delay = 100 ms for combinations of 2 Age Groups (Young-Old) × 2 Maskers (Noise, Babble) × 2 waveforms (N1, P2). The right-hand panel of **Figure 2** plots the rectangles for the same eight combinations when the delay was 600 ms. The left-panel shows that there is considerable overlap at an onset delay of 100 ms among the four groups for both the N1 and P2 waves.

Given that we have adjusted the SNR to produce equivalent behavioural results in both groups and masker types for the N1 and P2 waves, this is what we would expect if the amount of release from masking (if any) at a delay of 100 ms were the same in all cases. The results for a word onset delay of

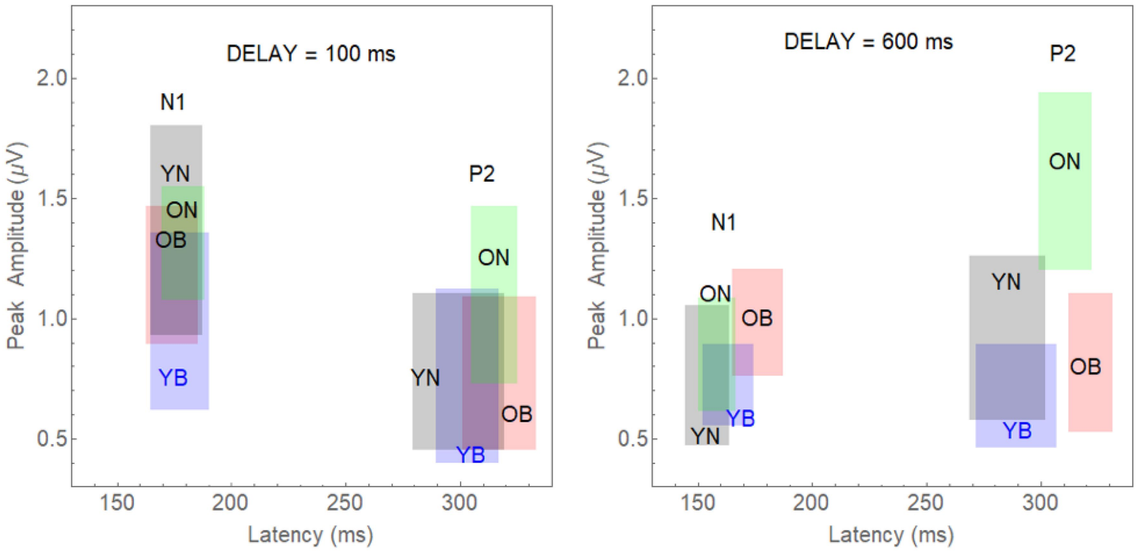


FIGURE 4 | The 95% joint CI rectangles for latency and amplitude of the N1 and P2 waves in Younger and Older Adults for the Noise and Babble Maskers and 100 and 600 ms Delays. **Left-panel**: 95% CI rectangles for a target word onset delay of 100 ms. Separate rectangles are shown for the two Age Groups reporting heard words in both Noise and Babble for the two different waves (N1 and P2). The grey rectangles depict the CIs for young adults in noise (YN); the blue rectangles are for young adults in babble (YB). The green rectangles are for old adults in noise (ON), and the red rectangles are for old adults in babble (OB). **Right-panel**: The equivalent rectangles for a target word onset delay of 600 ms.

600 ms differ substantially from those found for an onset delay of 100 ms. First, for the N1 wave, the CI rectangles for three of the rectangles (young and older adults in noise, and younger adults in babble) appear to overlap substantially among each other, with all three of them overlapping with the rectangle representing older adults in babble to a much lesser extent. Second, for the P2 wave, the CI rectangles for young adults in both babble and noise overlap with one another, with neither overlapping with either older adults in noise or in babble. In addition, the latter two rectangles (older adults in noise or in babble) also do not overlap with one another. **Table 2** quantifies the amount of overlap among the four rectangles comparing N1 and P2 outcome measures for both delays of 100 and 600 ms (see **Appendix** for how this was computed). For example, for N1, at an Onset Delay of 100 ms, the first entry in the cell defined by row YN and column YB (0.432) specifies the probability of finding both of the population means for the YN group within the CI rectangle corresponding to YB group ($p\left[\mu_{x,YN}\mu_{y,YN}\right]$ falling in the overlap of the confidence rectangle for the YN with the confidence interval for YB). The second entry in that cell (0.543) specifies the probability of finding both of the population means for the YB group within the confidence interval rectangle corresponding to YN group ($p\left[\mu_{x,YB}\mu_{y,YB}\right]$ falling in the overlap of the confidence rectangle for the YN with the confidence interval for YB). In general, the top entry in each of the cells in the table specifies the probability of the joint appearance of the population means for the row condition appearing in the confidence rectangle for the column condition ($p\left[\mu_{x,rowcondition}\mu_{y,rowcondition}\right]$ falling in the overlap of the confidence rectangle for the row condition with the confidence interval for the column condition). The bottom entry specifies the probability of the joint appearance of the population means for the column condition appearing in the CI for the row condition ($p\left[\mu_{x,columncondition}\mu_{y,columncondition}\right]$ falling in the overlap of the confidence rectangle for the row condition with the CI for the column condition). These two probabilities can range between 0 and 0.9025 (0 if there is no overlap between YN and YB, and 0.9025 if the CI rectangles are identical for YN and YB). The actual degree of overlap shown in **Figure 4** for these two rectangles is based on the confidence intervals constructed from the results of 23 younger adults tested in both Babble and Noise. As such this confidence interval rectangle will vary from experiment to experiment. In the **Appendix**, we derive the probability values expected under the null hypothesis when the population means for YN and YB are identical. In order to reject this null hypothesis at the $\alpha = 0.05$ level requires that these two probabilities are less than 0.059. Clearly, this is far from being the case in the present example.

An examination of **Table 2** confirms the visual impression that when the onset delay is 100 ms, we cannot reject the null hypothesis that the intersection of the confidence interval rectangles for any of the pairwise comparisons among the four rectangles occurs because the two groups have the same population mean latency and population mean peak amplitude. This holds for both N1 and P2.

However, for an onset delay of 600 ms, the null hypothesis is rejected in all of the comparisons involving older adults in babble for the N1 wave. For the P2 wave, the null hypothesis is rejected for all of the comparisons except for the comparison of the YN and YB CI rectangles. To identify the reasons for this result, we note that in **Figure 4**, going from on onset

TABLE 2 | The degree of overlap of the Pairs of CI Rectangles at two delays for two different waveforms.

	N1: Delay = 100				**P2: Delay = 100**		
	YB	**ON**	**OB**		**YB**	**ON**	**OB**
YN	0.432	0.601	0.607	YN	0.752	0.152	0.382
	0.543	0.888	0.863		0.877	0.386	0.552
YB		0.237	0.586	YB		0.206	0.563
		0.588	0.805			0.334	0.440
ON			0.782	ON			0.435
			0.563				0.453
N1: Delay = 600				P2: Delay = 600			
	YB	ON	OB		YB	ON	OB
YN	0.469	0.642	0.000***	YN	0.377	0.001***	0.000***
	0.486	0.820	0.000***		0.718	0.001***	0.000***
YB		0.618	0.095*	YB		0.000***	0.000***
		0.557	0.057**			0.000***	0.000***
ON			0.016**	ON			0.000**
			0.012**				0.000**

YN is Young-Noise, YB is Young-Babble, ON is Old-Noise, and OB is Old-Babble. *$\alpha = 0.10$; **$\alpha = 0.05$; ***$\alpha = 0.025$.

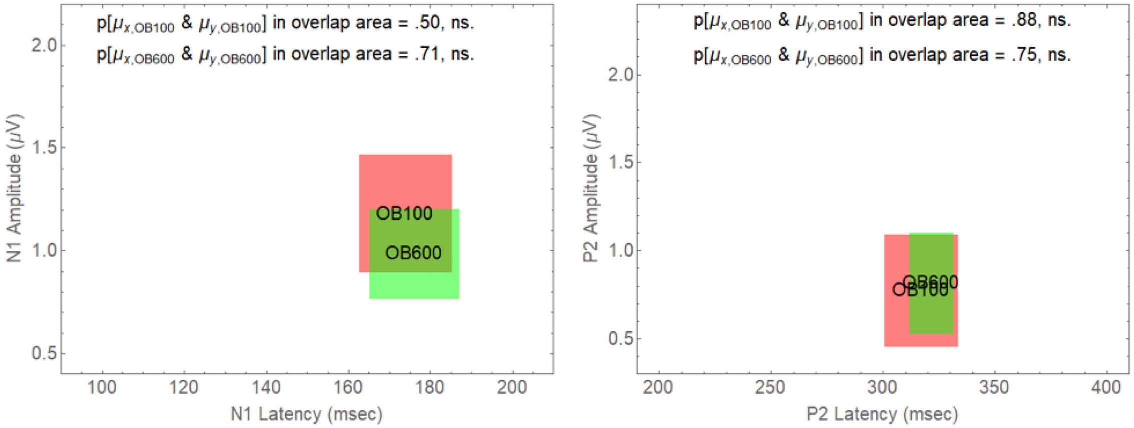


FIGURE 5 | Overlap between the CI rectangles for the 100 and 600 ms delay conditions for older adults in Babble. **Left Panel**: CI rectangles for older adults in the Babble condition for onset delays of 100 and 600 ms for the N1 wave. **Right Panel**: CI rectangles for older adults in the Babble condition for onset delays of 100 and 600 ms for the P2 wave.

delay of 100–600 ms appears to shift three of the rectangles (YN, YB, and ON) away from the position occupied by the CI rectangle for older adults in Babble, which appears to maintain its position for both the N1 and P2 waves. To confirm this visual impression from **Figure 4**, we plotted, in **Figure 5**, the locations for the rectangles for older adults in babble for onset delays of 100 and 600 ms for N1 (left panel) and P2 (right panel). This figure indicates that we cannot reject the null hypothesis that the CI rectangle for older adults in babble is independent of onset delay for both for the N1 and P2 waves.

To further explore how the difference in onset delay differentially affects younger and older adults, in **Figure 6**, we have plotted how age and onset delay affect the confidence rectangles of older adults for the N1 wave (panels A–D) and the P2 wave (panels E–H). Panel A and B shows that for both younger and older adults we cannot reject the null hypothesis that the population means in Noise are the same as they are in Babble when the onset delay is 100 ms. We also cannot reject this null hypothesis for younger adults for an onset delay of 600 ms (panel D). However, the null hypothesis is rejected for older adults at a delay of 600 ms (panel C). Panels E–H show that this same pattern holds for younger and older when the P2 wave is considered.

Figures 5, 6 considered together are consistent with the following hypothesis derived from the behavioural results relating thresholds to delay in younger and older adults. If increases in onset delay do not release older adults from masking when the masker is Babble, then early cortical processing (the N1 and P2 waves) of target words in Babble for older adults should be unaffected by delay as is found in **Figure 5**. In

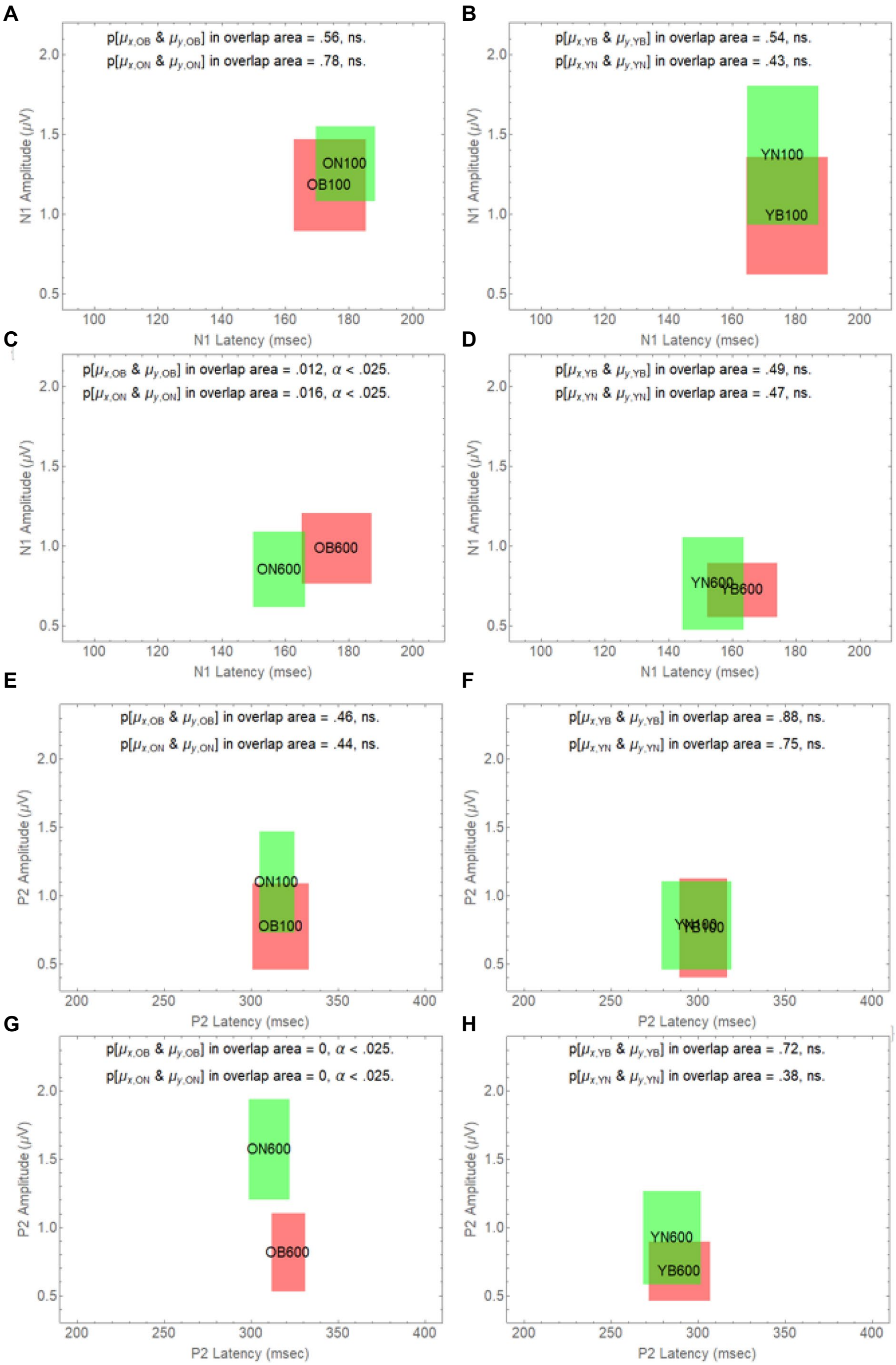


FIGURE 6 | Degree of overlap between Noise and Babble for Onset Delays of 100 and 600 ms for N1 and P2 waves in Younger and Older Adults. The overlap between masker condition (N vs. B) is depicted in separate panels **(A-H)** for all possible combinations of delay (100 vs. 600) and age (O vs. Y), for the two waveform components (N1 and P2).

addition, if increases in onset delay releases older adults in Noise but not in Babble, then the CI rectangle should be different for older adults in Noise than it is for them in Babble when the onset delay is 600 ms. However, the CI rectangles for older adults in noise should be equivalent to those in Babble when the delay is 100 ms because there is no release from masking for either Noise or Babble for a delay of that magnitude. Moreover, this prediction should hold true for both N1 and P2 waves. Conversely, because an increase in onset delay releases younger adults from masking for both Noise and Babble maskers, we would not expect CI rectangles to differ significantly for Noise and Babble maskers as a function of delay and ACC components (i.e., N1 and P2) in younger adults. **Figure 6** confirms this prediction.

DISCUSSION

Prediction Derived From a Release From a Masking Model for Early Cortical Processing of the Acoustic Signal

We have derived several predictions concerning the results of the present experiment based on the data from Ben-David et al. (2012). Those investigators measured 50% thresholds for the identification of words in two kinds of maskers (Babble and Steady-State Noise) as a function of the delay between the onset of the masker and the onset of the target word for both younger and older adults (**Figure 1**). Older adults require greater SNR at all Onset Delays tested for both types of maskers (see the section "Materials and Methods"). Thresholds decrease with increasing Onset Delay for young adults in both kinds of maskers, but only in Noise maskers for older adults. For older adults in a Babble masker, thresholds do not change as a function of Onset Delay.

Recall that, in the present experiment, SNR were adjusted to produce a word identification rate of approximately 95% in all four conditions in both younger and older adults. Now suppose that in the Onset Delay = 100 ms conditions, this SNR manipulation resulted in word identification being equally difficult in the babble and noise maskers for both younger and older adults for both waveforms as suggested by the Ben-David et al. (2012) results. In that instance, we would expect the early cortical processing of the acoustic signal (the N1 and P2 waves) to be identical. However, given the Ben-David et al. (2012) results, we would expect that the word-identification for younger adults should be easier for an Onset Delay of 600 ms than for an Onset Delay of 100 ms for both Noise and Babble Maskers, and for older adults only when the masker is Noise. Hence, word-identification difficulty should be equivalent in the following conditions: (1) Young adults in Noise, Onset Delay = 100 ms; (2) Young adults in Babble, Onset Delay = 100 ms; (3) Older adults in Noise, Onset Delay = 100 ms; (4) Older adults in Babble, Onset Delay = 100 ms; and (5) Older adults in Babble, Onset Delay = 600 ms. Moreover, these predictions should hold true for both N1 and P2 waves.

To evaluate these predictions, we constructed 95% CIs for the population means for latency and peak amplitude in each of the eight combinations of Age Group (Young, Old), Onset Delay (100 vs. 600 ms), and Type of Masker (Noise Babble) for both N1 and P2 waves. The two confidence intervals for a Group were used to define CI rectangles for that group in a two-dimensional space where the abscissa (*x*-axis) is the latency of the peak cortical response (either N1 or P2 waves) from word onset, and the ordinate (*y*-axis) is the amplitude associated with the peak in question. In this two-dimensional space, the width and location of the rectangle along the latency dimension corresponded to the CI for latency, and the extent and location of the rectangle along the amplitude dimension corresponded to the CI for peak amplitude.

Figure 4 plotted these CI rectangles in the two-dimensional space where the *x*-axis is the latency of the response (the time from word onset to the peaks of either the N1 and P2 wave), and the *y*-axis is the amplitude of the respective peaks. The degree of overlap among the four rectangles (YN, YB, ON, and OB) in **Figure 4** (delay = 100 ms) and the associated analysis supports the prediction that the early cortical responses (N1 and P2) are similar among these four rectangles. **Table 2** shows that the null hypothesis that the population means for latency and peak amplitude are the same for each of the six pairings of these rectangles cannot be rejected for either the N1 and P2 waves when Onset Delay = 100 ms. The fact that the CI rectangles for N1 and P2 waves for Older adults in Babble for a 100 ms Delay overlap those for the same rectangles for a 600 ms delay is consistent with the prediction that word identification is equally difficult for Older adults in Babble for a 600 ms Delay as it is for Older Adult in Babble when the Delay is 100 ms (see **Figure 5**).

Table 2 also indicates that, for the N1 wave at an onset delay of 600 ms, the overlap is significantly diminished when the OB group is compared to the other three groups, but that the null hypothesis cannot be rejected that the same population means for latency and amplitude can account for each pairing of the remaining three rectangles (YN, ON, and YB). Finally, the Ben-David et al. (2012) data predict that word identification difficulty should be the same for younger adults in both Noise and Babble when the Onset Delay = 100 ms because there is no release from masking with this Onset Delay. The Ben-David et al. (2012) results also predict that word identification difficulty should be the same for younger adults in both Noise and Babble when the Onset Delay is 600 ms because a 600 ms Onset Delay releases younger adults from masking for both Noise and Babble maskers. For older adults, the Ben-David et al. (2012) results predict the same for older adults when the Onset Delay is 100 ms because there is no release from masking for this Onset Delay for both types of maskers. However, when the Onset Delay = 600 ms, word identification should be easier for a Noise masker than for a Babble masker because there is a release from masking for older adults when the masker is Noise, but not when the masker is Babble. The CI rectangles in **Figure 6** support this prediction. For younger adults, there is significant overlap for Babble and Noise maskers for both Onset Delays. However, for older adults, while there is significant overlap between CI rectangles for Noise and Babble maskers for Onset Delay = 100 ms, the null hypothesis that the same population means can account for both Noise and Babble maskers when the Onset Delay is 600 ms is rejected, indicating that cortical processes for these two maskers is not the same when the Onset Delay is 600 ms. For older adults at Onset Delay 600 ms, the CI rectangle for a noise masker when examining the N1 wave occurs at a shorter latency than that that for the Babble masker. When examining the P2 wave, older adults at Onset Delay = 600 ms have higher amplitudes when the Masker is Noise than when it is Babble. This may reflect difference in attentional allocation or listening effort, with older adults paying more attention when the word is embedded in babble than in noise. This account is consistent with prior research indicating that the P2 amplitude is larger when words are familiar or could easily be identified (e.g., Faucette and Stuart, 2017).

Potential Cautions Associated With These Analyses

As noted in the section "Materials and Methods," the analyses related to Onset Delay = 100 ms are based on a difference waveform in that the first 600 ms of the ERP waveform when the Onset Delay was 600 ms was subtracted from the ERP waveform to when the Onset Delay was 100 ms. The failure to find any differences among the four groups in this difference waveform could be attributed to the increased variability in this difference waveform due to the subtraction process. In this experiment, two measures (latency and amplitude) were collected for each of two waveforms (N1 and P2) for two Age Groups attempting to identify words masked by either Noise or Babble when the Onset Delay was set to 100 ms. Hence, there were 16 measures of performance taken at an Onset Delay of 100 ms that were based on a difference waveform. There were also 16 measures of performance taken when the Onset Delay was set to 600 ms. For these 16 measures, we computed the ratio of the variance of each measure taken using an onset delay of 100 ms to its corresponding measure taken using a delay of 600 ms. If measures taken at an Onset Delay of 100 ms are more variable than comparable measures taken at an Onset Delay of 600 ms, their ratios of their variances should be greater than 1.0. Of the 16 measures, the ratio of variances was greater than 1 in 12 of them, which is significant at the 0.05 level (two-tail test). However, the average ratio of the 16 variances was only 1.66, suggesting that although there is increased variance for the measures taken with an Onset Delay of 100 ms when compared to the variance of measures taken with an Onset Delay of 600 ms, the increase in variance is not very large.

Nevertheless, the increased variance of the measures taken with an Onset Delay = 100 ms relative to the comparable measures taken with an Onset Delay = 600 ms could be responsible, in part, for the fact that no differences were found among the four groups in both N1 and P2 waves when the Onset Delay was 100 ms. However, a consideration of the pattern of results suggest that the effect of increased variance did not substantially affect the pattern of results found in this experiment.

First, if the failure to find differences among the conditions in the Onset Delay = 100 ms analyses were due to noisiness in the ERP difference wave, we would not expect to find correlations between Babble and Noise latencies, or between Babble and Noise peaks. **Figure 7**, however, shows that positive correlations are found in both younger and older adults between Noise and Babble latencies and Noise and Babble peaks for both age groups for both N1 and P2 waveforms. Moreover, three of these correlations were highly significant ($p < 0.005$ for all three correlations), while a fourth was marginally significant ($p = 0.076$). If the difference waveform was highly variable, we would not expect to find such correlations between Noise and Babble latencies and between Noise and Babble peak amplitudes.

Second, it is unlikely that we would find, as predicted by the Ben-David et al. (2012) study, that the CI rectangles for the Babble masker occupied the same position in the two-dimensional CI space for Onset Delays of 100 and 600 ms for both N1 and P2 waves (see **Figure 5**).

Third, it also unlikely that we would find, as predicted by the Ben-David et al. (2012) study, that the CI rectangles for Noise and Babble in younger adults would overlap for Onset Delays of 600 ms for both N1 and P2 waveforms (see **Figure 6**, panels D and H), but not as much for older adults in the same two conditions (**Figure 6**, panels C and G).

These three factors make it less likely that the failure to find differences among the four groups for an Onset Delay of 100 ms for both N1 and P2 waves is simply due to an increased variability in the difference wave that is used to determine N1 and P2 peak latencies and peak amplitude for an Onset Delay = 100 ms.

CONCLUSION

The N1 and P2 waves in the ACC are thought to be associated with early acoustical processing of the auditory target in a noisy background. The Ben-David et al. (2012) results indicate that increasing the Onset Delay between the masker and the target word from 100 to 600 ms results in a release from masking in both Noise and Babble for young adults, but only in Noise for older adults. If the release from masking due to Onset Delay occurs in the early stages of cortical processing, then we would expect N1 and P2 waves to be similar for both younger and older adults when the Onset Delay was too short (100 ms) to produce a release from masking—provided that the SNR was adjusted to produce equivalent per cent correct word identification in both younger and older adults, as they were in all of the conditions of this experiment. The results of this experiment confirmed this prediction. If a longer Onset Delay (600 ms) resulted in a release from masking for younger and older adults in Noise, but only for younger adults in Babble, and if the release from masking occurred in the early stages of cortical processing of the target word, we would expect to see evidence for this in both the N1 and P2 waves. This prediction was also confirmed in that we found both the N1 and P2 waveforms of older adults in Babble did not change when the Onset Delay was increased from 100 to 600 ms (see **Figure 5**). In addition, for younger adults, the two waveforms for Noise and Babble maskers were same when the Onset Delay was 600 ms, whereas they were different in older adults (see **Figure 6**). We would expect the waveforms for Babble and Noise to be the same in younger adults if they were experiencing the same amount of release from masking at this delay. Conversely, in older adults, we would expect the waveforms to be different at this delay if they were experiencing a release from masking in Noise but not in Babble. Finally, we would expect the waveforms in those conditions in which there was a release from masking at an Onset Delay of 600 ms (Young-Noise,

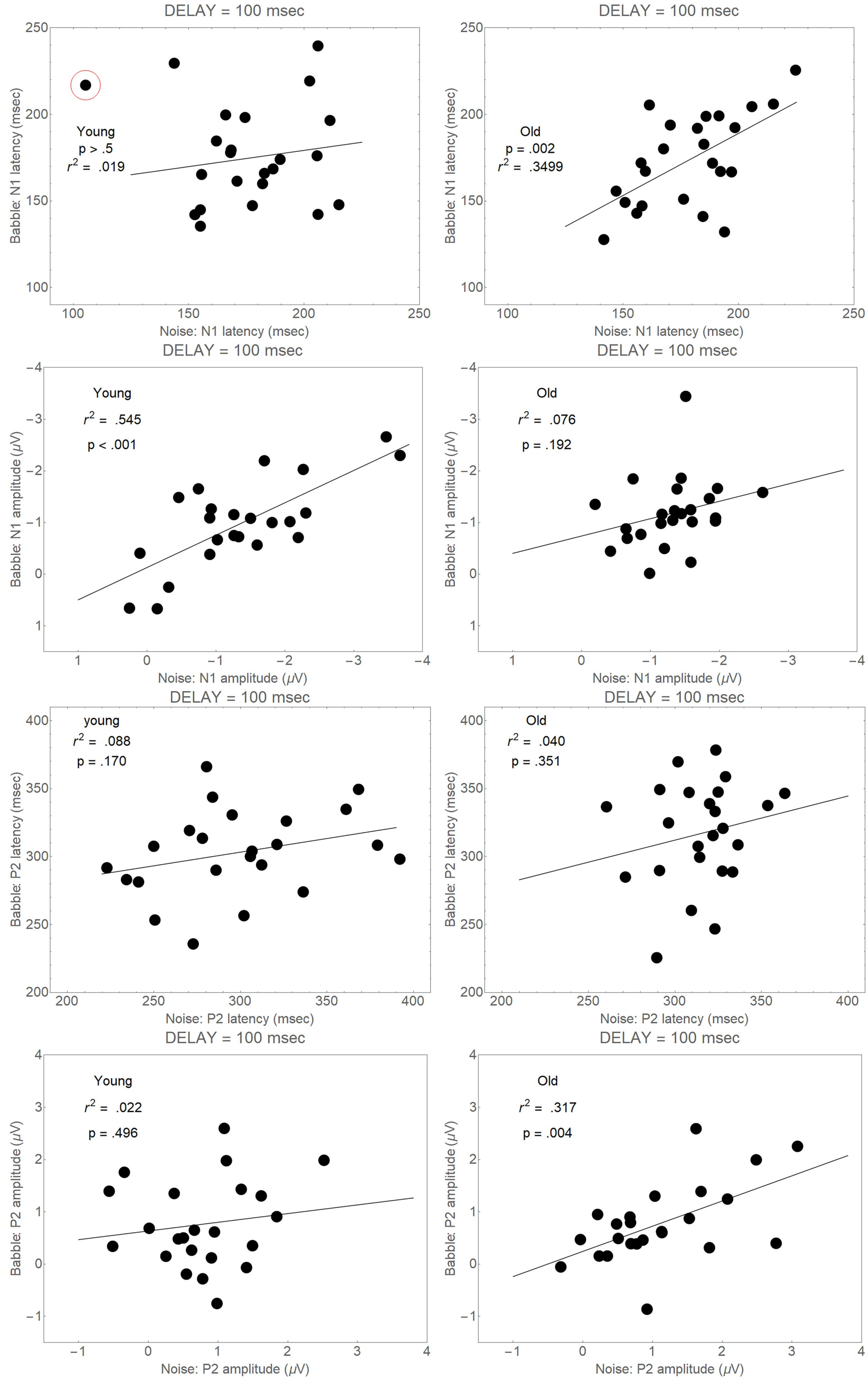


FIGURE 7 | Correlations between Noise and Babble latencies and between Noise and Babble peaks for Onset Delay = 100 ms for both N1 and P2 peaks. The circled point in the top left-hand panel was not included when the correlations were conducted because it was too close to the lower boundary for latency and the upper boundary for amplitude.

Young-Babble, and Old-Noise) to be distinct (have different latencies and peak amplitude) from those in which there was no release from masking (Old Babble). **Figure 4** (right panel) and **Table 2** indicate that the results of this experiment support this prediction. Hence, the electrophysiological results strongly indicate that the automatic release from masking (the segregation of the target word from the masker background) due to an Onset-Delay between Masker and Word Target, occurs in the early stages of acoustic processing of the word.

Because this release from speech masking takes place in the early stages of cortical processing and is likely to be automatic, the only remedy available to health-care practitioners to help older adults compensate for this specific age-related deficit is to improve the SNR. This can be accomplished by (1) manipulating the acoustic scene to shield older adults from competing speech, and/or (2) using directional microphones and/or noise reduction technology to improve the SNR. Otherwise, older adults are likely to continue struggling with speech recognition in the presence of multiple competing talkers, in part, because of their limited ability to use temporal onset cues to facilitate stream segregation so that they can more readily focus their attention on processing the targeted speech.

