Dissertations, 2020-current

The Graduate School

8-12-2023

The directional effect of target position on spatial selective auditory attention

Heesung Park
James Madison University

Recommended Citation

Park, Heesung, "The directional effect of target position on spatial selective auditory attention" (2023). *Dissertations, 2020-current*. 105.
https://commons.lib.jmu.edu/diss202029/105

The Directional Effect of Target Position on Spatial Selective Auditory Attention

Heesung Park

A dissertation submitted to the Graduate Faculty of
JAMES MADISON UNIVERSITY
In
Partial Fulfillment of the Requirements
for the degree of
Doctor of Philosophy

Communication Sciences and Disorders

August 2023

FACULTY COMMITTEE:
Committee Chair: Christopher G. Clinard, Ph.D.
Committee Members:
Ayasakanta Rout, Ph.D.
Yingjiu Nie, Ph.D.

DEDICATION

This work is lovingly dedicated to my wife and daughter, whose unwavering love and continuous support have been my strength and root throughout this journey. My wife, Mina Choi, and my beautiful daughter, Seeun Park have been my comfort, my joy, and my biggest cheerleaders. Their belief in me has been a constant source of inspiration. I am eternally grateful for their patience, understanding, and love. Their presence in my life has made this achievement possible.

ACKNOWLEDGEMENT

First and foremost, I would like to express my deepest gratitude to my academic advisor and the chair of the dissertation committee, Dr. Christopher Clinard. His intellectual guidance, insightful feedback, and relentless patience have been instrumental in shaping this work. His meticulous mentorship has been invaluable to my academic growth. I am continuously inspired and motivated by his constructive critiques. I am also incredibly grateful to Dr. Ayasakanta Rout and Dr. Yingjiu Nie, members of my dissertation committee, for their research support and continuous encouragement. Their macroscopic viewpoint and scholarly expertise greatly enriched this research. Their dedication, not only to this research but to my personal and professional development, has left a lasting impact. I deeply appreciate their time, effort, and dedication to guiding me toward the completion of this dissertation. I wish to extend my sincere gratitude to Dr. Inyong Choi. His guidance and expertise were invaluable in navigating the complex intricacies of our experimental setup. Furthermore, my sincere appreciation goes out to Raghav, Preeti, Husna, and Nicole. Their assistance in conducting the pilot study and in participant recruitment was integral to the completion of this project.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Spatial selective auditory attention plays a crucial role in listening in a mixture of competing speech sounds. Previous neuroimaging studies have reported alpha band neural activity modulated by auditory attention, along with the alpha lateralization corresponding to attentional focus. A greater cortical representation of the attended speech envelope compared to the ignored speech envelope was also found, a phenomenon known as 'neural speech tracking’. However, little is known about the neural activities when attentional focus is directed on speech sounds from behind the listener, even though understanding speech from behind is a common and essential aspect of daily life. The objectives of this study are to investigate the impact of four distinct target positions (left, right, front, and particularly, behind) on spatial selective auditory attention by concurrently assessing 1) spatial selective speech identification, 2) oscillatory alpha-band power, and 3) neural speech tracking.

Fifteen young adults with normal hearing (NH) were enrolled in this study (M = 21.40, ages 18-29; 10 females). The selective speech identification task indicated that the target position presented at back was the most challenging condition, followed by the front condition, with the lateral condition being the least demanding. The normalized alpha power was modulated by target position and the power was significantly lateralized to either the left or right side, not the front and back. The parieto-occipital alpha power in front-back configuration was significantly lower than the results for left-right listening configuration and the normalized alpha power in the back condition was significantly higher than in the front condition. The speech tracking function of to-be-attended speech envelope was affected by the direction of target stream. The behavioral outcome (selective speech identification) was correlated with parieto-occipital alpha power and neural speech tracking correlation coefficient as neural correlates of auditory attention, but there was no significant correlation between alpha power and neural speech tracking.

The results suggest that in addition to existing mechanism theories, it might be necessary to consider how our brain responds depending on the location of the sound in order to interpret the neural correlates and behavioral consequences in a meaningful way as well as a potential application of neural speech tracking in studies on spatial selective hearing.

1. INTRODUCTION

One of the most challenging circumstances in communication is to spatially and selectively listen to speech of interest when surrounded by multiple people conversing simultaneously. This situation is popularly referred to as the 'cocktail party problem' (Cherry, 1953), a prevalent and crucial conundrum in everyday life. Two prime factors affect speech intelligibility and comprehension - 'Informational Masking', resulting from the semantic similarity between competing verbal sounds, and 'Energetic Masking', resulting from the overlap of frequency energy bands that these speech sounds similarly possess (Brungart, 2001). Additionally, the rapid and constant topic changes and turn-taking in multi-speaker dialogues often cause listeners to lose track of some information from a newly attended stream after switching focus (Shinn-Cunningham & Best, 2008). As indicated by Kidd et al. (2005), even individuals with normal hearing exhibited variable listening capabilities in the presence of multiple speakers, with the accuracy ranging from 67% to above 90%, depending on the level of certainty about the target location.

The cocktail party problem worsens in elderly listeners (Singh et al., 2008), hearing-impaired groups, and pediatric populations due to degraded and immature peripheral auditory representation (Shinn-Cunningham & Best, 2008). Individuals with hearing loss often rely on hearing aids to improve their audibility and speech intelligibility. However, a recent hearing aid usage survey report, ‘MarkeTrak 10’ revealed that nearly one-third of respondents expressed dissatisfaction with their hearing aid’s performance in a large group conversation (Picou, 2020). This finding implies that the latest technologies have not yet succeeded in substantively addressing the cocktail party problem. This issue might stem from inherent limitation: currently, no hearing aids are capable of discerning the wearer’s desired auditory object selection (Geirnaert et al., 2021; Kidd, 2017; Marrone et al., 2008; Schwartz & Shinn-Cunningham, 2013).

Spatially selective auditory attention is crucial in navigating the challenges of cocktail-party listening. Auditory attention is known to modulate the neural representation of the auditory scene, boosting the representation of a target sound while suppressing distractors. This modulation facilitates effective auditory stream segregation and the retention of pertinent information (Shinn-Cunningham & Best, 2008, 2015). The well-formation of auditory objects is vital to selective attention, as these perceptual objects serve as the fundamental units of attention. The efficiency of auditory object formation and streaming is intimately linked to auditory selection. For instance, a slowdown in selective auditory processing due to indistinct object formation may result in a listener overlooking portions of a sound source of interest, especially in dynamic conversations where attention rapidly switches between objects. Conversely, while hazy object formation may disrupt effective auditory selective attention, the resulting ambiguity can draw more auditory attention to enhance stream formation, improve the within-stream event processing, refine scene analysis, and promote task goals (Sussman, 2017).

The selective focus of auditory attention can be driven by conscious and volitional direction to a set of stimulus attributes (top-down attention), and by stimulus salience (stimulus-driven attention) which may be determined by aspects like sound intensity, familiarity, emergency, contextual or social relevance and so on

(Larson & Lee, 2013). Moreover, a listener's prior auditory experiences (experience-driven auditory attention) also play a part in shaping auditory selective attention, aiding in the effective allocation of attention to specific time points or frequencies aligned with the listener's interests (Addleman & Jiang, 2019). These factors are always in competition with each other during auditory object selection. For example, if the stimulus salience is too strong, it can become more difficult to intentionally switch away from it. To enable top-down attention to prevail over the inherent salience of competing sounds and to select a desired target, the listener employs higher-order perceptual features (e.g., spatial location, pitch, timbre, and semantic and lexical structures) that set the target apart from its competitors. The location of the perceived target, more so than fundamental spatial cues such as interaural time differences, greatly influences our ability to connect short-term auditory objects into a comprehensive stream (Darwin & Hukin, 2000; Sach & Bailey, 2004).

Numerous neurological studies employing various methodologies have been conducted in an attempt to decode the mechanisms of spatial selective auditory attention, alongside behavioral and perceptual approaches. Notably, oscillatory alpha-band (8-13Hz) activity, primarily elicited in the parietal and occipital lobes, has been identified as a neural correlate of auditory attention (Banerjee et al., 2011; Deng, Reinhart, et al., 2019; Kerlin et al., 2010; Strauss et al., 2014; Wostmann et al., 2019). A prior study has demonstrated that alpha power is more activated during attentive listening compared to passive listening (Dimitrijevic et al., 2019). However, the relationship between the level of alpha power and listening difficulty appears to be equivocal. While some studies indicate a direct relationship, with increased difficulty leading to increased alpha power (Obleser et al., 2012; Wisniewski et al., 2017;

Wostmann et al., 2015; Wostmann et al., 2017), others suggest an inverse relationship, where increased listening difficulty results in attenuated alpha power (Hjortkjaer et al., 2020; Miles et al., 2017; Seifi Ala et al., 2020). Furthermore, Petersen et al. (2015) showed an inverted U-shape form of alpha power, implying listeners tend to “give up” under increasingly demanding situations. In addition to variation in alpha power magnitude modulated by auditory attention, hemispheric lateralization of alpha power occurs according to the direction of their auditory attention (Bonacci et al., 2020; Bonacci et al., 2019; Deng et al., 2020; Deng, Choi, et al., 2019; Deng, Reinhart, et al., 2019; Ikkai et al., 2016; Klatt et al., 2018; Mehraei et al., 2018). This lateralization involves an increase in alpha power in the hemisphere ipsilateral to the direction of attention and dynamically shifts with the direction of attentional focus (Deng et al., 2020). Given that left and right auditory cortices predominantly receive input from the contralateral ear (Tervaniemi & Hugdahl, 2003), it could be logically inferred that higher alpha power in one hemisphere might serve to suppress processing of the contralateral ear's input, while lower alpha power could enhance such processing (Poch et al., 2017; Wostmann et al., 2021). Lateralized alpha power has largely been interpreted as the result of the distractor suppression mechanism, more so than the amplification of the neural representation of attended signals (Banerjee et al., 2011; Foxe & Snyder, 2011; Jensen & Mazaheri, 2010; Wostmann et al., 2017). However, research by Wostmann et al. (2019) found alpha lateralization traced the target location independent of the distractor, suggesting that both mechanisms independently induce lateralization of alpha power.

Neuroimaging literature also found an increase in cortical representation for an attended speech envelope versus an ignored one, a phenomenon known as 'neural speech tracking' (Aiken & Picton, 2008; Brodbeck et al., 2020; Brodbeck & Simon, 2020; Kurthen et al., 2021; Petersen et al., 2017; Reetzke et al., 2021; Vanthornhout et al., 2019). The speech envelope refers to the slow variations in the amplitude of the incoming speech signal over time, recognized as a crucial auditory cue for understanding speech (Rosen, 1992; Shannon et al., 1995). When humans listen to spoken language, it leads to neural entrainment, a rhythmic brain activity, which consistently follows the attended speech envelope with low-frequency cortical activities (Obleser & Kayser, 2019; Reetzke et al., 2021). This entrainment, synonymous with 'neural speech tracking', can be quantified through a straightforward cross-correlation between the extracted speech envelope and the filtered EEG response (Aiken & Picton, 2008; Holtze et al., 2021; Jaeger et al., 2020; Petersen et al., 2017), or through modeling techniques such as a backward model (reconstructing speech from actual EEG) and a forward model (predicting EEG from the speech envelope) (Brodbeck & Simon, 2020; Crosse et al., 2016; Vanthornhout et al., 2019; Zion Golumbic et al., 2013). Importantly, it has been found that the extent of neural speech tracking correlates with behaviorally measured speech intelligibility in noisy conditions with varying signal-to-noise ratios (SNR)(Ding & Simon, 2012; Lesenfants et al., 2019; Petersen et al., 2017; Vanthornhout et al., 2018), as well as the extent of spectral speech degradation (Kong et al., 2015).

Communication is largely facilitated using a front-azimuthal angle, however, understanding speech presented behind listeners is also essential in our real-life such as vehicle conversations and sports activities. Despite this, research regarding the neural activities associated with attention specifically directed towards the rear(back)-azimuth of an auditory scene is limited, as most prior studies focus on frontal azimuth target sounds. It is plausible that listener performance deteriorates when attending to a target speech from the rear in the presence of a distractor at the front due to an unfavorable signal-to-noise ratio of approximately 5 dB for frequencies between 2 and 8 kHz (Blauert, 1996; Kuk et al., 2013), and front-back confusion (Carette et al., 2014; Wightman & Kistler, 1999). In addition to the energetic masking from the unfavorable SNR, the front-back confusion may exacerbate the informational masking effect as it can create an illusion of irrelevant information appearing in the perceived target location or of the target sound presenting in an irrelevant location. Furthermore, research suggest we have better ability to explore auditory spatial representation in frontal space than back space (Aggius-Vella et al., 2018; 2020). Consequently, it can be speculated that auditory attention processing may function differently in front-back spatial hearing compared to left-right lateral hearing. Thus, the research question under investigation is: How do neural activities change when listeners attend to sounds originating from behind them? To address this, the objective of this study is to examine through an EEG test the directional effect of four different target positions (left, right, front, and especially behind/back) on spatial selective auditory attention. This examination will simultaneously assess 1) behavioral performance in spatially selective hearing, 2) scalp topography of alpha power, and 3) neural speech tracking.

Specifically, the first aim is to compare the hit response rates for target words within speech sentences originating from four different positions (front, back, left, and right). We anticipate that the front-back spatial conditions would yield lower hit rates than the left-right lateral conditions and the hit rate would be lowest in the back condition, which is considered the most demanding situation. The second aim is to identify the scalp-topographic distribution of alpha power when attending to four different target positions. According to the suppression mechanism of to-be-ignored information, we hypothesize that under the front-back spatial hearing condition, alpha power would either be more enhanced or attenuated in both hemispheres than in the lateral hearing condition. This would indicate no alpha lateralization because the target signal and distractor are routed to the listener in a bilaterally symmetrical manner. The third aim is to examine the differences in amplitude of neural speech tracking for the to-be-attended speech envelope across the four different target positions. Our working hypothesis is that neural speech tracking would be more robust for attended target streams in the left and right hemifield, while being weakest in the back position. The final aim is to examine the correlation between the behavioral test outcomes, neural speech tracking, and alpha power. We also expect that the intelligibility of target speech in the presence of a competing distractor would correlate with both neural speech tracking and alpha power. This study explores neural signatures of spatial selective auditory attention, not only focusing on lateral aspects but also on front-back auditory scenes, with the aim of deepening our understanding of auditory attention functions in spatial hearing.

2. REVIEW OF LITERATURE

2.1. Front-back spatial hearing

Compared to other attentional directions, spatially selective hearing can be more challenging when the target sound is projected directly from behind a listener. In addition to disadvantages like the absence of visual cues such as lip-reading and gestures common in real-life listening scenarios, the pinna shadow effect might be a factor. The pinna primarily helps us distinguish monoaural cues we use as a spectral cue for front-back hearing. Specifically, the spectral variations influenced by the pinnae enable highly accurate front/back and up/down localization of sound. Nevertheless, to effectively utilize these cues provided by the pinnae, the sound must possess energy that spans across a diverse range of frequencies within the region of 4 to 12 kHz. In situations where acoustic energy exceeding 4 kHz is missing, one can anticipate localization inaccuracies which are consistent with the principles of a cone of confusion (Perrett & Noble, 1997). However, while making spectral cues, our pinna also physically impedes the transmission of acoustic energy coming from behind by about 5 dB especially for frequencies between 2 kHz and 8 kHz, resulting in adverse/unfavorable SNR (Blauert, 1996; Kuk et al., 2013).

A cone of confusion refers to a set of points from which a sound source is perceived as being in the same direction due to the ambiguity of certain spatial cues. This is because, for sound sources located anywhere along a particular cone-shaped space emanating from the listener's ear (the apex of the cone being at the ear), the interaural level differences (ILDs) and interaural time differences (ITDs) are the same. As a result, without the aid of other cues, these sounds are all perceived as coming from the same direction, creating a 'confusion' about their exact location. For example, an azimuth angle of 60° to the left would imply a 'cone of confusion' along which a sound source could potentially be located at any point - not just directly to the left, but also above, below, in front, or behind - and still produce the same interaural cues. This is one of the reasons why humans have difficulty accurately perceiving the elevation of a sound source, and its forward or rearward direction, especially in the absence of visual or other additional cues.

Interaural cues are crucial for sound localization, which take the form of time differences for low-frequency acoustic energy and level differences for high-frequency energy. At lower frequencies, there is a slight difference in the time it takes for a sound to reach one ear compared to the other due to the distance between our ears. Our brain uses this interaural time difference (ITD) to discern the direction of the sound. Conversely, at higher frequencies, our head acts as a barrier and blocks the level of sound reaching one ear depending on the direction of the sound. This creates a difference in the intensity of sound, interaural level differences (ILD), heard in one ear compared to the other, which our brain uses to determine the direction of the sound (Perrett & Noble, 1997). Because binaural cues such as ITD and ILD are often indistinct in front-back spatial hearing, the spatial ambiguity often occurs in the presence of competing sounds at the opposite site to the target sound (Carette et al., 2014; Wightman & Kistler, 1999). However, Head movements significantly affect our auditory spatial perception, particularly in resolving front-back confusions or ambiguities. When we move our heads, the dynamic changes in the sounds reaching our ears provide additional, valuable information that can help disambiguate these confusions. Essentially, head movements change the position of our ears relative to the sound source, and thus change the way the sound interacts with our head and ears. The head movement results in changes in the ITDs and ILDs over time and space, which our auditory system can interpret to better infer the sound's location. (Brimijoin & Akeroyd, 2012) showed that head movements improve localization accuracy and are crucial in resolving front-back ambiguities, and this was particularly significant in complex listening environments. When listeners were encouraged to move their heads, they made fewer mistakes in telling if sounds were coming from the front or back. In another experiment, listeners weren't allowed to move their heads, but the sound source was moved either by the person conducting the experiment or the listener. Interestingly, only when the listener moved the sound source themselves were they able to correctly tell if the sound was coming from the front or back (Wightman & Kistler, 1999). Furthermore, front-back confusion did not happen with a 0.5-second signal if a slight movement, as minimal as 5 degrees, was made before the signal ended. While the minimal degree of rotation appeared to be adequate to circumvent confusion between front and back (Perrett & Noble, 1997), head movements of a reasonable extent (specifically, spanning approximately 32 degrees of azimuth) may be necessary to ensure that spatial information is delivered with a high degree of accuracy (McAnally & Martin, 2014).

Prior studies also revealed that auditory spatial bisection, a measure of auditory spatial representation exploration, performs better in the frontal space than in the back space (Aggius-Vella et al., 2018; 2020). Auditory spatial bisection refers to a process where a listener attempts to determine the middle point between two sounds in space. This concept is often used in experiments studying spatial hearing and sound localization. The bisection task involves presenting two sounds from different locations (say, from one’s extreme left and extreme right) and then presenting a third sound. The listener's task is to determine whether the third sound is closer to the first or the second sound to bisect the auditory space. Interestingly, a prior study demonstrated that audiovisual training significantly improved sound localization accuracy and decreased front-back confusion in a virtual acoustic environment, using non-individualized HRTF. These enhancements, which extended to untrained locations, were tied to alterations in ERP components, suggesting they are the result of advanced cognitive learning during multiple late processing stages (Cai et al., 2018).

2.2. Auditory scene analysis: an attention perspective

The auditory system possesses an essential skill known as auditory scene analysis, which allows us to perceive and identify the various sound events within our surroundings. In this process, auditory selective attention plays a crucial role. It is noteworthy that auditory attention is not necessary for auditory scene analysis until there's an establishment of an attentional control setting, such as the initial segregation of sounds into different streams. Upon receiving input signals, this initial stream segregation automatically happens based on the bottom-up spectro-temporal attributes of the input, indicating a stimulus-driven process (bottom-up) that happens independently of attentional manipulation. In other words, stream segregation precedes, and within-stream events are subsequently established on the already divided streams (Sussman, 2005, 2017). Auditory attention can then be utilized to concentrate on and refine the process of stream formation and within-stream events to bolster scene analysis and achieve task goals (Sussman, 2017). In essence, attention is object-oriented, but it also influences object formation, implying that both automatic and attention-driven procedures can affect stream formation.

The formation of objects depends on various factors, from basic stimulus attributes to aspects like familiarity and expectations. Sound elements that possess common spectro-temporal regularities tend to be perceived as one stream, a process facilitated by the bottom-up process. However, sounds emanating from a single stream, like speech, often have temporal variations, discontinuities, and transient silences. The concept of ‘streaming’ is applied when grouping temporally disjoint sound elements over time (Shinn-Cunningham & Best, 2008). To decide which sounds should be grouped together amid spectro-temporal discontinuities, listeners may utilize numerous higher-order perceptual features, including perceived location, pitch, intensity, timbre, rhythm, and even the structure of the signal as learned through experience, such as the phonetic, semantic, and lexical structure of speech and language. Nonetheless, a definitive list of sound attributes or statistics that can be used for focusing auditory attention remains elusive (Alain et al., 2001; Sussman, 2017).

In the context of auditory attention, time is a crucial factor for establishing and refining object formation as it continuously interacts with stimulus-driven processes. An extended time for object formation and attention direction may result in the slowing of selective processing, potentially causing listeners to miss out on parts of a sound source of interest. This omission is likely to be particularly significant when attention is required to quickly alternate between objects, as is common in dynamic and engaging conversations. Each attention shift requires a resetting of object formation, which in turn lows down object selection (Shinn-Cunningham & Best, 2008, 2015).