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 Office of Research Ethics (ORE) University of Toronto, Mississauga. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

BS, MA-R, and DK conceived and planned the experiments. CR, DK, and MA-R carried out the experiments. SA was responsible for setting up and calibrating the EEG apparatus. CA and CR processed the EEG recordings. BS took the lead in statistically analysing the data. BS, CR, MA-R, and CA contributed to the writing of the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This research was supported by a grant from the Natural Sciences and Engineering Research Council of Canada (ROPIN9952-13 and RGPIN-2021-02721).

ACKNOWLEDGMENTS

The authors would like to thank Brendan Fifield for his assistance in data collection and Jane Carey for her assistance in participant recruitment.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2022.935475/full#supplementary-material

REFERENCES

Alain, C. (2004). "Speech separation: further insights from recordings of event-related brain potentials in human," in *Speech Separation by Humans and Machines*. ed. P. Divenyi (New York: Kluwer Academic Publishers), 13–30.

Alain, C., McDonald, K. L., Ostroff, J. M., and Schneider, B. (2004). Aging: a switch from automatic to controlled processing of sounds? *Psychol. Aging* 19:125. doi: 10.1037/0882-7974.19.1.125

Anderson, S., Gordon-Salant, S., and Dubno, J. R. (2018). Hearing and aging effects on speech understanding: challenges and solutions. *Acoustics Today* 14, 10–18. doi: 10.1121/AT.2018.14.4.12

Androulidakis, A. G., and Jones, S. J. (2006). Detection of signals in modulated and unmodulated noise observed using auditory evoked potentials. *Clin. Neurophysiol.* 117, 1783–1793. doi: 10.1016/j.clinph.2006.04.011

Arbogast, T. L., Mason, C. R., and Kidd, G. Jr. (2002). The effect of spatial separation on informational and energetic masking of speech. *J. Acoust. 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., 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. Psychophysiol.* 80, 242–261. doi: 10.3758/s13414-017-1423-5

Ben-David, B. M., Tse, V. 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. 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 Lang. Hear. Res.* 27, 32–48. doi: 10.1044/jshr.2701.32

Billings, C. J., Bennett, K. O., Molis, M. R., and Leek, &, M. R. (2011). Cortical encoding of signals in noise: effects of stimulus type and recording paradigm. *Ear Hear.* 32, 53–60. doi: 10.1097/AUD.0b013e3181ec5c46

Billings, C. J., Tremblay, K. L., Stecker, G. C., and Tolin, W. M. (2009). Human evoked cortical activity to signal-to-noise ratio and absolute signal level. *Hear. Res.* 254, 15–24. doi: 10.1016/j.heares.2009.04.002

Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound*. Cambridge, MA: MIT Press.

Bregman, A. S., and Campbell, J. (1971). Primary auditory stream segregation and perception of order in rapid sequences of tones. *J. Exp. Psychol.* 89:244. doi: 10.1037/h0031163

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

Dubno, J. R., Horwitz, A. R., and Ahlstrom, J. B. (2002). Benefit of modulated maskers for speech recognition by younger and older adults with normal hearing. *J. Acoust. Soc. Am.* 111, 2897–2907. doi: 10.1121/1.1480421

Durlach, N. I., Mason, C. R., Kidd, G. Jr., Arbogast, T. L., Colburn, H. S., and Shinn-Cunningham, B. G. (2003). Note on informational masking (L). *J. Acoust. Soc. Am.* 113, 2984–2987. doi: 10.1121/1.1570435

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

Faucette, S. P., and Stuart, A. (2017). Evidence of a speech evoked electrophysiological release from masking in noise. *J. Acoust. Soc. Am.* 142, EL218–EL223. doi: 10.1121/1.4998151

Faucette, S. P., and Stuart, A. (2020). An examination of electrophysiological release from masking in young and older adults. *J. Acoust. Soc. Am.* 148:1786. doi: 10.1121/10.0002010

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

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

Helfer, K. S., and Freyman, R. L. (2008). Aging and speech-on-speech masking. *Ear Hear.* 29:87. doi: 10.1097/AUD.0b013e31815d638b

Humes, L. E. (2020). What is "normal hearing" for older adults and can "normal-hearing older adults" benefit from hearing care intervention? *Hear. Rev.* 27, 12–18.

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

Jagadeesh, A. B., and Uppunda, A. K. (2021). Speech-on-speech masking: effect of maskers with different degrees of linguistic information. *Can. J. Speech Lang. Pathol. Audiol.* 45, 143–156.

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.

King, G., Corbin, N. E., Leibold, L. J., and Buss, E. (2020). Spatial release from masking using clinical corpora: sentence recognition in a colocated or spatially separated speech masker. *J. Am. Acad. Audiol.* 31, 271–276. doi: 10.3766/jaaa.19018

McFadden, D., and Wright, B. A. (1990). Temporal decline of masking and comodulation detection differences. *J. Acoust. Soc. Am.* 88, 711–724. doi: 10.1121/1.399774

Murphy, D. R., Craik, F. I. M., 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

Niemczak, C. E., and Vander Werff, K. R. (2019). Informational masking effects on neural encoding of stimulus onset and acoustic change. *Ear Hear.* 40, 156–167. doi: 10.1097/AUD.0000000000000604

Ostroff, J. M., Martin, B. A., and Boothroyd, A. (1998). Cortical evoked response to acoustic change within a syllable. *Ear Hear.* 19, 290–297. doi: 10.1097/00003446-199808000-00004

Ragert, M., Fairhurst, M. T., and Keller, P. E. (2014). Segregation and integration of auditory streams when listening to multi-part music. *PLoS One* 9:e84085. doi: 10.1371/journal.pone.0084085

Rajan, R., and Cainer, K. E. (2008). Ageing without hearing loss or cognitive impairment causes a decrease in speech intelligibility only in informational maskers. *Neuroscience* 154, 784–795. doi: 10.1016/j.neuroscience.2008.03.067

Rajasingam, S. L., Summers, R. J., and Roberts, B. (2021). The dynamics of auditory stream segregation: effects of sudden changes in frequency, level, or modulation. *J. Acoust. Soc. Am.* 149, 3769–3784. doi: 10.1121/10.0005049

Raven, J. C. (1965) The Mill Hill Vocabulary Scale. London: H. K. Lewis.

Roque, L., Karawani, H., Gordon-Salant, S., and Anderson, S. (2019). Effects of age, cognition, and neural encoding on the perception of temporal speech cues. *Front. Neurosci.* 13:749. doi: 10.3389/fnins.2019.00749

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, 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.

Sherbecoe, R. L., and Studebaker, G. A. (2004). Supplementary formulas and tables for calculating and interconverting speech recognition scores in transformed arcsine units. *Int. J. Audiol.* 43, 442–448. doi: 10.1080/14992020400050056

Shinn-Cunningham, B. G., and Best, V. (2008). Selective attention in normal and impaired hearing. *Trends Amplif.* 12, 283–299. doi: 10.1177/1084713808325306

Snyder, J. S., Alain, C., and Picton, T. W. (2006). Effects of attention on neuroelectric correlates of auditory stream segregation. *J. Cogn. Neurosci.* 18, 1–13. doi: 10.1162/089892906775250021

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

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

Tanner, A. M., Spitzer, E. R., Hyzy, J. P., and Grose, J. H. (2019). Masking release for speech in modulated maskers: electrophysiological and behavioral measures. *Ear Hear.* 40, 1009–1015. doi: 10.1097/AUD.0000000000000683

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. Ser. B Psychol. Sci. Soc. Sci.* 54, P317–P327. doi: 10.1093/geronb/54b.5.p317

Vander Werff, K. R., Niemczak, C. E., and Morse, K. (2021). Informational masking effects of speech versus nonspeech noise on cortical auditory evoked potentials. *J. Speech Lang. Hear. Res.* 64, 4014–4029. doi: 10.1044/2021_JSLHR-21-00048

Velia 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

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)

Wagener, K. C., and Brand, T. (2005). Sentence intelligibility in noise for listeners with normal hearing and hearing impairment: Influence of measurement procedure and masking parameters La inteligibilidad de frases en silencio para sujetos con audición normal y con hipoacusia: la influencia del procedimiento de medición y de los parámetros de enmascaramiento. *Int. J. Audiol.* 44, 144–156. doi: 10.1080/14992020500057517

Whiting, K. A., Martin, B. A., and Stapells, D. R. (1998). The effects of broadband noise masking on cortical event-related potentials to speech sounds/Ba/and/da. *Ear Hear.* 19, 218–231. doi: 10.1097/00003446-199806000-00005

Wong, P. C., Jin, J. X., Gunasekera, G. M., Abel, R., Lee, E. R., and Dhar, S. (2009). Aging and cortical mechanisms of speech perception in noise. *Neuropsychologia* 47:693. doi: 10.1016/j.neuropsychologia.2008.11.032

Wright, B. A. (1997). Detectability of simultaneously masked signals as a function of masker bandwidth and configuration for different delays. *J. Acoust. Soc. Am.* 101, 420–429.

Zendel, B. R., Tremblay, C. D., Belleville, S., and Peretz, I. (2015). The impact of musicianship on the cortical mechanisms related to separating speech from background noise. *J. Cogn. Neurosci.* 27, 1044–1059. doi: 10.1162/jocn_a_00758

Zwicker, E. (1965). Temporal effects in simultaneous masking by white-noise bursts. *J. Acoust. Soc. Am.* 37, 653–663.

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

TYPE Original Research
PUBLISHED 09 August 2022
DOI 10.3389/fnins.2022.932959

EDITED BY
Carine Signoret,
Linköping University, Sweden

REVIEWED BY
Paul Reinhart,
Cochlear, United States
Marcus Jeschke,
Deutsches Primatenzentrum, Germany

*CORRESPONDENCE
Payam Shahsavari Baboukani
pasba@es.aau.dk

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

RECEIVED 30 April 2022
ACCEPTED 29 June 2022
PUBLISHED 09 August 2022

CITATION
Shahsavari Baboukani P, Graversen C,
Alickovic E and Østergaard J (2022)
Speech to noise ratio improvement
induces nonlinear parietal phase
synchrony in hearing aid users.
Front. Neurosci. 16:932959.
doi: 10.3389/fnins.2022.932959

Speech to noise ratio improvement induces nonlinear parietal phase synchrony in hearing aid users

Payam Shahsavari Baboukani[1]*, Carina Graversen[2,3],
Emina Alickovic[4,5] and Jan Østergaard[1]

[1]Department of Electronic Systems, Aalborg University, Aalborg, Denmark, [2]Integrative Neuroscience, Department of Health Science and Technology, Aalborg University, Aalborg, Denmark, [3]Department of Health Science and Technology, Center for Neuroplasticity and Pain (CNAP), Aalborg University, Aalborg, Denmark, [4]Eriksholm Research Centre, Snekkersten, Denmark, [5]Department of Electrical Engineering, Linköping University, Linköping, Sweden

Objectives: Comprehension of speech in adverse listening conditions is challenging for hearing-impaired (HI) individuals. Noise reduction (NR) schemes in hearing aids (HAs) have demonstrated the capability to help HI to overcome these challenges. The objective of this study was to investigate the effect of NR processing (inactive, where the NR feature was switched off, vs. active, where the NR feature was switched on) on correlates of listening effort across two different background noise levels [+3 dB signal-to-noise ratio (SNR) and +8 dB SNR] by using a phase synchrony analysis of electroencephalogram (EEG) signals.

Design: The EEG was recorded while 22 HI participants fitted with HAs performed a continuous speech in noise (SiN) task in the presence of background noise and a competing talker. The phase synchrony within eight regions of interest (ROIs) and four conventional EEG bands was computed by using a multivariate phase synchrony measure.

Results: The results demonstrated that the activation of NR in HAs affects the EEG phase synchrony in the parietal ROI at low SNR differently than that at high SNR. The relationship between conditions of the listening task and phase synchrony in the parietal ROI was nonlinear.

Conclusion: We showed that the activation of NR schemes in HAs can non-linearly reduce correlates of listening effort as estimated by EEG-based phase synchrony. We contend that investigation of the phase synchrony within ROIs can reflect the effects of HAs in HI individuals in ecological listening conditions.

KEYWORDS

listening effort, electroencephalography, noise reduction, phase synchrony, local connectivity, hearing impaired

1. Introduction

Hearing impaired (HI) individuals often report that listening to speech in noisy environments such as competing talkers and background noise demands greater effort, which can lead to negative effects such as increased incidence of fatigue (Kramer et al., 2006; Mattys et al., 2012; Wang et al., 2018), disengagement from conversations (Jaworski and Stephens, 1998) and social withdrawal (Weinstein and Ventry, 1982). However, current measurements which are used to examine the performance of a listening task (e.g., speech reception threshold) do not typically consider the cognitive factors related to effortful listening (Sarampalis et al., 2009; Houben et al., 2013).

Pichora-Fuller et al. (2016) defined the concept of listening effort in a conceptual model as "the deliberate allocation of mental resources to overcome obstacles in goal pursuit when carrying out a task, with listening effort applying more specifically when tasks involve listening." The obstacles include acoustic challenges experienced by the listener, which is the combination of cognitive factors (e.g., linguistic ability and memory capacity) and acoustic characteristics (e.g., level of background noise and competing talker) (Peelle, 2018). Listening effort can also be modulated by the listener's motivation (Pichora-Fuller et al., 2016; Peelle, 2018). The goal of studying listening effort is to develop a reliable measurement tool, which can be simultaneously utilized with speech recognition tests and improve the assessment of hearing disability (Paul et al., 2021) and enhance the rehabilitation strategy (Miles et al., 2017).

A wide variety of methods and tools have been used to find correlates of listening effort. This includes subjective ratings such as scales (Krueger et al., 2017) and questioners (Hart and Staveland, 1988), dual tasks (Gagne et al., 2017), and physiological measures such as pupillometry (Zekveld et al., 2018), skin conductance (Mackersie and Calderon-Moultrie, 2016), heart rate (Mackersie and Calderon-Moultrie, 2016), and neuroimaging (Paul et al., 2021). Neuroimaging measures tend to reflect changes in the brain activity underlying listening effort. In particular, electroencephalography (EEG) is becoming popular for measuring correlates of listening effort due to its non-invasiveness and high temporal resolution (Dimitrijevic et al., 2019; Seifi Ala et al., 2020; Fiedler et al., 2021; Wisniewski et al., 2021).

A diverse range of signal processing and information theoretic methods have been used to analyze the EEG and extract correlates of listening effort. Some examples include time-frequency analysis, speech tracking, and functional connectivity. The change in power in the alpha (8-12 Hz) frequency band at the parietal region (Petersen et al., 2015; Dimitrijevic et al., 2017) and theta (4-8 Hz) frequency band at the frontal region (Wisniewski et al., 2015, 2018) have been reported by using time-frequency analysis. The coherence between the speech envelope and the corresponding brain signal at the left frontal cortex in the 2–5 Hz frequency range has also been demonstrated that it can be used for predicting correlates of listening effort in speech tracking analysis (Dimitrijevic et al., 2019).

Functional connectivity describes the statistical dependencies between neural data and can give some information about how the brain functions (Bidelman et al., 2018). Functional connectivity analysis in EEG signals has been extensively used to investigate cognitive functions of auditory processing in the brain. Some examples include perceived audio quality assessment (Mehta and Kliewer, 2017) and semantic processing (Zhang et al., 2019). The effect of acoustic challenges, age-related hearing loss, and comprehension of speech on functional connectivity were also investigated in Bidelman et al. (2018, 2019), and Zhu et al. (2020), respectively.

Functional connectivity can be extracted by using several approaches such as phase synchrony (Bernarding et al., 2013), transfer entropy (Mehta and Kliewer, 2017; Baboukani et al., 2020, 2021b), and Pearson correlation (Bidelman et al., 2019). The phase of neural data tends to reflect the timing of neural activity, and phase synchrony describes the interaction between or within brain regions in the neural networks (Wöstmann et al., 2017a). Correlates of listening effort have been estimated using phase synchrony analysis. Several methods have been used to extract the phase synchrony, such as wavelet phase synchronization stability (Bernarding et al., 2010, 2013), the distribution of the mapped phase mean vector on the unit circle (Bernarding et al., 2014) and the entropy of instantaneous phase of EEG signals (Bernarding et al., 2012, 2017).

The functional connectivity within a localized region of the brain is called local connectivity. It has been utilized to classify different motor imagery movements (Baboukani et al., 2017), estimate the cognitive workload (Zarjam et al., 2013), investigate schizophrenia (Jalili et al., 2007), and Alzheimer's disease (Jalili et al., 2013). Phase synchrony has also been used to assess the local connectivity. Phase synchrony assessment across multivariate signals (or channels of EEG) in a localized region of the brain by averaging over all possible traditional bi-variate phase synchrony values (e.g., phase coherence, phase locking value) may not provide a full picture of the global synchrony within the signals (Oshima et al., 2006; Canolty et al., 2011; Omidvarnia et al., 2013; Al-Khassaweneh et al., 2016; Baboukani et al., 2019). Alternatively, multivariate measures generalized the traditional bi-variate ones to evaluate phase synchrony within multichannel data (Omidvarnia et al., 2013; Al-Khassaweneh et al., 2016; Baboukani et al., 2019). Local connectivity estimated by multivariate phase synchrony has been used in several studies (Jalili et al., 2013; Al-Khassaweneh et al., 2016; Baboukani et al., 2017). A new multivariate phase synchrony measure called circular omega complexity (COC) was recently proposed and has shown better performance than conventional multivariate phase synchrony techniques in simulated and real EEG data (Baboukani et al., 2019). Recently, we showed that local connectivity estimated by the COC measure can be used to estimate the correlates of listening effort (Baboukani et al., 2021a).

Although HI individuals in real-life encounter listening situations which involve continuous speech and long sentences, most of the studies (some exceptions include Alickovic et al., 2020, 2021; Fiedler et al., 2021) investigate the effect of NR processing and SNR when the stimuli are single words or short sentences (Bernarding et al., 2014; Dimitrijevic et al., 2017; Miles et al., 2017). However, in this study, continuous long stimuli are used in a speech in noise (SiN) task.

Modern hearing aid (HA) devices can help HI individuals through various advanced signal processing approaches (Bernarding et al., 2014, 2017; Winneke et al., 2018). In particular, noise reduction (NR) processing intends to reduce the effect of background noise and enhance the signal-to-noise ratio (SNR). It has shown the capability to free up cognitive resources for other tasks during listening and reduce the listening effort (Sarampalis et al., 2009; Ohlenforst et al., 2018). The activation of NR processing can improve speech understanding at low SNRs. Furthermore, it has been also shown that activation of the NR schemes in HAs provides a listening effort enhancement in addition to any effect associated with improved intelligibility (Ohlenforst et al., 2018). In addition to that, NR schemes can improve the performance of the HA users during a selective attention task (Alickovic et al., 2020, 2021). Alickovic et al. (2020) also showed that the improvement of selective attention tasks due to NR was different at low SNR than that at high SNR. Another study on the same data showed that NR changed correlates of listening effort estimated by pupil size differently at the two SNR values, while a time-frequency analysis of EEG signals showed no statistical change due to SNR, NR, and the interaction between them (Fiedler et al., 2021). This inspired us to replace conventional power analysis in the time-frequency domain with local connectivity estimates based on multivariate phase synchrony to investigate the effect of NR processing at two SNR values. Inspired by the results in Alickovic et al. (2020) and Fiedler et al. (2021), we hypothesized:

H1: The use of NR in hearing aids affects the EEG phase synchrony within localized regions of the brain (i.e., local connectivity) at low SNR differently than that at high SNR.

H2: The effect of the NR scheme on EEG phase synchrony within localized regions of the brain at different SNR values shows a nonlinear (inverted U-shape) trend.

2. Materials and methods

In this section, the EEG data utilized in this study will be briefly explained. It will be followed by the description of the COC measure and the steps required to assess local connectivity. Finally, the statistical test used in this study will be described.

2.1. EEG data

The EEG data of this study has been utilized for other analyses by Alickovic et al. (2020) and Fiedler et al. (2021), which focused on neural tracking and pupil dilation, respectively. The EEG analysis of listening effort recruited by Fiedler et al. (2021) was based on alpha power and did not show significant results.

2.1.1. Participants

We recruited 22 (11 men), native Danish-speaking participants, with hearing loss. The stimuli used in this study were based on participants-centered language (Nicks et al., 2022) and consisted of Danish news clips of neutral content. They aged between 40 and 80 years with the mean and standard deviation (SD) ages of 69 and 11.2, respectively. The participants were experienced HA users with more than 3 months of HA usage. Participants had mild-to-moderate sensorineural hearing loss. The audiometric thresholds at 500, 1,000, 2,000, and 4,000 Hz ranged from 33 to 58 dB hearing level, with an average threshold of 45 dB hearing level. The maximum difference between the left and the right ear' averaged audiometric thresholds was less than 8 dB. The experimental procedure was approved by the ethics committee for the capital region of Denmark (journal number H-1-2011-033) and all participants signed written consent before the experiment.

2.1.2. Hearing aid fitting and signal processing

All participants were fitted with identical HA models during the experiment. Two pairs of HAs were adapted for each participant: NR inactive and NR active. Rather than NR, all other signal processing settings did not change between the conditions. The Voice Aligned Compression (VAC) rationale (Le Goff, 2015) based on each individual's hearing threshold was applied to amplify the sound in both pairs of HAs. In the NR inactive condition, the omnidirectional setting was applied with an added natural slight forward effect of the pinna. In the other pairs, NR active, the combination of minimum variance distortionless response beamformer and a single channel Wiener post filter was applied before the VAC. The articulation-index-weighted SNR improvements (Ohlenforst et al., 2018) were 6.24 dB and 5.17 dB at +3 dB SNR and +8 dB SNR, respectively, for NR active than that for NR inactive (Alickovic et al., 2020; Fiedler et al., 2021).

2.1.3. Experimental design

The experiment took place in an acoustically shielded listening booth with controlled light conditions. Participants were seated on a chair positioned in the middle of six loudspeakers (Genelec 8040A; Genelec Oy, Iisalmi, Finland) with a distance of 1.2 m from each loudspeaker (refer

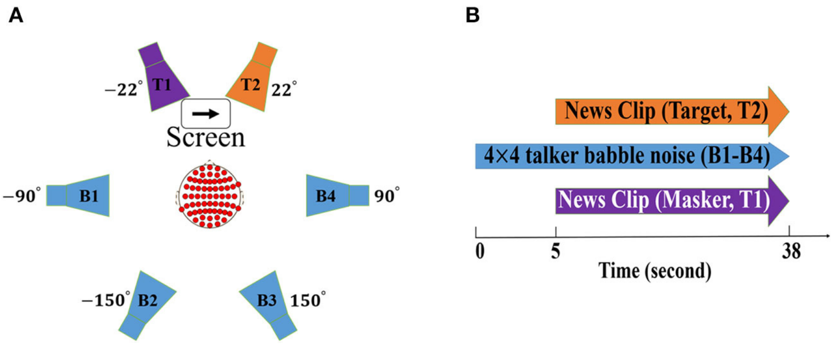


FIGURE 1
Schematic illustration of **(A)** experiment design including six loudspeakers. The target streams (colored as purple and orange) and background noise (colored as blue) were delivered by the two foreground and four background loudspeakers, respectively. The screen located in the middle of the two foreground speakers shows the to-be-attended talker (colored as orange). **(B)** trial design in which 5 s of only background noise and 33 s of simultaneous target, masker, and background noise stimuli were delivered in each trial.

to Figure 1A), two loudspeakers in the front (at ±22° azimuth) and four loudspeakers in the back (at ±90° and ±150° azimuth). Each of the background loudspeakers (B1-B4 in Figure 1A) played four-talker babble. The two foreground speakers played the target streams which were spoken by talkers of a different gender. Participants were asked to gaze at the screen positioned between the two frontal loudspeakers and were instructed to attend to one of the talkers in the foreground speakers while ignoring the other talker on the contralateral side and background noise. To-be-attended talkers (either the right or the left side) was indicated by an arrow on the screen (refer to Figure 1A).

2.1.4. Stimuli

Continuous 33 s long Danish news clips of neutral content were utilized for talker streams. The organization of the location (left or right) and gender (male or female) of the target stream was randomized. Each of the four-talker babble noises delivered by the background loudspeakers consisted of four unique single talkers, two women and two men, speaking different news giving the impression of the 16 talkers active in the background.

The experiment was a 2 × 2 design: the first factor was NR (active vs. inactive) and the second factor was the SNR level (+3 vs. + 8 dB). The SNR in our setup was defined as the ratio between the signal power of the attended talker and the total signal power of the background noise, similar to that in Das et al. (2018) and Alickovic et al. (2020). In order to create real-life listening conditions at two levels of difficulty (SNR values of +3 and +8 dB), the maskers were set at either 53 or 48 dB, leading to a total of 59 or 54 dB background Sound Pressure Level (SPL). Each of the two foreground loudspeakers was always set at a fixed level of 62 dB SPL, leading to a fixed level of 65 dB for the foreground talkers.

2.1.5. Procedure

A total of 84 trials were recorded for each participant, organized in a block design. For each of the four blocks (experimental conditions including +3 dB NR inactive, +3 dB NR active, +8 dB NR inactive, and +8 dB NR active), 20 trials were conducted. The remaining four trials were used for training. Each trial consisted of a short period of silence, 5 s of only background noise (delivered by background loudspeakers), and 33 s of the simultaneous target, masker, and background noise stimuli (refer to Figure 1B). After each trial, participants were asked to answer a two-choice question about the content of the attended speech which was displayed on the screen. The HAs were always removed and replaced again between the blocks. The participants were given a rest period after five trials.

2.1.6. EEG data acquisition and pre-processing

The BioSemi ActiveTwo amplifier system (Biosemi, Amsterdam, Netherlands) was used to record EEG data. A total of 64 electrodes on a cap were mounted on the scalp according to the 10-20 international system. The driven right leg and common mode sense electrode were used as a reference for all other recording electrodes. The EEG data were sampled at 1,024 Hz. All electrodes were mounted by applying conductive gel to obtain a stable connection and below 50 mv offset voltage.

The pre-processing includes a 0.5 Hz high-pass filter, 95 Hz low-pass filter, and downsampling to 512 Hz. Then, The EEG channels with excessive noise were visually identified (on average, 3.1 ± 0.8 channels were rejected) and interpolated from the surrounding clean EEG channels by using the nearest neighbor method in Fieldtrip (Oostenveld et al., 2011). The logistic Infomax independent component analysis algorithm was applied to reduce artifacts caused by eye movements, eye blinks, muscle activity, heartbeats, and single-channel noise, as implemented in Fieldtrip

(Bell and Sejnowski, 1995; Delorme and Makeig, 2004). The components were visually inspected and rejected if clearly reflected as artifacts, on average, 7.9 ± 3.6 of the components were rejected. Finally, the EEG data were epoched from 8 s before to 33 s after the onset of the target loudspeakers. The EEG data for one subject was removed from further analysis due to being excessively noisy. In addition to that, no data for one block of one participant was recorded due to technical problems.

2.2. Circular omega complexity

The COC measure is used in this study to extract local connectivity. This is a multivariate phase synchrony measure that was recently proposed in Baboukani et al. (2019) and was shown to perform better than the conventional measures in a particular setup. The COC assesses the level of phase dependency within multivariate signals by quantifying the dimensionality of the state-space formed by their corresponding instantaneous phases (Baboukani et al., 2019).

Estimating the instantaneous phase of a real valued mono-component discrete signal $X \in \mathbb{R}^{1\times N}$ is the first step to calculate the COC. A Hilbert transform-based approach is commonly used whereby the instantaneous phase is estimated as Baboukani et al. (2019):

$$\phi_X[n] = \tan^{-1}\left(\frac{\hat{X}[n]}{X[n]}\right), \tag{1}$$

where $\hat{X}[n]$ and $\phi_X[n]$ are the Hilbert transform and instantaneous phase of $X[n]$, respectively. The next step is calculating the circular correlation matrix. Considering a K-channels signal $\mathbf{X} \in \mathbb{R}^{K\times N}$ and its corresponding instantaneous phase signal $\phi_{\mathbf{X}} \in \mathbb{R}^{K\times N}$, the circular correlation matrix $C^{\mathbf{X}} \in \mathbb{R}^{K\times K}$ is defined as Baboukani et al. (2019):

$$C^{\mathbf{X}} = [C^{(m,l)}], \tag{2}$$

where $m, l \in \{1, 2, \dots K\}$. The circular correlation between the instantaneous phase ϕ_m and ϕ_l is noted by $C^{(m,l)}$ where ϕ_m and ϕ_l are the m^{th} and l^{th} rows of $\phi_{\mathbf{X}}$, respectively. The circular correlation $C^{(m,l)}$ is given as Baboukani et al. (2019):

$$C^{(m,l)} = \frac{\sum_{n=1}^{N} \sin\left(\phi_m[n] - \bar{\phi}_m\right) \sin\left(\phi_l[n] - \bar{\phi}_l\right)}{\sqrt{\sum_{n=1}^{N} \sin^2\left(\phi_n[n] - \bar{\phi}_m\right) \sin^2\left(\phi_l[n] - \bar{\phi}_l\right)}}, \tag{3}$$

where the circular mean $\bar{\phi}_m$ is given by Baboukani et al. (2019):

$$\bar{\phi}_m = \arg\left(\sum_{n=1}^{N} \exp^{i\phi_m[n]}\right). \tag{4}$$

It was shown by Baboukani et al. (2019) that the eigenvalues of $C^{\mathbf{X}}$ can be used as an index for the dimensionality of the state-space formed by $\phi_{\mathbf{X}}$ whereby the level of phase synchronization can be determined. The COC was then defined as Baboukani et al. (2019):

$$COC = 1 + \frac{\sum_{i=1}^{K} \bar{\lambda}_i \log \bar{\lambda}_i}{\log K}, \tag{5}$$

where $\bar{\lambda}_i = \frac{\lambda_m}{\sum_{j=1}^{K} \lambda_j}$ and $\lambda_i; i = 1, \dots, K$ are the eigenvalues of $C^{\mathbf{X}}$. The COC varies between 0 and 1 where higher values show that the channels within $C^{\mathbf{X}}$ are more phase synchronized, which means that only fewer eigenvalues of the $C^{\mathbf{X}}$ are significant (Baboukani et al., 2019).

2.3. Local connectivity assessment

The effect of NR at two SNR values on local connectivity will be explored in this article. Listening in adverse conditions can possibly engage multiple cognitive processes such as working memory and distractor inhibition (Wisniewski et al., 2021), which can possibly change the local connectivity within different brain regions and frequency bands. We, therefore, did not restrict our analysis to one specific band or ROI and local connectivity within eight ROIs and four conventional EEG bands will be examined. The ROIs include left frontal, frontal, right frontal, left temporal, central, right temporal, parietal, and occipital, similar to that in Mehta and Kliewer (2017). Figure 2 and Table 1 show the electrodes and their corresponding positions of different ROIs, respectively [1]. The EEG bands consist of Delta (0.5–4 Hz), Theta (4–8 Hz), Alpha (8–12.5 Hz), and Beta (12.5–25 Hz). The EEG channels were filtered by using Window-based FIR band-pass filters. Filtering to narrow frequency bands can also reduce the multi-component nature of EEG signals and can improve the estimation of the instantaneous phase signals by using the Hilbert transform technique (Boashash and Aïssa-El-Bey, 2018). The EEG channels were common average re-referenced to reduce the effect of volume conduction.

The local connectivity of each trial of the experiment was estimated during the time interval that frontal loudspeakers were presenting the target streams (33 s). We also omitted the first second of the delivering target streams to minimize the effects of event-related potential which led to 32 s (1 to 33 s relative to the onset of the target streams). The 32 s time span was then divided into 16 non-overlapping 2 s windows[2]. The COC was then quantified for each of the windows and the average over all windows was considered as an indicator of the strength

---

[1] An alternative group of electrodes for frontal, central, parietal, and occipital ROIs, which include the midline electrodes located in the regions, such as Fz, Pz, Cz, and Oz produced the same trend of results.

[2] An alternative longer length (such as 10 s) of time windows leads to a better estimation of COC. However, the results produced by longer or equal to 2 s time windows were following the same trend.

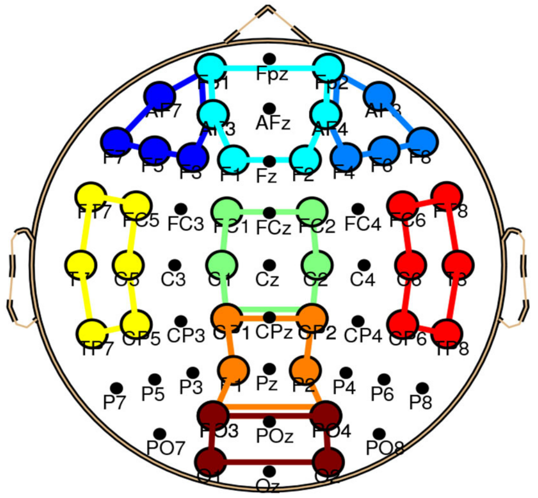


FIGURE 2
In the electrode position of ROIs, different colors show different ROIs. The lines in the figure show schematic presentations of the ROIs. The ROIs include left frontal (dark blue), frontal (aqua), right frontal (light blue), left temporal (yellow), central (green), right temporal (red), parietal (orange), and occipital (brown).

TABLE 1 Mapping electroencephalogram (EEG) electrodes to regions of interest (ROIs).

ROI	Electrodes	ROI	Electrodes
Left frontal	AF7, AF3, F3 F5, F7, Fp1	Frontal	Fp1, Fp2, AF4 AF3, F1, F2
Right frontal	AF4, AF8, F8 F6, F4, Fp2	Central	FC1, FC2, C1 CP1, C2, CP2
Left temporal	FT7, T7, TP7 CP5, FC5, C5	Parietal	CP1, CP2, P1 P2, PO4, PO3
Right temporal	FT8, T8, TP8 CP6, FC6, C6	Occipital	O1, O2, PO3 PO4

of local connectivity. The higher the average of COC values over time windows is, the more channels within the ROI are phase synchronized. The higher phase synchrony is considered as higher local connectivity in this study. The steps required to assess local connectivity in a specific band and ROI can be summarized as follows:

S1) Band-pass filters the EEG channels in the ROI to a conventional EEG band.
S2) Estimate the instantaneous phase of the filtered channels.
S3) Extract the COC value for each of the 2 s time windows.
S4) Average the COC values corresponding to time windows.

The aforementioned steps were repeated for eight ROIs and four EEG bands leading to 32 local connectivity values for each trial.

2.4. Statistical test

All the statistical analysis was performed in RStudio Team (2021). In order to investigate the effect of NR at two SNR values on local connectivity (our first Hypothesis H1), two-way Linear mixed effect (LMM) ANOVA was applied by using lme4 (Bates et al., 2014) and lmerTest (Kuznetsova et al., 2017) packages. We fitted separate LMM ANOVA models for local connectivity values estimated at each ROI and band. Local connectivity values were treated as a continuous variable and normalized by using $COC_{normalized} = \frac{COC-M}{S}$, where the M and S are the mean and SD of the local connectivity values at specific band and ROI calculated over all experimental conditions and all subjects. The experiment factor NR was treated as a factor variable with two levels, inactive and active. The experiment factor SNR was also treated as a factor variable with two levels, high (+8 dB) and low (+3 dB). The local connectivity was modeled as a function of fixed factors NR, SNR, and their interaction, and the participants were treated as a random effect. The analysis was conducted based on subject-averaged local connectivity values. We will also report the results based on single trial models for the statistically significant local connectivity, in which the interaction between participants and trials was treated as a random effect.

In order to investigate our second hypothesis which is about the relationship between local connectivity and the four experimental conditions- +3 dB active, +3 dB inactive, +8 dB active, and +8 dB inactive, we applied the measured SNR improvement of the NR processing, refer to Section 2.1.2 for more details. The SNR improvements of the active NR scheme were 6.24 dB and 5.17 dB at 3 dB SNR and 8 dB SNR, respectively. This process reduces the two factors SNR and NR of the experiment to only one factor SNR, with values of 3, 8, 9.24, and 13.17 dB. Then, one-way LMM ANOVA has applied to model local connectivity as a function of fixed factor SNR which was treated as a continuous variable. Two models were used for each local connectivity: the first model included the quadratic (nonlinear) term alongside the linear term for the fixed factor SNR and the second model only consisted of the linear term. The participants were treated as a random effect. Similar to the two-way LMM ANOVA, the results based on single trial models for the statistically significant trends will be reported.

Since a series (eight ROIs and four bands leading to 32 models) of LMM ANOVA models were applied, we used the Bonferroni correction to compensate for the multiple comparisons effect. The significance levels for all the two-way and one-way LMM ANOVA models were, therefore, chosen as $\alpha = \frac{0.05}{32} = 0.0016$.

TABLE 2 *P*-values of the two-way LMM ANOVA. (A) P-values for interaction between two factors SNR and NR. (B) *P*-values of the main factor SNR. (C) P-values of the main factor NR. The two factors are SNR values, +3 dB and +8 dB, and NR schemes, on and off. The boldface numbers show the rejection of the null hypothesis. The significance level was Bonferroni corrected, $\alpha = \frac{0.05}{32} = 0.0016$.

(A)

	Left frontal	Frontal	Right frontal	Central	Left temporal	Parietal	Right temporal	Occipital
Delta	0.3083	0.9577	0.9004	0.2891	0.7109	0.0086	0.5054	0.6982
Theta	0.5986	0.9038	0.8941	0.2419	0.8460	0.0098	0.1991	0.8890
Alpha	0.5811	0.7302	0.6898	0.1370	0.5116	**0.0009**	0.04	0.6041
Beta	0.6868	0.3757	0.5041	0.2295	0.8981	0.0033	0.1544	0.7814

(B)

	Left frontal	Frontal	Right frontal	Central	Left temporal	Parietal	Right temporal	Occipital
Delta	0.6430	0.1143	0.0535	0.6862	0.4565	0.7421	0.9075	0.7022
Theta	0.4882	0.0497	0.5245	0.3467	0.8811	0.4504	0.1736	0.9586
Alpha	0.6072	0.1096	0.6841	0.8241	0.8235	0.6635	0.0981	0.1719
Beta	0.2794	0.0757	0.9245	0.4805	0.6725	0.9951	0.0260	0.3742

(C)

	Left frontal	Frontal	Right frontal	Central	Left temporal	Parietal	Right temporal	Occipital
Delta	0.5015	0.8633	0.7418	0.4164	0.8241	0.6195	0.6881	0.5722
Theta	0.7668	0.9642	0.5383	0.4872	0.6733	0.3703	0.8001	0.8816
Alpha	0.6963	0.7806	0.2371	0.7766	0.6075	0.4634	0.5812	0.2326
Beta	0.7135	0.5241	0.2805	0.8134	0.7648	0.7150	0.8649	0.4089

3. Results

Participants were prompted with a two-choice question related to the content of the attended speech after each trial. They correctly answered 86% of the questions. This indicates that the participants followed the task as instructed. However, after applying a two-way LMM ANOVA on the behavioral performances, there was no statistical effect of NR, SNR, and their interaction, with the *p*-values of 0.25, 0.06, and 0.37, respectively.

To test Hypothesis H1, two-way LMM ANOVA was applied to local connectivity at each ROI and band, which modeled the normalized local connectivity as a function of fixed factors NR and SNR. Tables 2A–C summarized the *p*-values obtained from applying two-way LMM ANOVA on the average over trials for each subject local connectivity. As shown in able Table 2A, we found a significant interaction of SNR and NR on local connectivity at the parietal region alpha frequency band (will be referred to as parietal alpha hereinafter), $F_{(59.02)} = 12.28$, $p = 0.0009$. Note that, as mentioned in Section 2.1.6, the EEG data of one block for one subject was not recorded. Therefore, our data is unbalanced and the denominator degree of freedom (DF) is estimated by using Satterthwaite's method. As there is no *p*-value less than 0.0016 in Tables 2B,C, no significant main effect was found for SNR and NR. The results of applying LMM ANOVA on single trial data were in line with the average trial analysis. The interaction between SNR and NR was statistically significant at parietal alpha, $F_{(1228.5)} = 83.59, p < 0.0001$.

The one-way LMM ANOVA was applied to average trial data after applying the SNR improvement of the NR processing to investigate Hypothesis H2. The normalized local connectivity was modeled as a function of continuous fixed factor SNR by using two separate one-way LMM ANOVA to study the relationship between the normalized local connectivity and experiment conditions. The first model was based on including the quadratic term for the fixed factor SNR and the second model only consisted of the linear term. Table 3A shows the results of the first model in which quadratic terms alongside linear terms were included. As shown in Table 3A, the nonlinear trend between local connectivity and SNR at parietal alpha was statistically significant, $F_{(60.22)} = 11.92, p = 0.0010$. We found no linear relationship between the experimental conditions and local connectivity, as there is no *p*-value less than the significant

TABLE 3 P-values of the one-way LMM ANOVA. (A) P-values for the quadratic term. (B) P-values for the linear term. The local connectivity at different ROIs and bands is independently modeled by different listening conditions. The conditions are +3 dB inactive (+3 dB), +8 dB inactive (+8 dB), +3 dB active (9.24 dB), and +8 dB active (13.17 dB). The boldface numbers show the rejection of the null hypothesis. The significance level was Bonferroni corrected, $\alpha = \frac{0.05}{32} = 0.0016$.

(A)

	Left frontal	Frontal	Right frontal	Central	Left temporal	Parietal	Right temporal	Occipital
Delta	0.3358	0.9458	0.9256	0.3253	0.6948	0.0094	0.4790	0.6577
Theta	0.6181	0.9998	0.9405	0.2151	0.8761	0.0129	0.2118	0.8773
Alpha	0.6061	0.7128	0.6102	0.1282	0.5417	**0.0010**	0.0593	0.5358
Beta	0.7114	0.4093	0.5719	0.2195	0.9201	0.0035	0.1618	0.8435

(B)

	Left frontal	Frontal	Right frontal	Central	Left temporal	Parietal	Right temporal	Occipital
Delta	0.3699	0.2595	0.3415	0.6413	0.7490	0.4511	0.7277	0.5054
Theta	0.8699	0.2342	0.9215	0.2923	0.65551	0.6869	0.5770	0.8868
Alpha	0.9773	0.2288	0.5105	0.7980	0.7522	0.3106	0.6534	0.9599
Beta	0.3135	0.0942	0.3961	0.5875	0.9849	0.6326	0.2555	0.9071

level in Table 3B. The nonlinear trend between local connectivity and SNR at parietal alpha was also significant by single trial analysis, $F_{(1229.5)} = 76.36, p < 0.0001$.

For the purposes of visualization, (Figures 3, 4) plot regression line and 95% CI for the nonlinear trend between normalized local connectivity and experimental conditions at both individual and average trial analyses, respectively. The inverted U-shaped relationship shows that local connectivity at parietal alpha is higher for +3 dB active and +8 dB inactive and lower for +3 dB inactive and +8 dB active. The figures also show that NR processing at lower SNR (+3 dB) leads to an increase in the local connectivity at parietal alpha while NR processing at higher SNR (+8 dB) leads to a decrease.

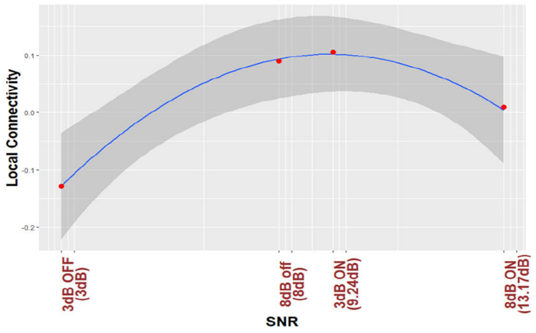


FIGURE 3
Parietal alpha local connectivity is regressed based on different listening conditions. The analysis is performed by single trial data and the red points show the average over all trials and all subjects for different listening conditions.

4. Discussion

4.1. Summary

In a sample of 22 HIs, we studied the effect of NR processing in HAs on the EEG local connectivity during a continuous SiN task. Inspired by the results reported in Alickovic et al. (2020) and Fiedler et al. (2021), we hypothesized that the effect of NR schemes on local connectivity differs at the two SNR values, +3 dB and +8 dB, of the experiment. Consistent with our Hypothesis (H1), we found a significant interaction between the factors of the experiment, SNR and NR, at parietal alpha by using both average-trial and single-trial analysis, which would suggest that NR processing affects the local connectivity at low SNR differently than that of at high SNR. It should be noted that the *p*-values corresponded to the interaction at the parietal, and all frequency bands are small. However, the dominant significant change due to the interaction between SNR and NR appears to be at the alpha band as only the parietal alpha band survives a correction for multiple comparisons.

The articulation-index-weighted SNR improvements (Ohlenforst et al., 2018) of the NR processing were applied, which reduces the two factors of the experiment to only one factor SNR with values +3 dB, +8 dB, +9.24, and +13.17 dB. We then investigated the relationship between the experimental conditions and local connectivity. We found a significant inverted U-shaped function at parietal alpha by both single

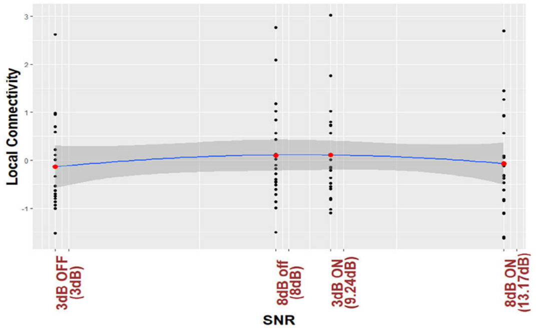


FIGURE 4
Parietal alpha local connectivity is regressed based on different listening conditions. The analysis is performed by average over trials data. The black points in the figure show the average over trials for subjects and the red points show the average over black points for different listening conditions.

and average trial analysis, which was in line with our second Hypothesis H2. Since this study is the first work, to our knowledge, that investigates the effect of different levels of listening effort and NR schemes in HA on local connectivity, the results will be discussed in terms of hypothesized functions in the following sections.

4.2. NR schemes in HAs reduce the listening effort

Recent studies have shown that the activation of advanced signal processing algorithms in HAs provides hearing benefits for HIs, particularly in adverse listening conditions (Sarampalis et al., 2009; Ohlenforst et al., 2018; Winneke et al., 2018). Studies focusing on changes or benefits in speech intelligibility may not provide a complete picture of the processes involved in speech recognition (Dillon and Lovegrove, 1993; Sarampalis et al., 2009; Ohlenforst et al., 2018). In particular, NR processing, which is the main focus of this study, can reduce listening effort and free up cognitive resources for other tasks while it may not have positive effects on speech reception threshold (Sarampalis et al., 2009). The effects of NR schemes have been investigated when the stimuli is a single word or sentence (Dimitrijevic et al., 2017; Miles et al., 2017; Ohlenforst et al., 2018). However, HI individuals encounter long speech in real ecological situations. For this reason, long continuous news clips were presented at different SNR levels. Our first finding based on speech tracking analysis of the EEG data published in Alickovic et al. (2020) showed that NR processing can improve the performance of HAs during a selective auditory attention task. Then, Fiedler et al. (2021) showed that the NR schemes can also reduce the listening effort estimated by pupillometry. However, the neural index of listening effort estimated by spectral power analysis of the EEG data did not show any statistical change. It inspired us to recruit a new correlate of listening effort estimated by local connectivity in EEG data to investigate the effect of NR schemes during a long continuous SiN task.

As shown in Table 2A, the interaction between SNR and NR on local connectivity is statistically significant. This suggests that NR processing affects the local connectivity differently at the two SNR values of the experiment, which is in line with pupillometry results reported in Ohlenforst et al. (2018) and Fiedler et al. (2021) where they also found a different effect of NR schemes at different SNR values. This result is also consistent with results published in Alickovic et al. (2020) in which they found that NR processing improved the performance of the selective attention task differently at the two SNR values. We also investigated the relationship between correlates of listening effort and the experimental conditions by applying the SNR-improvement of NR processing. As demonstrated in Figures 3, 4 and Table 3A, we found an inverted U-shaped trend. We believe that this study is the first work that showed the nonlinear trend of neural estimation of correlates of listening effort as a result of NR processing at different SNR values during a continuous long SiN task. This result is consistent with pupil dilation analysis in Ohlenforst et al. (2018) and EEG analysis in Marsella et al. (2017), Wisniewski et al. (2017), Decruy et al. (2020), and Paul et al. (2021) where nonlinear relationship due to NR processing at different SNR values and different levels of listening difficulty were found, respectively.

4.3. Local connectivity is modulated by top-down cognitive functions or changes of brain networks

Most of the existing studies in the literature which investigated the listening effort by using EEG signals tend to focus on spectral power features and particularly event-related spectral perturbation (ERSP) (c.f Section 4.5). Finding a relationship between local connectivity estimated at scalp level, which is the case in this study, and power change can be controversial and even two features can be significantly uncorrelated, as is the case in Jalili et al. (2007) study where no significant correlation was found between power change and local connectivity in Schizophrenia EEG data analysis. There might be possibilities to discuss the relationship based on the Firefly model presented in Burgess (2012) or the model presented in Jirsa (2009). However, we believe that the local connectivity estimated in this study can violate the required assumptions of these models. Nonetheless, the studies which investigate spectral power changes during effortful listening described the possible top-down cognitive functions or brain networks that can lead to the change in the power features. There are two theories that can connect top-down cognitive function or brain networks and local connectivity.

The first theory is based on the phase reset model in which phase locking of ongoing EEG activity can be a modulatory effect of top-down functions of the brain (Bernarding et al., 2017). Peelle et al. (2013) found that neural data and the envelope of the external acoustic stimuli become more phase-locked when linguistic information is available. They concluded that the phase-locking of the neural oscillations does not rely only on sensory cues and top-down cognitive function can also modulate phase locking. Dimitrijevic et al. (2019) also found that the phase-locked cortical representation can be modulated by top-down cognitive function related to different levels of listening effort. Bernarding et al. (2017) demonstrated that the distribution of the phase of the ongoing EEG signal can be modulated by the top-down cognitive functions related to different listening efforts. Considering these aspects, one interpretation can be that local connectivity estimated by multivariate phase synchrony is also modulated by the top-down cognitive functions related to the listening effort.

The second theory explains that change in local connectivity estimated by phase synchronization can be one of the mechanisms for coordinating the information transfer in brain networks (Helfrich et al., 2016; Olejarczyk et al., 2017). For example, Helfrich et al. (2016) showed that local parieto-occipital phase coupling at the alpha band controls the inter-hemispheric information transfer. Additionally, Olejarczyk et al. (2017) reported an increase in local phase coupling in closed eyes compared to open eyes in a resting-state EEG analysis and they concluded that fronto-parietal information transfer can be regulated by local phase synchrony. Regarding these aspects, it can be interpreted that local connectivity estimated by phase synchrony also coordinates the information transfer related to effortful listening and NR schemes in HAs.

4.4. Significant change at parietal alpha local connectivity

Most prior studies that investigated EEG correlates of listening effort have tended to restrict their analysis to a single EEG band and region (Wisniewski et al., 2015; Dimitrijevic et al., 2019; Seifi Ala et al., 2020). However, Wisniewski et al. (2021) conducted a comprehensive study in which they investigated a fuller range of the EEG power spectrum and independent source activities. They found several significant changes in different regions and bands. They concluded that listening in adverse conditions can possibly engage multiple cognitive processes. Consistent with Wisniewski et al. (2021), NR processing can also engage multiple cognitive processes which can possibly change the local connectivity. As the effect of NR processing in HAs on local connectivity in ecologically adverse conditions was investigated for the first time in this article, we did not restrict our analysis to one specific EEG band and region. Local connectivity at a total of eight ROIs and four conventional EEG bands were, therefore, examined.

Frontal theta and parietal alpha activity at the sensor level have been mainly reported in the literature as the regions and bands that can be used to estimate correlates of listening effort (Wisniewski et al., 2017, 2018; Dimitrijevic et al., 2019; Fiedler et al., 2021). The change in the frontal theta activity is mostly observed in experiments in which non-speech stimuli were used (Wisniewski et al., 2017, 2018). This tends to reflect the internal attention and it does not show general endogenously exerted effort related to externally generated object representations (e.g., competing speech streams and background noise) (Wisniewski et al., 2018). As the change of listening effort in this study is mostly due to changes in externally represented speech stimuli, we did not observe any significant change in frontal theta local connectivity, which is in line with the results reported in Seifi Ala et al. (2020) that significant change of frontal theta change was not observed as a results changes in the speech stimuli characteristics. On the other hand, the change of parietal alpha activity has been widely observed when listening effort was examined in experiments with speech stimuli (Petersen et al., 2015; Wöstmann et al., 2015, 2017b; McMahon et al., 2016; Dimitrijevic et al., 2017, 2019; Marsella et al., 2017; Miles et al., 2017; Seifi Ala et al., 2020; Fiedler et al., 2021; Paul et al., 2021), which is line with our results where we only found significant change at the parietal alpha activity.

4.5. Top-down cognitive functions in listening effort

The direction (i.e., increase, decrease, or inverted U-shape) of the parietal alpha activity modulation found in the literature has been controversial. Some studies reported that higher listening effort leads to an increase in parietal alpha power (relative to the baseline) arguing that it reflects the inhibition of neural activity in task-irrelevant brain area (Petersen et al., 2015; Wöstmann et al., 2015, 2017b; McMahon et al., 2016; Dimitrijevic et al., 2017, 2019; Marsella et al., 2017; Miles et al., 2017; Paul et al., 2021). In contrast, other studies showed that more demanding conditions lead to a decrease in parietal alpha power (Seifi Ala et al., 2020; Fiedler et al., 2021). The first explanation for the contradictory results is that multiple sources of alpha power contribute to parietal alpha power and the balance between suppression and enhancement can be determined by the stimuli and task design (Dimitrijevic et al., 2017; Seifi Ala et al., 2020; Fiedler et al., 2021; Wisniewski et al., 2021). Seifi Ala et al. (2020) observed lower parietal alpha power related to more difficult conditions during long speech listening. It was discussed in Seifi Ala et al. (2020)

that sustained attention and constant update of information in working memory is required when the stimuli are long, which would lead to contradictory results. Another explanation for the opposite direction can be related to the inverted U-shape relationship. Depending on the level of difficulties of the experiment, estimated correlates of listening effort can be on one or the other side of the inverted U's maximum, which would result in an increase or decrease in parietal alpha power, respectively (Fiedler et al., 2021).

The last relationship between listening conditions and parietal alpha power reported in the literature is an inverted U-shape (Marsella et al., 2017; Wisniewski et al., 2017; Decruy et al., 2020; Paul et al., 2021). There are two explanations for the observed nonlinear trend. One theory is that during difficult conditions subjects disengaged and gave up to perform the task, which can influence the parietal alpha changes (Marsella et al., 2017). The second explanation is that at very high levels of difficulty, other sensory networks might be activated to help speech understanding which leads to an inverse direction of parietal alpha modulation compared to that at lower difficulty levels (Paul et al., 2021). The supportive sensory networks under very hard conditions can be related to sustained attention and constant update of information in working memory as reported in Seifi Ala et al. (2020).

4.6. Parietal alpha local connectivity is modulated by listening effort

Referring to Figure 3 and Table 3A, we also found a significant nonlinear trend in local connectivity at parietal alpha. As shown in Figure 3, an increase in levels of difficulty in listening (decrease in SNR values) from the condition +8 dB ON (NR: active) to +8 dB off (NR: inactive) leads to an increase in local connectivity. Consistent with results reported in Petersen et al. (2015), Wöstmann et al. (2015, 2017b), McMahon et al. (2016), Dimitrijevic et al. (2017), Marsella et al. (2017), Miles et al. (2017), Dimitrijevic et al. (2019) and Paul et al. (2021), the increase of the parietal alpha power can be due to inhibition cognitive function. Considering the first theory mentioned in Section 4.3, the inhibition of top-down cognitive function can lead to the modulation of local connectivity. This interpretation is in line with the results of Mathewson et al. (2009) and Paul et al. (2021) where the authors also found that phase synchronization in parietal alpha increases due to inhibition of cognitive function. The change of local connectivity due to inhibition function can also be supported by the second theory mentioned in Section 4.3. The inhibition function mostly engages the fronto-parietal network. We interpreted that the local connectivity at parietal alpha can also coordinate the fronto-parietal information transfer. This interpretation is in line with the results reported in Olejarczyk et al. (2017) where the authors also found that phase synchrony in parietal alpha coordinates the fronto-parietal information transfer in rest-state EEG analysis.

As shown in Figure 3, in the more difficult condition (i.e., at +3 dB) the local connectivity decreases, whereas in the easier condition (i.e., at + 8dB) the local connectivity increases. In line with EEG band power analysis results, this change can be due to either giving up during more difficult conditions (Marsella et al., 2017) or other sensory networks that might be activated to help speech understanding during such listening conditions (Paul et al., 2021). Our findings provide evidence that the change from increase to decrease in local connectivity under more difficult conditions could be due to the activation of other networks at the lowest SNR value. Considering the second theory mentioned in Section 4.3 which describes that local connectivity estimated by phase synchrony can coordinate the information transfer in brain networks, the change of direction in local connectivity modulation at +3 dB OFF condition can also be due to activation of other sensory networks which can be coordinated by local connectivity at parietal alpha. One possible sensory network can be due to sustained attention and constant update of information in working memory which is in line with the results reported in Seifi Ala et al. (2020). It was discussed in Seifi Ala et al. (2020) that sustained attention and constant update of information in working memory are required when the stimuli are long. This was also observed during a Stenberg task in which encoding and retention phases were entangled and a contradictory increase in parietal alpha power was reported as a result of higher working memory loads (Jensen et al., 2002; Hjortkjær et al., 2020; Seifi Ala et al., 2020). Kim et al. found that the brain network involved in updating function engaged in an n-back level experimental paradigm mostly includes the parietal cortex which is served as the main hub for the cognitive network (Kim et al., 2017). They also found a substantially different pattern during the most demanding condition compared to easier conditions. Considering the second theory in Section 4.3, the change of direction of the local connectivity modulation at the hardest condition in our experiment can also be due to substantially different networks involved in updating function.

4.7. Limitations

The local connectivity at eight ROIs and four EEG frequency bands were investigated in this study. The selection of the ROIs was similar to that in Mehta and Kliewer (2017) where they used 128 electrodes, and we adapted their ROI selections with 64 electrodes setup. However, there might be a better selection of ROIs, which can lead to different results. Additionally, considering that we had two states of NR processing and two SNR values, our experiment had four listening conditions. We checked the relationship between local connectivity and listening conditions, and we found a nonlinear trend. Examination with more SNR values is required, which can provide more insights, and we expect to observe a complete inverted U-shape relationship with more SNR values.

Obleser and Kayser (2019) showed that the phase locking between neural data and the envelope of the speech can be modulated by the behavioral performance of the task. There is a possibility that local connectivity is also modulated by the performance of the task or subjective rating of listening effort (often referred to as self-report or experienced listening effort) (Paul et al., 2021), similar to the first theory in Section 4.3. The behavioral performance evaluation was accomplished by asking a two-choice question about the to-be-attended speech stream at the end of each trial in our experiment. Our investigation of the effect of the experimental factors on behavioral performance published in Fiedler et al. (2021) did not show any significant effect of NR, SNR, or their interactions on the behavioral performance. The behavioral performance was though well above the chance level (50%) and the participants followed the task as instructed. However, the lack of valuable behavioral performance or subjective ratings of listening effort prevented us from checking the possibility that local connectivity is modulated by them.

The second theory in Section 4.3 explained that local connectivity can coordinate the information transfer in brain networks. We interpreted that local connectivity at parietal alpha can also coordinate the large-scale connectivity engaged in inhibition function and constant update of the working memory and referred to the studies in which these information transfers were studied. There is a possibility that other brain networks are also engaged during a continuous long SiN task, which could be provided by a large-scale connectivity investigation.

5. Conclusion

We investigated the effect of activation of NR processing on EEG-based phase synchrony measure within localized regions of the brain at eight regions of interest and four conventional EEG frequency bands during a longer continuous speech in noise (SiN) task with two SNR levels. We demonstrated that the effect of noise reduction (NR) processing algorithms on EEG-based phase synchrony have a non-linear trend in the parietal region of interest, specifically in the alpha band. The interpretation of the phase synchrony modulation is in line with the literature. These results confirmed that the EEG-based phase synchrony within localized regions of the brain contains informative features which can reflect the effects of HA signal processing algorithms in HA users. Taken together, our study provided further evidence that the NR processing algorithms in HAs positively affect HA users in their everyday natural listening environments.

Data availability statement

MATLAB and R codes of this article can be found in the Supplementary material, further inquiries can be directed to the corresponding author/s.

Ethics statement

The studies involving human participants were reviewed and approved by Ethics Committee for the capital region of Denmark (journal number H-1-2011-033). The patients/participants provided their written informed consent to participate in this study.