2.3. Obstructive factors to auditory attention

The struggle with auditory object formation and selection can lead to failures of selective auditory attention. Since object selection is based on object formation, the quality of the formation of auditory objects is inevitably tied to selection with attention. In other words, imprecise object formation would result in imprecise object selection. Internal and external factors can influence auditory object formation (perception). First, a compromised or immature peripheral processing in the auditory system can result in ambiguous and indistinct formation, even when given additional processing time (Hopkins & Moore, 2011; Shinn-Cunningham & Best, 2008). Consequently, individuals with impaired hearing may experience a decline in the ability to rapidly direct and switch attention or may require greater effort to achieve performance comparable to those with normal hearing. Moreover, aging is undeniably associated with broad changes in many cognitive processes, including a decline in executive function and an increased difficulty in filtering out unwanted distractions, along with a gradual deterioration in spectro-temporal processing (Decruy et al., 2019; Mesik et al., 2021; Singh et al., 2008).

Additionally, when competing sounds are too numerous, greater energetic masking is likely to occur, causing the target to become blurred and obscured amidst the acoustic mixture. Even when the target and masker are perceptually separated into distinct entities, their acoustic, syntactic, and semantic similarities can disrupt the selection of the correct object within the auditory scene. These similarities may cause the target and masker to be perceived as elements of a larger, unified perceptual object, diminishing sensitivity to the target's content. High-reverberation conditions can cause reverberant energy to distort spatial and pitch cues reaching the ears, leading to disturbances in object formation and selection, and consequently interfering with selective attention (Ruggles & Shinn-Cunningham, 2011). In addition, target sounds occurring at unexpected times or in unpredictable locations are poorly detected. This uncertainty contributes to suboptimal selective attention. For example, a serial auditory search may be triggered when the listener can't identify a feature distinguishing the object from competing sources, and each additional interfering object present in the scene during such a search can further impair performance (Shinn-Cunningham & Best, 2015).

The often rapid and frequent exchanges in multi-speaker conversations can lead to the omission of some information in a newly attended stream, as the process of shifting attention requires about 100-200 ms, and sensory memory degrades over time. Such a situation can reduce performance due to the direct costs associated with attention switching. However, any missing sections from a desired stream can be compensated for by employing expectations formed from the phonetic, semantic, and linguistic content of the gathered fragments, as well as knowledge about the signal's spectro-temporal continuity. This process of automatic filling-in is referred to as "phonemic restoration"(Shinn-Cunningham & Best, 2008). Moreover, the selection of a particular auditory object may fail when a competing object is inherently more prominent (e.g., considerably louder) or unique/special (e.g., your name being called out during a conversation or in an emergency). In such cases, the top-down bias of attention might not be enough to override the bottom-up salience, resulting in a loss in the competition for attention.

On the surface, hearing aids seem to provide an effective solution for hearing-impaired individuals by enhancing audibility through dynamic range compression and subsequently facilitating stimulus-driven processing. However, this approach is not without its drawbacks. Independent binaural compression, when applied separately to each ear, can disrupt interaural level differences (ILDs) and interaural time differences (ITDs). Additionally, the noise reduction and directional beam-forming algorithms of hearing aids often prove ineffective during multi-person conversations, in which disturbances are nonstationary, unpredictable, and similar to the target speech. Above all, existing hearing aids lack the capacity to discern the user's intentions regarding auditory object selection (Geirnaert et al., 2021; Kidd, 2017; Marrone et al., 2008; Schwartz & Shinn-Cunningham, 2013).

2.4. Neural source location of auditory selective attention

The specific cortical regions involved in auditory selective attention are not clearly defined, yet it is known that it spans multiple and broad areas. Attention to a specific location provokes changes in neural activity throughout several brain regions. A substantial body of evidence suggests that a network of frontal and parietal regions is in control of attention, with the prefrontal cortex primarily linked to attentional control and selection. Studies utilizing functional magnetic resonance imaging (fMRI) have found that segments of the prefrontal cortex (PFC) and the posterior parietal cortex (PPC) participate in controlling spatial attention, in both the visual and auditory fields (Deng, Reinhart, et al., 2019; Jerde et al., 2012). Specifically, greater activation was observed in the dorsal prefrontal cortex when attention was paid to voice location, whereas more considerable activation was noted in the ventral prefrontal cortex when the focus was on voice identity (Baars & Gage, 2010).

Besides the prefrontal cortex, a majority of functional imaging studies have identified a marked increase in auditory cortex activity due to selective attention to sound (Fritz et al., 2007). This finding is further bolstered by MEG, fMRI, and PET data demonstrating an increase in early cortical responses in the contralateral temporal lobe when focusing on tones and speech, and this occurs within both primary and secondary auditory areas (Caporello Bluvas & Gentner, 2013). Additionally, selective auditory attention seems to modulate neural activity in subcortical sensory areas like the inferior colliculus and possibly the dorsal cochlear nucleus, presumably through corticofugal projections descending from the auditory cortex. While the existence of selective attention effects in the medial olivocochlear bundle remains a point of contention, recent investigations have confirmed the presence of attention-dependent changes in the amplitude of evoked otoacoustic emissions (Caporello Bluvas & Gentner, 2013; de Boer & Thornton, 2007; Maison et al., 2001). Therefore, it's reasonable to conclude that most components of the central auditory system can be modulated by auditory attention.

Significantly, the brain regions that are activated can vary depending on whether the attention processing is volitional or driven by stimulus features. The 'top-down' attention process appears to originate in the left frontal lobes, while the 'bottom-up' attention process might be activated by the sensory and parietal regions (Baars & Gage, 2010). When attention is directed toward locations, there's an increase in activity in brain areas associated with spatial processing, like the posterior temporal cortex and posterior parietal regions (Deng et al., 2020). In contrast, when attention is focused on the spectral features of speech, there is an increase in activity in regions related to auditory and speech, such as the anterior and superior temporal cortex (Hill & Miller, 2010).

Recent research on spatial attention using high-density electroencephalography revealed alpha band activity (8-12 Hz) from posterior parietal regions when pinpointing an attended speech source amidst two distracters (Kerlin et al., 2010). Similar activation of the superior parietal cortex was observed in fMRI data during attentional switching in visual tasks, suggesting an important role for the parietal cortex in spatial attention across different modalities. Additionally, different cortical networks have been suggested by other studies. The dorsal frontoparietal network, which includes the bilateral intraparietal sulcus (IPS) and frontal eye fields (FEF), is engaged in voluntary (endogenous) switching of auditory spatial attention (Ptak, 2012). Meanwhile, a right-lateralized ventral network encompassing the right-FEF and right temporo-parietal junction (RTPJ) is involved in stimulus-driven (exogenous) switching (Corbetta & Shulman, 2002; Larson & Lee, 2013).

When attention is directed toward speech, as opposed to non-linguistic stimuli, it activates areas like the superior temporal sulcus and gyri, which are recognized for their role in language processing, often more predominantly within the left hemisphere. As such, the modulation of different brain regions by auditory attention may depend on specific stimuli, task requirements, and the structure of the experiment. Conversely, other studies have identified consistent attention-dependent modulation of identical regions, regardless of task parameters, including the superior parietal cortex, dorsolateral frontal cortex, pre-central sulcus, and middle and superior temporal cortex (Caporello Bluvas & Gentner, 2013). Several fMRI studies have indicated that various areas of the cortex are engaged in both visual and auditory tasks. This suggests that the control of attention could be overseen by a network that transcends individual sensory modalities (Larson & Lee, 2013; Shomstein & Yantis, 2006; Wu et al., 2007). In sum, one term, “attention”, may operate at a global network level across multiple regions of the brain to modulate the modality-specific or modality-general processing capabilities.

2.5. Auditory attention measurement

While sensory processes such as visual clarity and hearing sensitivity are well-defined, the concept of attention presents a more intangible variable that has yet to be directly measured (Sussman, 2017). This presents a challenge in determining the degree to which both attended and unattended sounds are processed. For the reliable measurement, it is necessary to consider the impact of various cognitive factors like motivation, arousal, reward anticipation, working memory, and motor controls on spatial hearing and selective listening. Separating the influence of auditory attention from these variables can be tricky because auditory attention doesn't operate in isolation. Nonetheless, task complexity, as indicated by behavioral measures like accuracy percentage or error rates, is frequently used as an indication that attention is necessary for task completion. Additionally, it has been observed that the effects associated with attention escalate with increased difficulty. If participants are unable to accurately account for the distracter stream but can for the target stream, we can infer they are selectively paying attention to the target (Caporello Bluvas & Gentner, 2013). Importantly, several studies have identified significant individual variations in auditory selective attention capacity. This suggests that listeners with audiometrically normal pure tone thresholds could vary in their capacity to control the direction of their attention within auditory dimensions (Holt et al., 2018; Laffere et al., 2020; Ruggles & Shinn-Cunningham, 2011).

2.5a. Behavioral and perceptual approaches

The capabilities of individual auditory selective attention have been extrapolated using behavioral measures. These measures quantify and compare the effects of attention versus inattention on behavioral performance (such as reaction time and sensitivities) in auditory attention tasks, even if they do not directly measure attention processing. To investigate the cognitive mechanisms underlying selective auditory attention, the dichotic listening paradigm has been a long-standing tradition. In dichotic listening, participants are instructed to attend to information presented to one ear while disregarding information presented to the other ear. The task entails selectively listening and immediately repeating the speech presented to one ear while ignoring task-irrelevant speech presented to the other ear. Traditionally, the dichotic-listening paradigm requires maintaining selective attention to the specified source, and the research question revolves around how much of the nominally unattended information is processed. Additionally, a task-cuing paradigm presents a cue before the stimulus onset, indicating the upcoming task (such as target source or target location) and the moment to switch attention (Lawo & Koch, 2014). The task-switching paradigm integrates the dichotic-listening and task-cuing approaches to gauge the cost of switching compared to task repetition (Koch et al., 2011).

In these paradigms, participants are asked to recall the last few words of each sentence (a speech recall task) or answer multiple-choice questions pertaining to the target material (a speech comprehension task) (Lin & Carlile, 2015). Throughout these tasks, reaction time (RT) and accuracy have been used as indirect measures of auditory attention (Koch et al., 2011). Several studies have found that spatial attention influences reaction times and sensitivities in spatial discrimination tasks, but it does not affect reaction times in detection tasks (Buchtel & Butter, 1988; Scharf et al., 1987). Rhodes (1987) pointed out that the auditory detection tasks performed by the previous studies could be carried out via non-spatial representation at initial stages of the auditory system, leading to a performance that remains uninfluenced by auditory spatial attention. However, despite the extensive use of reaction time to investigate the impact of auditory spatial attention, it may not necessarily be an appropriate measure due to the involvement of motor response processing as well (Kanai et al., 2007).

2.5b. Broad-scale neural correlates of auditory attention

Event-related brain potentials (ERPs) are synchronized with specific stimulus events and can be extracted from the ongoing electroencephalography (EEG) record. They provide a direct and measurable representation of brain activity reacting to both attended and unattended information during spatial selective listening. When coupled with behavioral measures, ERPs act as a potent tool in evaluating the impact of human attention on auditory scene perception. Early research on dichotic listening uncovered that selective attention could alter auditory ERPs. Specifically, the N1 negative component (occurring 80-110 ms after sound onset), the P2 component (150-200 ms after onset), and the P3 component (250-450 ms after onset) of the ERP waveform expanded when auditory stimuli were attended to rather than ignored

(Hillyard et al., 1973). The N1 component is believed to partially originate within the primary auditory cortex, whereas the P2 component is considered to initiate in secondary cortices anterior to the primary auditory cortex. The P3 component, on the other hand, is thought to originate in parietal regions and is associated with modality-general decision-making, arousal, or motor response (Caporello Bluvas & Gentner, 2013). Interestingly, a subject's own name has been found to elicit a negative deflection approximately 170–220 ms after stimulus onset, referred to as the subject's own negativity (SON) (Wetzel & Schro¨ger, 2018).

Mismatch negativity (MMN), a distinct ERP component, is exceptionally valuable in assessing processes associated with auditory scene analysis. MMN is induced by the identification of sound anomalies. The repeated occurrence of a sound or sequence of sounds establishes the foundation for the detection of sound deviance. Any sound input breaking this repetitive pattern elicits an MMN. This process of identifying deviance hinges on the standard representation retained in auditory memory. It can be conceptualized as the disparity between the ERPs evoked by the standard sound and those elicited by the deviating sounds. The utilization of MMN as an investigatory tool for sound trace offers numerous advantages. With the use of MMN, we could delve into how disregarded sounds are processed, represented, and arranged in auditory memory, and compare this understanding with the impact of attention given to identical sounds (Sussman, 2017).

2.5c. Neural speech tracking

Speech envelope refers to the slow fluctuations in the amplitude of the incoming speech signal over time, a factor known to be a critical acoustic cue for speech comprehension (Rosen, 1992; Shannon et al., 1995). When people engage in listening to spoken words, it triggers neural entrainment—or rhythmic brain activity—with the low-frequency cortical activity consistently mirroring the speech envelope (Obleser & Kayser, 2019; Reetzke et al., 2021). Indeed, the low-frequency neural activity in the delta (2-3 Hz), theta (4-7 Hz), alpha band (8-12 Hz), and beta band (13-20 Hz) appears to have a strong correlation with the slow (2-20 Hz) acoustic envelope of speech (Aiken & Picton, 2008). This neural entrainment is alternatively termed 'neural speech tracking', which can be garnered from a simple cross-correlation between the speech envelope and the filtered EEG response, or through modeling techniques such as a backward model (reconstructing speech from actual EEG) and a forward model (predicting EEG from speech envelope) (Brodbeck & Simon, 2020; Crosse et al., 2016; Vanthornhout et al., 2019; Zion Golumbic et al., 2013). To be specific, forward modeling seeks to predict EEG data from a given set of stimulus features by adjusting a specific set of weights. These weights, known as the Temporal Response Function (TRF), have biological interpretability, comparable to a standard event-related potential (ERP). On the other hand, backward modeling adjusts a set of weights, identified as a decoder, in the opposite direction with the goal of reconstructing a set of stimulus features using EEG data. Although these coefficients provide important information, their neurophysiological interpretability is not equivalent to that of a TRF. (Crosse et al., 2021).

Past studies have uncovered that when the speech stream is being attended to, low-frequency delta and theta bands exhibit greater speech-EEG coherence than when it is overlooked. Additionally, a strong association has also been identified between the degree of neural speech tracking and behaviorally gauged speech intelligibility, such as significant speech reception threshold (SRT) predictions derived from a forward model (Lesenfants et al., 2019; Vanthornhout et al., 2018), and the stimulus signal-to-noise ratio (SNR) (Petersen et al., 2017; Vanthornhout et al., 2019). Cortical tracking in the theta frequency band is primarily correlated with clarity, while the delta band contributes to speech comprehension (Etard & Reichenbach, 2019). An increased number of background talkers has been shown to decrease performance and attenuate EEG tracking of target speech (Hambrook & Tata, 2019).

2.5d. Oscillatory alpha-band power

Oscillatory alpha band power refers to the amplitude or strength of alpha band waves in the brain's electrical activity, which typically oscillate between 8 to 14 Hz. Alpha waves are associated with a state of relaxation, and they tend to be dominant when a person is calm, relaxed, and alert but not actively processing information. In the context of neuroscience, oscillatory alpha band power is thought to play a crucial role in a range of cognitive processes, particularly those involving attention and perception. Numerous studies have indicated that alpha oscillations might play a significant role in inhibitory control and the regulation of sensory information processing, thus effectively controlling the allocation of attentional resources (Klimesch, 2012).

The power of induced alpha-oscillations is modulated by the degree of auditory selective attention to a target sound (Wostmann et al., 2016). Nevertheless, the association between the peak of alpha power and listening difficulty appears to be ambiguous. For instance, Dimitrijevic et al. (2017) reported a negative correlation between alpha power and digit-in-noise (DIN) performance. Furthermore, alpha power decreased with the increase in acoustic details (temporal fine structure) and predictiveness (Wostmann et al., 2015). On the contrary, Miles et al. (2017) observed a reduction in alpha power in demanding listening circumstances, where they manipulated the spectral content of the signal using noise vocoding and assessed speech intelligibility. Also, Seifi Ala et al. (2020) showed a decrease in EEG alpha power in the parietal lobe under conditions with a lower signal-to-noise ratio (SNR).

The scalp-topography of parietal alpha activity changes with the location of attention. Specifically, parietal alpha power lateralization signifies a topographically biased power toward the direction of attention. When focusing on each of five different target locations (approximately corresponding to angles of -60°, -25°, 0°, 25°, and 60°) amid distracting noise, parietal alpha power increased over the hemisphere ipsilateral to the corresponding attentional focus relative to the contralateral hemisphere. Furthermore, this changed systematically as the direction of attention shifted from far left to far right (Deng et al., 2020). Lateralized alpha power has largely been interpreted as the result of the distractor suppression mechanism, more so than the amplification of the neural representation of attended signals (Banerjee et al., 2011; Foxe & Snyder, 2011; Jensen & Mazaheri, 2010; Wostmann et al., 2017). This mechanism of selective attention, often referred to as "gating by inhibition," implies that an increase in alpha power in a specific brain region signifies the inhibition of that area, diverting attention away from potentially distracting information (Jensen & Mazaheri, 2010). On the other hand, a decrease in alpha power, referred to as event-related desynchronization (ERD), indicates the active processing or engagement of that particular region. In the realm of auditory attention, oscillatory alpha band power is highly pertinent. Research has found that top-down modulation of alpha power can help regulate the processing of auditory stimuli (Foxe & Snyder, 2011). For instance, when a listener is trying to attend to a sound source in a noisy environment, alpha power might increase in auditory areas processing distracting or irrelevant sounds, effectively "turning down the volume" of these distractions. Conversely, areas processing the attended sound source might exhibit decreased alpha power, representing enhanced processing of the relevant auditory stimulus (Klimesch et al., 2007).

However, research by Wostmann et al. (2019) found that alpha lateralization traced the target location independent of the distractor, suggesting that both mechanisms independently induce lateralization of alpha power. The right hemisphere typically exhibits more alpha power activity during spatial attention. Notably, auditory attention to nonspatial features such as pitch and timbre does not produce this lateralization and signature of attentional focus in the parietal cortex (Bonacci et al., 2020). The alpha lateralization is quantified by an attentional modulation index of alpha power, defined as the difference in alpha power between the attended left and attended right trials divided by the overall alpha power (Wostmann et al., 2016).

2.6. Influential variables to neural correlates of auditory attention

In hearing-impaired (HI) listeners, alpha power did not exhibit lateralization across parietal sensors, indicating a reduced capability to employ spatial features for selective attention (Bonacci et al., 2019). Attentional modulation of the N1 component was less pronounced in HI listeners than in those with normal hearing (NH) (Dai et al., 2018). However, on the contrary, the neural tracking of attended speech versus ignored speech was robust in both HI and NH listeners. This suggests that adults with hearing impairment may exhibit an enhanced sensitivity to envelope modulations or require a greater differential in neural tracking of target versus competing talker to segregate speech from noise (Decruy, Vanthornhout, et al., 2020; Fuglsang et al., 2020). Perceptual discontinuity of the talker by switching disrupted attentional modulation of cortical responses, reducing Alpha oscillations' power. (Mehraei et al., 2018). Furthermore, concerning the aging effect, older adults typically display a more potent neural speech tracking response than younger adults on average (Decruy et al., 2019; Presacco et al., 2016).

Neural tracking of attended speech deteriorates with an increase in background noise (from 4 to -4 dB signal-to-noise ratio) (Petersen et al., 2017). Considering intrusion errors—responses where listeners incorrectly report numbers from a distractor stream—it is evident that informational masking significantly influences the outcome of speech envelope tracking. (Hambrook & Tata, 2019; Kurthen et al., 2021). Moreover, the listener's familiarity with the language of the stimulus affects the quality of speech tracking amidst noise; cortical tracking of English (the listener's native language) was significantly higher than that of Dutch (a non-native language) in the delta band in the presence of background noise (Etard & Reichenbach, 2019). On the other hand, in a prior comparative study between native and non-native speakers, the non-native group demonstrated increased neural tracking of the speech envelope. This finding implies that those with limited command of the language being listened to often depend more on acoustic cues. Moreover, several studies have discovered substantial individual variances in auditory selective attention capacity. These findings indicate that even listeners with audiometrically normal pure tone thresholds demonstrate differences in their capacity to regulate their focus within auditory dimensions (Holt et al., 2018; Laffere et al., 2020; Ruggles & Shinn-Cunningham, 2011).

2.7. The training effect of auditory attention

Laffere et al. (2020) focused on individual differences in auditory selective attention, demonstrating variability in individuals' ability to control their attentional direction. These differences have been noted across studies, yet their understanding remains limited. A part of this variability was found to be explained by the degree of musical training, suggesting an 'experience-driven auditory attention' link between long-term auditory experience and auditory selective attention. This aligns with previous studies comparing musicians and non-musicians, where musicians typically performed better on selective attention tasks. Furthermore, the study showed that short-term training has the potential to enhance attention specifically directed toward spatial information and the understanding of speech amid background noise. Consequently, long-term experience may shape attentional control and top-down adjustments of the mean neural phase angle, both of which are likely capable of rapid short-term enhancements.

3. METHODS

3.1. Participants

Fifteen young adults with normal hearing (NH) were enrolled in this study (*M* = 21.40 yrs, *SD* = 2.97 yrs, 18-29 ages; 10 Females). A power analysis conducted using the G-power software program (version 3.1.9.7) determined this sample size, assuming a medium effect size (Cohen's f statistic = 0.40) for repeated measures ANOVA, an alpha level of 0.05, and a power of 0.95. All participants were right-handed native speakers of American English, with one exception of a left-handed female. Audiometric assessments, including pure-tone and speech audiometry, revealed that all participants had pure-tone thresholds equal to or better than 25 dB HL at octave frequencies between 250 and 8000 Hz. Moreover, word recognition scores from a monosyllabic word test using the Northwestern University Auditory Test No.6 (NU-6) word list exceeded 95% at 40 dBSL relative to the individual Pure Tone Average (PTA). The subjects reported no history of audiological or neurological pathologies, cognitive complaints, or current prescription medication for seizures, attention, memory, or mood disorders. The tests were conducted in a double-walled, sound-attenuated booth, under protocols reviewed and approved by the IRB of James Madison University, protocol#22-3348. All participants provided written informed consent prior to participation and were compensated at a rate of $12 per hour. The entire test battery took approximately three hours to complete.