Author contributions

PS, CG, EA, and JØ contributed to conception and design of the study. EA organized the database. PS performed the software and statistical analysis and wrote the first draft of the manuscript. All authors contributed to manuscript revision, read, 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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2022.932959/full#supplementary-material

References

Alickovic, E., Lunner, T., Wendt, D., Fiedler, L., Hietkamp, R., Ng, E. H. N., et al. (2020). Neural representation enhanced for speech and reduced for background noise with a hearing aid noise reduction scheme during a selective attention task. *Front. Neurosci.* 14, 846. doi: 10.3389/fnins.2020.00846

Alickovic, E., Ng, E. H. N., Fiedler, L., Santurette, S., Innes-Brown, H., and Graversen, C. (2021). Effects of hearing aid noise reduction on early and late cortical representations of competing talkers in noise. *Front. Neurosci.* 15, 636060. doi: 10.3389/fnins.2021.636060

Al-Khassaweneh, M., Villafane-Delgado, M., Mutlu, A. Y., and Aviyente, S. (2016). A measure of multivariate phase synchrony using hyperdimensional geometry. *IEEE Trans. Signal Process.* 64, 2774–2787. doi: 10.1109/TSP.2016.2529586

Baboukani, P. S., Azemi, G., Boashash, B., Colditz, P., and Omidvarnia, A. (2019). A novel multivariate phase synchrony measure: application to multichannel newborn eeg analysis. *Digit. Signal Process.* 84, 59–68. doi: 10.1016/j.dsp.2018.08.019

Baboukani, P. S., Graversen, C., Alickovic, E., and Østergaard, J. (2020). Estimating conditional transfer entropy in time series using mutual information and nonlinear prediction. *Entropy* 22, 1124. doi: 10.3390/e22101124

Baboukani, P. S., Graversen, C., Alickovic, E., and Østergaard, J. (2021a). "EEG phase synchrony reflects snr levels during continuous speech-in-noise tasks," in *2021 43rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (Mexico: IEEE), 531–534.

Baboukani, P. S., Graversen, C., and Østergaard, J. (2021b). "Estimation of directed dependencies in time series using conditional mutual information and non-linear prediction," in *2020 28th European Signal Processing Conference (EUSIPCO)* (Amsterdam: IEEE), 2388–2392.

Baboukani, P. S., Mohammadi, S., and Azemi, G. (2017). "Classifying single-trial eeg during motor imagery using a multivariate mutual information based phase synchrony measure," in *2017 24th National and 2nd International Iranian Conference on Biomedical Engineering (ICBME)* (Tehran: IEEE), 1–4.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2014). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01

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

Bernarding, C., Corona-Strauss, F. I., Latzel, M., and Strauss, D. J. (2010). "Auditory streaming and listening effort: an event related potential study," in *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology* (Buenos Aires: IEEE), 6817–6820.

Bernarding, C., Strauss, D. J., Hannemann, R., and Corona-Strauss, F. I. (2012). "Quantification of listening effort correlates in the oscillatory eeg activity: a feasibility study," in *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (San Diego, CA: IEEE), 4615–4618.

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

Bernarding, C., Strauss, D. J., Hannemann, R., Seidler, H., and Corona-Strauss, F. I. (2014). "Objective assessment of listening effort in the oscillatory eeg: Comparison of different hearing aid configurations," in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (Chicago, IL: IEEE), 2653–2656.

Bernarding, C., Strauss, D. J., Hannemann, R., Seidler, H., and Corona-Strauss, F. I. (2017). Neurodynamic evaluation of hearing aid features using eeg correlates of listening effort. *Cogn. Neurodyn.* 11, 203–215. doi: 10.1007/s11571-017-9425-5

Bidelman, G. M., Davis, M. K., and Pridgen, M. H. (2018). Brainstem-cortical functional connectivity for speech is differentially challenged by noise and reverberation. *Hear. Res.* 367, 149–160. doi: 10.1016/j.heares.2018.05.018

Bidelman, G. M., Mahmud, M. S., Yeasin, M., Shen, D., Arnott, S. R., and Alain, C. (2019). Age-related hearing loss increases full-brain connectivity while reversing directed signaling within the dorsal-ventral pathway for speech. *Brain Struct. Funct.* 224, 2661–2676. doi: 10.1007/s00429-019-01922-9

Boashash, B., and Aïssa-El-Bey, A. (2018). Robust multisensor time-frequency signal processing: a tutorial review with illustrations of performance enhancement in selected application areas. *Digit. Signal Process.* 77, 153–186. doi: 10.1016/j.dsp.2017.11.017

Burgess, A. P. (2012). Towards a unified understanding of event-related changes in the eeg: the firefly model of synchronization through cross-frequency phase modulation. *PLoS ONE* 7, 1–21. doi: 10.1371/journal.pone.0045630

Canolty, R. T., Cadieu, C. F., Koepsell, K., Knight, R. T., and Carmena, J. M. (2011). Multivariate phase-amplitude cross-frequency coupling in neurophysiological signals. *IEEE Trans. Biomed. En.* 59, 8–11. doi: 10.1109/TBME.2011.2172439

Das, N., Bertrand, A., and Francart, T. (2018). Eeg-based auditory attention detection: boundary conditions for background noise and speaker positions. *J. Neural Eng.* 15, 066017. doi: 10.1088/1741-2552/aae0a6

Decruy, L., Lesenfants, D., Vanthornhout, J., and Francart, T. (2020). Top-down modulation of neural envelope tracking: the interplay with behavioral, self-report and neural measures of listening effort. *Eur. J. Neurosci.* 52, 3375–3393. doi: 10.1111/ejn.14753

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

Dillon, H., and Lovegrove, R. (1993). Single microphone noise reduction systems for hearing aids: a review and an evaluation. *Acoust. Factors Affect. Hear. Perform.* 20, 353–370.

Dimitrijevic, A., Smith, M. L., Kadis, D. S., and Moore, D. R. (2017). Cortical alpha oscillations predict speech intelligibility. *Front. Hum. Neurosci.* 11, 88. doi: 10.3389/fnhum.2017.00088

Dimitrijevic, A., Smith, M. L., Kadis, D. S., and Moore, D. R. (2019). Neural indices of listening effort in noisy environments. *Sci. Rep.* 9, 1–10. doi: 10.1038/s41598-019-47643-1

Fiedler, L., Ala, T. S., Graversen, C., Alickovic, E., Lunner, T., and Wendt, D. (2021). Hearing aid noise reduction lowers the sustained listening effort during continuous speech in noise–a combined pupillometry and eeg study. *Ear. Hear.* 42, 1590–1601. doi: 10.1097/AUD.0000000000001050

Gagne, 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

Hart, S. G., and Staveland, L. E. (1988). Development of nasa-tlx (task load index): results of empirical and theoretical research. *Adv. Psychol.* 52, 139–183. doi: 10.1016/S0166-4115(08)62386-9

Helfrich, R. F., Knepper, H., Nolte, G., Sengelmann, M., König, P., Schneider, T. R., et al. (2016). Spectral fingerprints of large-scale cortical dynamics during ambiguous motion perception. *Hum. Brain Mapp.* 37, 4099–4111. doi: 10.1002/hbm.23298

Hjortkjær, J., Märcher-Rørsted, J., Fuglsang, S. A., and Dau, T. (2020). Cortical oscillations and entrainment in speech processing during working memory load. *Eur. J. Neurosci.* 51, 1279–1289. doi: 10.1111/ejn.13855

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

Jalili, M., Barzegaran, E., and Knyazeva, M. G. (2013). Synchronization of eeg: bivariate and multivariate measures. *IEEE Trans. Neural Syst. Rehabil. Eng.* 22, 212–221. doi: 10.1109/TNSRE.2013.2289899

Jalili, M., Lavoie, S., Deppen, P., Meuli, R., Do, K. Q., Cuénod, M., et al. (2007). Dysconnection topography in schizophrenia revealed with state-space analysis of eeg. *PLoS ONE* 2, e1059. doi: 10.1371/journal.pone.0001059

Jaworski, A., and Stephens, D. (1998). Self-reports on silence as a face-saving strategy by people with hearing impairment. *Int. J. Appl. Linguist.* 8, 61–80. doi: 10.1111/j.1473-4192.1998.tb00121.x

Jensen, O., Gelfand, J., Kounios, J., and Lisman, J. E. (2002). Oscillations in the alpha band (9-12 hz) increase with memory load during retention in a short-term memory task. *Cereb. Cortex* 12, 877–882. doi: 10.1093/cercor/12.8.877

Jirsa, V. K. (2009). Neural field dynamics with local and global connectivity and time delay. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* 367, 1131–1143. doi: 10.1098/rsta.2008.0260

Kim, N. Y., Wittenberg, E., and Nam, C. S. (2017). Behavioral and neural correlates of executive function: interplay between inhibition and updating processes. *Front. Neurosci.* 11, 378. doi: 10.3389/fnins.2017.00378

Kramer, S. E., Kapteyn, T. S., and Houtgast, T. (2006). Occupational performance: comparing normally-hearing and hearing-impaired employees using the amsterdam checklist for hearing and work: desempeño laboral: comparación de empleados con audición normal o alterada usando el listado amsterdam para audición y trabajo. *Int. J. Audiol.* 45, 503–512. doi: 10.1080/14992020600754583

Krueger, M., Schulte, M., Brand, T., and Holube, I. (2017). Development of an adaptive scaling method for subjective listening effort. *J. Acoust. Soc. Am.* 141, 4680–4693. doi: 10.1121/1.4986938

Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. (2017). lmertest package: tests in linear mixed effects models. *J. Stat. Softw.* 82, 1–26. doi: 10.18637/jss.v082.i13

Le Goff, N. (2015). *Amplifying Soft Sounds–A Personal Matter*. Oticon Whitepaper.

Mackersie, C. L., and Calderon-Moultrie, N. (2016). Autonomic nervous system reactivity during speech repetition tasks: heart rate variability and skin conductance. *Ear Hear.* 37, 118S-125S. doi: 10.1097/AUD.0000000000000305

Marsella, P., Scorpecci, A., Cartocci, G., Giannantonio, S., Maglione, A. G., Venuti, I., et al. (2017). Eeg activity as an objective measure of cognitive load during effortful listening: a study on pediatric subjects with bilateral, asymmetric sensorineural hearing loss. *Int. J. Pediatr. Otorhinolaryngol.* 99, 1–7. doi: 10.1016/j.ijporl.2017.05.006

Mathewson, K. E., Gratton, G., Fabiani, M., Beck, D. M., and Ro, T. (2009). To see or not to see: prestimulus α phase predicts visual awareness. *J. Neurosci.* 29, 2725–2732. doi: 10.1523/JNEUROSCI.3963-08.2009

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. doi: 10.1080/01690965.2012.705006

McMahon, C. M., Boisvert, I., de Lissa, P., Granger, L., Ibrahim, R., Lo, C. Y., et al. (2016). Monitoring alpha oscillations and pupil dilation across a performance-intensity function. *Front. Psychol.* 7, 745. doi: 10.3389/fpsyg.2016.00745

Mehta, K., and Kliewer, J. (2017). Directional and causal information flow in eeg for assessing perceived audio quality. *IEEE Trans. Mol. Biol. Multiscale Commun.* 3, 150–165. doi: 10.1109/TMBMC.2018.2806454

Miles, K., McMahon, C., Boisvert, I., Ibrahim, R., De Lissa, P., Graham, P., et al. (2017). Objective assessment of listening effort: coregistration of pupillometry and EEG. *Trends Hear.* 21, 2331216517706396. doi: 10.1177/2331216517706396

Nicks, S., Johnson, A. L., Traxler, B., Bush, M. L., Brame, L., Hamilton, T., et al. (2022). The use of person-centered language in medical research articles focusing on hearing loss or deafness. *Ear. Hear.* 43, 703–711. doi: 10.1097/AUD.0000000000001168

Obleser, J., and Kayser, C. (2019). Neural entrainment and attentional selection in the listening brain. *Trends Cogn. Sci.* 23, 913–926. doi: 10.1016/j.tics.2019.08.004

Ohlenforst, B., Wendt, D., Kramer, S. E., Naylor, G., Zekveld, A. A., and Lunner, T. (2018). Impact of snr, masker type and noise reduction processing on sentence recognition performance and listening effort as indicated by the pupil dilation response. *Hear. Res.* 365, 90–99. doi: 10.1016/j.heares.2018.05.003

Olejarczyk, E., Marzetti, L., Pizzella, V., and Zappasodi, F. (2017). Comparison of connectivity analyses for resting state eeg data. *J. Neural Eng.* 14, 036017. doi: 10.1088/1741-2552/aa6401

Omidvarnia, A., Azemi, G., Colditz, P. B., and Boashash, B. (2013). A time-frequency based approach for generalized phase synchrony assessment in nonstationary multivariate signals. *Digit. Signal Process.* 23, 780–790. doi: 10.1016/j.dsp.2013.01.002

Oostenveld, R., Fries, P., Maris, E., and Schoffelen, J.-M. (2011). Fieldtrip: open source software for advanced analysis of meg, eeg, and invasive electrophysiological data. *Comput. Intell. Neurosci.* 2011, 156869. doi: 10.1155/2011/156869

Oshima, K., Carmeli, C., and Hasler, M. (2006). State change detection using multivariate synchronization measure from physiological signals. *J. Signal Process.* 10, 223–226.

Paul, B. T., Chen, J., Le, T., Lin, V., and Dimitrijevic, A. (2021). Cortical alpha oscillations in cochlear implant users reflect subjective listening effort during speech-in-noise perception. *PLoS ONE* 16, e0254162. doi: 10.1371/journal.pone.0254162

Peelle, J. E. (2018). Listening effort: How the cognitive consequences of acoustic challenge are reflected in brain and behavior. *Ear. Hear.* 39, 204. doi: 10.1097/AUD.0000000000000494

Peelle, J. E., Gross, J., and Davis, M. H. (2013). Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. *Cereb. Cortex* 23, 1378–1387. doi: 10.1093/cercor/bhs118

Petersen, E. B., Wöstmann, M., Obleser, J., Stenfelt, S., and Lunner, T. (2015). Hearing loss impacts neural alpha oscillations under adverse listening conditions. *Front. Psychol.* 6, 177. doi: 10.3389/fpsyg.2015.00177

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

RStudio Team (2021). *RStudio: Integrated Development Environment for R*. Boston, MA: RStudio, PBC.

Sarampalis, A., Kalluri, S., Edwards, B., and Hafter, E. (2009). Objective measures of listening effort: effects of background noise and noise reduction. *J. Speech Lang. Hear. Res.* 52, 1230–1240. doi: 10.1044/1092-4388(2009/08-0111)

Seifi Ala, T., Graversen, C., Wendt, D., Alickovic, E., Whitmer, W. M., and Lunner, T. (2020). An exploratory study of eeg alpha oscillation and pupil dilation in hearing-aid users during effortful listening to continuous speech. *PLoS ONE* 15, e0235782. doi: 10.1371/journal.pone.0235782

Wang, Y., Naylor, G., Kramer, S. E., Zekveld, A. A., Wendt, D., Ohlenforst, B., et al. (2018). Relations between self-reported daily-life fatigue, hearing status, and pupil dilation during a speech perception in noise task. *Ear Hear.* 39, 573. doi: 10.1097/AUD.0000000000000512

Weinstein, B. E., and Ventry, I. M. (1982). Hearing impairment and social isolation in the elderly. *J. Speech Lang. Hear. Res.* 25, 593–599. doi: 10.1044/jshr.2504.593

Winneke, A., De Vos, M., Wagener, K. C., Derleth, P., Latzel, M., Appell, J., et al. (2018). Listening effort and eeg as measures of performance of modern hearing aid algorithms. *Audiol. Online* 24198, 1–13.

Wisniewski, M. G., Iyer, N., Thompson, E. R., and Simpson, B. D. (2018). Sustained frontal midline theta enhancements during effortful listening track working memory demands. *Hear. Res.* 358, 37–41. doi: 10.1016/j.heares.2017.11.009

Wisniewski, M. G., Thompson, E. R., and Iyer, N. (2017). Theta-and alpha-power enhancements in the electroencephalogram as an auditory delayed match-to-sample task becomes impossibly difficult. *Psychophysiology* 54, 1916–1928. doi: 10.1111/psyp.12968

Wisniewski, M. G., Thompson, E. R., Iyer, N., Estepp, J. R., Goder-Reiser, M. N., and Sullivan, S. C. (2015). Frontal midline θ power as an index of listening effort. *Neuroreport* 26, 94–99. doi: 10.1097/WNR.0000000000000306

Wisniewski, M. G., Zakrzewski, A. C., Bell, D. R., and Wheeler, M. (2021). Eeg power spectral dynamics associated with listening in adverse conditions. *Psychophysiology* 58, e13877. doi: 10.1111/psyp.13877

Wöstmann, M., Fiedler, L., and Obleser, J. (2017a). Tracking the signal, cracking the code: speech and speech comprehension in non-invasive human electrophysiology. *Lang. Cogn. Neurosci.* 32, 855–869. doi: 10.1080/23273798.2016.1262051

Wöstmann, M., Herrmann, B., Wilsch, A., and Obleser, J. (2015). Neural alpha dynamics in younger and older listeners reflect acoustic challenges and predictive benefits. *J. Neurosci.* 35, 1458–1467. doi: 10.1523/JNEUROSCI.3250-14.2015

Wöstmann, M., Lim, S.-J., and Obleser, J. (2017b). The human neural alpha response to speech is a proxy of attentional control. *Cereb. Cortex* 27, 3307–3317. doi: 10.1093/cercor/bhx074

Zarjam, P., Epps, J., Chen, F., and Lovell, N. H. (2013). Estimating cognitive workload using wavelet entropy-based features during an arithmetic task. *Comput. Biol. Med.* 43, 2186–2195. doi: 10.1016/j.compbiomed.2013.08.021

Zekveld, A. A., Koelewijn, T., and Kramer, S. E. (2018). The pupil dilation response to auditory stimuli: current state of knowledge. *Trends Hear.* 22, 2331216518777174. doi: 10.1177/2331216518777174

Zhang, G., Si, Y., and Dang, J. (2019). Revealing the dynamic brain connectivity from perception of speech sound to semantic processing by EEG. *Neuroscience* 415, 70–76. doi: 10.1016/j.neuroscience.2019.07.023

Zhu, Y., Liu, J., Ristaniemi, T., and Cong, F. (2020). Distinct patterns of functional connectivity during the comprehension of natural, narrative speech. *Int. J. Neural Syst.* 30, 2050007. doi: 10.1142/S0129065720500070

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 06 September 2022
DOI 10.3389/fnins.2022.972730

OPEN ACCESS

EDITED BY
Julia Stephen,
Mind Research Network (MRN),
United States

REVIEWED BY
Marcel Schulze,
University Hospital Bonn, Germany
Yuanyue Zhou,
First Affiliated Hospital of Hainan
Medical University, China

*CORRESPONDENCE
Rina Blomberg
rina.blomberg@liu.se

†These authors share last authorship

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

RECEIVED 18 June 2022
ACCEPTED 12 August 2022
PUBLISHED 06 September 2022

CITATION
Blomberg R, Signoret C, Danielsson H, Perini I, Rönnberg J and Capusan AJ (2022) Aberrant resting-state connectivity of auditory, ventral attention/salience and default-mode networks in adults with attention deficit hyperactivity disorder. *Front. Neurosci.* 16:972730.. doi: 10.3389/fnins.2022.972730

Aberrant resting-state connectivity of auditory, ventral attention/salience and default-mode networks in adults with attention deficit hyperactivity disorder

Rina Blomberg[1,2,3]*, Carine Signoret[1,2,3], Henrik Danielsson[1,2,3], Irene Perini[4,5], Jerker Rönnberg[1,2,3,5]† and Andrea Johansson Capusan[4,6]†

[1]Department of Behavioral Sciences and Learning, Linköping University, Linköping, Sweden, [2]Linnaeus Center HEAD, Linköping University, Linköping, Sweden, [3]Swedish Institute for Disability Research, Linköping University, Linköping, Sweden, [4]Center for Social and Affective Neuroscience, Department of Biomedical and Clinical Sciences, Linköping University, Linköping, Sweden, [5]Center for Medical Image Science and Visualization, Linköping University, Linköping, Sweden, [6]Department of Psychiatry and Biomedical and Clinical Sciences, Linköping University, Linköping, Sweden

Background: Numerous resting-state studies on attention deficit hyperactivity disorder (ADHD) have reported aberrant functional connectivity (FC) between the default-mode network (DMN) and the ventral attention/salience network (VA/SN). This finding has commonly been interpreted as an index of poorer DMN regulation associated with symptoms of mind wandering in ADHD literature. However, a competing perspective suggests that dysfunctional organization of the DMN and VA/SN may additionally index increased sensitivity to the external environment. The goal of the current study was to test this latter perspective in relation to auditory distraction by investigating whether ADHD-adults exhibit aberrant FC between DMN, VA/SN, and auditory networks.

Methods: Twelve minutes of resting-state fMRI data was collected from two adult groups: ADHD (n = 17) and controls (n = 17); from which the FC between predefined regions comprising the DMN, VA/SN, and auditory networks were analyzed.

Results: A weaker anticorrelation between the VA/SN and DMN was observed in ADHD. DMN and VA/SN hubs also exhibited aberrant FC with the auditory network in ADHD. Additionally, participants who displayed a stronger anticorrelation between the VA/SN and auditory network at rest, also performed better on a cognitively demanding behavioral task that involved ignoring a distracting auditory stimulus.

Conclusion: Results are consistent with the hypothesis that auditory distraction in ADHD is linked to aberrant interactions between DMN, VA/SN, and auditory systems. Our findings support models that implicate dysfunctional organization of the DMN and VA/SN in the disorder and encourage more research into sensory interactions with these major networks.

KEYWORDS

attention deficit hyperactivity disorder, adults, resting state, functional connectivity, default mode network, salience network, auditory network

Introduction

Current etiological models of attention deficit hyperactivity disorder (ADHD) emphasize dysfunctional interactions between intrinsic brain networks, rather than regional brain abnormalities, for explanations of behavioral and clinical symptoms in the disorder (for reviews see: Konrad and Eickhoff, 2010; Castellanos and Proal, 2012; Posner et al., 2014; Castellanos and Aoki, 2016; Sutcubasi et al., 2020). Two particular intrinsic brain networks: the default-mode network (DMN; **Figure 1A**), and the ventral attention/salience network (VA/SN; **Figure 1B**); are hypothesized to play a pivotal role in clinical aspects of inattention in ADHD. The DMN is generally more active when attention is directed internally—i.e., to introspective, self-referential thought in the absence of stimulus-driven tasks. The VA/SN is more active when attention is directed externally, and is heavily implicated in the vigilant anticipation, detection, and response-mediation of behaviorally salient stimuli. In healthy individuals, the DMN and VA/SN show robust anticorrelated (i.e., phasically negatively correlated) functional connectivity (FC) at rest (Anticevic et al., 2012). And this anticorrelated relationship is considered an inherent representation of the opposing resource demands and attentional states between these networks during goal directed cognitive tasks (Fox et al., 2005, 2009). A variety of resting-state fMRI studies on ADHD have reported a weaker anticorrelated relationship between core DMN and VA/SN regions compared to controls (e.g., Castellanos et al., 2008; Sun et al., 2012; Sripada et al., 2014; Lin et al., 2018; Mills et al., 2018; c.f. also Sato et al., 2012). And these findings have had a notable influence over our conceptualization of inattention in the disorder.

One prevailing perspective for instance suggests that a weaker anticorrelation between core regions of the DMN and VA/SN at rest is an index of poorer DMN regulatory capacity associated with symptoms of mind wandering in the disorder (see Posner et al., 2014 for a review). The logic here being that the DMN is more active during introspective, task-unrelated thought and suppressed during stimulus driven tasks. Hence observations of aberrant resting-state anticorrelation is thought to reflect an inherent susceptibility for attentional lapses resulting from obtruding interoceptive thought (Castellanos et al., 2008; Bozhilova et al., 2018). Alternative perspectives have suggested that the VA/SN plays a pivotal role in the manifestation of attentional deficits in the disorder. Aboitiz et al. (2014) for instance, hypothesized that a loss of anticorrelation impacts the sensitivity of the VA/SN which can lead to a bias toward irrelevant/salient stimuli and therein an increased susceptibility to environmental distraction. Similarly, Menon (2011) proposed that aberrant intrinsic VA/SN organization can result in the inappropriate assignment of saliency to either exogenous stimuli or internal mental events—underpinning, *inter alia*, clinical aspects of inattention in many psychiatric and neurological disorders, including ADHD. Hence, aside from mind wandering, a disrupted antagonistic balance between the DMN and VA/SN in the disorder may also be indicative of an increased sensitivity to the external environment, and it is this notion which founds the basis of inquiry for the current study.

Heightened sensory sensitivity is a frequently reported symptom in adults with ADHD particularly in the auditory modality (Schulze et al., 2020). However, very few resting-state analyses have studied auditory network FC in ADHD-adults. Hence little is known as to whether deficits in auditory attention are associable to the aberrant, intrinsic FC of the DMN and VA/SN implicated in the disorder (Castellanos and Aoki, 2016). Following through on previous work, the current study provides us with a unique opportunity to explore resting-state FC in relation to auditory distraction in this patient group. Here we analyze resting-state data from the same sample of adult participants (ADHD and healthy controls) from Blomberg et al.'s (2021) task-based fMRI study that utilized a cross modal attention paradigm to analyze the effects of working memory load on auditory distraction in adults with ADHD. The paradigm involved two main tasks: an auditory task where the goal was to actively detect salient oddball tones amidst a stream of standard tones, and a visual *n*-back task consisting of 0-, 1-, and 2-back working memory conditions. In all working memory conditions, participants were instructed to detect the visual *n*-back target and ignore the streaming tonal signal from the auditory task, which continued to play.

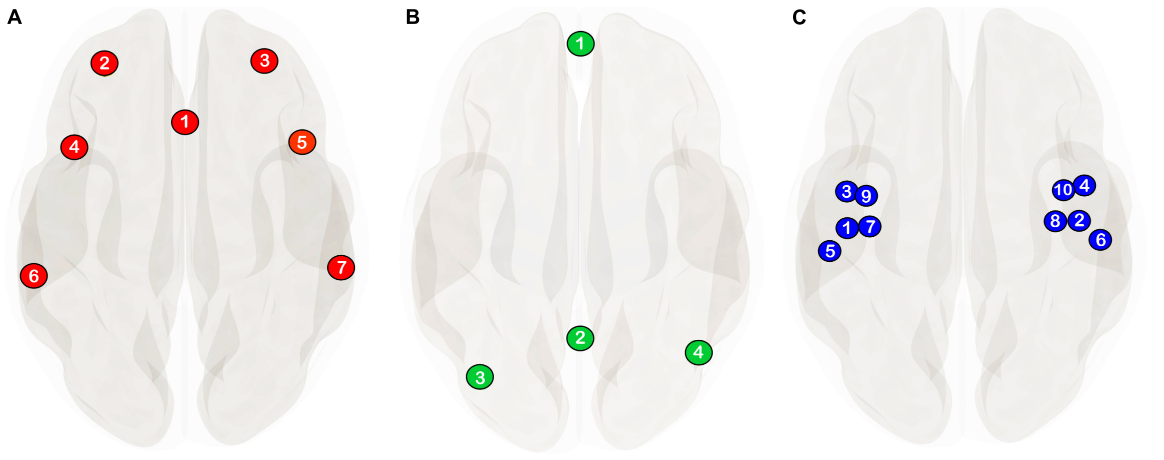


FIGURE 1
The network delineations used in the current study are from the CONN-toolbox's v.20b (Whitfield-Gabrieli and Nieto-Castanon, 2012; Nieto-Castanon, 2020) network atlas derived from an independent component analysis of 497 individuals from the human connectome project. **(A)** Ventral attention/salience network (see Uddin et al., 2019 for a discussion on network nomenclature) consisting of the anterior cingulate (1) and bilateral rostral prefrontal (2, 3), anterior insula (4, 5) and supramarginal cortices (6, 7). **(B)** Default-mode network definition comprising of the medial prefrontal cortex (1), the posterior cingulate cortex, (2) and bilateral angular gyri (3, 4). **(C)** Auditory seed regions used in the current study are the same anatomically defined regions of interest from Blomberg et al. (2021) comprising of the bilateral Heschl's gyrus (1, 2), planum polare (3, 4), planum temporale (5, 6), granular posterior insulae (7, 8), and dysgranular posterior insulae (9, 10).

When participants' attention was focused on the auditory task, auditory cortical activity was enhanced relative to a resting baseline; and when attention was instead directed toward the visual task, auditory processing was attenuated. The degree of attenuation in auditory regions was relative to the cognitive demands of the visual task—the greater the working memory load and attentional engagement in the visual modality, the greater the neural attenuation to task-irrelevant sound in auditory regions. Importantly, for ADHD participants, the relationship between attentional engagement and auditory attenuation proved less efficient than controls. In particular, under the most cognitively demanding visual condition, the 2-back task (2-bT), ADHD participants showed significantly poorer attenuation in auditory regions. Further, this heightened auditory activity was correlated with individual differences in symptomatic inattentiveness and 2-bT performance—for which the ADHD group were significantly inferior to controls. In the current study, we pose the question as to whether these aforementioned outcomes are related to both individual, and group level variances in intrinsic FC between the DMN, VA/SN, and auditory regions.

Our first objective (O1) in this regard was to determine if a weakener DMN–VA/SN anticorrelation was evident in the ADHD group. Our second objective (O2) was to test for group differences in FC between auditory seeds (**Figure 1C**) and regions pertaining to the VA/SN and DMN. Given the functional antagonism of the VA/SN and DMN with respect to externally and internally directed attention, here we hypothesized that ADHD participants would show increased FC between auditory seeds and VA/SN regions and reduced anticorrelated FC between auditory seeds and DMN regions. Our third objective (O3) was to explore individual differences in brain-behavioral relationships between DMN, VA/SN, and auditory FC. Here the goal was to determine if ADHD-symptom severity and performance on the 2b-T from Blomberg et al. (2021) was associated with increases/decreases in DMN, VA/SN and auditory FC.

Materials and methods

Participants

The resting-state data for this study was obtained from the sample of participants from Blomberg et al. (2021) which included 17 clinically stable adults with ADHD (age: $M = 28$, $SD = 6.8$) and 17 healthy controls (age: $M = 25$, $SD = 5.1$). Of the 17 ADHD participants, 15 were prescribed stimulant medication and abstained from their medication for at least 48 h prior to testing. The 18-item adult ADHD self-report scale (ASRS) v.1.1 (Kessler et al., 2005; Rodriguez et al., 2007) was used to assess ADHD-symptom severity associated with inattention, hyperactivity/impulsivity, and combined subtypes in both groups. See **Supplementary Methods** for further demographic and clinical details of the sample.

Image acquisition and preprocessing

Immediately prior to the task-based scan published in Blomberg et al. (2021), a ∼12 min, eyes-closed resting-state scan and an anatomical scan was acquired (Siemens Prisma 3T scanner). The functional resting state scan consisted of 940 echo planar imaging (EPI) whole-brain volumes (TR = 761 ms; TE = 24 ms; FA = 53°; FOV = 204 mm × 204 mm; acquisition matrix = 68 × 68; no. of slices = 45; slice thickness = 3 mm; voxel size = 3 mm × 3 mm × 3 mm). The anatomical scan consisted of 3D, T1-weighted MPRAGE (magnetization prepared rapid gradient echo) images (TR = 2300 ms; TE = 2.36 ms; FA = 8°; FOV = 250 mm × 250 mm × 225 mm; acquisition matrix = 288 × 288 × 208; slice orientation = sagittal; slice thickness = 0.9 mm; no. of slices = 208; voxel size = 0.87 mm × 0.68 mm × 0.9 mm) and double-echo spoiled gradient echo sequence field maps (TR = 520 ms; TE = 4.92/7.38 ms; FA = 60°; total EPI readout time = 16.415 ms; blip direction = 1).

Participants were instructed to lie as still as possible, let their mind's wander, and not to fall asleep. In order to make the acoustic environment as quiet as possible, the external auditory meatus of each ear was first protected with a self-hardening, moldable wax; next, participants were fitted with active noise canceling headphones (OptoAcoustics Ltd., Tel Aviv, Israel) which further attenuated the background EPI gradient noise to ∼58 dB SPL. The headphones were kept in place via inflatable positioning pads (Pearltec MRI/CT Multipad Plus, MagMedix, MA, United States) that also worked to minimize head movements within the 64-channel head coil.

Preprocessing was performed in MATLAB R2020B software using the CONN toolbox v.20.b (Whitfield-Gabrieli and Nieto-Castanon, 2012, 2017)[1] and included an additional denoising procedure in order to remove confounds of physiological noise (e.g., cerebral white matter, ventricles, large vessels, and cerebrospinal areas), head movement, outlier scans, as well as constant, and first-order linear session effects. We applied the software's default preprocessing pipeline for volume-based analyses but with indirect normalization to standard stereotactic (MNI) space as we had obtained gradient field maps during image acquisition (Nieto-Castanon, 2020). This particular procedure included: functional realignment and unwarp with the use of fieldmaps for susceptibility distortion correction; slice-timing correction; outlier identification in which framewise displacements greater than 0.9 mm or global BOLD signal changes above five SD were flagged as potential outliers; indirect segmentation and normalization; and spatial smoothing with CONN toolbox's default Gaussian kernel recommendation of 8 mm FWHM (full width half maximum).

The denoising pipeline consisted of the following two steps:

> Nuisance covariates derived from CONN-toolbox's implementation of anatomical component-based correction (aCompCor) were entered into an ordinary least squares regression in order to remove confounding effects on the estimated BOLD signal in each voxel per subject and run. The covariates included five noise components from cerebral white matter; five noise components from cerebrospinal areas; 12 subject motion components (three translation, three rotation, and their first-order temporal derivatives), outlier scans identified in the preprocessing procedure and components representing the effect of each task-condition convolved with the canonical hemodynamic response function in order to reduce the influence of slow trends, initial magnetization transients as well as constant task-related effects.
>
> Temporal band pass filtering (high pass: 0.008 Hz, low pass: 0.09 Hz) on the BOLD signal was applied in order to minimize the influence of physiological head motion and other noise sources.

Subsequent quality control analysis of preprocessing outcomes indicated that the mean framewise displacement (disregarding outlier scans) associated with micro-head movements (Controls: $M = 0.08$, $SD = 0.02$; ADHD: $M = 0.10$, $SD = 0.03$) was not significantly different between groups, $F_{(1,32)} = 3.8$, $p = 0.06$; nor was the mean number of valid (i.e., non-outlier) scans (Controls: $M = 916$, $SD = 14$ ADHD: $M = 912$, $SD = 17$) significantly different between groups $F_{(1,32)} = 0.5$, $p = 0.49$.

Statistical analysis

All statistical analyses were performed in the CONN-toolbox. The denoised, voxel-wise BOLD time series data was first averaged within each auditory, VA/SN and DMN predefined regions of interest (ROI; **Figure 1**) and then entered into a first level analysis wherein the correlation coefficient of each ROI to all other ROIs was calculated. Resulting correlation coefficients were Fisher z-transformed. Each participant's first level ROI-to-ROI connectivity matrices were then entered into a second level GLM to obtain group level estimates for connection-based and network-based inferences.

Functional network connectivity (FNC) analysis (Jafri et al., 2008) was used to ascertain if the expected anticorrelation between the VA/SN and DMN was weaker in the ADHD group relative to controls (O1). FNC analysis outputs an F-statistic representing the difference in network-level

---

[1] www.nitrc.org/projects/conn, RRID:SCR_009550

connectivity between groups and the significance of the F-statistic was corrected for by way of a false-discovery rate (FDR) cluster-level threshold of $p < 0.05$. Eventual *post hoc*, ROI-level exploration of the hypothesized reduced anticorrelation (ADHD > Controls) between the two networks utilized an uncorrected, one-tailed $p < 0.05$ connection-level threshold. In addition, for each network, an FDR corrected ($p < 0.05$) ROI-to-ROI, 2-tailed univariate analysis (Benjamini and Hochberg, 1995) was used to assess if there were differences in within-network FC between groups.

Seed-based ROI-to-ROI analysis explored the hypothesis that auditory regions would be more positively coupled with ROIs of the SN and less anticorrelated with ROIs of the DMN if ADHD participants were more inherently sensitive to their auditory environment (O2). To this end, a separate one-way MANOVA for each target ROI (VA/SN = 7 targets, **Figure 1A**; DMN = 4 targets; **Figure 1B**) was conducted to determine if there were group differences (ADHD > Controls) in FC with the 10 auditory seeds (**Figure 1C**). Thus, for each target ROI, the connectivity values for all 10 seed-to-target pairs were entered as dependent variables in the MANOVA. To correct for multiple analyses, we utilized an FDR-adjusted significance threshold of $p < 0.05$. We additionally performed an FDR corrected ($p < 0.05$) ROI-to-ROI, 2-tailed univariate analysis (Benjamini and Hochberg, 1995) to assess whether group differences in auditory FC alone, were evident in our sample.

To explore the brain-behavior relationship between individual differences (collapsed across groups) in connectivity and 2b-T accuracy and ADHD-symptom severity (O3), a threshold free cluster enhancement (TFCE) procedure (Smith and Nichols, 2009) was used to correct for multiple comparisons. First, the connectivity maps, made up of the 210 possible ROI-to-ROI pairs pertaining to the 21 ROIs of the DMN, VA/SN and auditory regions collectively, were sorted using a hierarchical optimal leaf ordering procedure (Bar-Joseph et al., 2001) embedded in the CONN-toolbox. CONN-toolbox's default statistical settings for TFCE analysis were then applied to identify significant clusters of ROI-to-ROI connections associated with 2b-T accuracy as well as symptom severity. This resulted in a TFCE score for each cluster and a family-wise error (FWE) corrected threshold of $p < 0.05$ (estimated using 1,000 permutation iterations of the data) was used to determine the significance of the TFCE scores. Eventual *post hoc* analysis utilized an uncorrected, two-tailed $p < 0.05$ connection-level threshold to identify the individual within-cluster ROI-to-ROI connections associated with the brain-behavior relationship. 2b-T data was missing for one of the ADHD participants, so this TFCE analysis included only 16 of the 17 ADHD participants. Similarly, impulsivity scores were missing for one of the control participants so TFCE analysis of the relationship between FC and impulsivity as well as total ASRS scores included only 16 of the 17 control participants.

Results

Group differences in default-mode network and ventral attention/salience network connectivity

Results of the FNC analysis confirmed that the anticorrelation between the DMN and VA/SN was significantly weaker $F_{(2,31)} = 7.2$, p-FDR = 0.008, in ADHD than controls (O1). *Post hoc* analysis of ROI-to-ROI connections indicated that the weaker anticorrelation (p-uncorrected) was most strongly associated with the medial prefrontal cortex (PFC) of the DMN and was evident across all VA/SN regions (**Table 1A**). The posterior cingulate cortex (PCC) of the DMN also contributed to the weaker anticorrelation of which coupling included the left anterior insula, the anterior cingulate and bilateral rostral PFC of the VA/SN (**Table 1A**). As an additional, *post hoc* explorative step, we tested to see if any of these DMN-VA/SN connections correlated positively (one-tailed, Spearman's rho) with individual differences (collapsed across groups) in ADHD-symptom severity. Results indicated that the weaker anticorrelated DMN–VA/SN FC was mostly associated with the severity of inattentive and combined symptoms across participants (see **Table 1B** for details). Tests for group differences in within-network FC were not significant.

Group differences in auditory connectivity

Between-group, seed-based analysis indicated that the phasic resting-state activity in auditory regions was significantly more positively correlated with the right supramarginal gyrus (SMG) of the VA/SN, $F_{(10,23)} = 4.7$, p-FDR = 0.006; and significantly less anticorrelated with the medial PFC of the DMN $F_{(10,23)} = 4.8$, p-FDR = 0.006, in the ADHD group; providing support to the hypothesis that adults with ADHD may be more inherently sensitive to their acoustic environment (O2). No differences between groups were observed for any of the other target ROIs. Nor did we observe groups differences in FC between auditory regions alone.

Individual differences in brain-behavior associations

The TFCE clustering procedure (O3), identified a significant cluster of connections between auditory and VA/SN ROIs that were negatively associated with 2b-T accuracy. TFCE = 34.9, p-FWE = 0.040. *Post hoc* connection-level analysis (see **Table 2** for detailed statistics) indicated that the negative brain-behavior relationship mostly involved FC between early

TABLE 1 **(A)** Results of the *post hoc* analysis (one-tailed, independent t-tests) characterizing the individual default-mode network–ventral attention/salience network (DMN–VA/SN) connections that were significantly ($p < 0.05$, uncorrected) less anticorrelated in attention deficit hyperactivity disorder (ADHD) participants compared to controls. **(B)** Results of the explorative *post hoc* correlation analysis (Spearman's rho) assessing the relationship of ADHD-symptom severity (inattentive, impulsive/hyperactive, combined) with the degree of reduced anticorrelated VA/SN–DMN functional connectivity (FC) across participants.

		A		**B**		
DMN –	**VA/SN**	***t* (32)**	***p***	**Inattentive**	**Impulsive**	**Combined**
Medial PFC	Anterior insula R	3.5	0.001	0.51**	0.41**	0.50**
	Rostral PFC R	3.0	0.006	0.43**	0.39*	0.48**
	Supramarginal gyrus R	2.9	0.007	0.33*	0.22	0.31*
	Anterior cingulate	2.8	0.009	0.37*	0.22	0.37*
	Supramarginal gyrus L	2.6	0.015	0.38*	0.36*	0.39*
	Anterior insula L	2.3	0.029	0.35*	0.24	0.31*
	Rostral PFC L	2.0	0.030	0.28	0.19	0.29
Posterior cingulate	Anterior insula L	2.7	0.006	0.45**	0.11	0.29
	Anterior cingulate	2.4	0.012	0.37*	0.19	0.30*
	Rostral PFC L	2.2	0.017	0.39*	0.17	0.32*
	Rostral PFC R	1.8	0.043	0.32*	0.15	0.29

Asterik indicate significant rho-values (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$). L, Left; R, right; PFC, prefrontal cortex.

auditory processing regions and core hubs of the VA/SN: the anterior cingulate, the anterior insulae and the right SMG (**Figure 2A**). Taken together, these results suggest that individuals who could perform well on the cognitively demanding working memory task whilst ignoring distracting acoustic stimulation also had more intrinsically segregated auditory–VA/SN connectivity at rest. This pattern of FC was also shown to have an inverse relationship with ADHD-symptom severity. TFCE analysis identified two significant clusters of auditory-VA/SN connections that were positively associated with symptom severity (**Figure 2B**). Cluster one, TFCE = 58.3, p-FWE = 0.006, consisted of increased FC between the right SMG and left lateralized auditory ROIs; and cluster two, TFCE = 40.07, p-FWE = 0.035, consisted of increased FC between the right SMG and right lateralized auditory ROIs (see **Table 2** for detailed connection-level statistics). TFCE analysis additionally identified one FC cluster, TFCE = 44.8, p-FWE = 0.032 that correlated with inattentive scores (**Figure 2C**) and one FC cluster, TFCE = 50.5, p-FWE = 0.014 that correlated with impulsivity scores (**Figure 2D**) wherein increases in inattentiveness and impulsivity were both associated with increased FC between the right SMG and left lateralized auditory ROIs.

Discussion

The purpose of the current study was to explore whether adults with ADHD would show aberrant FC compared to healthy controls between DMN, VA/SN, and auditory regions. Our combined results were rather striking. First, as per expectations, a weaker anticorrelation between the VA/SN and DMN was observed in ADHD participants. Second, this aberrant connectivity was underscored by an enhanced coupling between auditory ROIs and the right SMG of the VA/SN and a reduced anticorrelation between the medial PFC of the DMN and auditory ROIs. Third, it was shown that participants who displayed strong intrinsic segregation of the VA/SN and auditory ROIs at rest, were also better at performing well on a cognitively demanding visual working memory task that concurrently required participants to ignore a streaming acoustic signal. We discuss these results in more detail over the proceeding subsections.

Group differences in functional network connectivity

We observed a reduced anticorrelation between regions of the VA/SN and DMN in our sample of ADHD-adults relative to controls. This finding is in line with several previous studies (e.g., Castellanos et al., 2008; Sun et al., 2012; Sripada et al., 2014; Lin et al., 2018; Mills et al., 2018; c.f. also Sato et al., 2012) and contributes further evidence for an impaired functional organization between these two networks in the disorder. Particularly noteworthy, is that our results replicate the findings of Lin et al. (2018, **Supplementary Figure 4**), who used the same FNC analysis (Jafri et al., 2008) we employed here in a large medication-naive sample (n = 80) of adults with ADHD. However, unlike some reports from previous studies in ADHD (see Castellanos and Aoki, 2016 for a review), we did not observe reduced FC-strength between ROIs of the DMN in ADHD participants. Although it is difficult to speculate why we did not replicate this finding our sample, the fact that we did

TABLE 2 Table lists connection-level results for the clusters of VN/SN–auditory connections that were negatively associated with 2b-T accuracy and positively associated with attention deficit hyperactivity disorder (ADHD)-symptom severity (combined, inattentiveness and impulsivity). One-sample *t*-values, indicate that the correlation was significantly ($p < 0.05$, uncorrected, two-tailed) different from zero. Pearson's correlation coefficient (*r*) indicates the strength of the relationship with the behavioral variable for each ROI-to-ROI connection.

2-back task accuracy	**VA/SN -**	**Auditory**	***t* (31)†**	***p***	***r***
	Anterior cingulate	Heschl's gyrus L	−2.6	0.014	−0.42
		Granular posterior insula L	−2.1	0.045	−0.37
		Heschl's gyrus R	−2.1	0.045	−0.35
		Granular posterior insula R	−3.1	0.004	−0.48
	Anterior insula L	Heschl's gyrus L	−2.6	0.014	−0.42
		Heschl's gyrus R	−3.3	0.003	−0.51
		Granular posterior insula R	−3.5	0.001	−0.53
	Anterior insula R	Heschl's gyrus L	−2.6	0.014	−0.42
		Granular posterior insula L	−2.2	0.034	−0.37
		Granular posterior insula R	−2.8	0.009	−0.44
	Supramarginal gyrus R	Heschl's gyrus L	−2.5	0.018	−0.41
		Granular posterior insula L	−2.7	0.013	−0.43
		Heschl's gyrus R	−2.2	0.035	−0.37
		Planum temporale R	−2.1	0.045	−0.35
		Granular posterior insula R	−2.8	0.010	−0.44
ADHD combined	**VA/SN -**	**Auditory**	***t* (31)‡**	***p***	***r***
Cluster 1:	Supramarginal gyrus R	Heschl's gyrus L	3.7	0.001	0.55
		Granular posterior insula L	3.4	0.002	0.53
		Planum temporale L	4.7	0.000	0.65
Cluster 2:	Supramarginal gyrus R	Granular posterior insula R	4.1	0.000	0.59
		Heschl's gyrus R	3.0	0.006	0.47
ADHD combined	**VA/SN -**	**Auditory**	***t* (31)‡**	***p***	***r***
	Supramarginal gyrus R	Heschl's gyrus L	3.8	0.000	0.55
		Planum temporale L	4.2	0.001	0.59
ADHD combined	**VA/SN -**	**Auditory**	***t* (31)‡**	***p***	***r***
	Supramarginal gyrus R	Planum temporale L	4.7	0.000	0.64

† Analysis conducted on 33 of the 34 participants because 2b-T data was missing for one of the ADHD participants.
‡ Analysis conducted on 33 of the 34 participants because impulsivity scores were missing for one the control participants.
L, left; R, right.

not find group differences in within-network FC for either the DMN or the VA/SN suggests that the weaker between network anticorrelation was unlikely being driven by a single, abnormally regulated network.

Individual differences in the frequency of inattentive and combined symptoms also correlated with the magnitude of reduced anticorrelation between the VA/SN and DMN across participants. Interestingly, the *default-mode interference* hypothesis (Sonuga-Barke and Castellanos, 2007) has tended to dominate interpretations of this atypical connectivity associated with symptomatic inattentiveness in ADHD, wherein the weaker anticorrelation is conceived as an index for spontaneous mind wandering (cf, Kelly et al., 2008). However, an alternative hypothesis suggests that a reduced anticorrelation may additionally index an inherent susceptibility to environmental distraction (Menon, 2011; Aboitiz et al., 2014). Consistent with this latter hypothesis, we also observed in the ADHD group, an enhanced coupling between auditory regions and the right SMG. The SMG is an integrative hub of the VA/SN, and the right lateralized region has been particularly implicated in the mediation of exogenous attention toward visual, tactile and auditory modalities (Corbetta and Shulman, 2002; Corbetta et al., 2008; Cabeza et al., 2012; Vossel et al., 2014). The increased FC between the VA/SN and auditory regions may mean that the right SMG is intrinsically biased to the auditory modality in ADHD and indicative of a symptomatic, heightened sensitivity to the acoustic environment. Alternatively, the increased connectivity

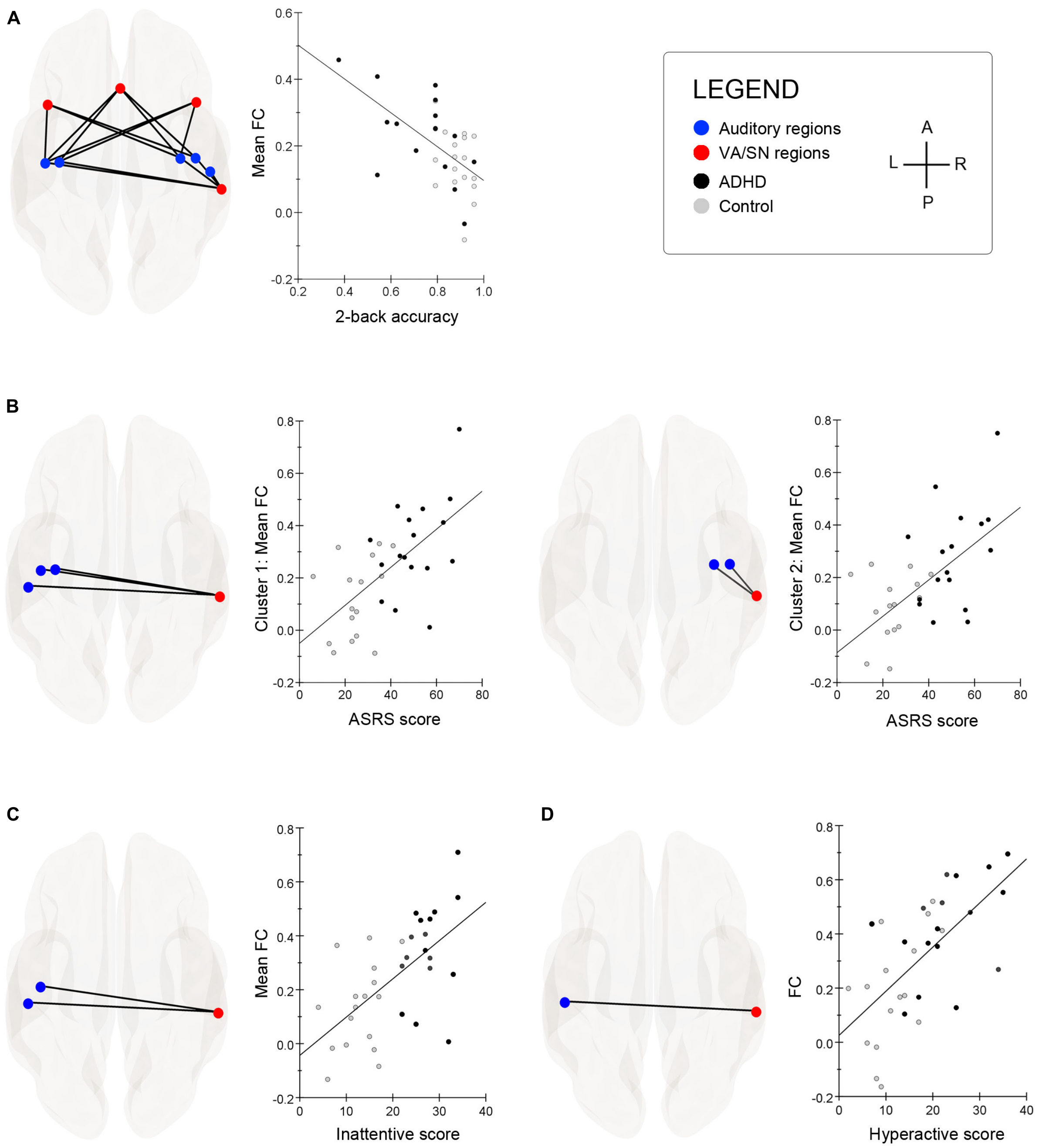


FIGURE 2

(A) Superior glass-brain images display the auditory–ventral attention/salience network (VA/SN) connections that were significantly associated with 2-back task (2b-T) accuracy. Scatter-chart shows that the more segregated the resting-state VA/SN–auditory functional connectivity (FC), the better the individual performed on a 2b-T whilst ignoring a distracting acoustic signal (for visualization purposes, *y*-axis represents the average of all significant FC values associated with 2b-T accuracy). **(B)** The threshold free cluster enhancement (TFCE) procedure identified two significant clusters of connections between the right supramarginal gyrus (SMG) and auditory ROIs that were positively associated with combined ADHD-symptom severity scores (ASRS). Cluster 1: left glass-brain; Cluster 2: right-glass brain. Associated scatter-charts show that increased resting-state FC between the right SMG and auditory ROIs was correlated with attention deficit hyperactivity disorder (ADHD)-symptom severity (*y*-axis represents the average of all significant FC values within each cluster). **(C)** Hyperconnectivity between the right SMG and left lateralized Heschl's gyrus and planum temporale was positively associated with participants' inattentive scores. Scatter chart depicts the strength of relationship collapsed across groups (*y*-axis represents the average of all significant FC values within each cluster). **(D)** Hyperconnectivity between the right SMG and left lateralized planum temporale was positively associated with participants' hyperactivity/impulsivity scores. Scatter chart depicts the strength of relationship collapsed across groups. A, anterior; P, posterior; L, left; R, right.

may be indicative of increased functional communications between the right SMG and auditory ROIs which would suggest that participants with ADHD were more aware of their acoustic environment throughout the resting-state. Further experimental research is needed to disentangle these two alternatives, but either way, both alternatives lend support to the hypothesis that adults with ADHD are more susceptible to auditory distraction.

Interestingly, a similar notion was underscored in an original study by Schulze et al. (2021) that explored whether aberrant intrinsic network FC was associated with performance deficits in multisensory integration in adult ADHD. By way of the McGurk illusion, the authors found that ADHD participants experienced significantly fewer illusions to that of healthy controls due to a sensory bias for auditory stimuli. And how well participants integrated the McGurk illusion was negatively associated with childhood symptom severity and self-rated inattentiveness in adulthood. Resting-state FC in ADHD participants was increased between the planum temporale and anterior insula for which the authors speculated that this auditory–VA/SN hyperconnectivity may be indicative of an increased susceptibility to auditory distraction. Interestingly, performance on the McGurk illusion was inversely associated with hyperconnectivity between the Heschl's gyrus and the middle temporal gyrus (a convergence zone for e.g., audiovisual integration). Taken together, Schulze et al. (2021) concluded that communication between sensory areas to integrative hub regions might be disrupted in ADHD, impacting the appropriate assignment of top-down attentional allocation.

The current study also observed a reduced anticorrelation between auditory regions and the medial PFC in our ADHD sample relative to controls. The medial PFC is a highly metabolically demanding hub of the DMN and similarly to the VA/SN, has been shown to have an anticorrelated relationship with sensory systems (e.g., Greicius and Menon, 2004). Interestingly, our finding is consistent with a study by Cocchi et al. (2012) who observed a reduced anticorrelation between the medial PFC of the DMN and an associative auditory region in the left superior temporal lobe in their non-clinical sample of drug-naive young adult students with ADHD. Our sample was similar to that of Cocchi et al.'s (2012) in that the majority of participants from both control and clinical groups were young adult students. The fact that Cocchi et al. (2012) observed similarly altered auditory–DMN FC in their drug-naive sample, that additionally correlated with ADHD-symptom severity scores, reinforces the hypothesis that the DMN is a locus of dysfunction in ADHD (see for example: Castellanos et al., 2009). However, Cocchi et al.'s (2012) findings, in combination with our own, are also an indication that the behavioral implication of DMN dysfunction in ADHD is not limited to attentional lapses associated with spontaneous mind wandering. Indeed, although heavily implicated in internal mentation (Gusnard et al., 2001), additional documented functions for the medial PFC include passive monitoring of the environment (Buckner et al., 2008; Dohmatob et al., 2017), perceptual binding (Martínez-Sanchis, 2014) and top-down modulation of sensory interference (Kucyi et al., 2013; Chadick et al., 2014; Martin-Cortecero and Nuñez, 2016). Collectively, these additional functions suggest that the DMN also plays an important role in both sensory and attentional processing and provide important clues about the implications that aberrant auditory–DMN functional organization may have with respect to auditory distraction in ADHD.

Brain-behavior relationships

Our most striking finding was a resting-state relationship with performance from a demanding visual 2-back task (2b-T) that concurrently required participants to ignore a streaming acoustic signal (∼75 dB SPL). Throughout this task, controls were more proficient than ADHD participants at attenuating the evoked auditory activity from the distracting acoustic signal and performed overall better than ADHD participants. In addition, the more attenuated the auditory evoked responses were across participants, the lower their symptom severity scores of inattentiveness. In the current study, we showed that participants who had the capacity to perform well on the 2b-T whilst ignoring the distracting acoustic signal, also tended to have more intrinsically segregated VA/SN–auditory connectivity at rest. This pattern of VA/SN–auditory FC was also shown to have an inverse relationship with ADHD-symptom severity. Specifically, increased connectivity between the right SMG of the VA/SN and auditory cortices (Heschl's gyrus, posterior insula, and planum temporale) was positively associated with the severity of ADHD-symptoms across participants. In sum, the resting-state brain-behavior relationships observed in the current study are supportive of theories that suggest that the VA/SN plays a pivotal role in the manifestation of attentional deficits in neuropsychiatric disorders (Menon, 2011), wherein an impaired intrinsic organization of the VA/SN can lead to an inappropriate bias toward irrelevant/salient stimuli, and behaviorally, result in an increased susceptibility for environmental distraction.

Limitations

Because we have already addressed in detail the limitations regarding our sample in Blomberg et al. (2021), these will only be briefly listed here. First, our ADHD-sample included more females than males. Although childhood-ADHD is more commonly diagnosed in boys, the differences in prevalence between sexes diminishes almost completely in adulthood (Faraone et al., 2015; Matte et al., 2015), so we should not expect the gender imbalance in our sample to dramatically affect more general conclusions of our results. Second, our ADHD-sample included medicated individuals, half of which were also on stable SSRI medication, which is indicative of earlier problems with anxiety and depression. However, both anxiety and depression are at the lower end of the spectrum of expected psychiatric comorbidities in adults with ADHD (Katzman et al., 2017). If individuals with ADHD are expected to have more severe symptoms and functional impairment than our sample, then the group differences reported here are likely to underestimate the overall differences between groups in the general population, rendering our results conservative. In addition, it is worth noting that similar findings of aberrant resting-state connectivity associated with ADHD-symptoms reported in the current study, have also been observed in other studies with medication-naive adult samples (Cocchi et al., 2012; Lin et al., 2018).

In addition to the aforementioned limitations regarding our sample, two methodological limitations specific to this study are worthy of addressing. First, we did not implement any physiological methods that would allow us to monitor the arousal levels of the participants throughout the resting-state period. The functional resting-state duration was ∼12 min—prior to which participants also underwent ∼6 min of anatomical scans—and it is known that some participants can fall asleep in the scanner under conditions where they are left to rely on mentation as their sole source of stimulation. We can however, state that only two accepted applicants for this study reported falling asleep in the scanner (one ADHD and one control), and these participants were thereby excluded from the sample used here and in Blomberg et al. (2021). Because our main purpose was to explore whether adults with ADHD are inherently more sensitive to their acoustic environment—which to some degree involves a heightened level of awareness to auditory stimulation—our findings are still interesting, even if participants waned in their levels of arousal throughout the resting-state period.

A second limitation is that we did not have a means of monitoring eyes-open versus eyes-closed in participants. Participants wore MRI goggles in the scanner throughout the anatomical and functional scan. Because the anatomical scan did not require participants to have their eyes closed, the goggles presented a dimly lit word (dark gray on black background) reading: *REST* (*VILA* in Swedish), which slowly moved (figure-eight animation) within participants' field of view. This animation continued throughout the duration of the resting-state period. Had participants opted to open their eyes during the resting-state scan, then this was the only visual stimulation they received. Compared with a visually salient fixation cross—which many studies use under these circumstances—we tentatively suggest that our choice of visual stimulation, albeit novel, may have at least circumvented unwanted externally-directed frontal eye field activity associated with ocular fixation (Vernet et al., 2014), even though visual cortical activity was imminent. Eyes-open resting-state is known to result in increased visual network connectivity (Yang et al., 2007). However, the modality of interest in the current study was the auditory modality, so we should not expect that eventual periodic states of eyes-open to impact the study's overall conclusions.

Conclusion

In accord with previous studies, a reduced resting-state anticorrelation between the VA/SN and DMN was observed in our sample of adults with ADHD. Moreover, core hubs of the DMN and VA/SN, which respectively have been implicated in top-down and bottom-up regulation of attention to sensory events, exhibited aberrant FC with the auditory network in ADHD participants. Additionally, it was shown that participants who displayed stronger intrinsic segregation of the VA/SN and auditory network at rest, were also better at performing well on a cognitively demanding visual working memory task whilst attenuating distracting auditory stimulation (task conditions where ADHD participants proved inferior to controls). Overall, our collective results are consistent with the hypothesis that auditory distraction and more generally inattentiveness in ADHD is linked to aberrant interactions between DMN, VA/SN, and auditory systems. Importantly, our findings contribute further evidence to current etiological models of ADHD that implicate dysfunctional organization of DMN, VA/SN and other major intrinsic networks in behavioral and clinical symptoms in the disorder. Our findings also encourage more research into sensory interactions with these major intrinsic networks so that we can refine our theories of inattention and better understand factors that impact symptoms of sensory distraction in the disorder.