3.2. Materials and Stimulus

The behavioral auditory listening task involved the presentation of speech sentences spoken by two different male speakers from the Coordinate Response

Measure (CRM) corpus devised by Bolia et al. (2000). The rationale for using talkers of the same gender was to minimize the chance that participants could rely on nonspatial features such as variations in fundamental frequency (F0) to track the target stream, rather than relying on spatial cues while facilitating informational masking (Allen et al., 2008; Bonacci et al., 2020). The CRM corpus is widely recognized and used as a valuable tool for studying speech comprehension amidst competing background sounds, proving particularly beneficial in research about informational masking. The CRM corpus is beneficial because it has a limited range of language use, low predictability due to context, and a small vocabulary size. Its closed-set format makes it easy to use and score. Moreover, if the same sentences come up again in the test, previous memory of those words doesn't significantly affect the results. Among the two speakers, one of these was a target speech, and the other acted as a distractor (competing speech) (Eddins & Liu, 2012). The CRM corpus comprises predictably structured sentences: “Ready [8 call signs, e.g., Charlie] go to [4 colors] [8 numbers] now” (Bolia et al., 2000; Eddins & Liu, 2012). Each speaker has 256 sentences from the possible combination. Both the color and number in the to-be-attended target stream were the target words that participants were asked to listen to and identify.

These stimuli had an approximate duration of 1.72 seconds (Min = 1.58, Max = 1.86, SD = 0.05 sec) on average. Due to the semantic and syntactic similarity between target and competing speech, CRM can induce substantial informational masking (Eddins & Liu, 2012). These materials were already used in a prior study to examine the effect of distractor set size on neural tracking of attended speech (Hambrook & Tata, 2019). Previous studies utilized either transient stimuli such as tones (Choi et al., 2013) and syllables (Deng, Choi, et al., 2019) or long narrative stories to elicit the ERPs or the neural entrainment to the attended speech (Fuglsang et al., 2020; Holtze et al., 2021; Jaeger et al., 2020; 2021; Reetzke et al., 2021). However, the utilization of sentence materials in this study allows for a more comprehensive speech intelligibility test.

3.3. Experimental Design and Procedures

The experiment setup comprised four distinct target positions (front, back, left, and right), paired into two listening configurations: front-back and left-right. In the sound-treated booth, four loudspeakers (FXi5, Polkaudio, CA) were positioned 1.4 meters from listener's center, with azimuths of ± 90º for the lateral condition and 0º and 180º for the front-back condition (See Fig. 1-A). Participants comfortably occupied a chair in the center that did not obstruct sound transmission from the rear side. The loudspeakers were set at a height aligned with the individual's pinna. During the listening task, the distractor was consistently located opposite the target position (180 degrees away). Both speech sounds were presented at the same level of 60 dBA, leading to a 0 dB target-to-masker broadband energy ratio (TMR). The level of TMR was selected because the difference in intensity between the target and the masker might provide listeners with a clue for distinguishing the target other than a spatial cue (Ihlefeld & Shinn-Cunningham, 2008). Call signs, colors, and numbers were kept distinct in both to-be-attended and to-be-ignored speech streams, and these elements randomly varied across trials. The target position also pseudo-randomly shifted across trials within the left-right or front-back configuration and it was counter-balanced. EEG data were concurrently collected during the auditory attention listening task. Both EEG and behavioral tests were managed and controlled through a custom GUIbased MATLAB (Mathworks, Natick, MA) system, facilitating sound generation, triggering, feedback response, data recording, and more. A monitor, mounted on the frontal loudspeaker, facilitated interaction with participants.

Table 1 illustrates the experimental trial timeline. First, a fixation cross mark appears at the center of the screen to instruct listeners to fix their gaze, minimizing potential eye movement artifacts. Prior to stimulus onset, the upcoming target position was indicated by the target talker's spoken word 'ready,' which lasted for a total of 600 ms (260 ms, plus zero padding of 300 ms) from one of four spatial locations, chosen pseudo-randomly for each trial. A pause (preparatory) period of 1 second took place between the auditory cue offset and stimulus onset. Following the stimulus presentation period with EEG triggering inserted at the onset, a response screen appeared, prompting the listener for their response. Participants indicated the given target words by clicking with a computer mouse. The experiment proceeded to the next trial after a response time of 3.5 seconds, irrespective of whether an individual response had been made. Instant feedback was provided on each trial, with a green light indicating 'correct' and a red light for 'incorrect changing from a gray light,' as shown in Fig. 1-B. A single trial spanned 9 seconds, therefore, a session lasted 15 minutes, containing 100 trials. Every configuration condition included two sessions of 100 trials each, making a total of 400 trials for the entire experiment (2 configurations x 2 sessions x 100 trials = total 400 trials). Before the start of each configuration, a practice session was conducted. To mitigate the test order effect, the order of each configuration was alternated for each participant.

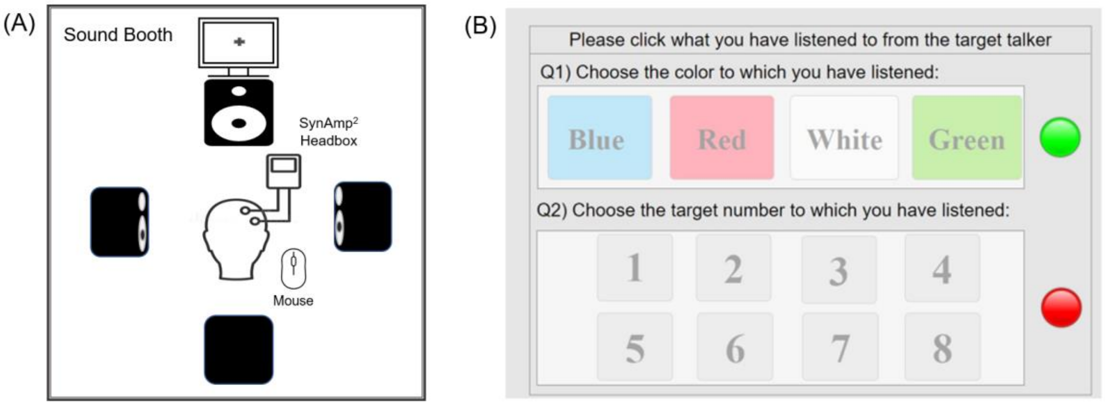

Fig. 1. (A) Schematic illustration of the experimental setup for the front-back and left-right spatial hearing configurations. (B) MATLAB-based Graphical User Interface for the behavioral test. The colored circles serve as feedback indicators (green signifies 'correct', while red denotes 'incorrect').

Table 1. Experimental trial timeline.

	Step #1	Step #2	Step #3	Step #4	Step #5	Total
Sequence	Eye Fixation (+)	Cue “Ready” (+)	Pause (+)	Stimulus & Trigger	Response & Feedback	#1+2+3+4+5
Duration	2.0 sec	0.6 sec	1.0 sec	1.9 sec	3.5 sec	9 sec

3.4. EEG measurement

EEG data were recorded using a Neuroscan system (Compumedics Neuroscan, Victoria, Australia) with 64 scalp electrodes at a sampling rate of 1000 Hz while the subjects were performing the spatial selective listening task. The light in the test room was dimmed slightly. Impedances were maintained below 5 KΩ and online filtering (0.1 – 100 Hz) was applied. Scalp electrode positions were arranged following the international 10–20 system, with the reference electrode placed between the Cz and CPz electrodes. Event triggers generated by MATLAB interfaced with a USB parallel

24-bit trigger I/O interface, SD-MS-TCPUA (Cortech Solutions, Wilmington, NC) were systematically sent to the recording computer. The stimulus and trigger timing were controlled by the custom MATLAB coding. During the task, subjects were instructed to keep their eyes naturally open and to fixate on the cross mark on the test monitor.

The EEG data were processed and analyzed using the EEGLAB toolbox in MATLAB (Delorme & Makeig, 2004). The raw EEG data were first filtered between 1 to 41 Hz cutoff frequency using a Hamming windowed sinc FIR band-pass filter, and unwanted electrode channels such as electromyography (EMG), electrocardiogram (EKG), mastoid (M1 & M2), two cerebellar electrodes (CB1 & CB2), and electrooculogram (EOG1 & EOG2) were removed. Channels of poor quality were visually inspected for the exclusion before the data were down-sampled to 256 Hz. Artifact rejection was carried out using independent component analysis (ICA) with the logistic infomax algorithm to remove components related to eye movements, blinks, cardiac signals, and muscle artifacts. On average, 15.87 components (SD = 4.18, range = 5-23) were excluded based on their topographical and spectral representations. The EEG data were then re-referenced from Cz to an average reference of all electrodes and epoched from -3.6 to 5 sec relative to the onset of the target speech sound for further analysis.

3.5. Data Analysis

3.5a. Behavioral spatial selective auditory attention task

Individual performance in the behavioral spatial selective listening task was gauged by ‘hit rates (%)’ that refer to the percentage of correctly identified target words of color and number, respectively, out of 100 trials. Considering the possible numbers of each target word in the test, the chance level for color is 25% and for number is 12.5%. This test encountered three types of errors. When listeners reported color or number from an interfering stream, these responses were classified as ‘intrusion errors’. On the other hand, when listeners reported words not present in any of the streams, these responses were identified as ‘insertion errors’. 'Omission errors' were noted when the subject did not report within the given response time. Each error type's underlying causes differ markedly: intrusion errors arise from the intrusion of distractor stream information on target perception, whereas insertion errors typically transpire when the listener is bereft of information and resorts to guessing from the limited pool of potential number words (Hambrook & Tata, 2019). This study mainly analyzed the hit rate, assuming a lower hit rate indicates a more demanding and effortful condition.

3.5b. Oscillatory Alpha-band Power

The induced alpha power during the stimulus period (from the onset, 0 to 1.8 sec) was computed in the decibel (dB) scale using the EEGLAB toolbox. This was accomplished through Welch's Power Spectral Density estimate method (Seifi Ala et al., 2020; Welch, 1967), which offers a measure of how the energy of a signal is distributed in the frequency domain. Welch's method breaks the signal in the time domain into smaller segments, computing the Fast Fourier Transform (FFT) for each. The final spectral estimate is then obtained by averaging these individual spectra. The advantage of Welch's method lies in its ability to reduce noise in the power spectrum estimation by averaging the individual periodograms, albeit at the cost of a reduced frequency resolution. Here, a fixed, predefined frequency range (8–13 Hz) of alpha activity was considered, rather than the individual alpha peak frequency (IAF) (Bonacci et al., 2020; Deng et al., 2020). This approach was chosen as alpha power can exhibit multiple peak amplitudes at different frequencies between participants (Seifi Ala et al., 2020).

The absolute alpha power for each trial was computed and subsequently averaged across trials to yield the individual mean alpha power for each electrode across target positions. The EEG alpha power data were then normalized using a z-score transformation, achieved by subtracting the mean alpha power (averaged across all electrodes for all conditions) from each data point and dividing the result by the standard deviation of the sample. Z-scores, which measure a value's relationship to the mean of a group of values, are expressed in terms of standard deviations from the mean. For instance, a z-score of 1 indicates that a data point is one standard deviation above the mean, while a z-score of -1 implies that a data point is one standard deviation below the mean. Z-scores are beneficial for standardizing data and for comparing scores between different groups or variables. Consequently, the normalized alpha power values (or 'spatial z-scores of alpha powers') for each channel help illustrate the relative strength of alpha oscillations on the scalp (the 'relative topographical distribution of alpha power'). The steps involved in this post-processing procedure are depicted in Fig. 2.

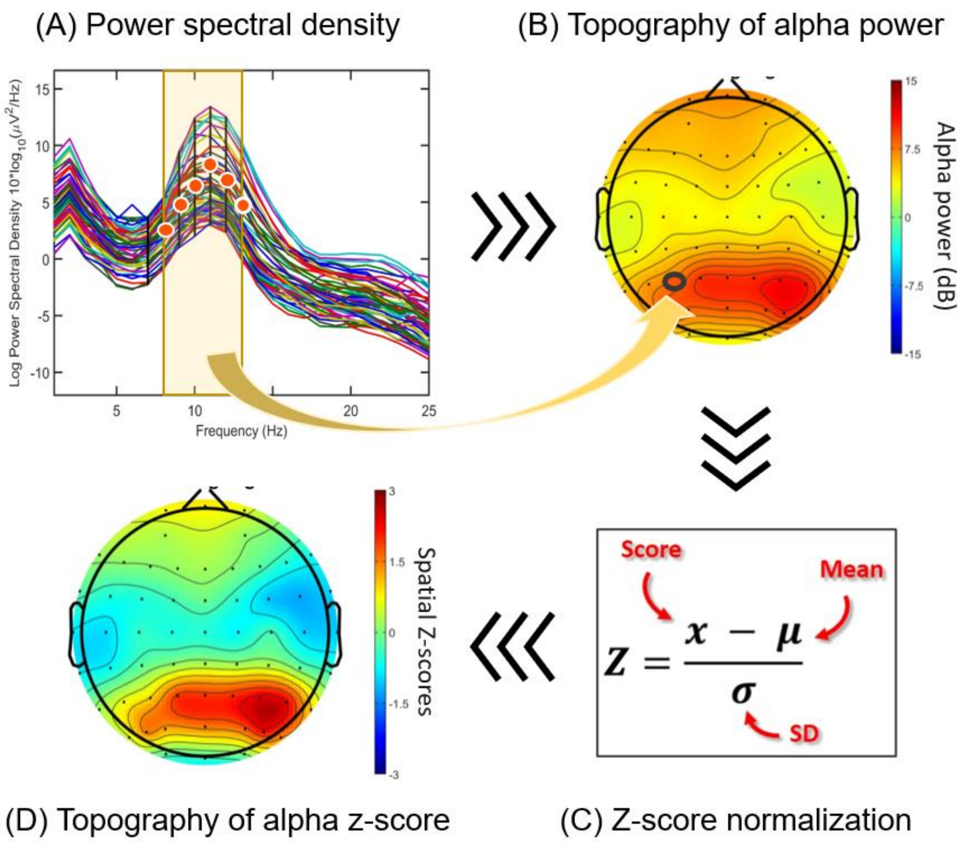

Fig. 2. A schematic illustration of the EEG post-processing analysis procedure. (A) Individual average alpha power spectral density over the stimulus period (0 – 1800 ms) across all trials. Each colored line represents an individual electrode channel. The boxed region highlights the predefined alpha frequency range from 8 to 13 Hz. EEG data at red points within alpha band frequencies are averaged to yield averaged alpha power across trials for a single electrode. (B) Topography of alpha power, showing how absolute alpha power for each channel is distributed on the scalp. (3) The formula used for calculating a z-score. (4) Topography of the normalized alpha power (spatial alpha z-scores).

For the cluster analysis aimed at comparing alpha power magnitude across four distinct target positions, we selected two clusters: parieto-occipital and frontal electrodes for further investigation. This selection was influenced by neuroimaging studies that highlight the frontoparietal attention network (Baars & Gage, 2010; Larson & Lee, 2013; Misselhorn et al., 2019), coupled with our own observations of neural alpha band activities within the frontal cortex and the parieto-occipital cortex, despite the parietal region being the major focus of many studies. The individual mean alpha power was computed across each cluster: 19 parieto-occipital electrodes (P7, P5, P3, P1, PZ, P2, P4, P6, P8, PO7, PO5, PO3, POZ, PO4, PO6, PO8, O1, OZ, and O2) and 14 frontal electrodes (FP1, FPZ, FP2, AF3, AF4, F7, F5, F3, F1, FZ, F2, F4, F6, and F8). Grand averages for frontal clusters and parieto-occipital clusters were obtained for each condition by averaging the normalized alpha power (spatial z-scores) across subjects. The resulting grand mean was subsequently compared across four different target positions.

Moreover, the study employed the Attentional Modulation Index (AMI) to ascertain whether hemispheric lateralization of alpha power was present in both front-back and left-right conditions. To achieve this, the normalized alpha power for each electrode when the participant directed their attention to the right (or back) target position was subtracted from the alpha power when attention was focused on the left target (or front) position, according to the formula [AMI = Xa -Xb], where X denotes the normalized alpha power, 'a' is attention directed to the left (or front), and 'b' is attention directed to the right (or back). A positive AMI suggests more pronounced neural responses for Xa, and a negative AMI indicates more substantial responses for Xb. In contrast, a zero AMI implies no discernible difference between the conditions (Bonacci et al., 2020; Deng et al., 2020; Seifi Ala et al., 2020; Wostmann et al., 2016). Subsequently, the averaged AMI across 11 left-hemispheric parieto-occipital electrodes (TP7, CP5, CP3, CP1, P7, P5, P3, P1, PO7, PO3, and O1) was compared with the averaged AMI across 11 right-hemispheric parieto-occipital electrodes (CP2,

CP4, CP6, TP8, P2, P4, P6, P8, PO4, PO8, O2) to evaluate the presence of alpha lateralization (Wostmann et al., 2019).

3.5c. Neural speech tracking

For the analysis of neural speech tracking, the duration of each sentence was standardized to precisely 1.8 seconds, ensuring a uniform length across all sentence groups. Any sentence longer than 1.8 seconds was truncated, while those that fell short were extended with trailing silence. The acoustic envelopes of both the target and distractor speech stimuli were extracted by applying the absolute value of the Hilbert transform. This transform was subsequently low-pass filtered at 25 Hz using a finite impulse response (FIR) filter, and the first derivative was computed before half-wave rectification and down-sampling to a sampling frequency of 128 Hz (Jaeger et al., 2020; Mirkovic et al., 2019; Petersen et al., 2017). By computing the first derivative of the speech envelope (referred to as the speech onset envelope), salient changes in the speech signal were accentuated, particularly at the commencement of syllables. In practice, using the first derivative of the speech envelope eliminates potential drift in the correlation between EEG and speech envelope (Petersen et al., 2017). The process of speech signal processing for the neural speech tracking analysis is illustrated in Fig. 3.

The preprocessed EEG underwent a bandpass filter between 2 and 25 Hz and was subsequently down-sampled to align with the sample rate of the speech envelope (128 Hz). Each channel's filtered EEG was cross-correlated with the envelopes of the to-be-attended and to-be-ignored speech (See Fig 4), generating a cross-correlation function that depicts activity phase-locked to the acoustic dynamics of each speech stream. The output coefficients range from -1 to +1, with a value of +1 (or -1) denoting a completely positive (or negative) linear relationship between the two time series. The neural speech tracking response, averaged on central clusters (FC1, FCZ, FC2, C1, CZ, C2, CP1, CPZ, and CP2) around the Cz electrode, was selected for further analysis (Jaeger et al., 2020; Mirkovic et al., 2019). The effect of auditory attention on neural speech tracking can be quantified by the cross-correlation coefficient of only the to-be-attended speech, or by comparing the to-be-attended speech with the to-be-ignored speech for each participant and target position condition (Petersen et al., 2017). The time course of the cross-correlation function revealed the varying coefficients over the time lag. In this study, the peak correlation coefficients of the first negative (N1), second positive (P2), second negative (N2), and third positive (P3) deflections were compared across target positions.

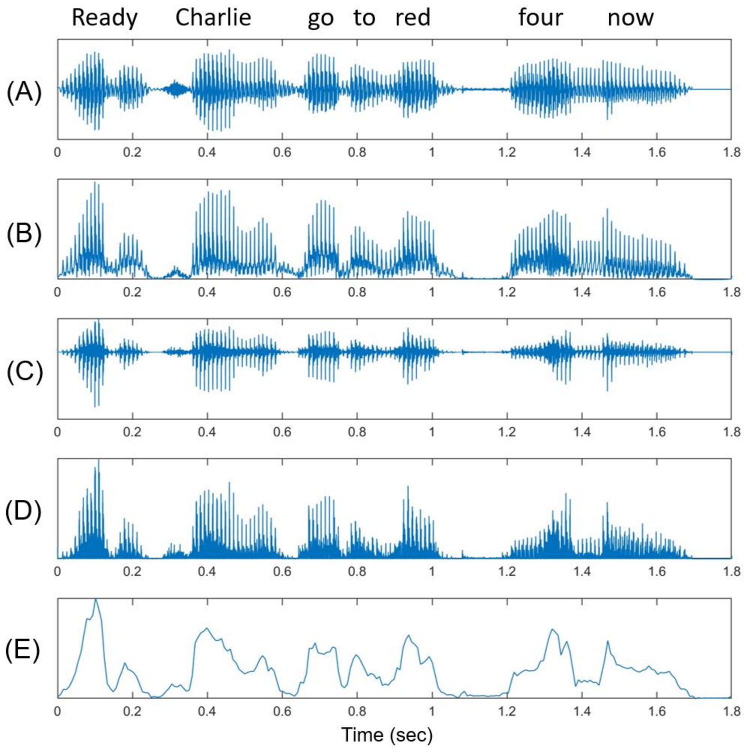

Fig. 3. The process of speech signal processing for the neural speech tracking analysis. (A) Waveform of the CRM speech sentence (‘Ready Charlie go to red four now’). (B) The low-pass-filtered speech envelope obtained via Hilbert transform. (C) The first derivative of the filtered envelope. (D) The half-wave rectification of the processed signal. € The resampled speech envelope at 128 Hz.

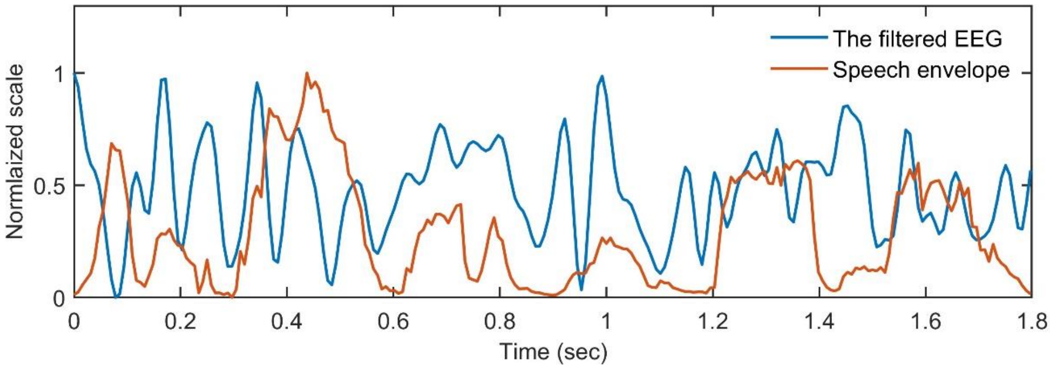

Fig. 4. Normalized waveforms of the filtered EEG response and the to-be-attended speech envelope, both rescaled within a 0 - 1 interval. Red indicates the time course of the speech envelope, while blue represents the neural activity response recorded at the Cz channel from subject #22, captured while the subject listened to the speech.

3.5d. Statistical analysis

Data analysis was conducted using SPSS, version 28.0 (IBM SPSS Statistics for Windows, Armonk, New York). When the data satisfied the normality assumption ($p > 0.05$), parametric tests were deployed. However, in cases where the data deviated from this assumption, appropriate non-parametric tests were used instead. Mauchly's sphericity test was applied to all repeated-measures ANOVA models. Should there be any significant sphericity violations ($p < 0.05$), the Huynh-Feldt correction was introduced. The threshold for significance was set at an alpha level of 0.05.

To contrast the behavioral listening task outcomes (hit response rates for target words of given speech sentences) among the four different target positions, a two-way repeated measures ANOVA was utilized with within-factors comprising target positions (front, back, left, and right) and target words (colors and numbers), performed on the hit rates (%). This was also succeeded by a post-hoc test using a Bonferroni adjustment for pairwise comparison of hit rates for each target word across target positions, and of hit rates for each target position across target words. The interaction between the two factors was tested.

To contrast the normalized alpha power (spatial alpha z-scores) in both frontal and parieto-occipital clusters among different target positions, a one-way repeated measures ANOVA was employed with a within-factor of target positions. This was performed on the normalized alpha power for both clusters respectively, and it was followed by a post-hoc analysis using Bonferroni's adjustment for pairwise comparison between target positions. Additionally, a simple paired t-test was conducted to compare alpha power between the front-back and left-right configurations. In order to assess alpha lateralization in front-back and left-right conditions, paired-sample t-tests were conducted to compare the left hemispheric AMI and right hemispheric AMI for both conditions, respectively.

To examine the main effect of target position on neural speech tracking, a one-way repeated measures ANOVA was conducted on the four peak coefficient components (N1, P2, N2, and P3), respectively. Following this, post-hoc analysis with Bonferroni's adjustment was used for pairwise comparison between target positions. Additionally, we employed a two-way repeated measures ANOVA with stimuli (to-be-attended and to-be-ignored) and target position as within-subject factors to evaluate the principal impact of the stimulus, followed by post-hoc analysis for pairwise comparison. We also assessed the interaction between these two factors. To determine a correlation between the outcomes of the behavioral test and the neural correlates of auditory attention, as well as between the normalized alpha power and cross-correlation coefficients, we implemented Pearson's correlation analysis.

4. RESULTS

4.1. Spatial selective auditory attention task results

In a two-way repeated measures ANOVA with two within-factors of target position and target words, all variables satisfied the normality assumption except for the hit rate for the target number in the right condition. Despite this, ANOVA was carried out, disregarding non-normality, as ANOVA is robust against violations of the normality assumption, only one variable failed to meet the assumption and the within-group analysis has an equal number of data points. Due to the violation of sphericity ($X^2(5) = 39.16$, $p < 0.001$), the Huynh-Feldt correction was applied. A two-way ANOVA demonstrated a statistically significant main effect of target position ($F(1.33, 18.68) = 22.98$, $p < 0.001$, $\eta_p^2 = 0.62$), but not of target words ($p > 0.05$), indicating hit rates varied with target positions. The post hoc analysis, which collapsed across the factor of target words, showed that the hit rate was the lowest in the back condition compared to other conditions (all $p < 0.01$), and the results for the front condition were significantly less than those for the left and right conditions (all $p < 0.01$), but no difference in the hit rate was observed between the left and right conditions, as seen in Fig 5. The pairwise comparison indicates each hit rate for target words significantly followed the same pattern: [Left = Right > Front > Back condition]. Furthermore, no significant interaction was found between the target word and the target position. Notably, approximately 96% and 93% of the errors made for the color and number over all trials, respectively, were identified as intrusion errors. Table 2 outlines the results of the behavioral spatial selective listening task. These results demonstrate a distinct listening difficulty as a function of the target position. These findings align perfectly with our first hypothesis, where the front-back spatial conditions would yield lower hit rates than the left-right lateral conditions and the hit rate would be lowest in the back condition, which is considered the most demanding situation.

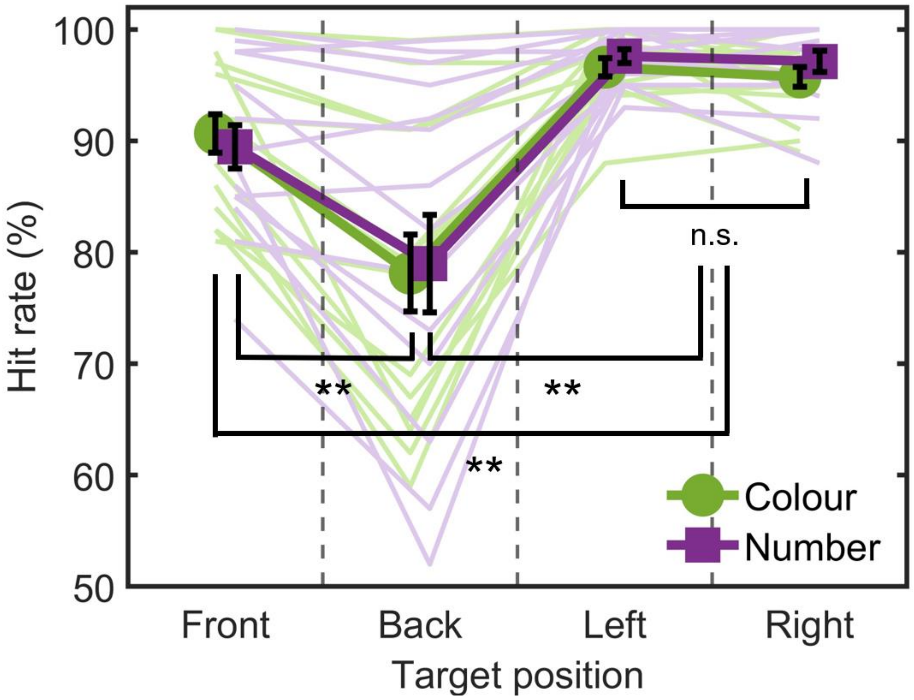

Fig. 5. Grand average hit response rate (%) for target colors (represented by the color green) and target numbers (shown in purple) across the four different target positions. The light green and purple lines depict individual data. Error bars indicate ±1 SE. n.s. = not significant, *p < .05, **p < .01.

Table 2. Behavioral spatial selective listening task results.

Words	Direction	Behavioral response (%): Mean (SD)			
		Hit rate	Intrusion error	Insertion error	Omission error
Color	Front	90.67 (6.65)	9.07 (6.47)	0.20 (0.41)	0.07 (0.26)
	Back	78.13 (13.37)	21.07 (13.47)	0.80 (1.15)	0.00 (0.00)
	Left	96.60 (3.22)	3.33 (3.04)	0.07 (0.26)	0.00 (0.00)
	Right	95.73 (3.43)	3.87 (2.97)	0.27 (0.59)	0.13 (0.52)
	Total	90.28 (6.67)	9.33 (6.49)	0.33 (0.60)	0.05 (0.19)
Number	Front	89.47 (7.47)	9.93 (7.51)	0.40 (0.63)	0.20 (0.56)
	Back	79.00 (16.96)	19.47 (16.50)	0.67 (0.90)	0.87 (0.99)
	Left	97.60 (2.32)	2.07 (2.37)	0.27 (0.80)	0.07 (0.26)
	Right	97.13 (3.62)	2.60 (3.33)	0.20 (0.41)	0.07 (0.26)
	Total	90.80 (7.60)	8.52 (7.43)	0.38 (0.69)	0.30 (0.52)

4.2. Magnitude and topography of alpha power

Alpha frequency power spectra over the stimulus period implied that alpha power appeared to be successfully induced during the behavioral listening task as evidenced by the inverted U-shape curved responses identified within the predefined alpha frequency range (8 – 13 Hz). Moreover, every individual alpha peak frequency was identifiable (See Fig. 6), even though some individual activity seemed weak. The absolute alpha power in decibels was normally distributed, with a mean of 3.26 dB (SD = 3.88). The 95% confidence interval for the mean was [1.12, 5.42]. The Individual Alpha Peak Frequency (IAF) was not normally distributed, with a mean of 9.93 Hz (SD = 1.33, CI = [9.19, 10.67]).

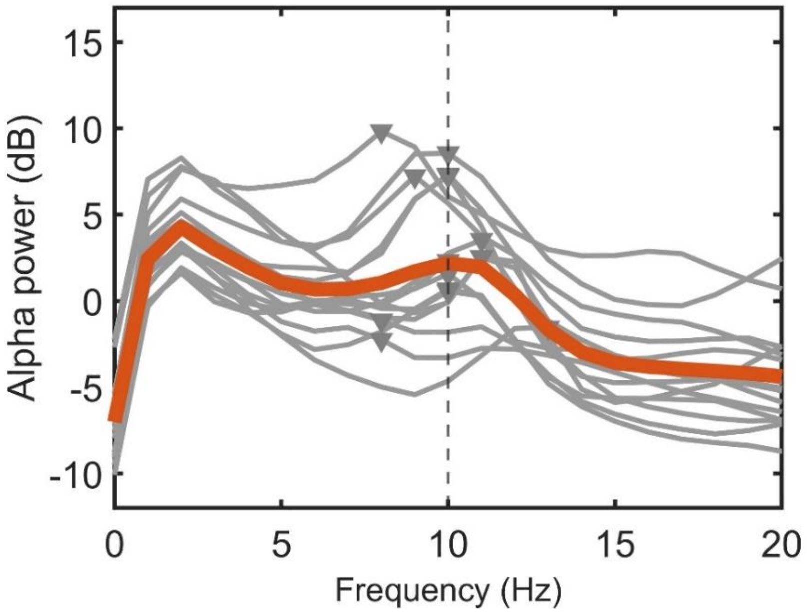

Fig. 6. The Individual Alpha Frequency (IAF) power spectra during the stimulus period (0 – 1800 ms). The orange line represents the averaged alpha power spectra across all subjects and all electrodes, whereas the gray lines denote the response spectra of each participant. The gray triangle marks indicate the individual alpha peak frequency.

Regarding the topographical distribution of normalized alpha power across different target positions as illustrated in Fig. 7, alpha power was remarkably induced and intensively distributed in parieto-occipital electrodes in all conditions. There also appeared to be slight alpha activity in the frontal lobe. However, there were no visible changes in the topographical distribution of alpha power across the target position based on the grand average data. Upon detailed observation, we noticed a relatively weaker alpha power in the frontal target scenario compared to others, and a slight enhancement of alpha activity in the left hemisphere under the left target scenario, relative to other target positions.

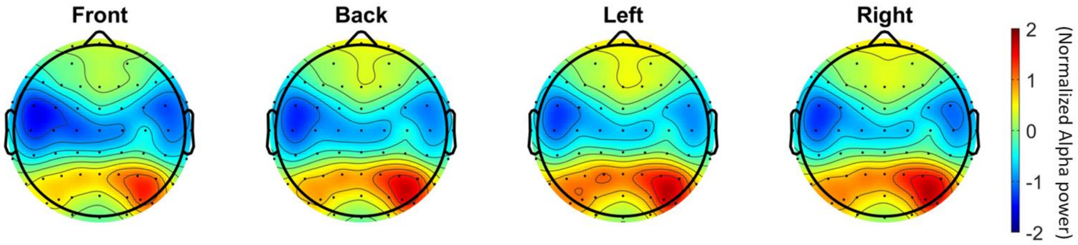

Fig. 7. Grand average topographies of alpha power spatial z-scores (normalized alpha power) for the four different target positions. The spatial z-score of ‘0’ refers to the grand mean alpha power over all conditions, which is presented with a green color in this topography. A z-score of 1 corresponds to a data point that is one standard deviation greater than the mean, while a z-score of -1 points to a data point that is one standard deviation less than the mean. As the values increase positively, they are displayed in a deeper red, and as they increase negatively, they are shown in a deeper blue on the scalp topography.

A one-way repeated measures ANOVA was conducted on the normalized alpha in both parieto-occipital (PO) and frontal electrode clusters. While all normality assumptions were met, the sphericity assumption was not satisfied according to Mauchly's test ($X^2(5) = 29.72$, $p < 0.001$); hence, degrees of freedom were corrected using the Huynh-Feldt method. This showed a marginal yet significant main effect of target position in the parieto-occipital cluster ($F(1.73, 24.21) = 3.573$, $p = 0.0495$, $\eta_p^2 = 0.203$), suggesting that the target position significantly influenced the amplitude of normalized alpha power. However, the post hoc analysis using a Bonferroni adjustment didn't spot any significant differences in normalized alpha power (spatial alpha power z-score) between target positions (See Fig. 8). Notwithstanding, the individual data trend in Fig. 8-(B) hinted that the averaged alpha power was attenuated in the front condition compared to other conditions. Indeed, when excluding the left and right target conditions, a simple direct comparison between the front and back conditions using a paired t-test revealed the significantly greater normalized alpha power when attending to the back than to the front ($t(14) = -2.17$, $p < 0.05$, $d = -0.56$). The study also revealed that the data obtained from the front-back configuration was significantly lower than the results from the left-right configuration ($t(29) = -2.38$, $p < 0.05$, $d = -0.43$). There was no significant main effect of target position on averaged alpha power in the frontal lobe ($p = 0.20$). These findings are partially consistent with our second hypothesis, which posited that alpha power in the front-back conditions would differ from the power in the left-right condition.

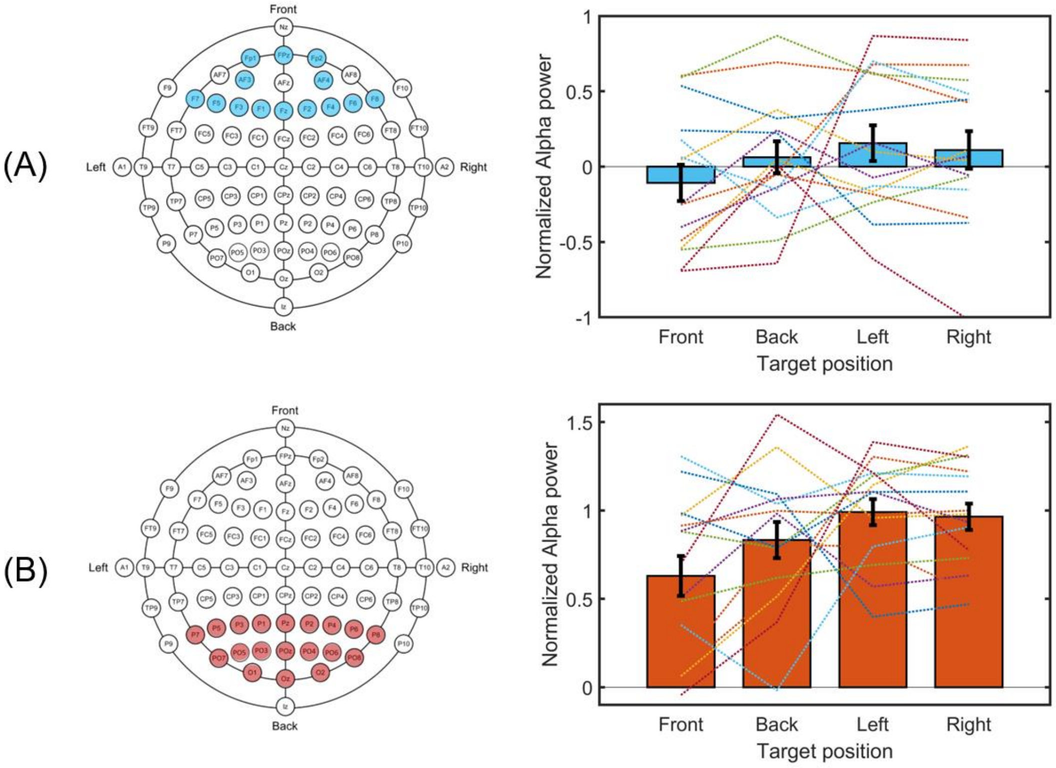

Fig. 8. Cluster analysis of normalized alpha power on (A) frontal electrodes (shown in cyan) and (B) parieto-occipital electrodes (shown in orange) across different target positions. Cyan and orange bars represent spatial alpha z-scores for both clusters. Thin-dashed color lines illustrate individual data sets. Error bars indicate ±1 standard error (SE).

4.3. Alpha power modulation indices

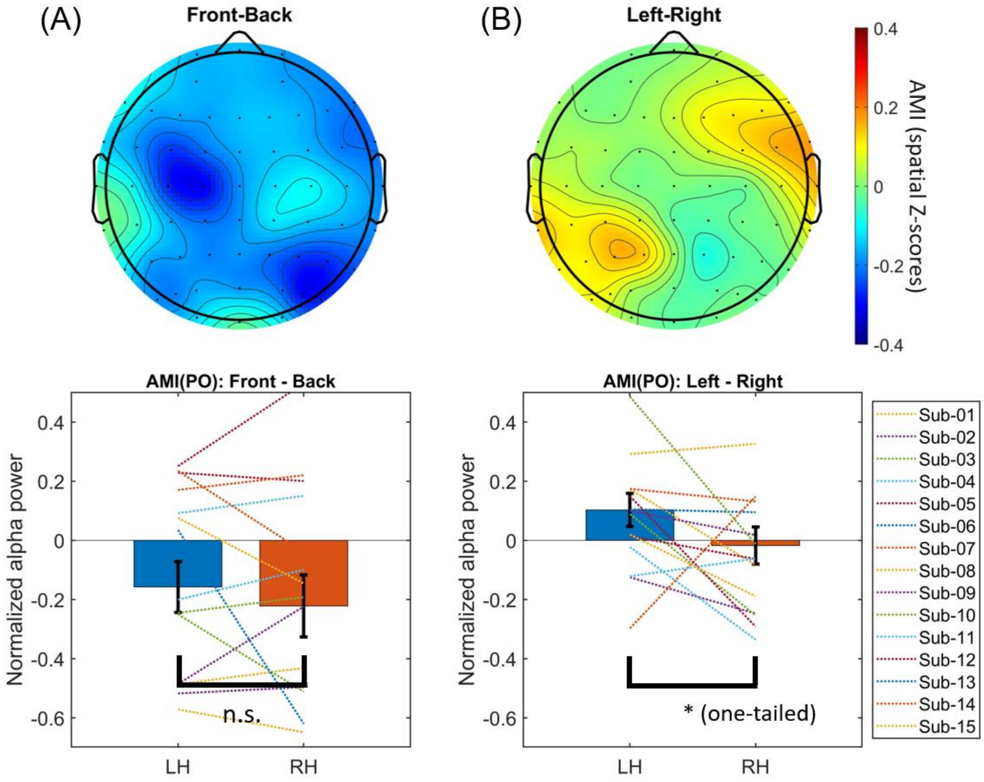

Fig. 9. Topographic maps of the attentional modulation index (AMI) for (A) front-back and (B) left-right conditions over the speech stimulus period. The green color on the scalp topography refers to a zero difference in normalized alpha power between two AMI topographies. The bar graphs in (B) depict the mean AMI across the parieto-occipital (PO) sensors on the left hemisphere (LH) and right hemisphere (RH). Thin-dashed color lines represent individual data trends. Error bars represent ±1 standard error (SE). n.s. = not significant, *p < .05.

For alpha power contrast analysis using AMI within the parieto-occipital lobe, with the normality assumption being non-significant for all variables (all $p > 0.05$), a paired samples t-test showed that the average left-hemispheric AMI for the left-right configuration (M = 0.10, SD = 0.22) was significantly greater than the right-hemispheric AMI (M = -0.02, SD = 0.24) (t(14) = 1.970, $p = 0.034$, one-tailed, $d = 0.51$). This indicated the presence of hemispheric alpha lateralization in the lateral listening condition (See Fig. 9). In contrast, the t-test revealed no significant difference in AMI for the front-back configuration between both hemispheres ($p > 0.05$), suggesting no alpha lateralization in the front-back condition. This finding aligns with our hypothesis, which proposed no observable alpha lateralization in the case of front-back spatial hearing.