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 Regional Ethical Review Board in Linköping, Sweden (DNR 2019-06158). The patients/participants provided their written informed consent to participate in this study.

Author contributions

JR, RB, AC, and HD contributed to the conception and design of the study. AC and RB were responsible for participant recruitment and data collection. RB was responsible for data analysis as well as manuscript drafting and was assisted by IP, AC, and CS. All authors scrutinized the statistical analysis, contributed to the manuscript's revision, and approved submitted version.

Funding

This research was funded by the Swedish Research Council (2015-01917 and 2017-06092).

Acknowledgments

We would like to extend a special thank you to Paul Hamilton and Robin Kämpe for the MRI expertise necessary for this study. We also thank Per Gustafsson for his collaborative recommendations and theoretical advice. Thank you to Albin Pettersson and Michaela Socher for their piloting assistance.

Conflict of interest

AC had received speaker's fees, and/or scientific advisory board compensation from Lundbeck, Indivior, DNE Pharma, and Camurus, all outside the scope of the current project.

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.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2022.972730/full#supplementary-material

References

Aboitiz, F., Ossandon, T., Zamorano, F., Palma, B., and Carrasco, X. (2014). Irrelevant stimulus processing in ADHD: Catecholamine dynamics and attentional networks. *Front. Psychol.* 5:183. doi: 10.3389/fpsyg.2014.00183

Anticevic, A., Cole, M. W., Murray, J. D., Corlett, P. R., Wang, X.-J., and Krystal, J. H. (2012). The role of default network deactivation in cognition and disease. *Trends Cogn. Sci.* 16, 584–592. doi: 10.1016/j.tics.2012.10.008

Bar-Joseph, Z., Gifford, D. K., and Jaakkola, T. S. (2001). Fast optimal leaf ordering for hierarchical clustering. *Bioinformatics* 17, S22–S29. doi: 10.1093/bioinformatics/17.suppl_1.S22

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Lond. B Methodol.* 57, 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x

Blomberg, R., Johansson Capusan, A., Signoret, C., Danielsson, H., and Rönnberg, J. (2021). The effects of working memory load on auditory distraction in adults with attention deficit hyperactivity disorder. *Front. Hum. Neurosci.* 15:724. doi: 10.3389/fnhum.2021.771711

Bozhilova, N. S., Michelini, G., Kuntsi, J., and Asherson, P. (2018). Mind wandering perspective on attention-deficit/hyperactivity disorder. *Neurosci. Biobehav. Rev.* 92, 464–476. doi: 10.1016/j.neubiorev.2018.07.010

Buckner, R. L., Andrews-Hanna, J. R., and Schacter, D. L. (2008). The Brain's default network: Anatomy, function, and relevance to disease. *Ann. N. Y. Acad. Sci.* 1124, 1–38. doi: 10.1196/annals.1440.011

Cabeza, R., Ciaramelli, E., and Moscovitch, M. (2012). Cognitive contributions of the ventral parietal cortex: An integrative theoretical account. *Trends Cogn. Sci.* 16, 338–352. doi: 10.1016/j.tics.2012.04.008

Castellanos, F. X., and Aoki, Y. (2016). Intrinsic functional connectivity in attention-deficit/hyperactivity disorder: A science in development. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* 1, 253–261. doi: 10.1016/j.bpsc.2016.03.004

Castellanos, F. X., and Proal, E. (2012). Large-scale brain systems in ADHD: Beyond the prefrontal–striatal model. *Trends Cogn. Sci.* 16, 17–26. doi: 10.1016/j.tics.2011.11.007

Castellanos, F. X., Kelly, A. M. C., and Milham, M. P. (2009). The restless brain: Attention-deficit/hyperactivity disorder, resting state functional connectivity and intrasubject variability. *Can. J. Psychiatry* 54, 665–672. doi: 10.1177/070674370905401003

Castellanos, F. X., Margulies, D. S., Kelly, C., Uddin, L. Q., Ghaffari, M., Kirsch, A., et al. (2008). Cingulate-precuneus interactions: A new locus of dysfunction in adult attention-deficit/hyperactivity disorder. *Biol. Psychiatry* 63, 332–337. doi: 10.1016/j.biopsych.2007.06.025

Chadick, J. Z., Zanto, T. P., and Gazzaley, A. (2014). Structural and functional differences in medial prefrontal cortex underlie distractibility and suppression deficits in ageing. *Nat. Commun.* 5:4223. doi: 10.1038/ncomms5223

Cocchi, L., Bramati, I. E., Zalesky, A., Furukawa, E., Fontenelle, L. F., Moll, J., et al. (2012). Altered functional brain connectivity in a Non-clinical sample of young adults with attention-deficit/hyperactivity disorder. *J. Neurosci.* 32, 17753–17761. doi: 10.1523/JNEUROSCI.3272-12.2012

Corbetta, M., and Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nat. Rev. Neurosci.* 3, 201–215. doi: 10.1038/nrn755

Corbetta, M., Patel, G., and Shulman, G. L. (2008). The reorienting system of the human brain: From environment to theory of mind. *Neuron* 58, 306–324. doi: 10.1016/j.neuron.2008.04.017

Dohmatob, E., Dumas, G., and Bzdok, D. (2017). Dark control: A unified account of default mode function by control theory and reinforcement learning. *bioRxiv* [Preprint]. doi: 10.1101/148890

Faraone, S. V., Asherson, P., Banaschewski, T., Biederman, J., Buitelaar, J. K., Ramos-Quiroga, J. A., et al. (2015). Attention-deficit/hyperactivity disorder. *Nat. Rev. Dis. Primers* 1:15020. doi: 10.1038/nrdp.2015.20

Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Essen, D. C. V., and Raichle, M. E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proc. Natl. Acad. Sci. U.S.A.* 102, 9673–9678. doi: 10.1073/pnas.0504136102

Fox, M. D., Zhang, D., Snyder, A. Z., and Raichle, M. E. (2009). The global signal and observed anticorrelated resting state brain networks. *J. Neurophysiol.* 101, 3270–3283. doi: 10.1152/jn.90777.2008

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

Gusnard, D. A., Akbudak, E., Shulman, G. L., and Raichle, M. E. (2001). Medial prefrontal cortex and self-referential mental activity: Relation to a default mode of brain function. *Proc. Natl. Acad. Sci. U.S.A.* 98, 4259–4264. doi: 10.1073/pnas.071043098

Jafri, M. J., Pearlson, G. D., Stevens, M., and Calhoun, V. D. (2008). A method for functional network connectivity among spatially independent resting-state components in schizophrenia. *Neuroimage* 39, 1666–1681. doi: 10.1016/j.neuroimage.2007.11.001

Katzman, M. A., Bilkey, T. S., Chokka, P. R., Fallu, A., and Klassen, L. J. (2017). Adult ADHD and comorbid disorders: Clinical implications of a dimensional approach. *BMC Psychiatry* 17:302. doi: 10.1186/s12888-017-1463-3

Kelly, A. M. C., Uddin, L. Q., Biswal, B. B., Castellanos, F. X., and Milham, M. P. (2008). Competition between functional brain networks mediates behavioral variability. *NeuroImage* 39, 527–537. doi: 10.1016/j.neuroimage.2007.08.008

Kessler, R. C., Adler, L., Ames, M., Demler, O., Faraone, S., Hiripi, E., et al. (2005). The world health organization adult ADHD self-report scale (ASRS): A short screening scale for use in the general population. *Psychol. Med.* 35, 245–256. doi: 10.1017/S0033291704002892

Konrad, K., and Eickhoff, S. B. (2010). Is the ADHD brain wired differently? A review on structural and functional connectivity in attention deficit hyperactivity disorder. *Hum. Brain Mapp.* 31, 904–916. doi: 10.1002/hbm.21058

Kucyi, A., Salomons, T. V., and Davis, K. D. (2013). Mind wandering away from pain dynamically engages antinociceptive and default mode brain networks. *Proc. Natl. Acad. Sci. U.S.A.* 10, 18692–18697. doi: 10.1073/pnas.1312902110

Lin, H.-Y., Cocchi, L., Zalesky, A., Lv, J., Perry, A., Tseng, W.-Y. I., et al. (2018). Brain–behavior patterns define a dimensional biotype in medication-naïve adults with attention-deficit hyperactivity disorder. *Psychol. Med.* 48, 2399–2408. doi: 10.1017/S0033291718000028

Martin-Cortecero, J., and Nuñez, A. (2016). Sensory responses in the medial prefrontal cortex of anesthetized rats. Implications for sensory processing. *Neuroscience* 339, 109–123. doi: 10.1016/j.neuroscience.2016.09.045

Martínez-Sanchis, S. (2014). Neurobiological foundations of multisensory integration in people with autism spectrum disorders: The role of the medial prefrontal cortex. *Front Hum Neurosci* 8:970. doi: 10.3389/fnhum.2014.00970

Matte, B., Anselmi, L., Salum, G. A., Kieling, C., Gonçalves, H., Menezes, A., et al. (2015). ADHD in DSM-5: A field trial in a large, representative sample of 18- to 19-year-old adults. *Psychol. Med.* 45, 361–373. doi: 10.1017/S0033291714001470

Menon, V. (2011). Large-scale brain networks and psychopathology: A unifying triple network model. *Trends Cogn. Sci.* 15, 483–506. doi: 10.1016/j.tics.2011.08.003

Mills, B. D., Miranda-Dominguez, O., Mills, K. L., Earl, E., Cordova, M., Painter, J., et al. (2018). ADHD and attentional control: Impaired segregation of task positive and task negative brain networks. *Netw. Neurosci.* 2, 200–217. doi: 10.1162/netn_a_00034

Nieto-Castanon, A. (2020). *Handbook of functional connectivity magnetic resonance imaging methids in CONN*. La Salle, IL: Hilbert Press.

Posner, J., Park, C., and Wang, Z. (2014). Connecting the dots: A review of resting connectivity MRI studies in attention-deficit/hyperactivity disorder. *Neuropsychol. Rev.* 24, 3–15. doi: 10.1007/s11065-014-9251-z

Rodriguez, A., Ginsberg, Y., Fernholm, A., and Nyberg, L. (2007). ADHD difficult to diagnose in adults. ASRS v1.1 self-report scales valuable help–now translated to Swedish. *Lakartidningen* 104, 1398–1400.

Sato, J. R., Hoexter, M. Q., Castellanos, X. F., and Rohde, L. A. (2012). Abnormal brain connectivity patterns in adults with ADHD: A coherence study. *PLoS One* 7:e45671. doi: 10.1371/journal.pone.0045671

Schulze, M., Aslan, B., Stöcker, T., Stirnberg, R., Lux, S., and Philipsen, A. (2021). Disentangling early versus late audiovisual integration in adult adhd: A combined behavioural and resting-state connectivity study. *J. Psychiatry Neurosci.* 46, E528–E537. doi: 10.1503/jpn.210017

Schulze, M., Lux, S., and Philipsen, A. (2020). Sensory processing in Adult ADHD – A systematic review. *Res. Sq.* [Preprint]. doi: 10.21203/rs.3.rs-71514/v1

Smith, S. M., and Nichols, T. E. (2009). Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. *Neuroimage* 44, 83–98. doi: 10.1016/j.neuroimage.2008.03.061

Sonuga-Barke, E., and Castellanos, F. X. (2007). Spontaneous attentional fluctuations in impaired states and pathological conditions: A neurobiological hypothesis. *Neurosci. Biobehav. Rev.* 31, 977–986. doi: 10.1016/j.neubiorev.2007.02.005

Sripada, C., Kessler, D., Fang, Y., Welsh, R. C., Prem Kumar, K., and Angstadt, M. (2014). Disrupted network architecture of the resting brain in attention-deficit/hyperactivity disorder: Disrupted network architecture in ADHD. *Hum. Brain Mapp.* 35, 4693–4705. doi: 10.1002/hbm.22504

Sun, L., Cao, Q., Long, X., Sui, M., Cao, X., Zhu, C., et al. (2012). Abnormal functional connectivity between the anterior cingulate and the default mode network in drug-naïve boys with attention deficit hyperactivity disorder. *Psychiatry Res. Neuroimag.* 201, 120–127. doi: 10.1016/j.pscychresns.2011.07.001

Sutcubasi, B., Metin, B., Kurban, M. K., Metin, Z. E., Beser, B., and Sonuga-Barke, E. (2020). Resting-state network dysconnectivity in ADHD: A system-neuroscience-based meta-analysis. *World J. Biol. Psychiatry* 21, 662–672. doi: 10.1080/15622975.2020.1775889

Uddin, L. Q., Yeo, B. T. T., and Spreng, R. N. (2019). Towards a universal taxonomy of macro-scale functional human brain networks. *Brain Topogr.* 32, 926–942. doi: 10.1007/s10548-019-00744-6

Vernet, M., Quentin, R., Chanes, L., Mitsumasu, A., and Valero-Cabré, A. (2014). Frontal eye field, where art thou? Anatomy, function, and non-invasive manipulation of frontal regions involved in eye movements and associated cognitive operations. *Front. Integr. Neurosci.* 8:66. doi: 10.3389/fnint.2014.00066

Vossel, S., Geng, J. J., and Fink, G. R. (2014). Dorsal and ventral attention systems. *Neuroscientist* 20, 150–159. doi: 10.1177/1073858413494269

Whitfield-Gabrieli, S., and Nieto-Castanon, A. (2012). Conn: A functional connectivity toolbox for correlated and anticorrelated brain networks. *Brain Connectivity* 2, 125–141. doi: 10.1089/brain.2012.0073

Whitfield-Gabrieli, S., and Nieto-Castanon, A. (2017). *Functional connectivity SPM toolbox 2017*. Available online at: http://www.nitrc.org/projects/conn (accessed January 27, 2020).

Yang, H., Long, X.-Y., Yang, Y., Yan, H., Zhu, C.-Z., Zhou, X.-P., et al. (2007). Amplitude of low frequency fluctuation within visual areas revealed by resting-state functional MRI. *NeuroImage* 36, 144–152. doi: 10.1016/j.neuroimage.2007.01.054

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 28 September 2022
DOI 10.3389/fnins.2022.995119

Intention-based predictive information modulates auditory deviance processing

Andreas Widmann[1,2*†] and Erich Schröger[1*†]

[1]Wilhelm Wundt Institute for Psychology, Leipzig University, Leipzig, Germany, [2]Leibniz Institute for Neurobiology, Magdeburg, Germany

The human brain is highly responsive to (deviant) sounds violating an auditory regularity. Respective brain responses are usually investigated in situations when the sounds were produced by the experimenter. Acknowledging that humans also actively produce sounds, the present event-related potential study tested for differences in the brain responses to deviants that were produced by the listeners by pressing one of two buttons. In one condition, deviants were unpredictable with respect to the button-sound association. In another condition, deviants were predictable with high validity yielding correctly predicted deviants and incorrectly predicted (mispredicted) deviants. Temporal principal component analysis revealed deviant-specific N1 enhancement, mismatch negativity (MMN) and P3a. N1 enhancements were highly similar for each deviant type, indicating that the underlying neural mechanism is not affected by intention-based expectation about the self-produced forthcoming sound. The MMN was abolished for predictable deviants, suggesting that the intention-based prediction for a deviant can overwrite the prediction derived from the auditory regularity (predicting a standard). The P3a was present for each deviant type but was largest for mispredicted deviants. It is argued that the processes underlying P3a not only evaluate the deviant with respect to the fact that it violates an auditory regularity but also with respect to the intended sensorial effect of an action. Overall, our results specify current theories of auditory predictive processing, as they reveal that intention-based predictions exert different effects on different deviance-specific brain responses.

KEYWORDS

prediction, audition, intention, perception, action, predictive coding, mismatch negativity (MMN)

Check for updates

OPEN ACCESS

EDITED BY
Jerker Rönnberg,
Linköping University, Sweden

REVIEWED BY
Torge Dellert,
University of Münster, Germany
Manuel S. Malmierca,
University of Salamanca, Spain

*CORRESPONDENCE
Andreas Widmann
widmann@uni-leipzig.de
Erich Schröger
schroger@uni-leipzig.de

[†]These authors have contributed equally to this work

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

RECEIVED 15 July 2022
ACCEPTED 08 September 2022
PUBLISHED 28 September 2022

CITATION
Widmann A and Schröger E (2022) Intention-based predictive information modulates auditory deviance processing.
Front. Neurosci. 16:995119.
doi: 10.3389/fnins.2022.995119

Introduction

Sounds violating an auditory regularity trigger a cascade of deviance-specific brain responses, even when the auditory stimulation is task-irrelevant (Näätänen, 1990). The underlying mechanisms are in the service of detecting "new," unexpected, yet potentially relevant information. A phenomenological consequence of this deviance-specific processing can be attentional capture, while a behavioral consequence can be impaired performance in a primary task not related to the deviancy (Parmentier, 2014). Current theories of auditory predictive processing postulate that deviance processing is achieved on the basis of neural models representing detected auditory regularities that generate (implicit) predictions about what to expect next (Grimm and Schröger, 2007; Winkler et al., 2009; Escera et al., 2014; Schröger et al., 2015; Heilbron and Chait, 2018). The huge amount of research on this topic is based on experiments where the experimenter controls the delivery of the sounds. However, listeners are also active agents intentionally producing sounds by themselves. Predictive coding theory postulates that actions induce active inference to minimize sensory prediction errors (Friston and Stephan, 2007; Friston et al., 2010; Brown et al., 2013; Clark, 2013). In other words, active behavior should interact with sensory processing. Indeed, it has been shown that self-produced sounds are compared to the intended (predicted) sensorial consequence of the action (Waszak et al., 2012; Hughes et al., 2013), and auditory regularity-based and intention-based predictive processing of sounds interact (Korka et al., 2019; Darriba et al., 2021). The present event-related potential (ERP) study investigates whether and how deviance-specific processing based on auditory regularities is modulated for self-produced sounds.

Participants were asked to press one button frequently and a second button rarely. In one experimental condition the two buttons were not associated with a particular sound, but standard and deviant sounds were randomly and unpredictably presented irrespective of which of the two buttons was pressed ("unpredictable condition"; see **Figure 1**). In another condition standard and deviant sounds were predictably associated to the two buttons with high validity ("predictable condition"). One button produced a standard sound, and the other button produced a deviant sound ("predicted deviant") in most trials. However, there were also some incorrectly predicted deviant sounds produced when the button for a standard sound was pressed ("mispredicted deviant"). The present study considers three major deviance-specific ERP effects, namely, the N1 enhancement, the mismatch negativity (MMN) and the P3a.

The N1 (peaking around 100 ms after sound onset) is often reported to be enhanced for deviants relative to standards. This effect can be explained by standard sound-specific adaptation of the N1 eliciting neurons (Näätänen, 1990). When a deviant is presented, (partly) non-adapted neurons are activated resulting in relative enhancement of the N1. Such an effect on scalp-recorded ERPs can be explained by short-term synaptic depression of neurons in auditory cortex causing a transient weakening of synaptic connections (May, 2021). As this theory does not (explicitly) include top-down influence of intentional action, a modulation of the auditory-regularity based N1 enhancement to deviants is not to be expected. Indeed, Korka et al. (2019) did not find a N1 deviance effect for a deviant sound which violated an intention-based prediction. Similarly, Darriba et al. (2021) found no modulations for the Na and Tb subcomponents of the N1 for violations of an expected action effect. Note, however, that the auditory N1 *per se* can be modulated by top-down effects, for example, it is increased when the sound is attended and decreased when the sound is self-generated (for reviews see, Horvath, 2015; Schröger et al., 2015). Thus, one may possibly expect modulations of the N1 oddball effect by intention when perception and action are as strongly linked as in the present paradigm.

Subsequent to and partially overlapping with the N1, the MMN is elicited by violations of an auditory regularity. MMN is often explained as resulting from a mismatch process comparing the actual sound with a prediction derived from an internal model representing the regularity (Garrido et al., 2009; Winkler and Czigler, 2012). Many studies with externally generated sounds reported that the MMN is not modulated by attentional top-down predictive information (for review see e.g., Sussman et al., 2014). The MMN-system is of special interest for the present study because it can process different predictions concurrently and can generate respective MMN responses to violations of these predictions in parallel (Paavilainen et al., 2001, 2003; Wolff and Schröger, 2001; Pieszek et al., 2013). According to an extension of the "auditory event representation system (AERS)" framework (Korka et al., 2022), it is assumed that sound predictions generated by action intention are functionally equivalent to sound predictions generated by an extracted auditory regularity. This is evidenced by the finding that the violation of an intention-based prediction alone–in the absence of an auditory regularity-based prediction–can elicit MMN (Korka et al., 2019).

Do these MMNs for auditory-regularity violation and action-intention violation interact? In a study by Nittono (2006), the MMNs for self-generated sounds triggered by a button press and externally generated sounds did not differ from each other. As deviants were fully unpredictable in this pioneering study, an additional MMN modulation by action intention was not necessarily to be expected by predictive coding theories. In a study by Waszak and Herwig (2007), where two buttons (instead of one as in the Nittono, 2006 study) were associated with standard and deviant sounds in a training phase (but not during the actual experiment), an effect could have been expected by ideomotor theory stating that the perceptual idea of an action (i.e., its anticipated sensorial effect) initiates the selection and execution of that action (Greenwald, 1970; Hommel et al., 2001;

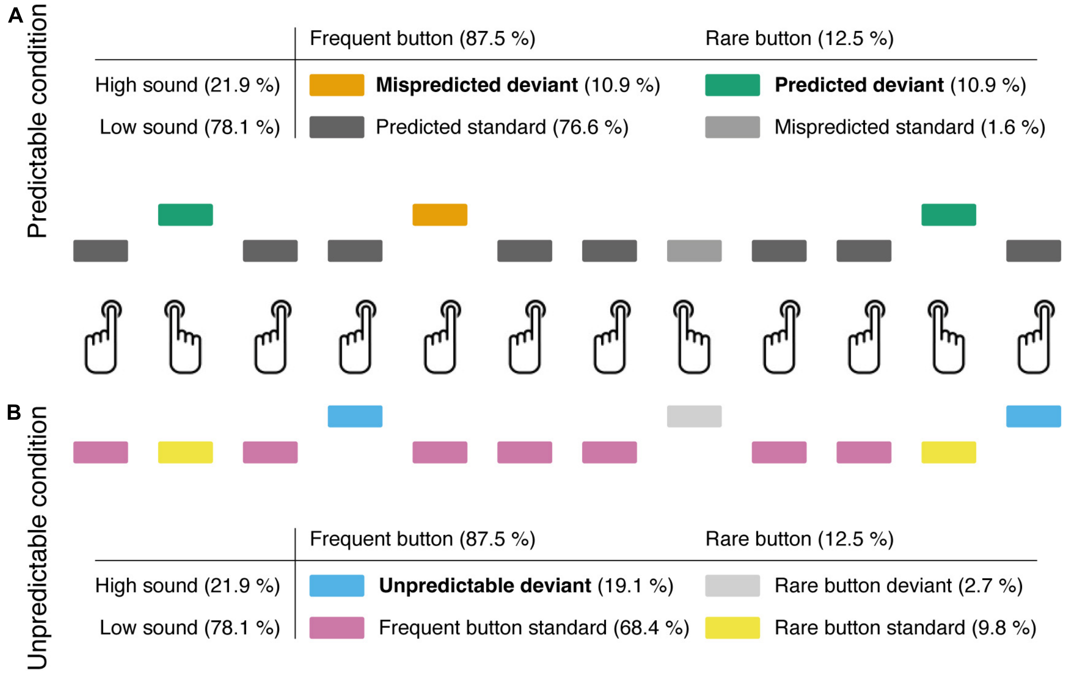


FIGURE 1
Participants pressed one button frequently and the other button rarely. Button presses generated a frequent low pitch ("standard") or a rare high pitch ("deviant") tone. **(A)** In the predictable condition, participants were instructed to generate standard and deviant sounds via the respective buttons. In addition to self-produced "predicted deviants," the frequently pressed button occasionally elicited a "mispredicted deviant" (instead of a standard). **(B)** In the unpredictable condition, the type of button-press (frequent, rare) and the type of sound (standard, deviant) were unrelated, so that "unpredictable deviants" were triggered.

Shin et al., 2010). However, Waszak and Herwig (2007) also did not observe a modulation of MMN depending on whether the sounds were elicited by the button associated with the deviant or the button associated with the standard during the training phase (but a modulation of P3a; see below). In a study by Rinne et al. (2001), self-generated deviants yielded a regular MMN even when they were fully predictable due the deterministic button-sound mapping. This suggests that the intention-based prediction of a deviant has no effect on the auditory-regularity-based MMN. In contrast to the study by Rinne et al. (2001), the present study emphasizes an intention-based action mode and included mispredicted deviants (triggered by the button-press that usually produced a standard), both manipulations presumably promoting the monitoring of action effects. Thus, it appears plausible that the auditory-regularity-based MMN might be attenuated when a deviant is predicted according to intended outcome of an action in this experimental scenario. However, if the present study still yields full-fledged MMN, this would be a strong case for a strictly modular organization of the MMN for the violation of an auditory regularity which cannot be accessed by top-down processing of intentional action.

Darriba et al. (2021) reported an enhancement of the auditory deviance effect in the MMN latency range in response to the violation of a learned sound pattern when the sound additionally violated an intended action effect. This possibly indicates two separate, additive rather than interactive routes of prediction violation. The authors labeled this effect peaking 148 ms after sound onset as an effect on the N1b rather than the MMN. As N1b and MMN share important characteristics in terms of latency and supratemporal generators, and as the MMN has also been proposed to be a subcomponent of the N1 wave (Näätänen and Picton, 1987; May and Tiitinen, 2010), the deviance N1b effect and MMN have possibly not been disentangled here. Anyway, unlike the Rinne et al. (2001) study, the studies by Korka et al. (2019) and by Darriba et al. (2021) show that the violation or confirmation of an intention-based prediction can modulate auditory deviance ERP effects in the MMN time range (and Le Bars et al., 2019 for related N2b).

The MMN is often followed by the P3a, which is regarded as indicating a switch of attention toward the deviant sound (Escera et al., 1998; for review see, Polich, 2007). It is assumed that it not only includes an evaluation of the mere physical difference between deviant and standard, but also an evaluation of the potential significance of the stimulus with respect to the aims of the listener (Winkler and Schröger, 2015). According to Nieuwenhuis et al. (2011), the P3a indicates activity of the locus coeruleus-norepinephrine system elicited by motivationally significant stimuli mobilizing resources for action. An increase of P3a has been reported by Nittono (2006; also see, Nittono and Ullsperger, 2000; Knolle et al., 2019) for self-generated sounds (without a specific button-sound association) compared to externally generated sounds presumably due to unequivocal stimulus timing and voluntary stimulus production enhancing orienting of attention explained with reference to the ideomotor theory (for review see, Hommel et al., 2001). The perceptual representation of the forthcoming stimulus is activated by action intention by means of associative learning. Furthermore, in case of established button-sound relationships, the P3a has been observed even by predicted deviants and enhanced for unpredictable deviants (Waszak and Herwig, 2007; Knolle et al., 2019; Darriba et al., 2021). Darriba et al. (2021) suggested that the P3a results indicate that auditory regularity-based and action intention-based sound predictions coexist simultaneously as independent predictions (i.e., parallel and additive). We expect to replicate this result in our experimental scenario.

Materials and methods

Participants

Data from 17 participants were recorded. The data from two participants had to be excluded from analysis due to technical problems during the recording. The mean age of the remaining 15 participants was 23.5 years (range 19–36 years). 14 of the participants were right-handed, one left-handed. Eight of the participants were female, seven male. For three participants, the two experimental conditions were recorded in two sessions on separate days. All of them reported normal hearing and normal or corrected to normal vision. None of the participants had a history of a neurological disease or injury. Participants received either course credit or payment (18 Euros) for their participation in the experiment and gave their written informed consent after the details of the procedure had been explained to them. The experiment was conducted according to the Declaration of Helsinki and the ethical guidelines of The German Psychological Society ("Deutsche Gesellschaft für Psychologie", DGPs[1]) and complied with all institutional and country-specific legal requirements.

Procedure

The experiment consisted of two conditions, a predictable and an unpredictable condition, each including 12 blocks of 128 trials. In both conditions participants were instructed to produce sounds by button presses and press one button 112 times (87.5%; frequent button) and another button 16 times per block (12.5%; rare button). Each button press was followed by a sound. In a "predictable" condition the type of button-press (frequent and rare) correctly predicted the type of sound (standard and deviant) in most trials, whereas in an "unpredictable" condition, the type of button press and type of sound were unrelated. In the predictable condition, 98 (87.5%) of the 112 frequent button presses were followed by a low sound (predicted standard) and 14 (12.5%) were followed by a high sound (mispredicted deviant). 14 (87.5%) of the 16 rare button presses were followed by a high sound (predicted deviant) and 2 (12.5%) were followed by a low sound (mispredicted standard). Participants were informed that frequent button presses were usually followed by a low sound and rare button presses usually were followed by a high sound and instructed not to care about rare, unexpected sounds. In the unpredictable condition, 87 or 88 (78.1%) of the frequent button presses were followed by a low sound (frequent standard) and 24 or 25 (21.9%) were followed by a high sound (frequent deviant). 12 or 13 (78.1%) of the 16 rare button presses were followed by a low sound (rare standard) and 3 or 4 (21.9%) followed by a high sound (rare deviant). Participants were informed that button presses were followed by either a low sound with higher probability or a high sound with lower probability irrespective whether the frequent or the rare button was pressed. In total, in both conditions, 100 low sounds (78.1%) and 28 high sounds (21.9%) were presented per block. Trials were pseudo-randomized with the constraint that never two mispredicted deviants in the predictable condition and never two frequent deviants in the unpredictable condition directly succeeded each other. We would like to note that sounds were not fully unpredictable in the unpredictable condition as standard sounds were presented with higher probability than deviant sounds. We chose this terminology to emphasize the contrast between conditions with actions (i.e., button presses) predictably associated with action effects (i.e., sound type) in the "predictable" and unpredictably and therefore independent of action selection in the "unpredictable" condition.