4.4. Neural speech tracking functions

4.4.a. Time course of grand average neural speech tracking function

The time course of the grand average cross-correlation (neural speech tracking) for the to-be-attended speech for the four different target positions, as depicted in Fig. 10-(A), displayed the first negative deflections (N1), second positive deflections (P2), second negative deflections (N2), and third positive deflections (P3) at approximate time lags of 301 ms (range: 280 – 310), 386 ms (range: 360 – 400), 563 ms (range: 560 – 570), and 656 ms (range: 650 – 670) respectively (refer to Table 3). Although negative peaks appeared to emerge prior to the current N1, we chose to designate the most pronounced peak as the first negative peak (N1) due to its significant amplitude in the following response. Similarly, the large N2 and P3 components were considered for this study. In the morphology of the grand average cross-correlation function, N2-P3 peaks for the to-be-attended envelope appeared more robust than N1-P2. Subsequently, the N2-P3 coefficients for the lateral condition were apparently stronger than the coefficients for the front-back condition.

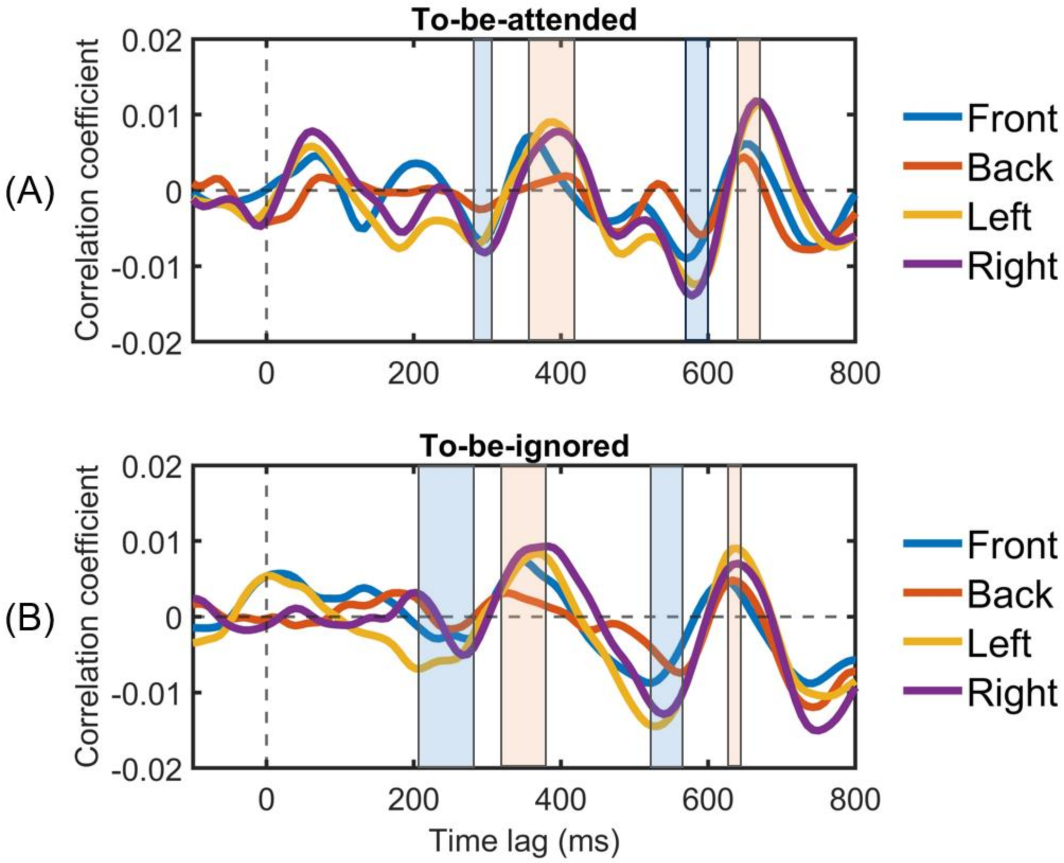

Fig. 10. Time course of the grand average neural speech tracking (cross-correlation) functions for (A) the to-be-attended speech envelope and (B) the to-be-ignored speech across target positions: front target position (blue line), back position (orange line), left position (yellow line), and right position (purple line). The blue and red transparent windows represent the ranges in which negative and positive peak correlation coefficients were identified, respectively.

Table 3. Mean values and standard deviations (SDs) of latencies for neural speech tracking cross-correlation components.

Stimulus	Position	Latency of Cross-correlation components (ms)							
		N1		P2		N2		P3	
		Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)
To-be-attended	Front	284.38	(58.30)	369.27	(55.12)	561.98	(33.34)	649.48	(28.46)
	Back	318.23	(69.03)	398.44	(56.10)	560.94	(55.10)	650.00	(39.75)
	Left	296.88	(64.76)	385.42	(50.43)	565.63	(44.67)	660.94	(26.71)
	Right	306.25	(41.26)	393.75	(47.50)	567.19	(36.02)	665.10	(31.92)
	Total	301.43	(58.34)	386.72	(52.29)	563.93	(42.28)	656.38	(31.71)
To-be-ignored	Front	261.46	(64.01)	361.46	(68.77)	519.27	(45.64)	616.67	(28.85)
	Back	266.67	(73.93)	360.42	(80.32)	533.85	(57.99)	619.79	(44.36)
	Left	269.27	(59.12)	359.38	(42.59)	528.13	(41.63)	632.29	(37.64)
	Right	269.27	(43.68)	360.42	(60.43)	526.56	(50.53)	629.17	(34.67)
	Total	266.67	(60.19)	360.42	(63.03)	526.95	(48.95)	624.48	(36.38)

Table 4. Mean values and standard deviations (SDs) of amplitudes for neural speech tracking cross-correlation components.

Stimulus	Position	Amplitude of Cross-correlation components (Coefficients)							
		N1		P2		N2		P3	
		Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)
To-be-attended	Front	-0.013	(0.009)	0.012	(0.008)	-0.016	(0.009)	0.009	(0.008)
	Back	-0.012	(0.011)	0.011	(0.012)	-0.016	(0.010)	0.008	(0.011)
	Left	-0.014	(0.013)	0.017	(0.015)	-0.022	(0.013)	0.018	(0.011)
	Right	-0.012	(0.010)	0.016	(0.012)	-0.021	(0.010)	0.017	(0.012)
	Total	-0.013	(0.011)	0.014	(0.011)	-0.019	(0.011)	0.013	(0.011)
To-be-ignored	Front	-0.008	(0.008)	0.014	(0.009)	-0.015	(0.010)	0.008	(0.007)
	Back	-0.008	(0.007)	0.008	(0.007)	-0.013	(0.010)	0.009	(0.009)
	Left	-0.009	(0.012)	0.013	(0.011)	-0.021	(0.015)	0.016	(0.007)
	Right	-0.008	(0.009)	0.016	(0.009)	-0.020	(0.011)	0.013	(0.011)
	Total	-0.008	(0.009)	0.013	(0.009)	-0.017	(0.011)	0.011	(0.009)

4.4.b. N1-P2 components of speech tracking function

For a one-way as well as two-way ANOVA analysis, N1 and P2 coefficient data for both to-be-attended and to-be-ignored speech were found to be normally distributed, and the sphericity assumptions were satisfied. A one-way repeated measures ANOVA revealed no significant main effect of target position on N1 and P2 peak coefficients of to-be-attended speech envelope, respectively (both $p > 0.05$). A two-way repeated measure ANOVA with two factors (target position and stimuli) revealed a significant main effect of stimuli (to-be-attended versus to-be-ignored) on N1 coefficient ($F(1, 14) = 7.06$, $p = 0.\ 05$, $\eta_p^2 = 0.335$), not on P2 coefficient ($p = 0.37$). Even though Fig. 11 illustrated an apparently stronger speech tracking response (N1) to the to-be-attended speech versus to-be-ignored speech envelope, the post hoc analysis using Bonferroni correction indicated that this difference was significant only in the front target condition. Here, the N1 coefficient for the to-be-attended speech envelope was significantly more pronounced than its to-be-ignored counterpart ($p < 0.05$). Furthermore, there was no significant interaction between target position and stimulus in both N1 and P2 component (both $p > 0.05$) (See Fig.12). Table 4 presents the means and standard deviations (SDs) of the amplitudes for the cross-correlation components of neural speech tracking.

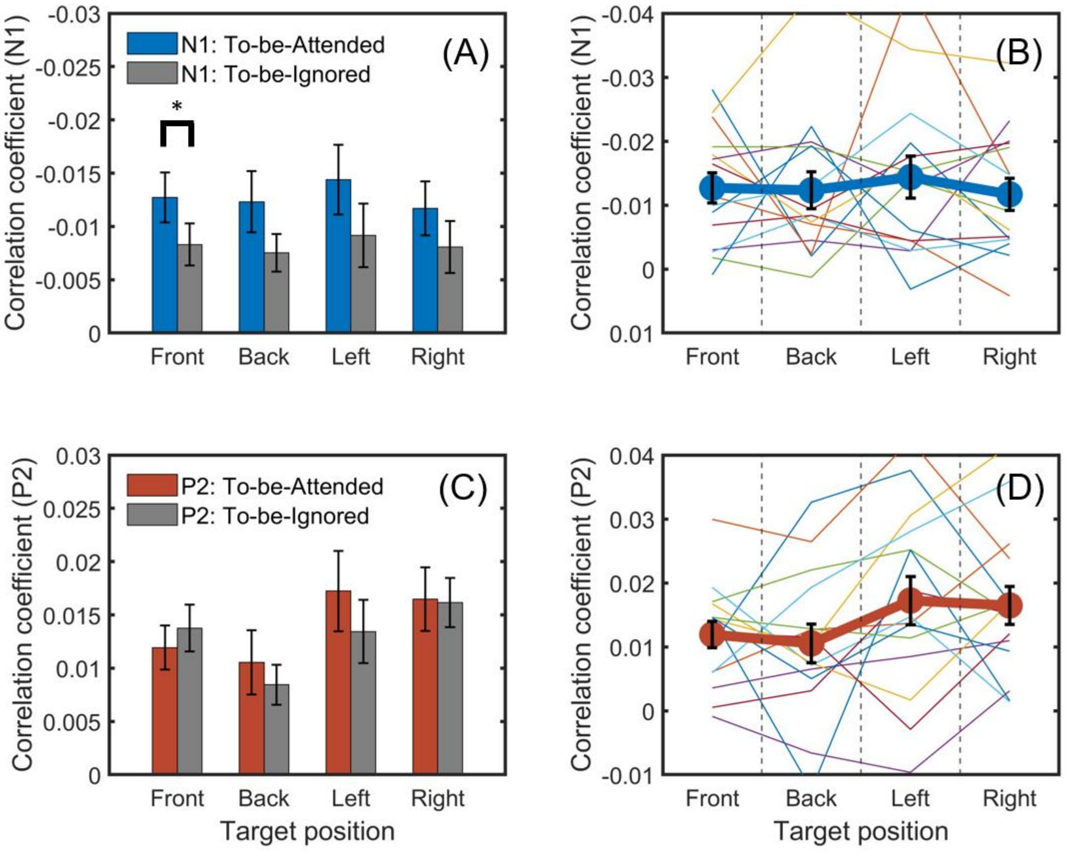

Fig. 11. Cross-correlation coefficients of neural speech tracking across target positions for N1 and P2 components. **(**A) Negative peak (N1) coefficient of neural speech tracking response to the to-be-attended speech envelope (blue bar) and to-be-ignored speech envelope (gray bar) across target positions. (B) Individual N1 coefficients of neural speech tracking response to the to-be-attended speech. (C) P2 coefficient for the to-be-attended and to-be-ignored speech. (D) Individual P2 coefficients. Thick solid blue and red lines in (B) and (D) denote the group means of N1 and P2. The Error bars denote ±1 SE. The asterisks indicate significant differences: * P < 0.05.

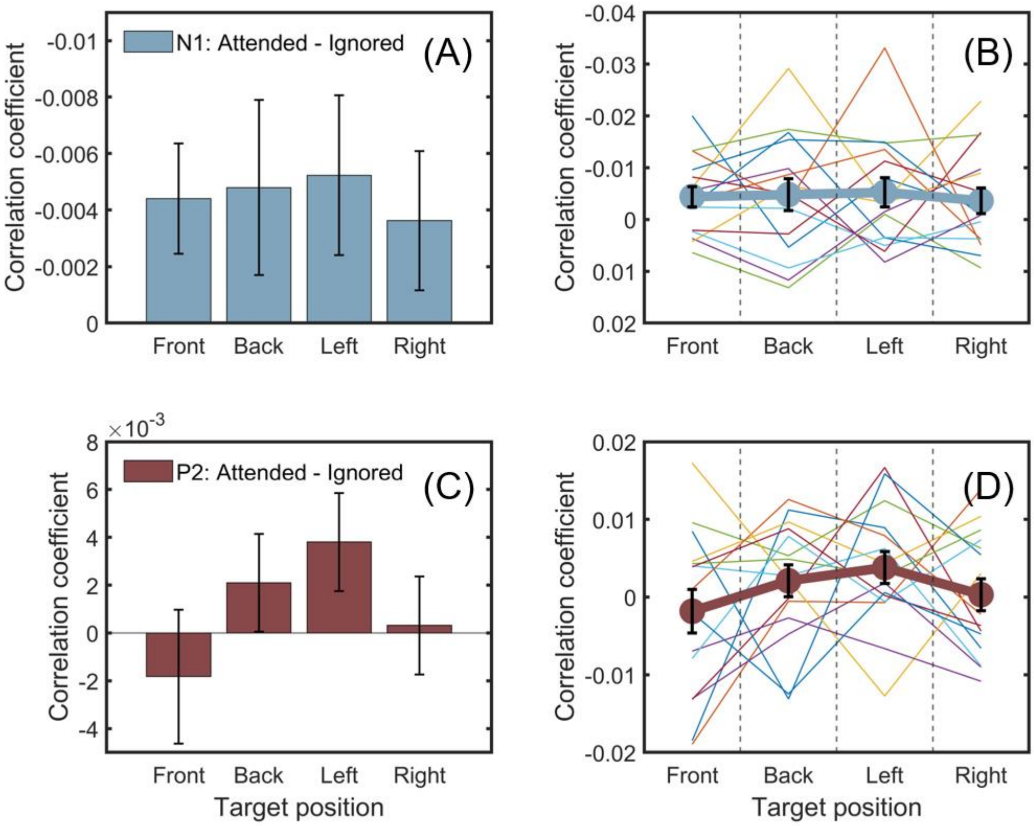

Fig. 12. Cross-correlation coefficient gap of neural speech tracking between the to-be-attended speech and to-be-ignored speech across target positions for the N1 and P2 components. (A) Negative peak (N1) coefficient gap (sky-blue bar) across different target positions. (B) Individual gap measurements of N1 coefficients between attended and ignored speech. (C) P2 coefficients for the to-be-attended speech in comparison to the to-be-ignored speech (reddish bar). (D) Individual data for the P2 coefficient gap. Error bars denote ±1 SE.

4.4.c. N2-P3 components of speech tracking function

Although 3 out of the 32 variables (4 components x 4 target positions x 2 stimuli) used for the analysis did not adhere to the normality assumption, one-way and two-way ANOVA were still executed. This could be permissible because ANOVA is robust to violations of the normality assumption. Additionally, only a minor fraction of the variables (3 out of 32) violated the assumption, and the number of data points in the within-group analysis was balanced. Sphericity was confirmed using Mauchly's test. A one-way repeated measures ANOVA revealed a significant main effect of target position on N2 ($F(3, 42) = 3.30$, $p < 0.029$, $\eta_p^2 = 0.191$), and P3 peak coefficients of to-be-attended speech envelope ($F(3, 42) = 5.34$, $p = 0.003$, $\eta_p^2 = 0.276$). The post hoc analysis using Bonferroni adjustment showed stronger N2 peak coefficients for the left condition than the back condition ($p = 0.048$, one-tailed), and lower P3 peak coefficients for front and back conditions than the left condition ($p = 0.47$ and 0.34, both one-tailed). The P3 for the back was slightly lower than the front but the difference was not significant.

A two-way ANOVA with two factors (target position and stimuli) revealed a significant main effect of stimuli (to-be-attended versus to-be-ignored) on P3 coefficient ($F(1, 14) = 5.22$, $p < 0.05$, $\eta_p^2 = 0.335$), but not on N2 coefficient ($p = 0.27$). The post hoc analysis using Bonferroni correction did not find any significant pairwise comparisons for P3 coefficients between the to-be-attended speech and to-be-ignored speech across four different target positions (all $p > 0.05$). Furthermore, no significant interactions between target position and stimulus were found in both N2 and P3 components (both $p > 0.05$) (See Fig.2). With the main effect of target position, the trend of P3 coefficient across target positions aligns closely with the behavioral test outcome and individual data for the P3 showed a more distinct pattern compared to other components (See Fig. 13-D). Fig. 14 shows the coefficient gap across target positions, and notably, the gap for the P3 resembles the hit rate outcome pattern. The gap seems to be larger in the easier right condition, but in the back - the most challenging condition, it showed negative scores indicating ‘reverse’. However, in the end, this trend didn't quite reach statistical significance, indicating no significant interaction between stimulus and target position. Although it did not show a statistically significant lowest correlation coefficient in the back condition, the P3 coefficient results partially support our third hypothesis.

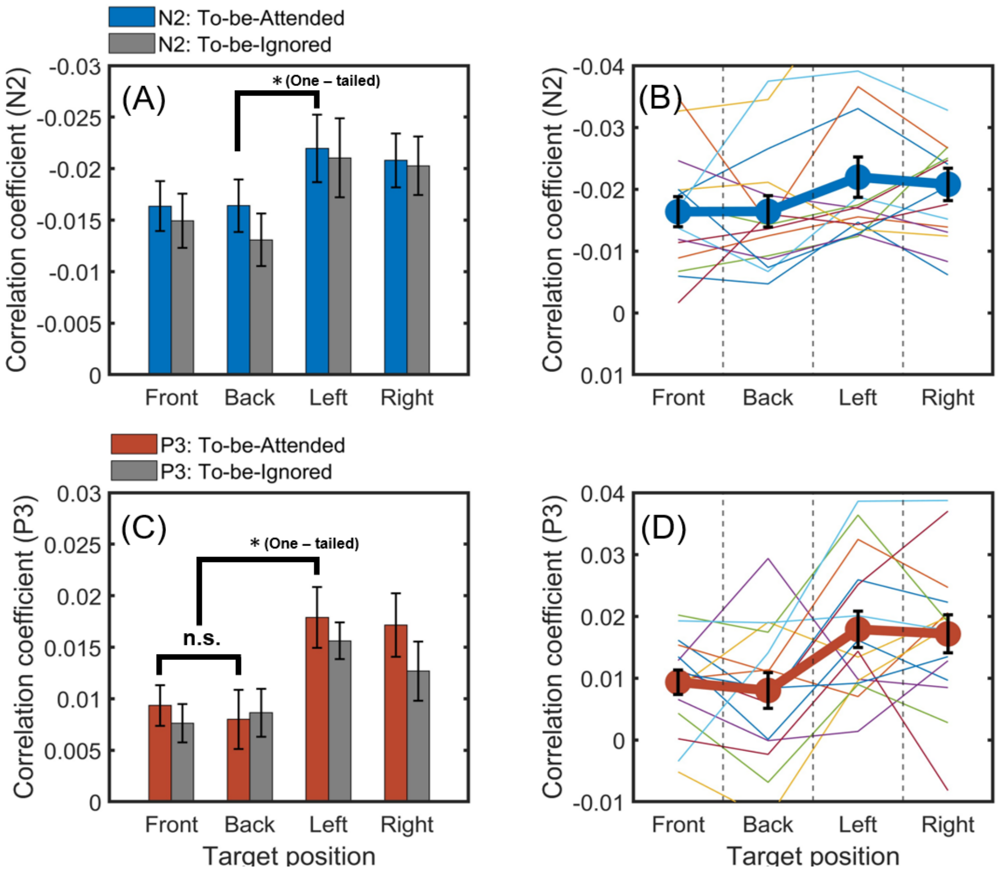

Fig. 13. Cross-correlation coefficients of neural speech tracking across target positions for N2 and P3 components. **(**A) The negative peak (N2) coefficient of the neural speech tracking response to the to-be-attended speech envelope (blue bar) and to-be-ignored speech envelope (gray bar) across target positions. (B) Individual N2 coefficients of the neural speech tracking response to the to-be-attended speech. (C) P3 coefficient for the to-be-attended and to-be-ignored speech. (D) Individual P3 coefficients. Thick solid blue and red lines in (B) and (D) denote the group means of N2 and P3. The Error bars denote ±1 SE. The asterisks indicate significant differences: * P < 0.05.

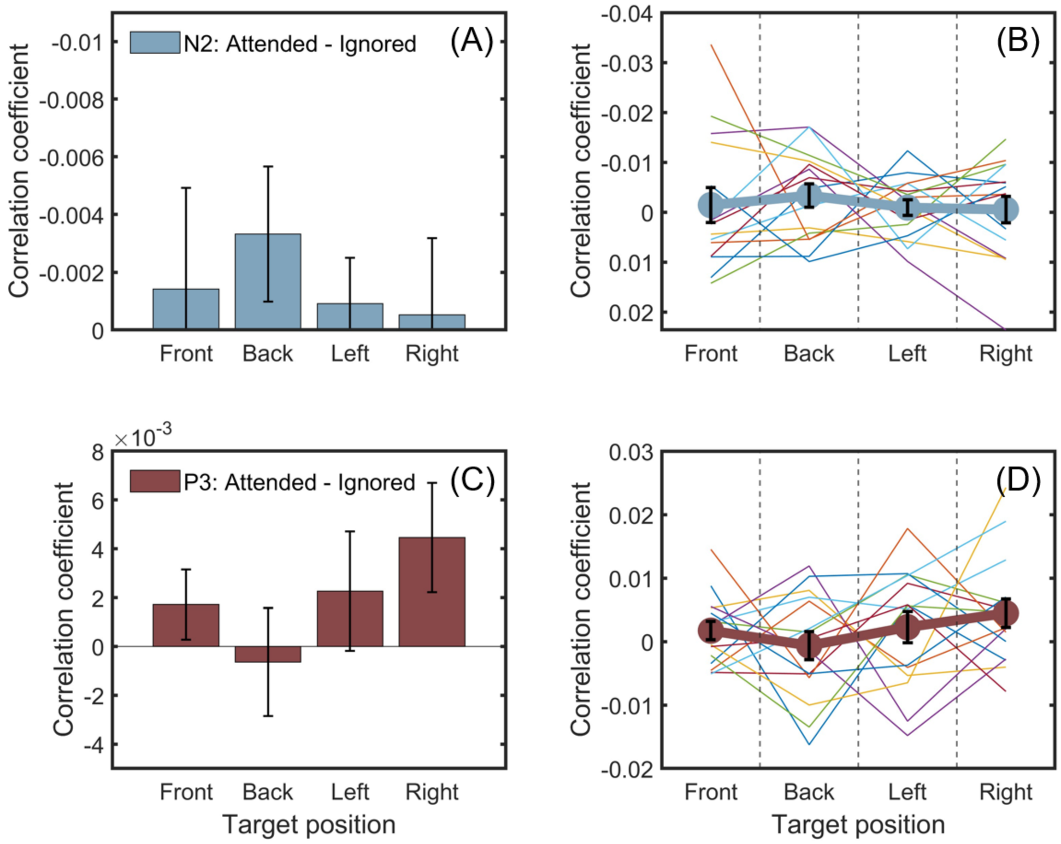

Fig. 14. Cross-correlation coefficient gap of neural speech tracking between the to-be-attended speech and to-be-ignored speech across target positions for the N2 and P3 components. (A) Negative peak (N2) coefficient gap (sky-blue bar) across different target positions. (B) Individual gap measurements of N2 coefficients between attended and ignored speech. (C) P3 coefficients for the to-be-attended speech in comparison to the to-be-ignored speech (reddish bar). (D) Individual data for the P3 coefficient gap. Error bars denote ±1 SE.

4.5. Correlation between behavioral and EEG data

When collapsing the factor of target position, the relationship between the measured behavioral data and neural activity in all conditions was examined. The total number of data points each variable could have is 60 samples, calculated by multiplying the number of participants (15) by the number of target conditions (4). Although the behavioral data violated the assumption of normality, Pearson's correlation analysis was conducted since it is acceptable to perform if at least one of the variables to be tested is normally distributed. Only the correlation between behavioral results and N2 coefficients was tested using Spearman's rank correlation.

Behavioral hit responses to the target color and number were positively correlated with normalized parieto-occipital alpha power (Pearson’s r(60) = 0.27 and 0.30 respectively, both $p < 0.05$) and with P3 cross-correlation coefficients (Pearson’s r(60) = 0.23 and 0.24, both $p < 0.05$, one-tailed). However, there was no significant correlation between alpha power and the neural speech tracking response ($p > 0.05$). A significant correlation was identified between frontal alpha activity and parieto-occipital alpha energies (Pearson’s r(60) = 0.27, $p < 0.05$). The relationship between neural speech tracking components was not the focus of this study. Table 5 describes the correlation statistics. Additional correlation analyses were conducted to investigate if the measured neural correlates of auditory attention could account for variance in the hit rates for the front (n = 15), back (n = 15), and both conditions combined (n = 30). However, these analyses yielded no significant correlations ($p > 0.05$). A simple linear regression analysis was conducted to predict the hit rates for both words, based on normalized alpha power and the P3 coefficient. When considering alpha power, the regression equations for the target color and number depicted in Fig. 13-(A) & (B) were found to be statistically significant (*F*(1, 58) = 4.57 and *F*(1,58) = 5.84, respectively, both $p < 0.05$), with an R-squared of 0.073 and 0.091. Likewise, the regression equations incorporating the P3 coefficient (See Fig. 15) were also statistically significant (*F*(1, 58) = 3.19 and *F*(1,58) = 3.613, respectively, both $p < 0.05$, one-tailed), with an R-squared of 0.073 and 0.091. These findings are consistent with our hypothesis where selective speech identification results would correlate with alpha power and neural speech tracking.

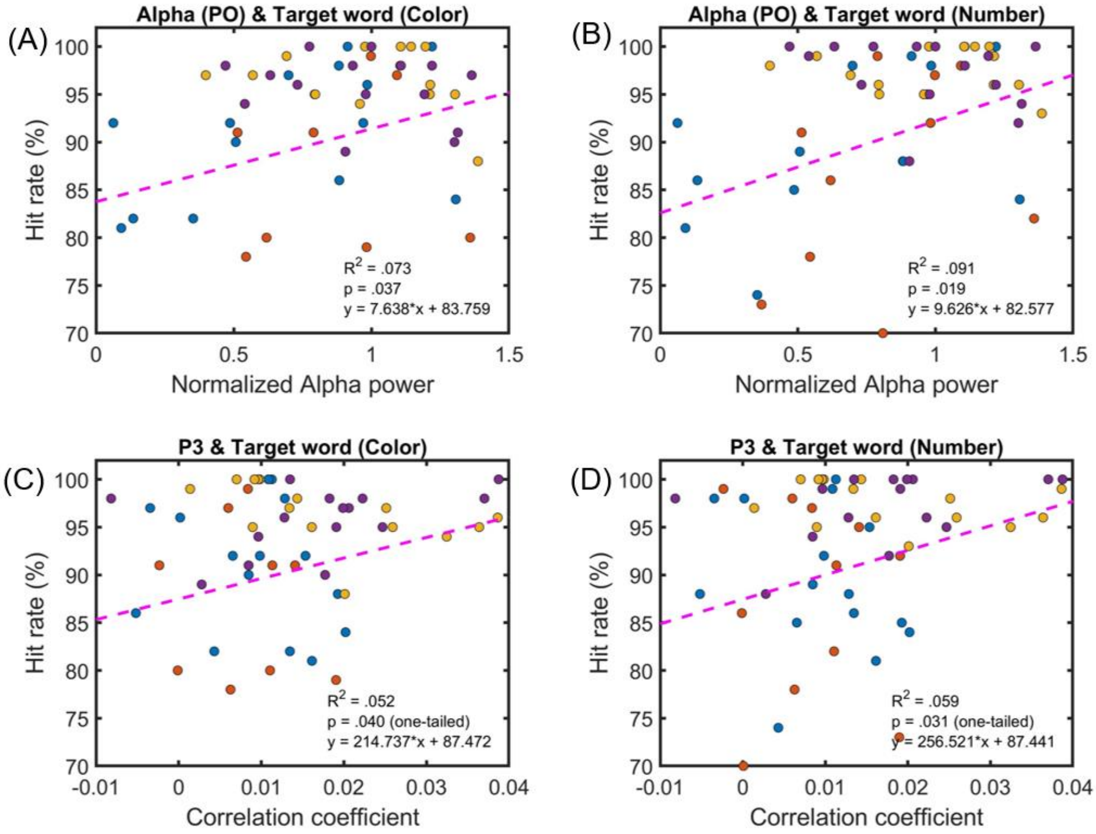

Fig. 15. Scatter plots of the neural correlates (normalized alpha power and P3 correlation coefficient) and auditory attention task results. The pink lines indicate a simple linear regression.

Table 5. Correlation between hit response rates and neural correlates of spatial selective auditory attention.

Variables		Hit_ Col	Hit_ Num	PSD_ PO	PSD_ FT	NST_ N1	NST_ P2	NST_ N2	NST _P3
HIt_ Col	r	1	.945**	.270*	0.048	-0.009	0.112	-0.171	0.228
	p		0.000	0.037	0.715	0.945	0.394	0.193	0.080
Hit_ Num	r		1	.302*	0.120	-0.078	0.141	-0.184	0.242
	p			0.019	0.362	0.556	0.282	0.159	0.062
PSD_ PO	r			1	.273*	0.093	0.106	-0.120	0.140
	p				0.035	0.478	0.422	0.362	0.287
PSD_ FT	r				1	-0.050	-0.240	0.222	-.260*
	p					0.706	0.064	0.088	0.045
NST_ N1	r					1	-0.155	.504**	-0.003
	p						0.237	0.000	0.981
NST_ P2	r						1	-.629**	.503**
	p							0.000	0.000
NST_ N2	r s							1	-.347**
	p								0.007
NST_ P3	r								1
	p								

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Hit_Col: Hit rate response to target color, Hit_Num: Hit rate response to target. number, PSD_PO: Parieto-occipital alpha power, PSD_FT: Frontal alpha power, NST_N1(P2,N2,P3): Neural speech tracking N1(P2, N2, P3) coefficient.

5. DISCUSSION

5.1. Behavioral measure: Selective speech identification task

Our behavioral results clearly demonstrated the impact of spatial listening configurations on listeners' ability to selectively identify target words in a target speech stream in the presence of competing speech. The significant finding that hit response rates were lowest when target speech originated from behind the listener, compared to other directions, and hit rates for the front condition were significantly lower than results for the left and right conditions aligns with our working assumption. In this context, a masker can diminish target intelligibility both by interfering with the target's peripheral representation (termed "energetic masking") and/or by causing more central interference (referred to as "informational masking"). An analysis of response errors can distinguish between effects caused by energetic and informational masking (Brungart & Simpson, 2007; Hambrook & Tata, 2019; Ihlefeld & Shinn-Cunningham, 2008). In selective speech identification tasks dominated by informational masking, subjects frequently report words from the masker instead of the target messages. Conversely, for selective-listening tasks primarily affected by energetic masking, errors tend to be more randomly distributed.

The current study revealed that approximately 95% of the errors that occurred were of the 'intrusion' type. This suggests that the selective identification of speech was predominantly hindered by informational masking. Typically, intelligibility, the clarity or comprehensibility of speech, improves as the spatial separation between two sources increases (Allen et al., 2008; Ihlefeld & Shinn-Cunningham, 2008). Pertaining to the spatial release from masking (SRM), our experiment employed both front-back and left-right configurations. In these configurations, the target and the masker were positioned at the same spatial separation of 180 degrees in terms of spatial angle. Despite this similar spatial separation, the ratio of intrusion errors significantly increased for the front-back condition in comparison to left-right configurations. Specifically, for the back condition, there was approximately a sixfold increase in intrusion errors in terms of target color and a 7.5-fold increase in the target number relative to lateral conditions.

One plausible explanation for this observation is the occurrence of front-back confusion. Front-back confusion is a perceptual phenomenon where a listener misidentifies a sound source as coming from the front when it is actually originating from the back, or vice versa (Brimijoin & Akeroyd, 2012; Perrett & Noble, 1997). This issue can interfere with the listener's ability to accurately identify and differentiate sounds in an environment, leading to an increased rate of intrusion errors. Therefore, our results seem to highlight the significant impact that front-back confusion can have on degraded discrimination of auditory stimuli. Additionally, the higher intrusion errors might explain the absence of differences in hit response rates for target color and number. Given the potential number of choices for both targets, the chance level for color is higher than for number. However, participants seemed to consider only the two words from the target and distracting stream regardless of the potential number of choices during the test.

On the other hand, despite the similarities between the two stimuli in terms of sentence structure, semantics, speech timing, and talker gender, these factors did not influence speech identification at 0 dB TMR when listeners had spatial certainty about the target location, as evidenced by the ceiling effect in the left and right conditions (See Table 2). This suggests that informational masking occurred due to failures in the top-down spatial selection of the target, rather than due to issues with the across-time linkage of target segment streaming.

In addition, the front result was significantly better than the back but worse than the lateral condition. There are a few possible explanations for why better performance might be observed when the target is located at the front rather than the back in a front-back confusion scenario. Humans typically have better auditory acuity for sounds originating from the front. This is because our ears, specifically the shape of the pinna, and auditory processing systems have evolved primarily to localize sounds in the forward-facing direction. The asymmetry in our ear shape can create acoustic cues that help us better determine the elevation and direction of sounds coming from the front (Ihlefeld & Shinn-Cunningham, 2008; Perrett & Noble, 1997; Wightman & Kistler, 1999). Furthermore, most of our daily activities occur in front of us. As a result, we are more accustomed to identifying sounds originating from the front as a sort of learning effect. This expectation or cognitive bias might contribute to better performance when the target sound is in front. In addition, our pinnae also physically impede the transmission of acoustic energy coming from behind by about 5 dB especially for frequencies between 2 kHz and 8 kHz, leading to a more favorable listening condition than the target back condition (Blauert, 1996; Kuk et al., 2013). Lastly, auditory spatial bisection, a measure of auditory spatial representation exploration, performs better in the frontal space than in the back space (Aggius-Vella et al., 2018; 2020). Auditory spatial bisection refers to a process where a listener attempts to determine the middle point between two sounds in space.

There was greater inter-subject variability in the back condition. The difference between the maximum and minimum hit rates for the target color and number was 40 (99 minus 59) and 47 (99 minus 52), respectively, showing a significant normal distribution. This variability could be attributed to some participants' head movements since it is known that front-back confusion can be resolved by such movements (Brimijoin & Akeroyd, 2012; Perrett & Noble, 1997; Wightman & Kistler, 1999). This ambiguity arises mainly because the interaural level differences (ILDs) and interaural time differences (ITDs) are identical for sound sources in front of and behind the listener. However, head movements can alleviate this confusion as they cause ILDs and ITDs to vary differently depending on whether the sound source is in front of or behind the listener. Essentially, by moving our heads, we alter the relative position of our ears to the sound source, which changes the sound's arrival time and level at each ear, aiding in a more accurate location discernment.

To prevent the head movement effect during the test, our experimental protocol required participants to fix their gaze on the central cross mark while performing the selective speech identification task. Although previous studies suggest that a small amount of head movement (5-35 degrees) would be sufficient to offset front-back ambiguity, it is unlikely in our case since participants' performance and activity were visually and electrophysiologically monitored. Notably, out of all the participants, 6 had hit rates above 90, and 4 of these were male participants, while the total number of male participants was 5. Therefore, any factor altering binaural and monaural cues, including head size, body size, etc., could potentially affect front-back confusion. Moreover, the alignment of the pinna height with the center of the loudspeaker could also influence results. For example, because of a non-adjustable chair used in this experiment, the position of participants' pinna might not be accurately aligned with the center of the loudspeaker.

5.2. Neural correlates of auditory attention: Alpha-band power

5.2.a. Reliable alpha peak frequency

The currently designed listening task allowed participants to direct their spatial selective attention to a given target position, eliciting the oscillatory alpha band activity with an averaged alpha peak frequency of 9.93 Hz as seen in Fig. 6. The grand average alpha frequency is very consistent with 10 Hz, known as the approximate mean of alpha frequency in neuroscience studies (Klimesch, 2012). Like the alpha peak frequency range from 9.32 to 11.19 Hz observed by Deng et al. (2020), it showed individual variances (95% CI: 9.19 – 10.67) in the peak alpha frequency, but all peaks from listeners were within the predefined alpha frequency range (8 – 13Hz).

Some prior studies used peak alpha frequency with ± 1 Hz for further analysis to avoid a skew estimate of alpha power (Bonacci et al., 2020; Deng et al., 2020), but we used the fixed alpha range to avoid inclusion of energy out of the alpha band if the peak frequency is at the edges of the band range as well as multiple peaks issue. Various frequency bands exhibit unique task-linked reactivity and topographical patterns. The lower alpha ERD, typically in the 7–10 Hz range, is diffusely distributed across the scalp. Although its functional interpretation, potentially associated with broad attentional demands, remains somewhat nebulous. In contrast, the upper alpha

ERD, within approximately 10 – 13.5 Hz, demonstrates a more localized distribution and a discernible relationship with semantic processing demands (Klimesch et al., 2007). Additionally, rises in the alpha peak frequency have been associated with cognitive burden (Haegens et al., 2014) and the functioning of working memory (Klimesch, 2012). Among the participants, all 11 exhibited a peak alpha frequency exceeding 10 Hz, which appears to be a consequence of executing the semantic processing demand, including the informational masking effect, while performing the selective speech identification task.

5.2.b. Topographical distribution of alpha band power across target position

The topographical map in Fig.7 showed robust alpha power distribution over parieto-occipital cortex across target positions. These enhanced alpha power with attentional focus has been considered a neural correlate of auditory attention by previous studies (Bonacci et al., 2020; Deng et al., 2020; Deng, Choi, et al., 2019; Dimitrijevic et al., 2019; Klatt et al., 2018; Wostmann et al., 2019). The present distribution of alpha power in the parietal area likely stems from the parietal cortex, part of what is known as the "dorsal attention network (Wostmann et al., 2015). It also may reflect the role of the frontoparietal attentional control network (also known as the frontoparietal control network or FPCN), a cognitive system within the brain that is involved in high-level cognitive functions, including attention regulation, decision-making, and integrating information from different sources. The dorsal surface of the human frontal and parietal lobes forms a network that plays a vital role in the selection of sensory content through attention (Ptak, 2012). In addition to the parieto-occipital lobe, we detected minor alpha activities around the frontal cortex. Consequently, we also included this cortex in our analysis, given its role as part of the attentional network. Misselhorn et al. (2019) contended that alpha oscillations in the frontal region denote the source of top-down control that modulates perceptual gains, while fluctuations in parietal alpha oscillations are linked with the shifting of sensory focus among different senses. Although it was expected that, contingent on the specific task demands, we might observe alpha suppression in the frontal brain regions, especially during the processing of semantic information (Klimesch et al., 2007), there was no significant difference in the frontal normalized alpha power between target positions. The pattern of the individual frontal alpha power is too complex to interpret. Given that Deng et al. (2020) discovered that the lateralization of parietal alpha monotonically varies with shifts in the focus of auditory spatial attention, we too expected a noticeable change in the topographical distribution of alpha power across four different target positions. However, only minor variations were observed within a similar pattern.

5.2.c. Parieto-occipital alpha power modulation.

Based on the results of our behavioral tests, the level of listening difficulty varied across different target positions. The target position at the back was found to be the most challenging condition, followed by the front condition, with the lateral condition being the least demanding. Our EEG test revealed that parieto-occipital alpha power was significantly modulated by target position even though it did not reach the statistical significance of the pairwise comparison. However, our alpha power modulation appeared not to vary with the listening difficulty across the target position, but to show mixed and contradictory results. It was not surprising there was no difference in PO alpha between the left target and the right target condition, aligning with the behavioral results. However, there was a contradiction where the alpha power for the front-back condition which was considered a more effortful condition than the lateral listening condition was lower than the power for the lateral condition while the alpha activities were more enhanced in the back condition regarded as the most challenging condition compared to the front condition. This implies that listening difficulty alone can't fully explain the changes in alpha power across target positions.

When it comes to the relationship between alpha oscillation power and listening difficulty, results have been equivocal since previous studies have shown contradictory outcomes with varying listening demand. Dimitrijevic et al. (2017) reported a negative correlation between alpha power and digit-in-noise (DIN) performance. Furthermore, alpha power decreased with the increase in acoustic details (temporal fine structure) and predictiveness (Wostmann et al., 2015). On the contrary, Miles et al. (2017) observed a reduction in alpha power in demanding listening circumstances, where they manipulated the spectral content of the signal using noise vocoding and assessed speech intelligibility. Also, Seifi Ala et al. (2020) showed a decrease in EEG alpha power in the parietal lobe under conditions with a lower signal-to-noise ratio (SNR). Furthermore, some studies have found an "inverted U-shaped" pattern in the alpha band, which has been linked with listeners' tendency to "give up" as situations become progressively more challenging. (Decruy, Lesenfants, et al., 2020; Petersen et al., 2015). These inconsistent observations accentuate the complexities to account for variations in alpha power while listening.

Several theories have been proposed to explain alpha power activity. The idea that large amplitudes of synchronized alpha activity reflect a brain state of reduced information processing is consistent with the concepts of 'idling' or 'nil working' (Klimesch, 2012; Klimesch et al., 2007). Synchronized alpha activity reflects a specific type of inhibition that can be described in terms of 'top-down control.' Alpha power increases in the brain regions responsible for processing distractors to suppress them. On the other hand, alpha power decreases in the brain areas dealing with relevant information, thereby facilitating their enhancement. This dual mechanism has been well supported by the hemispheric lateralization of alpha power in a dichotic listening task (Klatt et al., 2020; Poch et al., 2017). Furthermore, a comprehensive understanding of alpha modulation has been suggested, based on anatomically separate alpha oscillation (Schneider et al., 2022).

Based on this, in our front-back listening scenario, both the target and distracting signals reach and then go through the brain in the same way. So, there is likely to be an overlap between target enhancement and distractor suppression in similar brain regions. If so, which one ends up being dominant? Our data seemed to depend on what sound is being presented in the frontal azimuth, suggesting a dominant effect of auditory frontal space. Specifically, when the target signal was presented in front of listeners, it yielded decreases in alpha for the enhancement. On the other hand, when the distractor comes from the front, there's an increase in alpha power for suppression. The concept of the directional effect is different from the idea suggested by Wostmann et al. (2019) that both mechanisms independently induce lateralization of alpha power. Moreover, the cortical processes involved can vary depending on the nature of the speech material or the manner in which it is presented, and how a task is performed (Seifi Ala et al., 2020). The distinct patterns of alpha activity could be attributed to our cognitive and auditory systems responding differently to both informational masking and uncertainty about spatial target location, caused by front/back confusion. Taken together, in addition to existing mechanism theories, it might be necessary to consider how our brain responds depending on the location of the sound to interpret the neural correlates and behavioral consequences in a meaningful way.