Participants were instructed to distribute the infrequent button presses as randomly as possible across the whole block, to press buttons at a regular interval of 800–900 ms, not to press the rare button two times in a row, and to avoid fixed patterns (e.g., pressing the rare button every fifth time). The number of remaining button presses per button per block and the time between the last two button presses were displayed to the participants on a computer screen. If the interval between the last two button presses was shorter than 600 ms or longer than 1,200 ms, or the participant pressed the rare button two times in a row, or pressed buttons in a fixed pattern (if the number of frequent button presses between two rare button presses was identical three times in a row) a visual error message was presented ("Zu schnell" [Too fast], "Zu langsam" [Too slow], "Falsche Taste" [Wrong button], or "Festes Muster" [Fixed pattern]) and the button press was not followed by a sound.

---

[1] https://www.dgps.de/die-dgps/aufgaben-und-ziele/berufsethische-richtlinien/

Each condition was preceded by a detailed explanation including the task and the relation between button presses and sounds and a training block. Blocks were separated by short breaks. The order of conditions and the assignment of frequent and rare button to the participants' left and right hand was counterbalanced across participants.

Stimuli and apparatus

Participants were comfortably seated in a dimly lit, sound attenuated, and electrically shielded booth. They held a response pad with buttons under the index fingers of their left and right hand. Sounds consisted of triangle waves (containing only odd harmonics with an amplitude ratio proportional to the inverse square of the harmonic number) with a frequency of 352 Hz (low sound; F^4) or 440 Hz (high sound; A^4) with a duration of 200 ms including 5 ms rise and 5 ms fall time (raised cosine window). Sounds were presented 400 ms after a button press via headphones (Sennheiser HD 25) at an intensity of 65 dB SPL. An LCD-computer screen was placed about 130 cm in front of the participants' eyes so that visual stimuli appeared slightly below the horizontal line of sight. The visual display consisting of white digits on black background was separated into two rows. In the first row either the interval between the last two button presses in ms or an error message was displayed in case the button was pressed too fast or too slow or a wrong button was pressed. In the second row the number of remaining button presses per button per block and the ratio of remaining rare to frequent button presses in percent was displayed. The numbers of remaining button presses were presented spatially corresponding to the buttons. The visual display was updated immediately after a button press and subtended a visual angle of $2.5° \times 0.75°$ (W × H).

Data recording

The EEG was recorded with Ag-AgCl electrodes from 27 standard positions of the extended 10-20-system (Fp1, Fp2, F7, F3, Fz, F4, F8, FC5, FC1, FC2, FC6, T7, C3, Cz, C4, T8, CP5, CP1, CP2, CP6, P7, P3, Pz, P4, P8, O1, and O2) and from the left and right mastoids (M1 and M2). All electrodes were referenced to the tip of the nose. The vertical electrooculogram (EOG) was recorded between Fp1 and an infraorbitally placed electrode and the horizontal EOG between the outer canthi of the two eyes. Impedances of all electrodes were kept below 10 kΩ. EEG and EOG were filtered online with a bandpass of 0.1–250 Hz and sampled with a digitization rate of 500 Hz (BrainAmp, Brain Products, Gilching, Germany). Time was recorded for each button press.

Data analysis

The EEG data were pre-processed using EEGLAB (Delorme and Makeig, 2004). Data were filtered offline with a 48 Hz low-pass filter (415-point Hamming-windowed sinc FIR filter, transition band width = 4 Hz; Widmann et al., 2015) and a 0.1 Hz high-pass filter (8,251-point Hamming-windowed sinc FIR filter, transition band width = 0.2 Hz). Data were divided into epochs of 0.6 s time-locked to sound onset, including a pre-stimulus baseline of 0.1 s. Only trials where the previous trial consisted of a frequent button press followed by a standard sound were included in the analysis. We excluded all epochs with signals exceeding peak-to-peak amplitudes of 500 μV at any electrode (to remove large non-stereotypical artifacts but to keep stereotypical artifacts as blinks and eye-movements to be later removed using ICA). Channels (except Fp1, Fp2, M1, M2, or EOG channels) were excluded if they had a robust z score of the robust standard deviation greater than 3 (Bigdely-Shamlo et al., 2015; a single channel in two participants). Artifacts were corrected with an independent component analysis (ICA), using the AMICA algorithm (Delorme et al., 2012). For the ICA, the 48 Hz low-pass filtered data were filtered with a 1 Hz high-pass filter (827-point Hamming-windowed sinc FIR filter, transition band width = 2 Hz), and divided into epochs of 0.6 s (removing the same channels and trials as in the previous step) but not baseline-corrected (Groppe et al., 2009). We then applied the obtained de-mixing matrix to the 0.1-48 Hz filtered data (Klug and Gramann, 2021). Artifact ICs were detected with support of the ICLabel plugin (Pion-Tonachini et al., 2019). All eye-movement (horizontal and vertical movements of the corneo-retinal dipoles and pre-saccadic spike potentials; Plöchl et al., 2012) and blink related artifact IC activity was subtracted from the data. On average, 4.5 ICs were removed from the data per participant (*Mdn* = 4; *min* = 4; *max* = 6). Bad channels were interpolated using spherical spline interpolation. Data were baseline corrected using the 0.1 s window before stimulus presentation. Finally, epochs with signals exceeding peak-to-peak amplitudes of 150 μV were excluded. Individual average ERPs were computed per participant for mispredicted (*mean/min/max N* of included trials per participant 136.9/127/144), predicted (141.7/129/165), and unpredictable deviants (247.9/239/253), and frequent (629.1/607/646) and rare button standards (104.4/89/133). As previously reported by Rinne et al. (2001) we also observed slightly different ERPs to standard sounds following a frequent button press and standard sounds following a rare button press in the unpredictable condition. To exclude differences between mispredicted and predicted deviants being based on the different frequency of the related button press, difference waves were calculated subtracting the ERPs to rare button standards from the ERPs to predicted deviants and the ERPs to frequent button standards from the ERPs to mispredicted deviants (as done similarly by Rinne et al., 2001). Grand average waveforms were computed from the individual average ERPs per stimulus type.

Statistical analysis

There is no final consensus on the nomenclature for N1, MMN and P3a in the field. This is because each of these three components presumably consists of several subcomponents, which cannot easily be disentangled from each other, and because the components overlap in time (i.e., N1 with MMN, and MMN with P3a) with each other and also with other components (e.g., P2 and N2). In other words, the identification of ERP components in the measured ERPs is obscured because the measured ERPs are a mixture of latent underlying (sub-) components. Spatial and temporal overlap considerably biases the observed component peaks typically used to identify and label components (Scharf et al., 2022). Moreover, the practice of determining time windows for the respective components based on (peaks in) the observed ERP frequently suffers from the relatively arbitrary definition of time windows and double dipping (Kriegeskorte et al., 2009). Temporal PCA largely reduces these problems (e.g., Dien, 2012; Scharf et al., 2022). For that reason, we used temporal PCA to delineate the components in a straight-forward, data driven approach.

We conducted temporal principal component analysis (PCA) on the individual average ERP data of all channels and stimulus types using the tutorial code provided by Scharf et al. (2022). PCA was computed using Promax rotation ($\kappa = 3$) with a covariance relationship matrix (preferable over correlation matrix for ERP analyses as all sampling points are measured on the same scale; for discussion see, Dien et al., 2005; Scharf et al., 2022) and Kaiser weighting (to ensure that each variable has equal influence on the rotation process and therefore prevent that large factors dominate the results of the rotation step; for discussion see, Dien et al., 2005; Scharf et al., 2022). The number of components to be retained was determined using Horn's parallel test. A total of 10 components was extracted. We focused our analyses on three components of interest, N1, MMN, and P3a.

Mean component scores were computed within frontal (FC5 and FC6; N1 and MMN), mastoidal (M1 and M2; N1 and MMN), and fronto-central (Fz, FC1, FC2, and Cz; P3a) regions of interest (ROI) centered on the observed spatial peaks across components (N1/MMN) and conditions. To obtain difference scores we subtracted component scores for frequent button standards from mispredicted and unpredictable deviants and rare button standards from predicted deviants (note that we only used standards from the unpredictable condition to correct for the confound introduced by different button press frequencies; cf. the last paragraph of the data analysis section above for a more detailed justification). For each component, stimulus type, and ROI, we computed one-sided Bayesian t-tests on the difference component scores (to verify that the components were elicited) and two-sided Bayesian t-tests for difference component scores of mispredicted vs. predicted deviants, mispredicted vs. unpredictable deviants, and predicted vs. unpredictable deviants (minus standards, respectively; to examine whether the components were modulated by condition) in R using the BayesFactor package (Morey and Rouder, 2021). The null hypothesis corresponded to a standardized effect size $\delta = 0$, while the alternative hypothesis was defined as a Cauchy prior distribution centered around 0 with a scaling factor of $r = 0.707$ (the default "medium" effect size prior scaling). Data were interpreted as moderate evidence in favor of the alternative (or null) hypothesis if BF_{10} was larger than 3 (or lower than 0.33), or strong evidence if BF_{10} was larger than 10 (lower than 0.1). BF_{10} between 0.33 and 3 are considered as weak/anecdotal evidence (following Lee and Wagenmakers, 2013). In **Table 1**, we additionally report Cohen's d effect sizes and frequentist t-tests for the tests of difference scores against baseline per component, stimulus type, and ROI.

Results

In the following we will present results on the comparison of deviant vs. standard component scores per condition (the corresponding grand-average ERPs are displayed in **Figure 2**) and the comparison of deviant minus standard difference scores between conditions (the corresponding component loadings, difference scores and grand-average difference waves as well as topographies are displayed in **Figures 3**, **4**, respectively) separately for the N1, MMN, and P3a PCA components.

Component 4/ΔN1

N1 was reflected in PCA component 4 peaking 90 ms after stimulus onset. The data provided strong evidence for enhanced N1 component amplitudes at frontal (more negative) and mastoidal electrode locations (more positive) in response to all deviant types compared to standards (ΔN1; all $BF_{10} > 80$). The data provided moderate evidence against a difference of ΔN1 amplitudes between mispredicted and predicted deviants [frontal ROI: $BF_{10} = 0.28$, $d = -0.09$, $t(14) = -0.33$, $p = 0.744$; mastoidal ROI: $BF_{10} = 0.33$, $d = 0.19$, $t(14) = 0.74$, $p = 0.471$] and moderate evidence against a difference of ΔN1 amplitudes between mispredicted and unpredictable deviants [frontal ROI: $BF_{10} = 0.27$, $d = 0.06$, $t(14) = 0.22$, $p = 0.83$; mastoidal ROI: $BF_{10} = 0.27$, $d = -0.07$, $t(14) = -0.26$, $p = 0.799$] as well as moderate evidence against a difference between predicted and unpredictable deviants at frontal electrode locations and inconclusive evidence at mastoidal electrode locations [frontal ROI: $BF_{10} = 0.32$, $d = 0.18$, $t(14) = 0.69$, $p = 0.499$; mastoidal ROI: $BF_{10} = 0.94$, $d = -0.46$, $t(14) = -1.79$, $p = 0.096$].

TABLE 1 Deviant minus standard difference component scores for the PCA components N1, MMN, and P3a, effect sizes (Cohen's *d*), and results of one-sided Bayesian and frequentist *t*-tests against baseline separately per deviant type and ROI.

Comp.	Deviant	ROI	Diff. score	BF_{10}	d	$t(14)$	p
4/ΔN1	Mispredicted	Frontal	−0.61	84.75	−1.08	−4.20	< 0.001
		Mastoidal	0.26	167.24	1.19	4.60	< 0.001
	Predicted	Frontal	−0.56	974.65	−1.47	−5.70	< 0.001
		Mastoidal	0.22	174.46	1.20	4.63	< 0.001
	Unpredictable	Frontal	−0.63	3.95×10^4	−2.15	−8.31	< 0.001
		Mastoidal	0.27	3.63×10^3	1.70	6.57	< 0.001
3/MMN	Mispredicted	Frontal	−0.38	8.74	−0.73	−2.83	0.007
		Mastoidal	0.22	22.95	0.88	3.42	0.002
	Predicted	Frontal	0.21	0.12	0.36	1.41	0.910
		Mastoidal	−0.03	0.20	−0.10	−0.38	0.644
	Unpredictable	Frontal	−0.36	98.58	−1.11	−4.29	< 0.001
		Mastoidal	0.19	38.35	0.96	3.72	0.001
1/P3a	Mispredicted	Fronto-central	1.18	2.41×10^3	1.63	6.29	< 0.001
	Predicted	Fronto-central	0.63	48.74	1.00	3.87	< 0.001
	Unpredictable	Fronto-central	0.55	31.78	0.93	3.61	0.001

Component 3/mismatch negativity

Mismatch negativity was reflected in PCA component 3 peaking 138 ms after stimulus onset. The data provided moderate to strong evidence for the elicitation of a frontal MMN component inverting polarity over mastoidal electrode locations for mispredicted and unpredictable deviants (all $BF_{10} > 8$) and moderate to strong evidence against the elicitation of a MMN component for predicted deviants (all $BF_{10} < 0.25$). The data provide moderate to strong evidence for a difference of MMN amplitudes between mispredicted and predicted deviants [frontal ROI: $BF_{10} = 3.05$, $d = -0.67$, $t(14) = -2.6$, $p = 0.021$; mastoidal ROI: $BF_{10} = 14.18$, $d = 0.91$, $t(14) = 3.54$, $p = 0.003$] and moderate evidence against a difference of MMN amplitudes between mispredicted and unpredictable deviants [frontal ROI: $BF_{10} = 0.27$, $d = -0.04$, $t(14) = -0.16$, $p = 0.876$; mastoidal ROI: $BF_{10} = 0.31$, $d = 0.16$, $t(14) = 0.63$, $p = 0.537$] as well as strong evidence for a difference between predicted and unpredictable deviants [frontal ROI: $BF_{10} = 10.54$, $d = 0.87$, $t(14) = 3.36$, $p = 0.005$; mastoidal ROI: $BF_{10} = 29.46$, $d = -1.03$, $t(14) = -3.98$, $p = 0.001$].

Component 1/P3a

The P3a was reflected in PCA component 1 peaking 282 ms after stimulus onset. The data provided strong evidence for the elicitation of a fronto-central P3a component for all deviant types (all $BF_{10} > 30$). The data provide anecdotal/weak to moderate evidence for a difference of P3a amplitudes between mispredicted and predicted deviants [fronto-central ROI: $BF_{10} = 2.72$, $d = 0.65$, $t(14) = 2.53$, $p = 0.024$], strong evidence for a difference of P3a amplitudes between mispredicted and unpredictable deviants [fronto-central ROI: $BF_{10} = 18.89$, $d = 0.96$, $t(14) = 3.71$, $p = 0.002$], and moderate evidence against a difference between predicted and unpredictable deviants [fronto-central ROI: $BF_{10} = 0.28$, $d = 0.1$, $t(14) = 0.38$, $p = 0.709$].

Discussion

The present study aimed at determining the effects of action-effect intention on auditory oddball processing. Active listeners produced standard and deviant (oddball) sounds by pressing one of two buttons, one button frequently and the other button rarely. In an unpredictable condition the type of button to be pressed (frequent and rare) was unrelated to the type of sound produced (standard and deviant); standard and deviant sounds were "unpredictable" for the participant. In a predictable condition, the frequent button produced a standard sound and the rare button a deviant sound in most trials. Participants were asked to generate standards by pressing the one button frequently and deviants by pressing the other button rarely. Most deviants were correctly "predicted"; importantly however, occasionally a button press for a standard triggered a ("mispredicted") deviant. It turned out that (1) the deviance-specific N1 enhancements were highly similar between the three different deviant types (unpredictable, correctly predicted, and mispredicted), (2) that MMN was highly similar for mispredicted and unpredictable deviants, but no MMN was elicited for predicted deviants, (3) that predicted and unpredictable deviants elicited similar P3a, whereas the P3a for mispredicted deviants was

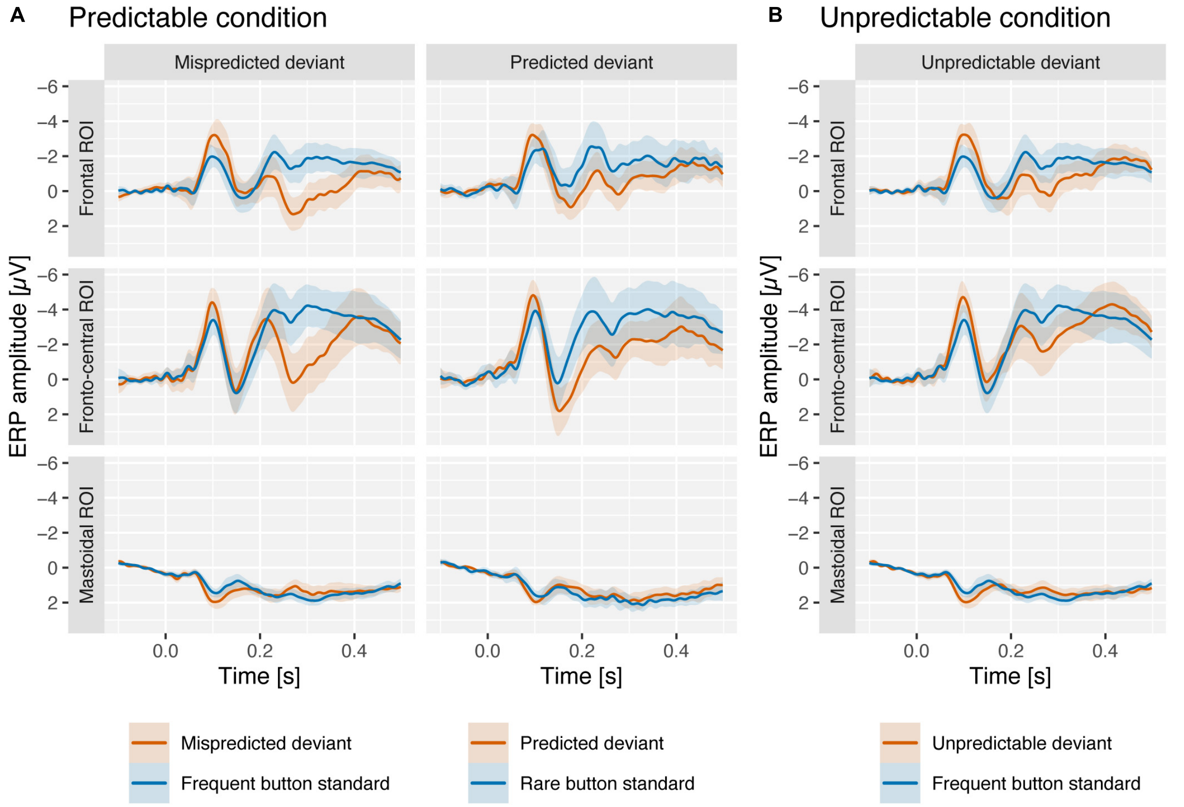


FIGURE 2
Grand-average ERPs at frontal ROI (FC5 and FC6), fronto-central ROI (Fz, FC1, FC2, and Cz), and mastoidal ROI (M1 and M2) from predictable **(A)** and unpredictable conditions **(B)** in response to mispredicted and predicted deviants (predictable condition; red) and unpredictable deviants (unpredictable condition; red). Deviants from both conditions are compared to frequent and rare button standards (blue) from the unpredictable condition (see "Materials and methods" section). Shaded areas reflect 95% confidence intervals. At around 100–150 ms ERPs are more negative for deviants than for standards at frontal and fronto-central regions, and more positive on mastoidal areas. At around 200–400 ms the ERPs for standards were more positive for deviants than for standards at fronto-central regions.

enlarged. Thus, the system underlying the N1 deviance processing was not modulated depending on whether an intended action effect did or did not occur. The MMN-system was modulated if the action intention was confirmed (MMN reduced or abolished for predicted deviants) but not if the action intention was violated. Mispredicted deviants violating both auditory regularity and action intention did not elicit an enhanced MMN compared to unpredictable deviants (violating auditory regularity only). In contrast, the P3a-system was affected if the action intention was violated (P3a enhanced for mispredicted deviants) but not if it was confirmed (P3 was not reduced or abolished for predicted deviants).

No impact of action intention on deviance-specific N1 enhancement

Many studies showed that the auditory N1 is attenuated for self-generated sounds supporting motor-to-sensory forward-modeling accounts of sound processing (for reviews see, Horvath, 2015; Schröger et al., 2015). If the N1 *per se* can be modulated by intentional action, it seems reasonable to assume that also the deviance-specific enhancement of the N1 can be attenuated for intended action effects. Moreover, according to predictive coding theory (Friston et al., 2010; Clark, 2013) such an effect would be expected. On the other hand, according to the adaptation model by May (2021) such an effect is not (necessarily) to be expected as the N1 enhancement to deviants can be explained by bottom-up driven short-term synaptic depression of neurons in auditory cortex, which does not involve top-down processing. Indeed, our study revealed deviance-specific N1 enhancement at around 90 ms which was highly similar for unpredictable deviants, correctly predicted deviants, and mispredicted deviants. That is, the N1 enhancement to violations of an auditory-regularity was not influenced by the intention-based sound predictions.

Complementary evidence for the functional independence of oddball processing from intentional action at the N1-level has been reported by Korka et al. (2019), who found that sounds

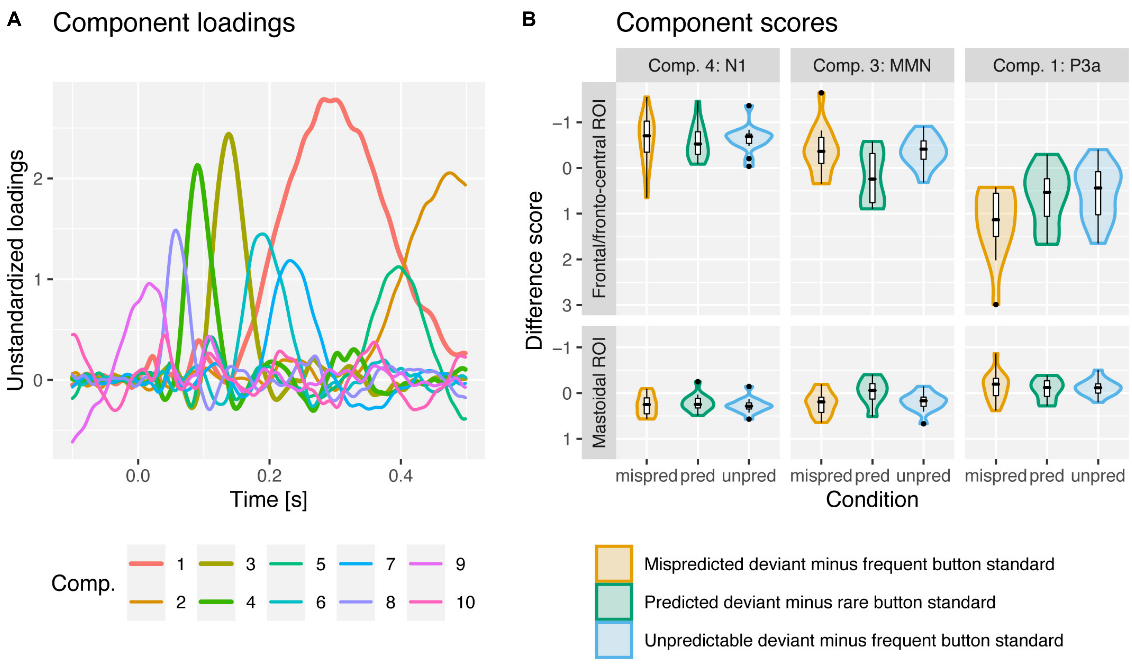


FIGURE 3
PCA component loadings **(A)** and violin and boxplots for deviant minus standard difference component scores **(B)** for mispredicted (orange; minus frequent button standards) and unpredictable deviants (blue; minus frequent button standards) and predicted deviants (green; minus rare button standards) for PCA components N1, MMN and P3a at frontal and mastoidal (N1 and MMN) and fronto-central ROIs (P3a). PCA components 4, 3, and 1 were associated with the N1, MMN and P3a ERP-components. N1 was enhanced (more negative at frontal, more positive at mastoidal electrode sites) for deviants compared to standards similarly in all conditions. MMN was observed for mispredicted and unpredictable deviants but abolished for predicted deviants. P3a was observed in all conditions but enhanced (more positive at fronto-central electrode sites) in response to mispredicted deviant compared to predicted and unpredictable deviants.

that violated the intention-based prediction did not cause an N1 enhancement (but MMN and P3a, see below). Correspondingly, Darriba et al. (2021) did not find an N1 effect in this time-range when an intention-based prediction was violated. Together, these studies suggest (though from different angles) that the N1-system is sensitive to auditory regularity violations, but apparently not to violations of intention-based predictions. If the system underlying N1 generation is not sensitive to violations of intention-based predictions, it seems possible that the N1 enhancement for violations of an auditory regularity is also not a direct expression of prediction error processing and may possibly better be explained more parsimoniously, without referring to prediction violation (May, 2021). It should be noted that adaptation (in the sense of repetition suppression) presumably underlying the auditory oddball N1 effect has been explained in terms of more precise, optimized predictions about sensory inputs (Auksztulewicz and Friston, 2016). In the light of this theory, it is somewhat surprising that the violation of an expected action effect did not matter for the oddball N1 effect but confirms the functional separation of N1 vs. MMN reflecting adaptation-driven vs. genuine prediction-driven deviance processing (Quiroga-Martinez et al., 2020; Schröger and Roeber, 2021).

Strong impact of action intention on mismatch negativity when the action intention is confirmed: Top-down influence on mismatch negativity

The finding that MMN was elicited for unpredictable deviants and for mispredicted deviants but not for predicted deviants shows that MMN is modulated by the top-down influence of the action intention prediction. Even though the deviant violated an auditory regularity, it did not elicit an MMN when the brain was informed by the intention-based prediction that a deviant sound will occur. At a first glance, the present results seem to be at odds with previous research suggesting that MMN cannot be modulated in a top-down manner by preceding visual or by action effect information. Previous studies (Ritter et al., 1999; Sussman et al., 2003) found no top-down modulation of MMN with visual cues informing about forthcoming deviants (though P3a was affected). This is evidence that this kind of visual cuing has no impact on the auditory regularity-based deviance detection system. On the other hand, the present results were to be expected on the basis of recent research showing that the violation of predictive information provided from non-auditory processing modules

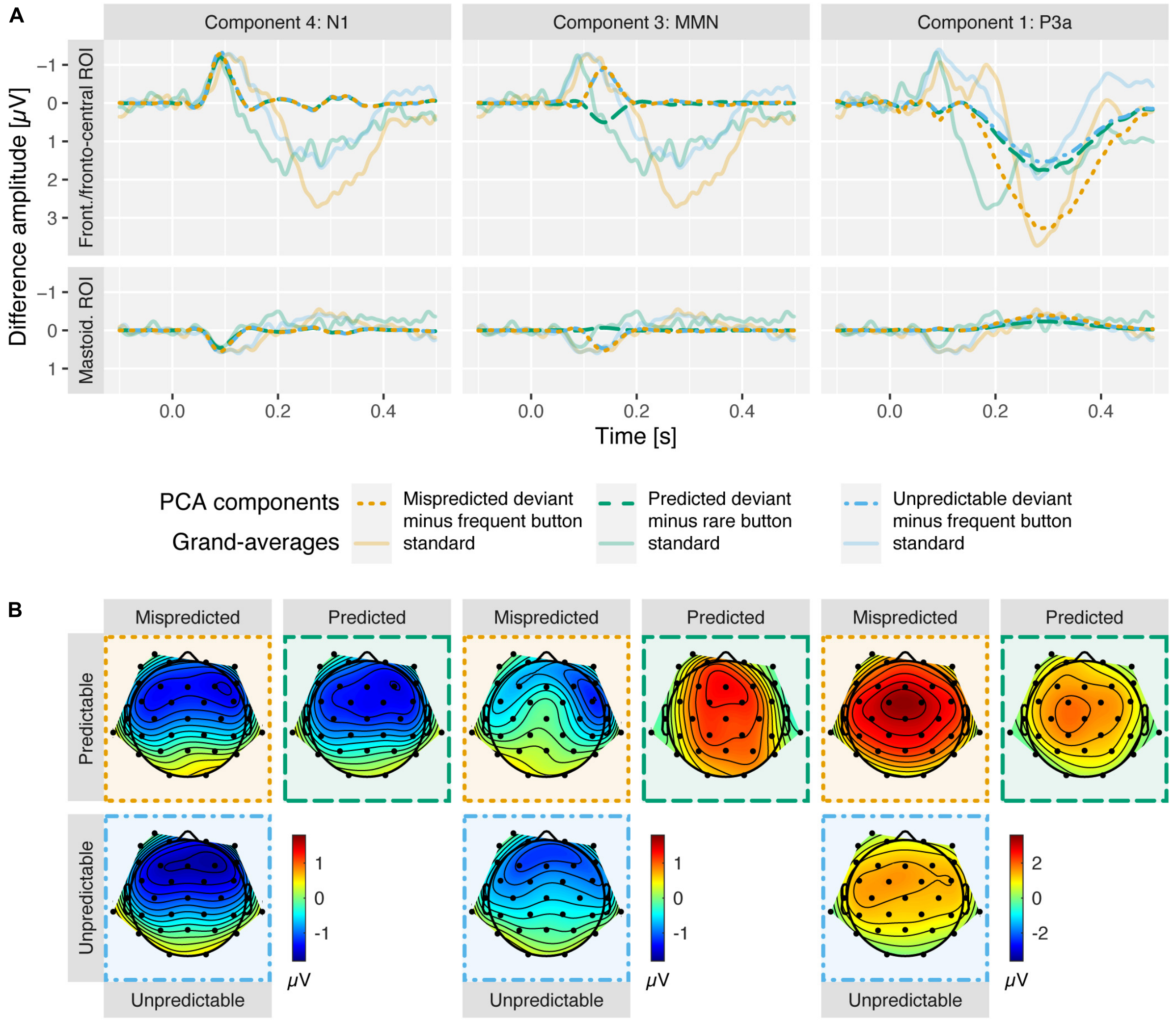


FIGURE 4
(A) Deviant minus standard difference waves, separately for N1, MMN, and P3a PCA components in columns one to three (opaque, dashed) at frontal and mastoidal (N1 and MMN) and fronto-central ROIs (P3a). For each column, the respective grand-average ERP differences waves are shown (transparent, solid) to enable a comparison between the time courses of the components scores and the ERPs. Note that N1 component traces overlap for all deviant types and MMN component traces overlap for mispredicted and unpredictable deviants. **(B)** Deviant minus standard difference topographies for N1 (90 ms), MMN (138 ms), and P3a (282 ms) PCA components at component peak latencies. In both panels mispredicted deviants (orange; minus frequent button standards) and predicted deviants (green; minus rare button standards) from predictable condition and unpredictable deviants (blue; minus frequent button standards) from unpredictable condition are displayed. N1 was enhanced (more negative at frontal, more positive at mastoidal electrode sites) for deviants compared to standards similarly in all conditions. MMN was observed for mispredicted and unpredictable deviants but abolished for predicted deviants. P3a was observed in all conditions but enhanced (more positive at fronto-central electrode sites) in response to mispredicted deviant compared to predicted and unpredictable deviants. Component score differences reveal topographies typical for N1, MMN and P3a.