5.2.d. Lateralization of Alpha power

This study detected a minor yet significant lateralized alpha power in the lateral listening condition. Differences in scalp topography for both the left and right target positions can be observed, referred to as AMI, unveiling disparities in positive and negative normalized alpha power between the two hemispheres in parieto-occipital electrode clusters. When calculating the AMI, prior studies subtract the values of two conditions and then divide by the overall alpha power (Deng, Choi, et al., 2019; Wostmann et al., 2019; Wostmann et al., 2016) . However, in this study, we simply calculated the difference using individual data already standardized as spatial z-scores. This approach aligns with the methodology employed in a research by Deng et al. (2020). Based on prior research demonstrating that parietal alpha power typically decreases in the hemisphere contralateral to the attended stimuli and/or increases in the ipsilateral hemisphere (Lim et al., 2015; Wostmann et al., 2016), the significance of a one-tailed test can be affirmed. Even though the lateralized response appears slight, individual data largely supported it with a similar pattern. This parietal alpha lateralization is presumed to represent the suppression of irrelevant information in the parietal lobe contralateral to the direction of distractor as well as the enhancement of relevant information in the lobe contralateral to the direction of attention (Ikkai et al., 2016; Schneider et al., 2022).

In contrast, the AMI did not exhibit hemispheric alpha lateralization in the parieto-occipital (PO) cluster for the front-back condition, showing considerable variance instead. This result aligns with one aspect of our hypothesis, where we posited the absence of hemispheric lateralization due to the symmetrically binaural stimulation routed from the front and back. The design of this study aimed to determine whether alpha power would rise to suppress noise or exhibit event-related desynchronization (ERD) to amplify the processed target signal when the target speech and distractor are concurrently presented from the front and back or vice versa. The observed results indicated that higher alpha power was induced when participants' attention was oriented towards the target sound emitted from the rear than the front, as depicted by the predominantly blue color on the topographical map without alpha lateralization (See Fig. 9).

Previous studies have reported a diminishing effect of aging on alpha lateralization (Dahl et al., 2019; Hong et al., 2015; Leenders et al., 2018). Furthermore, alpha was not lateralized across parietal sensors in the hearing-impaired group (Bonacci et al., 2019). Our participants who were young with normal hearing should be independent of these effects. Additionally, without a rear loudspeaker, Wostmann et al. (2019) utilized an experimental paradigm: 1) target selection in the front while ignoring a lateral distractor and 2) target selection in the left or right while ignoring a distractor in the front. They discovered that both inhibition of the underlying cortical tissue and the release of inhibition independently result in alpha power lateralization. While our experiment design shares similarities in the presentation of both target and distractor from the front, the involvement of the lateral side in their study ultimately induced alpha lateralization. As a result, our findings suggest that the front-back listening configuration does not lead to alpha lateralization.

5.3. Neural Correlates of Auditory Attention: Neural Speech Tracking

5.3.a. Time Course Analysis of Neural Speech Tracking Function

We obtained the neural speech tracking function through cross-correlation between the filtered EEG data and the first derivatives of the speech envelope. This function reflects activity that is phase-locked to acoustic transients in the attended speech stream, with the selectivity or strength of this tracking process being modulated by selective attention. Figure 10 depicts the time course of the grand average neural speech tracking, indicating both positive and negative deflections. The peak correlation coefficient represents the level of similarities between both variables.

When compared with peak latencies from previous studies: N1-P2 at 124 - 208 ms (Aiken & Picton, 2008), 150 – 250 ms (Petersen et al., 2017), and 136 – 240 ms (Mirkovic et al., 2019), our components' peak latencies on the time lag appeared to be delayed. Such a difference might result from experimental protocol discrepancies between studies. We utilized short sentences from the coordinate response measure corpus as stimuli, but ongoing narratives have been largely used in prior research on speech tracking. Indeed, there were minor peaks prior to the latencies of the defined peaks, but we chose the most prominent component peaks for the ultimate purpose of cross-correlation analysis. Our determined latency of N1-P2 corresponding to the to-be-attended speech envelope is 301 – 386 ms. In addition to N1-P2, the subsequent robust N2-P3 component was included for analysis. At the N1-P2-N2-P3 latency, peak components for the four different conditions occurred with similar temporal characteristics. Despite the predefined latencies ranges from the time course of the grand average neural speech tracking, it was complex to find individual component peaks because of individual peak variances. As alternatives, some studies used the mean of peak coefficients within a pre-defined range (Holtze et al., 2021), and another took an evoked spectral power from the cross-correlation function (Hambrook & Tata, 2019).

5.3.b. Neural speech tracking for to-be-attended speech envelope

It turns out that the P3 coefficient for the left target condition was significantly greater than P3 for both the front and back conditions, at least on a one-tailed test. The P3 for the back is slightly lower than the one for the front but the difference is not significant. With significant main effects of target position and stimuli as well as consistent individual data, we found that the trend of the P3 coefficient most closely reflected the behavioral test outcome among the peak components. In comparison with the P3 component, N1 and P2 did not show any significant main effect of target position, and N2 showed significant but smaller differences on average in cross-correlation coefficients between the front-back configuration and left-right configuration.

Our study indicated that the level of neural speech tracking (neural entrainment) varied with the target positions associated with listening difficulty. However, it revealed that there was no difference in neural speech tracking coefficients between the front and back conditions even though the behavioral listening task outcome showed a significant difference between them. When the target speech and distractor were presented at 0 dB TMR in our experiment, the front condition was expected to have a positive TMR while the back condition was assumed to have a negative TMR due to the pinna shadow effect. Petersen et al. (2017) showed significantly lower speech tracking at – 4 dB SNR of speech reception threshold (SRT) at 80% compared to 0 dB SRT80 and +4 dB SRT80. Hjortkjaer et al. (2020) also found that an increase in background noise from 10 dB SNR to 0 dB SNR decreased cortical speech envelope entrainment. In regard to the SNR, our data in front and back condition are not consistent with previous studies.

5.3.c. Neural speech tracking for to-be-attended versus to-be-ignored envelope

The influence of attention on neural speech tracking was quantified by computing the difference in the cross-correlation coefficient between the to-be-attended and to-be-ignored speech envelope (i.e., attended - ignored) for each participant and target condition (Petersen et al., 2017). We expected that the coefficient gap would reflect listening difficulties and intrusion errors across different target positions. However, the current speech tracking experiment revealed no significant difference in peak coefficients between the target speech and the distractor, resulting in a non-significant interaction (target position * stimulus) for all components. Nonetheless, the coefficient gaps for the P3 across four different target positions appeared to have a pattern similar to the hit rates outcome pattern (See Fig. 14-C). Particularly, the negative coefficient gap for the P3 in the back condition might reflect an increased number of intrusion errors compared to other conditions. This implies a potential applicability of the coefficient gap to studies on front-back spatial hearing, but further study is needed for the validity and reliability of the analysis.

The lack of difference in the coefficient gap between the target speech and distractor could stem from the similarity between them. The Coordinate Response Measure (CRM) corpus materials we used share the same sentence structure, temporal onset timing, speech rate, and number of syllables (Bolia et al., 2000; Eddins & Liu, 2012). This similarity is in fact why we employed CRM sentences, intending to effectively induce information masking in the task. However, their CRM speech envelopes are likely too similar to allow for any significant distinction in neural speech tracking analysis, irrespective of auditory attention and listening conditions.

5.4. Correlation between behavioral and EEG data

In this study, we examined the relationship between behavioral responses and EEG neural activity in a variety of conditions. The results highlighted some intriguing connections, demonstrating that the number of behavioral hit responses to target color and number were positively correlated with normalized parieto-occipital alpha power and with the P3 cross-correlation coefficients. This suggests that these specific EEG neural activities may reflect spatial auditory attention processing. Despite the relatively small coefficients of determination, regression analyses further illustrated the potential influence of normalized alpha power and the P3 coefficient on hit rates. Both these variables were significant predictors for hit rates related to both words, lending credence to the role of these specific EEG activities in cognitive hearing performance. However, given the low coefficients of determination and the lack of correlation in the front, back, and front plus back conditions, apart from the overall collapsed condition, we need to exercise caution to avoid overinterpreting these significant correlations.

Note that there was no significant correlation between alpha power and neural speech tracking responses, suggesting that these two cognitive phenomena might operate independently. Alpha power is modulated by auditory attention for the purpose of either enhancement or suppression over some relevant brain area whereas speech tracking acts as an entrainment to a sound of interest. Instead, a prior study proposed that the combination of neural tracking and alpha power lateralization may provide a benefit for auditory attention detection in cocktail party scenarios (Drgas et al., 2021). An intriguing observation was the significant correlation between frontal alpha activity and parieto-occipital alpha energies, which might implicate a concerted interplay between frontal and parieto-occipital regions during these tasks (Keitel et al., 2017; Ptak, 2012). In conclusion, these findings extend our understanding of the link between behavioral and EEG data, offering a foundation for subsequent investigations, particularly to understand the role of alpha power and neural speech tracking coefficients in the spatial selective speech identification task.

5.5. Limitation

As previously mentioned, the Coordinate Response Measure (CRM) sentences have a high degree of structural and informational similarity, which could make the distinction between target speech from the competing speech in neural speech tracking analysis challenging. While the similarity to the target amplifies informational masking and front-back confusion, the CRM sentences may not be suitable as a tool for speech tracking tests if they cannot result in dynamics between coefficients due to auditory attention. The second limitation pertains to the experimental protocol, in which sentences were randomly presented from a pool of 256 sentences for each talker during the tests. While this doesn't impact the alpha analysis, it introduces variance in the time lag latency of each trial's peak point in the cross-correlation analysis of neural speech tracking. In our study, we used short structured sentences instead of narratives, meaning that the averaging process was applied over multiple short trials, resulting in the generation of multiple peaks and making clear observation difficult. If all participants were to go through the same sentence trial set, responses could be more consistent, which would ease comparison. The third limitation is whether we were able to match precisely our pinna height with the center of the loudspeaker. As we used non-adjustable chairs for the experiment, we can't guarantee exact alignment to every participant. We attempted to adjust as much as possible, such as adjusting sitting postures, but there's a high probability that participants, especially taller males, were positioned differently than the ideal straight-line alignment. In the future, the exact alignment needs to be controlled. Lastly, for our alpha analysis, this study focused on topographical alpha power distribution. However, the EEG measurement has a relatively lower spatial resolution as compared to an MEG test while it is good at temporal resolution. There might be a limitation to examining the topographic map of alpha power and the distribution change.

5.6. Future direction

Understanding the underlying mechanism of spatial selective auditory attention through a single listening scenario is challenging. We could benefit from introducing more conditions necessary for a front-back configuration comparison. For instance, the inclusion of conditions where the target and distractor coexist at the front and the back, as well as conditions where only the target talker is presented, could provide deeper insights into the attentional modulation of auditory processing in the front-back spatial hearing by analyzing and comparing neural activities under each circumstance. Second, we could investigate the inter-subject variability in front-back listening. We could identify a neural signature such as anatomically-specific alpha oscillation in high performers compared to low performers in the front-back spatial hearing test. Conversely, we could study the significance of variance in neural correlates of auditory attention within ranges of behaviorally similar performance. Next, the subject group could be expanded to include musically trained individuals and clinical populations, such as the elderly and individuals with hearing impairments. Taken together, these measures would enhance our understanding of how auditory attention functions in spatial selective hearing including the rear space of auditory scene. Additionally, it could provide a foundational reference for developing neuro-steered hearing aids and cognitive hearing interventions.

6. CONCLUSION

This study addressed the inconsistent or contradictory alpha power modulation across the four different target positions in perspective of listening difficulty. We observed an increase in alpha power in both the most demanding (back condition) and the least demanding conditions (left and right). These findings cannot be simply interpreted using existing suppression and enhancement mechanism theories. Thus, in addition to existing mechanism theories, it might be necessary to consider how our brain responds depending on the location of the sound of interest in order to interpret the neural correlates and behavioral consequences in a meaningful way. This study also suggests the potential applicability of neural speech tracking in studies on spatial selective hearing and the need for more comprehensive experimental designs that are capable of capturing the full scope of brain processes involved in spatial selective hearing. In conclusion, while our findings may raise more questions than answers, they are instrumental in driving the discourse in the field forward. By redefining the framework of spatial selective hearing, we open a multitude of possibilities for further investigation and discovery.

7. REFERENCES

Addleman, D. A., & Jiang, Y. V. (2019). Experience-Driven Auditory Attention. *Trends Cogn Sci*, *23*(11), 927-937. https://doi.org/10.1016/j.tics.2019.08.002

Aggius-Vella, E., Campus, C., & Gori, M. (2018). Different audio spatial metric representation around the body. *Sci Rep*, *8*(1), 9383. https://doi.org/10.1038/s41598-018-27370-9

Aggius-Vella, E., Kolarik, A. J., Gori, M., Cirstea, S., Campus, C., Moore, B. C. J., & Pardhan, S. (2020). Comparison of auditory spatial bisection and minimum audible angle in front, lateral, and back space. *Sci Rep*, *10*(1), 6279. https://doi.org/10.1038/s41598-020-62983-z

Aiken, S. J., & Picton, T. W. (2008). Human cortical responses to the speech envelope. *Ear Hear*, *29*(2), 139-157. https://doi.org/10.1097/aud.0b013e31816453dc

Alain, C., Arnott, S. R., & Picton, T. W. (2001). Bottom-up and top-down influences on auditory scene analysis: evidence from event-related brain potentials. *J Exp Psychol Hum Percept Perform*, *27*(5), 1072-1089. https://doi.org/10.1037//0096-1523.27.5.1072

Allen, K., Carlile, S., & Alais, D. (2008). Contributions of talker characteristics and spatial location to auditory streaming. *J Acoust Soc Am*, *123*(3), 1562-1570. https://doi.org/10.1121/1.2831774

Baars, B. J., & Gage, N. M. (2010). Chapter 8 - Consciousness and attention. In B. J. Baars & N. M. Gage (Eds.), *Cognition, Brain, and Consciousness (Second Edition)* (pp. 238-303). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-375070-9.00008-5

Banerjee, S., Snyder, A. C., Molholm, S., & Foxe, J. J. (2011). Oscillatory alpha-band mechanisms and the deployment of spatial attention to anticipated auditory and visual target locations: supramodal or sensory-specific control mechanisms? *J Neurosci*, *31*(27), 9923-9932. https://doi.org/10.1523/JNEUROSCI.4660-10.2011

Blauert, J. (1996). *Spatial hearing : the psychophysics of human sound localization* (Rev. ed.). MIT Press. https://doi.org/https://doi.org/10.7551/mitpress/6391.001.0001

Bolia, R. S., Nelson, W. T., Ericson, M. A., & Simpson, B. D. (2000). A speech corpus for multitalker communications research. *J Acoust Soc Am*, *107*(2), 1065-1066. https://doi.org/10.1121/1.428288

Bonacci, L. M., Bressler, S., & Shinn-Cunningham, B. G. (2020). Nonspatial Features Reduce the Reliance on Sustained Spatial Auditory Attention. *Ear Hear*. https://doi.org/10.1097/AUD.0000000000000879

Bonacci, L. M., Dai, L., & Shinn-Cunningham, B. G. (2019). Weak neural signatures of spatial selective auditory attention in hearing-impaired listeners. *J Acoust Soc Am*, *146*(4), 2577. https://doi.org/10.1121/1.5129055

Brimijoin, W. O., & Akeroyd, M. A. (2012). The role of head movements and signal spectrum in an auditory front/back illusion. *Iperception*, *3*(3), 179-182. https://doi.org/10.1068/i7173sas

Brodbeck, C., Jiao, A., Hong, L. E., & Simon, J. Z. (2020). Neural speech restoration at the cocktail party: Auditory cortex recovers masked speech of both attended and ignored speakers. *PLoS Biol*, *18*(10), e3000883. https://doi.org/10.1371/journal.pbio.3000883

Brodbeck, C., & Simon, J. Z. (2020). Continuous speech processing. *Curr Opin Physiol*, *18*, 25-31. https://doi.org/10.1016/j.cophys.2020.07.014

Brungart, D. S. (2001). Informational and energetic masking effects in the perception of two simultaneous talkers. *J Acoust Soc Am*, *109*(3), 1101-1109. https://doi.org/10.1121/1.1345696

Brungart, D. S., & Simpson, B. D. (2007). Effect of target-masker similarity on across-ear interference in a dichotic cocktail-party listening task. *J Acoust Soc Am*, *122*(3), 1724. https://doi.org/10.1121/1.2756797

Buchtel, H. A., & Butter, C. M. (1988). Spatial attentional shifts: implications for the role of polysensory mechanisms. *Neuropsychologia*, *26*(4), 499-509. https://doi.org/10.1016/0028-3932(88)90107-8

Cai, Y., Chen, G., Zhong, X., Yu, G., Mo, H., Jiang, J., Chen, X., Zhao, F., & Zheng, Y. (2018). Influence of Audiovisual Training on Horizontal Sound Localization and Its Related ERP Response. *Front Hum Neurosci*, *12*, 423. https://doi.org/10.3389/fnhum.2018.00423

Caporello Bluvas, E., & Gentner, T. Q. (2013). Attention to natural auditory signals. *Hear Res*, *305*, 10-18. https://doi.org/10.1016/j.heares.2013.08.007

Carette, E., Van den Bogaert, T., Laureyns, M., & Wouters, J. (2014). Left-right and front-back spatial hearing with multiple directional microphone configurations in modern hearing aids. *J Am Acad Audiol*, *25*(9), 791-803. https://doi.org/10.3766/jaaa.25.9.2

Cherry, E. C. (1953). Some Experiments on the Recognition of Speech, with One and with Two Ears. *J Acoust Soc Am*, *25*(5), 975-979. https://doi.org/10.1121/1.1907229

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

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

Crosse, M. J., Di Liberto, G. M., Bednar, A., & 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. https://doi.org/10.3389/fnhum.2016.00604

Crosse, M. J., Zuk, N. J., Di Liberto, G. M., Nidiffer, A. R., Molholm, S., & Lalor, E. C. (2021). Linear Modeling of Neurophysiological Responses to Speech and Other Continuous Stimuli: Methodological Considerations for Applied Research. *Front Neurosci*, *15*, 705621. https://doi.org/10.3389/fnins.2021.705621

Dahl, M. J., Ilg, L., Li, S. C., Passow, S., & Werkle-Bergner, M. (2019). Diminished pre-stimulus alpha-lateralization suggests compromised self-initiated attentional control of auditory processing in old age. *Neuroimage*, *197*, 414-424. https://doi.org/10.1016/j.neuroimage.2019.04.080

Dai, L., Best, V., & Shinn-Cunningham, B. G. (2018). Sensorineural hearing loss degrades behavioral and physiological measures of human spatial selective auditory attention. *Proc Natl Acad Sci U S A*, *115*(14), E3286-E3295. https://doi.org/10.1073/pnas.1721226115

Darwin, C. J., & Hukin, R. W. (2000). Effectiveness of spatial cues, prosody, and talker characteristics in selective attention. *J Acoust Soc Am*, *107*(2), 970-977. https://doi.org/10.1121/1.428278 de Boer, J., & Thornton, A. R. (2007). Effect of subject task on contralateral suppression of click evoked otoacoustic emissions. *Hear Res*, *233*(1-2), 117-123. https://doi.org/10.1016/j.heares.2007.08.002

Decruy, L., Lesenfants, D., Vanthornhout, J., & 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*(5), 3375-3393. https://doi.org/10.1111/ejn.14753

Decruy, L., Vanthornhout, J., & Francart, T. (2019). Evidence for enhanced neural tracking of the speech envelope underlying age-related speech-in-noise difficulties. *J Neurophysiol*, *122*(2), 601-615. https://doi.org/10.1152/jn.00687.2018

Decruy, L., Vanthornhout, J., & Francart, T. (2020). Hearing impairment is associated with enhanced neural tracking of the speech envelope. *Hear Res*, *393*, 107961. https://doi.org/10.1016/j.heares.2020.107961

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J Neurosci Methods*, *134*(1), 9-21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Deng, Y., Choi, I., & Shinn-Cunningham, B. (2020). Topographic specificity of alpha power during auditory spatial attention. *Neuroimage*, *207*, 116360. https://doi.org/10.1016/j.neuroimage.2019.116360

Deng, Y., Choi, I., Shinn-Cunningham, B., & Baumgartner, R. (2019). Impoverished auditory cues limit engagement of brain networks controlling spatial selective attention. *Neuroimage*, *202*, 116151. https://doi.org/10.1016/j.neuroimage.2019.116151

Deng, Y., Reinhart, R. M., Choi, I., & Shinn-Cunningham, B. G. (2019). Causal links between parietal alpha activity and spatial auditory attention. *Elife*, *8*. https://doi.org/10.7554/eLife.51184

Dimitrijevic, A., Smith, M. L., Kadis, D. S., & Moore, D. R. (2017). Cortical Alpha Oscillations Predict Speech Intelligibility. *Front Hum Neurosci*, *11*, 88. https://doi.org/10.3389/fnhum.2017.00088

Dimitrijevic, A., Smith, M. L., Kadis, D. S., & Moore, D. R. (2019). Neural indices of listening effort in noisy environments. *Sci Rep*, *9*(1), 11278. https://doi.org/10.1038/s41598-019-47643-1

Ding, N., & Simon, J. Z. (2012). Emergence of neural encoding of auditory objects while listening to competing speakers. *Proc Natl Acad Sci U S A*, *109*(29), 11854-11859. https://doi.org/10.1073/pnas.1205381109

Drgas, S., Blaszak, M., & Przekoracka-Krawczyk, A. (2021). The Combination of Neural Tracking and Alpha Power Lateralization for Auditory Attention Detection. *J Speech Lang Hear Res*, *64*(9), 3603-3616. https://doi.org/10.1044/2021_JSLHR-20-00608

Eddins, D. A., & Liu, C. (2012). Psychometric properties of the coordinate response measure corpus with various types of background interference. *J Acoust Soc Am*, *131*(2), EL177-183. https://doi.org/10.1121/1.3678680

Etard, O., & Reichenbach, T. (2019). Neural Speech Tracking in the Theta and in the Delta Frequency Band Differentially Encode Clarity and Comprehension of Speech in Noise. *J Neurosci*, *39*(29), 5750-5759. https://doi.org/10.1523/JNEUROSCI.1828-18.2019

Foxe, J. J., & Snyder, A. C. (2011). The Role of Alpha-Band Brain Oscillations as a Sensory Suppression Mechanism during Selective Attention. *Front Psychol*, *2*, 154. https://doi.org/10.3389/fpsyg.2011.00154

Fritz, J. B., Elhilali, M., David, S. V., & Shamma, S. A. (2007). Auditory attention--focusing the searchlight on sound. *Curr Opin Neurobiol*, *17*(4), 437-455. https://doi.org/10.1016/j.conb.2007.07.011

Fuglsang, S. A., Marcher-Rorsted, J., Dau, T., & Hjortkjaer, J. (2020). Effects of Sensorineural Hearing Loss on Cortical Synchronization to Competing Speech during Selective Attention. *J Neurosci*, *40*(12), 2562-2572. https://doi.org/10.1523/JNEUROSCI.1936-19.2020

Geirnaert, S., Vandecappelle, S., Alickovic, E., de Cheveigne, A., Lalor, E., Meyer, B. T., Miran, S., Francart, T., & Bertrand, A. (2021). Electroencephalography-Based Auditory Attention Decoding: Toward Neurosteered Hearing Devices. *IEEE Signal Processing Magazine*, *38*(4), 89-102. https://doi.org/10.1109/msp.2021.3075932

Haegens, S., Cousijn, H., Wallis, G., Harrison, P. J., & Nobre, A. C. (2014). Inter- and intra-individual variability in alpha peak frequency. *Neuroimage*, *92*(100), 46-55. https://doi.org/10.1016/j.neuroimage.2014.01.049

Hambrook, D. A., & Tata, M. S. (2019). The effects of distractor set-size on neural tracking of attended speech. *Brain Lang*, *190*, 1-9. https://doi.org/10.1016/j.bandl.2018.12.005

Hill, K. T., & Miller, L. M. (2010). Auditory attentional control and selection during cocktail party listening. *Cereb Cortex*, *20*(3), 583-590. https://doi.org/10.1093/cercor/bhp124

Hillyard, S. A., Hink, R. F., Schwent, V. L., & Picton, T. W. (1973). Electrical signs of selective attention in the human brain. *Science*, *182*(4108), 177-180. https://doi.org/10.1126/science.182.4108.177

Hjortkjaer, J., Marcher-Rorsted, J., Fuglsang, S. A., & Dau, T. (2020). Cortical oscillations and entrainment in speech processing during working memory load. *Eur J Neurosci*, *51*(5), 1279-1289. https://doi.org/10.1111/ejn.13855

Holt, L. L., Tierney, A. T., Guerra, G., Laffere, A., & Dick, F. (2018). Dimension-selective attention as a possible driver of dynamic, context-dependent re-weighting in speech processing. *Hear Res*, *366*, 50-64. https://doi.org/10.1016/j.heares.2018.06.014

Holtze, B., Jaeger, M., Debener, S., Adiloglu, K., & Mirkovic, B. (2021). Are They Calling My Name? Attention Capture Is Reflected in the Neural Tracking of Attended and Ignored Speech. *Front Neurosci*, *15*, 643705. https://doi.org/10.3389/fnins.2021.643705

Hong, X., Sun, J., Bengson, J. J., Mangun, G. R., & Tong, S. (2015). Normal aging selectively diminishes alpha lateralization in visual spatial attention. *Neuroimage*, *106*, 353-363. https://doi.org/10.1016/j.neuroimage.2014.11.019

Hopkins, K., & Moore, B. C. (2011). The effects of age and cochlear hearing loss on temporal fine structure sensitivity, frequency selectivity, and speech reception in noise. *J Acoust Soc Am*, *130*(1), 334-349. https://doi.org/10.1121/1.3585848

Ihlefeld, A., & Shinn-Cunningham, B. (2008). Spatial release from energetic and informational masking in a selective speech identification task. *J Acoust Soc Am*, *123*(6), 4369-4379. https://doi.org/10.1121/1.2904826

Ikkai, A., Dandekar, S., & Curtis, C. E. (2016). Lateralization in Alpha-Band Oscillations Predicts the Locus and Spatial Distribution of Attention. *PLoS One*, *11*(5), e0154796. https://doi.org/10.1371/journal.pone.0154796

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

Jensen, O., & Mazaheri, A. (2010). Shaping functional architecture by oscillatory alpha activity: gating by inhibition. *Front Hum Neurosci*, *4*, 186. https://doi.org/10.3389/fnhum.2010.00186

Jerde, T. A., Merriam, E. P., Riggall, A. C., Hedges, J. H., & Curtis, C. E. (2012). Prioritized maps of space in human frontoparietal cortex. *J Neurosci*, *32*(48), 17382-17390. https://doi.org/10.1523/JNEUROSCI.3810-12.2012

Kanai, K., Ikeda, K., & Tayama, T. (2007). The effect of exogenous spatial attention on auditory information processing. *Psychol Res*, *71*(4), 418-426. https://doi.org/10.1007/s00426-005-0024-4

Keitel, A., Ince, R. A. A., Gross, J., & Kayser, C. (2017). Auditory cortical delta-entrainment interacts with oscillatory power in multiple fronto-parietal networks. *Neuroimage*, *147*, 32-42. https://doi.org/10.1016/j.neuroimage.2016.11.062

Kerlin, J. R., Shahin, A. J., & Miller, L. M. (2010). Attentional gain control of ongoing cortical speech representations in a "cocktail party". *J Neurosci*, *30*(2), 620-628. https://doi.org/10.1523/JNEUROSCI.3631-09.2010

Kidd, G., Jr. (2017). Enhancing Auditory Selective Attention Using a Visually Guided Hearing Aid. *J Speech Lang Hear Res*, *60*(10), 3027-3038. https://doi.org/10.1044/2017_JSLHR-H-17-0071

Kidd, G., Jr., Arbogast, T. L., Mason, C. R., & Gallun, F. J. (2005). The advantage of knowing where to listen. *J Acoust Soc Am*, *118*(6), 3804-3815. https://doi.org/10.1121/1.2109187

Klatt, L. I., Getzmann, S., Begau, A., & Schneider, D. (2020). A dual mechanism underlying retroactive shifts of auditory spatial attention: dissociating target- and distractor-related modulations of alpha lateralization. *Sci Rep*, *10*(1), 13860. https://doi.org/10.1038/s41598-020-70004-2

Klatt, L. I., Getzmann, S., Wascher, E., & Schneider, D. (2018). The contribution of selective spatial attention to sound detection and sound localization: Evidence from event-related potentials and lateralized alpha oscillations. *Biol Psychol*, *138*, 133-145. https://doi.org/10.1016/j.biopsycho.2018.08.019

Klimesch, W. (2012). alpha-band oscillations, attention, and controlled access to stored information. *Trends Cogn Sci*, *16*(12), 606-617. https://doi.org/10.1016/j.tics.2012.10.007

Klimesch, W., Sauseng, P., & Hanslmayr, S. (2007). EEG alpha oscillations: the inhibition-timing hypothesis. *Brain Res Rev*, *53*(1), 63-88. https://doi.org/10.1016/j.brainresrev.2006.06.003

Koch, I., Lawo, V., Fels, J., & Vorlander, M. (2011). Switching in the cocktail party: exploring intentional control of auditory selective attention. *J Exp Psychol Hum Percept Perform*, *37*(4), 1140-1147. https://doi.org/10.1037/a0022189

Kong, Y. Y., Somarowthu, A., & Ding, N. (2015). Effects of Spectral Degradation on Attentional Modulation of Cortical Auditory Responses to Continuous Speech. *J Assoc Res Otolaryngol*, *16*(6), 783-796. https://doi.org/10.1007/s10162-015-0540-x

Kuk, F., Korhonen, P., Lau, C., Keenan, D., & Norgaard, M. (2013). Evaluation of a pinna compensation algorithm for sound localization and speech perception in noise. *Am J Audiol*, *22*(1), 84-93. https://doi.org/10.1044/1059-0889(2012/12-0043)

Kurthen, I., Galbier, J., Jagoda, L., Neuschwander, P., Giroud, N., & Meyer, M. (2021). Selective attention modulates neural envelope tracking of informationally masked speech in healthy older adults. *Hum Brain Mapp, 42*(10), 3042-3057. https://doi.org/10.1002/hbm.25415

Laffere, A., Dick, F., & Tierney, A. (2020). Effects of auditory selective attention on neural phase: individual differences and short-term training. *Neuroimage, 213*, 116717. https://doi.org/10.1016/j.neuroimage.2020.116717

Larson, E., & Lee, A. K. (2013). The cortical dynamics underlying effective switching of auditory spatial attention. *Neuroimage, 64,* 365-370. https://doi.org/10.1016/j.neuroimage.2012.09.006

Lawo, V., & Koch, I. (2014). Dissociable effects of auditory attention switching and stimulus-response compatibility. *Psychol Res, 78*(3), 379-386. https://doi.org/10.1007/s00426-014-0545-9

Leenders, M. P., Lozano-Soldevilla, D., Roberts, M. J., Jensen, O., & De Weerd, P. (2018). Diminished Alpha Lateralization During Working Memory but Not During Attentional Cueing in Older Adults. *Cereb Cortex, 28*(1), 21-32. https://doi.org/10.1093/cercor/bhw345

Lesenfants, D., Vanthornhout, J., Verschueren, E., Decruy, L., & Francart, T. (2019). Predicting individual speech intelligibility from the cortical tracking of acoustic- and phonetic-level speech representations. *Hear Res, 380,* 1-9. https://doi.org/10.1016/j.heares.2019.05.006

Lim, S. J., Wostmann, M., & Obleser, J. (2015). Selective Attention to Auditory Memory Neurally Enhances Perceptual Precision. *J Neurosci, 35*(49), 16094-16104. https://doi.org/10.1523/JNEUROSCI.2674-15.2015

Lin, G., & Carlile, S. (2015). Costs of switching auditory spatial attention in following conversational turn-taking. *Front Neurosci, 9,* 124. https://doi.org/10.3389/fnins.2015.00124

Maison, S., Micheyl, C., & Collet, L. (2001). Influence of focused auditory attention on cochlear activity in humans. *Psychophysiology, 38*(1), 35-40. https://www.ncbi.nlm.nih.gov/pubmed/11321619

Marrone, N., Mason, C. R., & Kidd, G., Jr. (2008). Evaluating the benefit of hearing aids in solving the cocktail party problem. *Trends Amplif, 12*(4), 300-315. https://doi.org/10.1177/1084713808325880

McAnally, K. I., & Martin, R. L. (2014). Sound localization with head movement: implications for 3-d audio displays. *Front Neurosci, 8,* 210. https://doi.org/10.3389/fnins.2014.00210

Mehraei, G., Shinn-Cunningham, B., & Dau, T. (2018). Influence of talker discontinuity on cortical dynamics of auditory spatial attention. *Neuroimage, 179,* 548-556. https://doi.org/10.1016/j.neuroimage.2018.06.067

Mesik, J., Ray, L., & Wojtczak, M. (2021). Effects of Age on Cortical Tracking of Word-Level Features of Continuous Competing Speech. *Front Neurosci, 15*, 635126. https://doi.org/10.3389/fnins.2021.635126

Miles, K., McMahon, C., Boisvert, I., Ibrahim, R., de Lissa, P., Graham, P., & Lyxell, B. (2017). Objective Assessment of Listening Effort: Coregistration of Pupillometry and EEG. *Trends Hear, 21*, 2331216517706396. https://doi.org/10.1177/2331216517706396

Mirkovic, B., Debener, S., Schmidt, J., Jaeger, M., & 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. https://doi.org/10.1016/j.heares.2019.04.005

Misselhorn, J., Friese, U., & Engel, A. K. (2019). Frontal and parietal alpha oscillations reflect attentional modulation of cross-modal matching. *Sci Rep*, *9*(1), 5030. https://doi.org/10.1038/s41598-019-41636-w

Obleser, J., & Kayser, C. (2019). Neural Entrainment and Attentional Selection in the Listening Brain. *Trends Cogn Sci*, *23*(11), 913-926. https://doi.org/10.1016/j.tics.2019.08.004

Obleser, J., Wostmann, M., Hellbernd, N., Wilsch, A., & Maess, B. (2012). Adverse listening conditions and memory load drive a common alpha oscillatory network. *J Neurosci*, *32*(36), 12376-12383. https://doi.org/10.1523/JNEUROSCI.4908-11.2012

Perrett, S., & Noble, W. (1997). The contribution of head motion cues to localization of low-pass noise. *Percept Psychophys*, *59*(7), 1018-1026. https://doi.org/10.3758/bf03205517

Petersen, E. B., Wostmann, M., Obleser, J., & Lunner, T. (2017). Neural tracking of attended versus ignored speech is differentially affected by hearing loss. *J Neurophysiol*, *117*(1), 18-27. https://doi.org/10.1152/jn.00527.2016

Petersen, E. B., Wostmann, M., Obleser, J., Stenfelt, S., & Lunner, T. (2015). Hearing loss impacts neural alpha oscillations under adverse listening conditions. *Front Psychol*, *6*, 177. https://doi.org/10.3389/fpsyg.2015.00177

Picou, E. M. (2020). MarkeTrak 10 (MT10) Survey Results Demonstrate High Satisfaction with and Benefits from Hearing Aids. *Semin Hear*, *41*(1), 21-36. https://doi.org/10.1055/s-0040-1701243

Poch, C., Carretie, L., & Campo, P. (2017). A dual mechanism underlying alpha lateralization in attentional orienting to mental representation. *Biol Psychol*, *128*, 63-70. https://doi.org/10.1016/j.biopsycho.2017.07.015

Presacco, A., Simon, J. Z., & Anderson, S. (2016). Evidence of degraded representation of speech in noise, in the aging midbrain and cortex. *J Neurophysiol*, *116*(5), 2346-2355. https://doi.org/10.1152/jn.00372.2016

Ptak, R. (2012). The frontoparietal attention network of the human brain: action, saliency, and a priority map of the environment. *Neuroscientist*, *18*(5), 502-515. https://doi.org/10.1177/1073858411409051

Reetzke, R., Gnanateja, G. N., & Chandrasekaran, B. (2021). Neural tracking of the speech envelope is differentially modulated by attention and language experience. *Brain Lang*, *213*, 104891. https://doi.org/10.1016/j.bandl.2020.104891

Rhodes, G. (1987). Auditory attention and the representation of spatial information. *Percept Psychophys*, *42*(1), 1-14. https://doi.org/10.3758/bf03211508

Rosen, S. (1992). Temporal information in speech: acoustic, auditory and linguistic aspects. *Philos Trans R Soc Lond B Biol Sci*, *336*(1278), 367-373. https://doi.org/10.1098/rstb.1992.0070

Ruggles, D., & 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*(3), 395-405. https://doi.org/10.1007/s10162-010-0254-z

Sach, A. J., & Bailey, P. J. (2004). Some characteristics of auditory spatial attention revealed using rhythmic masking release. *Percept Psychophys*, *66*(8), 1379-1387. https://doi.org/10.3758/bf03195005

Scharf, B., Quigley, S., Aoki, C., Peachey, N., & Reeves, A. (1987). Focused auditory attention and frequency selectivity. *Percept Psychophys*, *42*(3), 215-223. https://doi.org/10.3758/bf03203073

Schneider, D., Herbst, S. K., Klatt, L. I., & Wostmann, M. (2022). Target enhancement or distractor suppression? Functionally distinct alpha oscillations form the basis of attention. *Eur J Neurosci*, *55*(11-12), 3256-3265. https://doi.org/10.1111/ejn.15309

Schwartz, A. H., & Shinn-Cunningham, B. G. (2013). Effects of dynamic range compression on spatial selective auditory attention in normal-hearing listeners. *J Acoust Soc Am*, *133*(4), 2329-2339. https://doi.org/10.1121/1.4794386

Seifi Ala, T., Graversen, C., Wendt, D., Alickovic, E., Whitmer, W. M., & 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*(7), e0235782. https://doi.org/10.1371/journal.pone.0235782

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science*, *270*(5234), 303-304. https://doi.org/10.1126/science.270.5234.303

Shinn-Cunningham, B. G., & Best, V. (2008). Selective attention in normal and impaired hearing. *Trends Amplif*, *12*(4), 283-299. https://doi.org/10.1177/1084713808325306

Shinn-Cunningham, B. G., & Best, V. (2015). Auditory Selective Attention. In J. Fawcett, E. Risko, & A. Kingstone (Eds.), *The Handbook of Attention:* (pp. 100-117). The MIT Press. https://doi.org/10.7551/mitpress/10033.001.0001

Shomstein, S., & Yantis, S. (2006). Parietal cortex mediates voluntary control of spatial and nonspatial auditory attention. *J Neurosci*, *26*(2), 435-439. https://doi.org/10.1523/JNEUROSCI.4408-05.2006

Singh, G., Pichora-Fuller, M. K., & Schneider, B. A. (2008). The effect of age on auditory spatial attention in conditions of real and simulated spatial separation. *J Acoust Soc Am*, *124*(2), 1294-1305. https://doi.org/10.1121/1.2949399

Strauss, A., Wostmann, M., & Obleser, J. (2014). Cortical alpha oscillations as a tool for auditory selective inhibition. *Front Hum Neurosci*, *8*, 350. https://doi.org/10.3389/fnhum.2014.00350

Sussman, E. S. (2005). Integration and segregation in auditory scene analysis. *J Acoust Soc Am*, *117*(3 Pt 1), 1285-1298. https://doi.org/10.1121/1.1854312

Sussman, E. S. (2017). Auditory Scene Analysis: An Attention Perspective. *J Speech Lang Hear Res*, *60*(10), 2989-3000. https://doi.org/10.1044/2017_JSLHR-H-17-0041

Tervaniemi, M., & Hugdahl, K. (2003). Lateralization of auditory-cortex functions. *Brain Res Brain Res Rev*, *43*(3), 231-246. https://doi.org/10.1016/j.brainresrev.2003.08.004

Vanthornhout, J., Decruy, L., & Francart, T. (2019). Effect of Task and Attention on Neural Tracking of Speech. *Front Neurosci*, *13*, 977. https://doi.org/10.3389/fnins.2019.00977

Vanthornhout, J., Decruy, L., Wouters, J., Simon, J. Z., & Francart, T. (2018). Speech Intelligibility Predicted from Neural Entrainment of the Speech Envelope. *J Assoc Res Otolaryngol*, *19*(2), 181-191. https://doi.org/10.1007/s10162-018-0654-z

Welch, P. (1967). The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio and Electroacoustics*, *15*(2), 70-73. https://doi.org/10.1109/TAU.1967.1161901

Wetzel, N., & Schro¨ger, E. (2018). Auditory Attention in Children and Adults: A Psychophysiological Approach. In T. Lachmann & T. Weis (Eds.), *Invariances in Human Information Processing* (1st Ed. ed., pp. 223-248). Routledge. https://doi.org/https://doi.org/10.4324/9781315169903

Wightman, F. L., & Kistler, D. J. (1999). Resolution of front-back ambiguity in spatial hearing by listener and source movement. *J Acoust Soc Am*, *105*(5), 2841-2853. https://doi.org/10.1121/1.426899

Wisniewski, M. G., Thompson, E. R., & Iyer, N. (2017). Theta- and alpha-power enhancements in the electroencephalogram as an auditory delayed match-to-sample task becomes impossibly difficult. *Psychophysiology*, *54*(12), 1916-1928. https://doi.org/10.1111/psyp.12968

Wostmann, M., Alavash, M., & Obleser, J. (2019). Alpha Oscillations in the Human Brain Implement Distractor Suppression Independent of Target Selection. *J Neurosci*, *39*(49), 9797-9805. https://doi.org/10.1523/JNEUROSCI.1954-19.2019

Wostmann, M., Herrmann, B., Maess, B., & Obleser, J. (2016). Spatiotemporal dynamics of auditory attention synchronize with speech. *Proc Natl Acad Sci U S A*, *113*(14), 3873-3878. https://doi.org/10.1073/pnas.1523357113

Wostmann, M., Herrmann, B., Wilsch, A., & Obleser, J. (2015). Neural alpha dynamics in younger and older listeners reflect acoustic challenges and predictive benefits. *J Neurosci*, *35*(4), 1458-1467. https://doi.org/10.1523/JNEUROSCI.3250-14.2015

Wostmann, M., Lim, S. J., & Obleser, J. (2017). The Human Neural Alpha Response to Speech is a Proxy of Attentional Control. *Cereb Cortex*, *27*(6), 3307-3317. https://doi.org/10.1093/cercor/bhx074

Wostmann, M., Maess, B., & Obleser, J. (2021). Orienting auditory attention in time: Lateralized alpha power reflects spatio-temporal filtering. *Neuroimage*, *228*, 117711. https://doi.org/10.1016/j.neuroimage.2020.117711

Wu, C. T., Weissman, D. H., Roberts, K. C., & Woldorff, M. G. (2007). The neural circuitry underlying the executive control of auditory spatial attention. *Brain Res*, *1134*(1), 187-198. https://doi.org/10.1016/j.brainres.2006.11.088

Zion Golumbic, E. M., Ding, N., Bickel, S., Lakatos, P., Schevon, C. A., McKhann, G. M., Goodman, R. R., Emerson, R., Mehta, A. D., Simon, J. Z., Poeppel, D., & Schroeder, C. E. (2013). Mechanisms underlying selective neuronal tracking of attended speech at a "cocktail party". *Neuron*, *77*(5), 980-991. https://doi.org/10.1016/j.neuron.2012.12.037