(vision and action intention) may elicit MMN in the absence of an auditory-regularity: First, sounds violating a prediction which has been induced by visual symbolic information (i.e., music notation) elicit a so-called visuo-auditory incongruency response (IR; e.g., Widmann et al., 2004; Aoyama et al., 2006). The IR shares essential features of MMN, namely, latency and generators in supratemporal areas (Pieszek et al., 2013), so that it might be interpreted as a top-down, non-oddball variant of MMN. Second, MMN can be elicited by the violation of an intention-based prediction for an upcoming sound, when there is no auditory regularity (Korka et al., 2019). If MMN can be elicited in the absence of an auditory regularity via predictive information delivered top-down from non-auditory modules, it seems likely that MMN for the violation of an auditory regularity can also be modulated by top-down predictive information of intentional action. Taken together, the finding that MMN can be elicited by sounds violating a visual-based prediction (Widmann et al., 2004) or an intended action effect only (Korka et al., 2019) and the finding that action intention can abolish the MMN for the violation of an auditory regularity (present study) reveal that intentional action exerts a strong impact on the MMN system: it can turn the MMN system on (Korka et al., 2019) or off (present study). In sum, the present finding is consistent with predictive coding theory, where the action system is attributed a privileged role in changing sensations and overriding sensory predictions (e.g., Friston et al., 2010; Brown et al., 2013).

No impact of action intention on mismatch negativity when the action intention is violated: No mismatch negativity amplitude increase for concurrent violations of regularity and intention

The present experiment was designed to enable the concurrent establishment of two generative models, the one considering the auditory regularity, the other considering the intended action effect. This poses the question what happens if both models either generate contradictory or congruent predictions about the forthcoming sound: In the case of mispredicted deviants the predictions are congruent, in the case of predicted deviants they are contradictory. Mispredicted deviants (violating the auditory regularity and the intention-based prediction) should elicit larger MMN than unpredictable deviants (only violating the auditory regularity). This was not the case. MMN (Paavilainen et al., 2001; Wolff and Schröger, 2001), IR and MMN (Pieszek et al., 2013), and N1b (Darriba et al., 2021) studies yielded enlarged MMN, IR, and N1b, respectively, when two regularity predictions were violated in parallel as compared to when only one regularity prediction was violated. The absence of an MMN increase for regularity plus action intention deviants relative to single, regularity only deviants in the present study points to the special role of action intention as outlined in the predictive coding theory (Brown et al., 2013). At a first glance, the additivity of prediction violation effects on the N1b reported by Darriba et al. (2021) for violations of the auditory regularity (established by the learned sound pattern) and the action intention (referring to the same sound feature) seem to contradict this interpretation. We propose that the difference in the results between the Darriba et al. (2021) and the present study are due to two differences in the experimental designs. (1) In Darriba et al. (2021) action intention was established before sensory regularity. The task cue was presented before the sound pattern. In the present study sensory regularity was established before action intention. (2) In Darriba et al. (2021) the sensory regularity was established independently of action intention; auditory regularity and action intention were manipulated orthogonally. Thus, prioritizing one over the other would not have resulted in better predictions. However, in the predictable condition in the present study, sensory regularity and action intention were correlated and mutually dependent. Selecting the rare button predicted the deviant sound action effect with high probability and therefore presumably gave rise to an adjustment of the regularity-based generative model. Prioritizing action intention overall resulted in better predictions. This interpretation is compatible with results demonstrating stronger expectations due to the intention to produce a specific auditory effect relative to the expectation due to stimulus-driven expectancy which has been reported during music performance (Maidhof et al., 2010).

In the context of auditory scene analysis it has been claimed that several auditory regularity-based predictive representations can coexist (Mill et al., 2013; Schröger et al., 2014; Szabo et al., 2016). This corresponds to the situation of parallel processing of the violation of concurrent regularities underlying MMN and IR-additivity and N1b-additivity. However, according to a computational model of auditory scene analysis these concurrent predictive representations compete with each other when it comes to the next level of processing, which is conscious perception in the context of auditory scene analysis (e.g., Mill et al., 2013). In the light of this model, it seems possible that a competition between the two predictive regularities happened in the present study and that intention-based violation detection processing took over, while the auditory regularity-based violation detection processing was attenuated. In other words, these two processing systems may not be organized in a modular fashion in a situation where the intention-based prediction system is in charge. From a more general view, this perspective is in line with studies showing that context is highly relevant for modulations of early auditory processing (e.g., Dercksen et al., 2021); and, vice-versa, the execution of a simple action (e.g., a right button-press) depends on the specific context, for example, whether the button-press denotes a "yes" or a "no"-answer (Aberbach-Goodman et al., 2022).

In view of the present and previous results we suggest that at the MMN-level (1) several predictions relating to the same or different features of a sound can be maintained and mismatched concurrently (MMN-additivity). If (2) congruent predictions result from different generative models (bottom-up extracted auditory regularity, top-down visual-auditory predictive association) functional independence (evidence accumulation) for prediction violations is achieved (IR/MMN/N1b-additivity). Importantly, (3) intention-based predictions may overwrite the auditory regularity-based prediction depending on context (note that this has been demonstrated also for the case of congruent predictions showing no additivity; Korka et al., 2019). Suggestion (3) is consistent with predictive coding theory according to which the prediction error is weighted by the confidence in the sensory data (Friston, 2005; Brown et al., 2013; Clark, 2013). Confidence (precision) can be modulated by attention (Feldman and Friston, 2010) and by active inference induced by actions (Friston et al., 2010; Brown et al., 2013). Active inference is involved in our task, where participants produced standards and deviants by intentional actions. Considering that "under active inference, perception tries to explain away prediction errors by changing predictions" (Friston et al., 2010, p. 235) the observed primacy of the intention-based prediction over the auditory regularity-based prediction at the MMN-level is to be expected according to the predictive coding theory. Our result is also supportive of Clark's (2015) provocative claim that "motor control is just more top-down sensory prediction".

Impact of action-intention on P3a when action-intention is violated, but not when it is confirmed

All three deviant types elicited a P3a. While unpredictable and predicted deviants elicited P3a of comparable amplitude, the P3a for mispredicted deviants was enlarged. The P3a increase for deviants that violated an auditory-regularity and an action-effect intention replicates previous findings (Nittono, 2006; Waszak and Herwig, 2007; Herwig and Waszak, 2009; Knolle et al., 2013; Darriba et al., 2021). Waszak and Herwig (2007) interpret the P3a increase to deviants when the action intention actually predicted a standard as an increase in the orienting response (Waszak and Herwig, 2007). Consistently, Darriba et al. (2021) argued that the auditory regularity-based and the intention-based predictions were not integrated but remained independent. Our results are compatible with this interpretation.

Interestingly, the P3a elicited by a sound violating an auditory regularity does not differ between predicted and unpredictable deviants. Metaphorically spoken, although the P3a-system does care when the action intention is violated (replicating previous findings, see above), it does not care when it conforms to the action intention (that is, it is enhanced for mispredicted but not reduced for correctly predicted deviants). On the one hand, this is not what one would expect based on the MMN results, characterized by an absence of MMN for predicted deviants. On the other hand, this result is compatible with the idea that the P3a-system evaluates stimuli with respect to their 'significance' by combining the stimulus information with its relevance within a wider context (here, additively integrating violations of both sensory regularity and action intention; Horvath et al., 2008; Wetzel et al., 2013) eventually activating the organism's resources for action (Nieuwenhuis et al., 2011). Thus, our results are compatible with the notion that prediction error increases and adaptation decreases with higher level within the cortical hierarchy obtained from human imaging studies (Schlossmacher et al., 2022) and electrophysiological animal (Parras et al., 2021) studies. However, the present results also reveal that the P3a-system still considers the information about a deviancy from the auditory regularity (which has been assessed already at the N1 level).

Limitations

Amongst the limitations of the present study is that we cannot be sure about the divergence of the MMN results between the Rinne et al. (2001) and the present study, with regular MMN for predicted deviants in the Rinne study but NO MMN for predicted deviants in the present study. We suspect that it is the difference between the instructions in these two studies resulting in quite contradictory effects. While in the Rinne study participants were instructed to press buttons, they were instructed to produce sounds in the present study. In the context of ideomotor theory, it has been argued that actions are only selected with respect to their anticipated sensory effects in a so-called "intention-based action mode" (Herwig and Waszak, 2009). If one assumes that the action performed by the participants were not sufficiently strongly associated to its effect, and–consequently–did not give rise to respective predictions for the forthcoming sounds, a modulation of the MMN is not necessarily to be expected in the study by Rinne et al. (2001). Such striking effects of a (presumably) minor change in instruction has, for example, be shown on the Simon effect, a stimulus-response spatial compatibility effect (Hommel, 1993). In this study by Hommel, it turned out that when participants intended to switch on a (left or right lateralized) light, rather than to press a (left or right) button as response to a lateralized sound, the Simon effect inverts. Though we believe that the difference in instruction is the cause for the striking difference in MMN results, there are two further differences between the studies, which could possibly play a role. In the Rinne et al. (2001) study, the auditory regularity-based and the action intention-based predictions were fully predictable. That is, unlike to the present study, the contingency relations in the Rinne study did not promote the need to monitor the outcome of the actions. Finally, the Rinne study utilized duration deviants, whereas the present study used pitch deviants. Also, this difference could, in principle, be responsible for the difference in MMN results.

Another limitation of the present study is that we cannot fully exclude that participants may occasionally have thought they made a mistake when an unexpected tone occurred. This, in turn, might have resulted in motor error-related ERPs (e.g., ERN). We have intentionally tried to prevent this by the instructing participants not to care about rare, unexpected sounds. Also, when performing this task, the occurrence of a mispredicted tone does not "feel like" that one has committed a motor error, but it rather sounds like an auditory deviant. Also, the topography of the N1 and MMN effects, with polarity reversal at mastoidal leads (**Figure 4**) pointing at generators in supratemporal areas argues against the possibility that we might misinterpret an ERN as an oddball-N1 or MMN. Anyway, we see no way to disentangle the two cases where participants did not think that they made a (motoric) mistake but noticed that an unpredicted sound occurred versus where participants noted the unexpected sound and ascribed it to a motoric mistake of their own. Thus, we decided to avoid speculations on possible (and interesting) relations between the present auditory oddball ERP effects and motor error-related ERPs in the present paper.

Conclusion

In sum, the impact of the violation (or confirmation) of an intention-based prediction on auditory-regularity-based deviant processing is (at least) threefold. (1) The pattern of results for the early-level (N1) processing is compatible with stimulus-driven neural adaptation mechanisms, which can be explained without referring to predictive processing (May, 2021), but which is also compatible with a predictive coding account (Auksztulewicz and Friston, 2016). (2) The pattern of results for the intermediate level (MMN) processing is supportive of generalized predictive coding theory that includes action (Friston et al., 2010; Clark, 2013). Although stimulus-driven and intention-driven effects take place at this level, intention-based predictive processing may be prioritized over the stimulus-driven effects depending on context. (3) Results for the late-level (P3a) processing support the idea that the P3a indicates an overall accumulation process considering the available information for deviants detected at the earlier levels (Winkler and Schröger, 2015).

Data availability statement

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

Ethics statement

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study.

Author contributions

AW and ES equally contributed to conception and design of the study, writing the first draft and manuscript revision, and read and approved the submitted version. AW implemented the study and performed data and statistical analysis. Both authors contributed to the article and approved the submitted version.

Funding

The authors acknowledge support from the German Research Foundation (DFG) and Leipzig University within the program of Open Access Publishing.

Acknowledgments

We are grateful to Betina Korka for valuable discussion and comments and proofreading of the manuscript and to Nicole Koburger and Caroline Max for their help in conducting the experiment.

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

Aberbach-Goodman, S., Buaron, B., Mudrik, L., and Mukamel, R. (2022). Same Action, Different Meaning: Neural Substrates of Action Semantic Meaning. *Cereb. Cortex* bhab483. doi: 10.1093/cercor/bhab483 [Epub ahead of print].

Aoyama, A., Endo, H., Honda, S., and Takeda, T. (2006). Modulation of early auditory processing by visually based sound prediction. *Brain Res.* 1068, 194–204. doi: 10.1016/j.brainres.2005.11.017

Auksztulewicz, R., and Friston, K. (2016). Repetition suppression and its contextual determinants in predictive coding. *Cortex* 80, 125–140. doi: 10.1016/j.cortex.2015.11.024

Bigdely-Shamlo, N., Mullen, T., Kothe, C., Su, K. M., and Robbins, K. A. (2015). The PREP pipeline: Standardized preprocessing for large-scale EEG analysis. *Front. in Neuroinf.* 9:16. doi: 10.3389/fninf.2015.00016

Brown, H., Adams, R. A., Parees, I., Edwards, M., and Friston, K. (2013). Active inference, sensory attenuation and illusions. *Cogn. Proc.* 14, 411–427. doi: 10.1007/s10339-013-0571-3

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavio. Brain Sci.* 36, 181–204. doi: 10.1017/S0140525X12000477

Clark, A. (2015). "Embodied prediction," in *Open MIND*, eds T. Metzinger and J. M. Windt (Frankfurt am Main: MIND Group), 1–21. doi: 10.15502/9783958570115

Darriba, A., Hsu, Y. F., Van Ommen, S., and Waszak, F. (2021). Intention-based and sensory-based predictions. *Sci. Rep.* 11:19899. doi: 10.1038/s41598-021-99445-z

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

Delorme, A., Palmer, J., Onton, J., Oostenveld, R., and Makeig, S. (2012). Independent EEG sources are dipolar. *PLoS One* 7:e30135. doi: 10.1371/journal.pone.0030135

Dercksen, T. T., Stuckenberg, M. V., Schröger, E., Wetzel, N., and Widmann, A. (2021). Cross-modal predictive processing depends on context rather than local contingencies. *Psychophysiology* 58:e13811. doi: 10.1111/psyp.13811

Dien, J. (2012). Applying principal components analysis to event-related potentials: A tutorial. *Dev. Neuropsychol.* 37, 497–517. doi: 10.1080/87565641.2012.697503

Dien, J., Beal, D. J., and Berg, P. (2005). Optimizing principal components analysis of event-related potentials: Matrix type, factor loading weighting, extraction, and rotations. *Clin. Neurophysiol.* 116, 1808–1825. doi: 10.1016/j.clinph.2004.11.025

Escera, C., Alho, K., Winkler, I., and Näätänen, R. (1998). Neural mechanisms of involuntary attention to acoustic novelty and change. *J. Cogn. Neurosci.* 10, 590–604. doi: 10.1162/089892998562997

Escera, C., Leung, S., and Grimm, S. (2014). Deviance detection based on regularity encoding along the auditory hierarchy: Electrophysiological evidence in humans. *Brain Topogr.* 27, 527–538. doi: 10.1007/s10548-013-0328-4

Feldman, H., and Friston, K. J. (2010). Attention, uncertainty, and free-energy. *Front. Hum. Neurosci.* 4:215. doi: 10.3389/fnhum.2010.00215

Friston, K. (2005). A theory of cortical responses. *Philosophical. Trans. R. Soc. B* 360, 815–836. doi: 10.1098/rstb.2005.1622

Friston, K. J., Daunizeau, J., Kilner, J., and Kiebel, S. J. (2010). Action and behavior: A free-energy formulation. *Biol. Cyber.* 102, 227–260. doi: 10.1007/s00422-010-0364-z

Friston, K. J., and Stephan, K. E. (2007). Free-energy and the brain. *Synthese* 159, 417–458. doi: 10.1007/s11229-007-9237-y

Garrido, M. I., Kilner, J. M., Stephan, K. E., and Friston, K. J. (2009). The mismatch negativity: A review of underlying mechanisms. *Clin. Neurophysiol.* 120, 453–463. doi: 10.1016/j.clinph.2008.11.029

Greenwald, A. G. (1970). Sensory feedback mechanisms in performance control: With special reference to the ideo-motor mechanism. *Psychol. Rev.* 77, 73–99. doi: 10.1037/h0028689

Grimm, S., and Schröger, E. (2007). The processing of frequency deviations within sounds: Evidence for the predictive nature of the Mismatch Negativity (MMN) system. *Restorative Neurol. Neurosci.* 25, 241–249.

Groppe, D. M., Makeig, S., and Kutas, M. (2009). Identifying reliable independent components *via* split-half comparisons. *Neuroimage* 45, 1199–1211. doi: 10.1016/j.neuroimage.2008.12.038

Heilbron, M., and Chait, M. (2018). Great Expectations: Is there Evidence for Predictive Coding in Auditory Cortex? *Neuroscience* 389, 54–73. doi: 10.1016/j.neuroscience.2017.07.061

Herwig, A., and Waszak, F. (2009). Intention and attention in ideomotor learning. *Quart. J. Exp. Psychol.* 62, 219–227. doi: 10.1080/17470210802373290

Hommel, B. (1993). Inverting the Simon effect by intention. *Psychol. Res.* 55, 270–279. doi: 10.1007/BF00419687

Hommel, B., Müsseler, J., Aschersleben, G., and Prinz, W. (2001). The theory of event coding (TEC): A framework for perception and action planning. *Behav. Brain Sci.* 24:849. doi: 10.1017/s0140525x01000103

Horvath, J. (2015). Action-related auditory ERP attenuation: Paradigms and hypotheses. *Brain Res.* 1626, 54–65. doi: 10.1016/j.brainres.2015.03.038

Horvath, J., Winkler, I., and Bendixen, A. (2008). Do N1/MMN, P3a, and RON form a strongly coupled chain reflecting the three stages of auditory distraction? *Biol. Psychol.* 79, 139–147. doi: 10.1016/j.biopsycho.2008.04.001

Hughes, G., Desantis, A., and Waszak, F. (2013). Mechanisms of intentional binding and sensory attenuation: The role of temporal prediction, temporal control, identity prediction, and motor prediction. *Psychol. Bull.* 139, 133–151. doi: 10.1037/a0028566

Klug, M., and Gramann, K. (2021). Identifying key factors for improving ICA-based decomposition of EEG data in mobile and stationary experiments. *Eur. J. Neurosci.* 54, 8406–8420. doi: 10.1111/ejn.14992

Knolle, F., Schröger, E., and Kotz, S. A. (2013). Prediction errors in self- and externally-generated deviants. *Biol. Psychol.* 92, 410–416. doi: 10.1016/j.biopsycho.2012.11.017

Knolle, F., Schwartze, M., Schröger, E., and Kotz, S. A. (2019). Auditory Predictions and Prediction Errors in Response to Self-Initiated Vowels. *Front. Neurosci.* 13:1146. doi: 10.3389/fnins.2019.01146

Korka, B., Schröger, E., and Widmann, A. (2019). Action Intention-based and Stimulus Regularity-based Predictions: Same or Different? *J. Cogn. Neurosci.* 31, 1917–1932. doi: 10.1162/jocn_a_01456

Korka, B., Widmann, A., Waszak, F., Darriba, A., and Schröger, E. (2022). The auditory brain in action: Intention determines predictive processing in the auditory system-A review of current paradigms and findings. *Psychonomic Bull. Rev.* 29, 321–342. doi: 10.3758/s13423-021-01992-z

Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S., and Baker, C. I. (2009). Circular analysis in systems neuroscience: The dangers of double dipping. *Nat. Neurosci.* 12, 535–540. doi: 10.1038/nn.2303

Le Bars, S., Darriba, A., and Waszak, F. (2019). Event-related brain potentials to self-triggered tones: Impact of action type and impulsivity traits. *Neuropsychologia* 125, 14–22. doi: 10.1016/j.neuropsychologia.2019.01.012

Lee, M. D., and Wagenmakers, E.-J. (2013). *Bayesian Cognitive Modeling: A Practical Course*. Cambridge: Cambridge University Press.

Maidhof, C., Vavatzanidis, N., Prinz, W., Rieger, M., and Koelsch, S. (2010). Processing expectancy violations during music performance and perception: An ERP study. *J. Cogn. Neurosci.* 22, 2401–2413. doi: 10.1162/jocn.2009.21332

May, P. J., and Tiitinen, H. (2010). Mismatch negativity (MMN), the deviance-elicited auditory deflection, explained. *Psychophysiology* 47, 66–122. doi: 10.1111/j.1469-8986.2009.00856.x

May, P. J. C. (2021). The Adaptation Model Offers a Challenge for the Predictive Coding Account of Mismatch Negativity. *Front. Hum. Neurosci.* 15:721574. doi: 10.3389/fnhum.2021.721574

Mill, R. W., Bohm, T. M., Bendixen, A., Winkler, I., and Denham, S. L. (2013). Modelling the emergence and dynamics of perceptual organisation in auditory streaming. *PLoS Comput. Biol.* 9:e1002925. doi: 10.1371/journal.pcbi.1002925

Morey, R. D., and Rouder, J. N. (2021). *BayesFactor: Computation of Bayes Factors for Common Designs". R package version 0.9.12-4*, 3 Edn.

Näätänen, R. (1990). The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive function. *Behav. Brain Sci.* 13, 201–288. doi: 10.1017/S0140525X00078407

Näätänen, R., and Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: A review and an analysis of the component structure. *Psychophysiology* 24, 375–425. doi: 10.1111/j.1469-8986.1987.tb00311.x

Nieuwenhuis, S., De Geus, E. J., and Aston-Jones, G. (2011). The anatomical and functional relationship between the P3 and autonomic components of the orienting response. *Psychophysiology* 48, 162–175. doi: 10.1111/j.1469-8986.2010.01057.x

Nittono, H. (2006). Voluntary stimulus production enhances deviance processing in the brain. *Int. J. Psychophysiol.* 59, 15–21. doi: 10.1016/j.ijpsycho.2005.06.008

Nittono, H., and Ullsperger, P. (2000). Event-related potentials in a self-paced novelty oddball task. *Neuroreport* 11, 1861–1864. doi: 10.1097/00001756-200006260-00012

Paavilainen, P., Mikkonen, M., Kilpelainen, M., Lehtinen, R., Saarela, M., and Tapola, L. (2003). Evidence for the different additivity of the temporal and frontal generators of mismatch negativity: A human auditory event-related potential study. *Neurosci. Lett.* 349, 79–82. doi: 10.1016/s0304-3940(03)00787-0

Paavilainen, P., Simola, J., Jaramillo, M., Näätänen, R., and Winkler, I. (2001). Preattentive extraction of abstract feature conjunctions from auditory stimulation as reflected by the mismatch negativity (MMN). *Psychophysiology* 38, 359–365. doi: 10.1017/S0048577201000920

Parmentier, F. B. (2014). The cognitive determinants of behavioral distraction by deviant auditory stimuli: A review. *Psychol. Res.* 78, 321–338. doi: 10.1007/s00426-013-0534-4

Parras, G. G., Casado-Roman, L., Schröger, E., and Malmierca, M. S. (2021). The posterior auditory field is the chief generator of prediction error signals in the auditory cortex. *Neuroimage* 242:118446. doi: 10.1016/j.neuroimage.2021.118446

Pieszek, M., Widmann, A., Gruber, T., and Schröger, E. (2013). The human brain maintains contradictory and redundant auditory sensory predictions. *PLoS One* 8:e53634. doi: 10.1371/journal.pone.0053634

Pion-Tonachini, L., Kreutz-Delgado, K., and Makeig, S. (2019). ICLabel: An automated electroencephalographic independent component classifier, dataset, and website. *Neuroimage* 198, 181–197. doi: 10.1016/j.neuroimage.2019.05.026

Plöchl, M., Ossandon, J. P., and König, P. (2012). Combining EEG and eye tracking: Identification, characterization, and correction of eye movement artifacts in electroencephalographic data. *Front. Hum. Neurosci.* 6:278. doi: 10.3389/fnhum.2012.00278

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

Quiroga-Martinez, D. R., Hansen, N. C., Hojlund, A., Pearce, M., Brattico, E., and Vuust, P. (2020). Decomposing neural responses to melodic surprise in musicians and non-musicians: Evidence for a hierarchy of predictions in the auditory system. *Neuroimage* 215:116816. doi: 10.1016/j.neuroimage.2020.116816

Rinne, T., Antila, S., and Winkler, I. (2001). Mismatch negativity is unaffected by top-down predictive information. *Neuroreport* 12, 2209–2213. doi: 10.1097/00001756-200107200-00033

Ritter, W., Sussman, E., Deacon, D., Cowan, N., and Vaughan, H. G. Jr. (1999). Two cognitive systems simultaneously prepared for opposite events. *Psychophysiology* 36, 835–838. doi: 10.1017/S0048577299990248

Scharf, F., Widmann, A., Bonmassar, C., and Wetzel, N. (2022). A tutorial on the use of temporal principal component analysis in developmental ERP research - Opportunities and challenges. *Dev. Cogn. Neurosci.* 54:101072. doi: 10.1016/j.dcn.2022.101072

Schlossmacher, I., Dilly, J., Protmann, I., Hofmann, D., Dellert, T., Roth-Paysen, M. L., et al. (2022). Differential effects of prediction error and adaptation along the auditory cortical hierarchy during deviance processing. *Neuroimage* 259:119445. doi: 10.1016/j.neuroimage.2022.119445

Schröger, E., Bendixen, A., Denham, S. L., Mill, R. W., Bohm, T. M., and Winkler, I. (2014). Predictive regularity representations in violation detection and auditory stream segregation: From conceptual to computational models. *Brain Topogr.* 27, 565–577. doi: 10.1007/s10548-013-0334-6

Schröger, E., Marzecova, A., and SanMiguel, I. (2015). Attention and prediction in human audition: A lesson from cognitive psychophysiology. *Eur. J. Neurosci.* 41, 641–664. doi: 10.1111/ejn.12816

Schröger, E., and Roeber, U. (2021). Encoding of deterministic and stochastic auditory rules in the human brain: The mismatch negativity mechanism does not reflect basic probability. *Hear. Res.* 399:107907. doi: 10.1016/j.heares.2020.107907

Shin, Y. K., Proctor, R. W., and Capaldi, E. J. (2010). A review of contemporary ideomotor theory. *Psychol. Bull.* 136, 943–974. doi: 10.1037/a0020541

Sussman, E., Winkler, I., and Schröger, E. (2003). Top-down control over involuntary attention switching in the auditory modality. *Psychonomic Bull. Rev.* 10, 630–637. doi: 10.3758/bf03196525

Sussman, E. S., Chen, S., Sussman-Fort, J., and Dinces, E. (2014). The five myths of MMN: Redefining how to use MMN in basic and clinical research. *Brain Topogr.* 27, 553–564. doi: 10.1007/s10548-013-0326-6

Szabo, B. T., Denham, S. L., and Winkler, I. (2016). Computational Models of Auditory Scene Analysis: A Review. *Front. Neurosci.* 10:524. doi: 10.3389/fnins.2016.00524

Waszak, F., Cardoso-Leite, P., and Hughes, G. (2012). Action effect anticipation: Neurophysiological basis and functional consequences. *Neurosci. Biobehav. Rev.* 36, 943–959. doi: 10.1016/j.neubiorev.2011.11.004

Waszak, F., and Herwig, A. (2007). Effect anticipation modulates deviance processing in the brain. *Brain Res.* 1183, 74–82. doi: 10.1016/j.brainres.2007.08.082

Wetzel, N., Schröger, E., and Widmann, A. (2013). The dissociation between the P3a event-related potential and behavioral distraction. *Psychophysiology* 50, 920–930. doi: 10.1111/psyp.12072

Widmann, A., Kujala, T., Tervaniemi, M., Kujala, A., and Schröger, E. (2004). From symbols to sounds: Visual symbolic information activates sound representations. *Psychophysiology* 41, 709–715. doi: 10.1111/j.1469-8986.2004.00208.x

Widmann, A., Schröger, E., and Maess, B. (2015). Digital filter design for electrophysiological data–a practical approach. *J. Neurosci. Methods* 250, 34–46. doi: 10.1016/j.jneumeth.2014.08.002

Winkler, I., and Czigler, I. (2012). Evidence from auditory and visual event-related potential (ERP) studies of deviance detection (MMN and vMMN) linking predictive coding theories and perceptual object representations. *Int. J. Psychophysiol.* 83, 132–143. doi: 10.1016/j.ijpsycho.2011.10.001

Winkler, I., Denham, S. L., and Nelken, I. (2009). Modeling the auditory scene: Predictive regularity representations and perceptual objects. *Trends Cogn. Sci.* 13, 532–540. doi: 10.1016/j.tics.2009.09.003

Winkler, I., and Schröger, E. (2015). Auditory perceptual objects as generative models: Setting the stage for communication by sound. *Brain Lang.* 148, 1–22. doi: 10.1016/j.bandl.2015.05.003

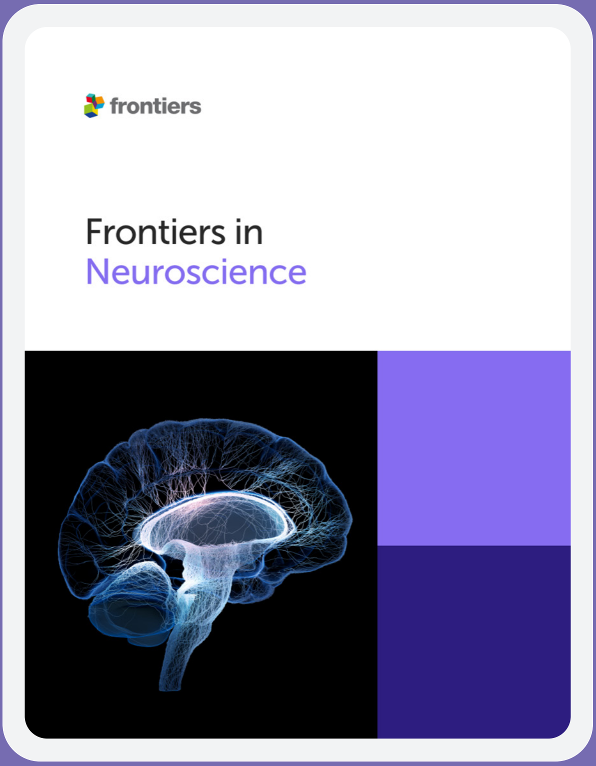
Wolff, C., and Schröger, E. (2001). Human pre-attentive auditory change-detection with single, double, and triple deviations as revealed by mismatch negativity additivity. *Neurosci. Lett.* 311, 37–40. doi: 10.1016/s0304-3940(01)02135-8

Frontiers in Neuroscience

Provides a holistic understanding of brain function from genes to behavior Part of the most cited neuroscience journal series which explores the brain - from the new eras of causation and anatomical neurosciences to neuroeconomics and neuroenergetics.

Discover the latest Research Topics

See more →



frontiers

Frontiers in
Neuroscience

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact