Check for updates

Trends in Hearing
Volume 30: 1–18
© The Author(s) 2026
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/23312165261464723
journals.sagepub.com/home/tia
Sage

Interaural Time and Level Differences Contribute Differently to Source Segregation and Spatial Selection

Benjamin N. Richardson[1], Sahil Luthra[2], Jana M. Kainerstorfer[3], Barbara G. Shinn-Cunningham[1], and Christopher A. Brown[4]

Abstract

Interaural time and level differences are crucial in sound localization, yet their contributions to sound source segregation and spatial selection remain underspecified. Here, participants completed a spatial auditory selective attention task while we measured hemodynamic activity in the prefrontal cortex and superior temporal gyrus using functional near-infrared spectroscopy. Participants listened to a target sound stream and a simultaneous spatially separated speech stream or white noise masker. Sound streams were spatialized with either 50 μs interaural time differences (ITDs), 500 μs ITDs, naturally occurring interaural level differences (ILDs) from a non-individualized head-related transfer function (HRTF), or broadband 10 dB ILDs. Behavioral results revealed a stronger effect of spatial cues when the masker was speech. Error patterns differed in the two difficult conditions, small ITDs and natural ILDs: Small ITDs produced lower hit rates, while naturally occurring ILDs produced higher false alarm rates. Small ITDs led to greater activity in prefrontal cortex and activity in superior temporal gyrus that was lateralized, greater in the hemisphere contralateral to attentional focus, consistent with previous reports. These results suggest that natural ILDs alone support source segregation even if they are insufficient to cause large shifts in perceived lateralization, explaining high false alarm rates (confusions between target and distractor words). In contrast, small ITDs alone may be insufficient to segregate competing sources, leading to low hit and false alarm rates. Together, these results reveal differences in how ITDs and ILDs contribute to auditory scene analysis and spatial attention.

Keywords

auditory attention, spatial hearing, binaural processing, functional near-infrared spectroscopy

Received: 30 October 2025; Revised: 28 May 2026; Accepted: 16 June 2026

Introduction

In many typical sound environments, the signals reaching a listener's ears comprise a rich mixture of concurrent, spectro-temporally overlapping but independent sounds, often arriving from different directions (Cherry, 1953). To understand a sound in such settings, the brain must achieve at least two interrelated tasks: 1) *segregate* the sources in the mixture, forming distinct perceptual objects by determining what acoustic energy belongs to what source and 2) *select* which object to process in detail by focusing selective attention to filter out competing sources (Noyce et al., 2023; Shinn-Cunningham, 2008a). Many

[1]Neuroscience Institute, Carnegie Mellon University, Pittsburgh, PA, USA
[2]Department of Psychology, Stony Brook University, Stony Brook, NY, USA
[3]Department of Biomedical Engineering, Carnegie Mellon University, Pittsburgh, PA, USA
[4]Department of Communication Sciences and Disorders, University of South Florida, Tampa, FL, USA

Corresponding author:
Christopher A. Brown, Department of Communication Sciences and Disorders, University of South Florida, 3711 USF Citrus Drive, Tampa, FL, 33612, USA.
Email: cb43@usf.edu

Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (https://creativecommons.org/licenses/by-nc/4.0/) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (https://us.sagepub.com/en-us/nam/open-access-at-sage).

acoustic factors influence both segregation and selection, including pitch, temporal envelope cues, and spatial features, among others (Middlebrooks & Waters, 2020; Moore & Gockel, 2002). The current work examines how spatial cues influence segregation and spatial selection.

Spatial cues are relatively weak compared to spectrotemporal cues in grouping together simultaneous sound elements, but they play a larger role in "streaming" sounds across silent gaps (Ihlefeld & Shinn-Cunningham, 2008; Middlebrooks, 2017; Moore & Gockel, 2002). Perceived spatial separation can also guide spatial attention to a target in a mixture (Arbogast et al., 2002; Freyman et al., 1999; Kidd et al., 2010). Yet spatial cues have little influence if competing sounds have different and distinct spectrotemporal structure, such as when a woman speaks over a background of radio static. When spectrotemporal features alone are enough to support segregation and selection, spatial separation between competing sources has little effect on psychophysical outcomes (Culling & Summerfield, 1995; Darwin, 2007; Kubovy, 1988; Kubovy & Van Valkenburg, 2001; Noyce et al., 2023). Thus, the impact of spatial separation on auditory attention depends critically on the spectrotemporal properties of competing sounds.

The influence of spatial cues on segregation and selection helps explain spatial release from masking (SRM), wherein listeners are better at detecting and understanding a target in the presence of a spatially separated masker compared to when they are co-located (Freyman et al., 1999; Middlebrooks & Waters, 2020). The most important and robust spatial cues for SRM are interaural time and level differences (ITDs and ILDs). These cues help listeners to selectively attend to a target, though their specific contributions to segregation versus selection remain underspecified.

ITDs and ILDs have distinct effects on the signals reaching the ears, which may impact how they contribute to SRM. Acoustically, ITDs do not alter the target-to-masker ratio (TMR) in the mixture of signals reaching the ears; thus, any benefits from ITDs must arise from explicit spatial computations within the auditory pathway. In contrast, ILDs affect the TMR in the mixture of signals reaching each ear; typically, when a target and a masker have different ILDs, the TMR will be larger at the ear nearer the target, producing a "better ear" effect (Bronkhorst & Plomp, 1988; Glyde, Buchholz, Dillon, Best, et al., 2013). Perceptually, this TMR boost can improve target understanding even in the absence of perceived spatial separation or in monaural listening conditions by simply reducing energetic masking. Physiologically, a source containing ILDs is represented asymmetrically in auditory cortex including the superior temporal gyrus or STG), with a stronger representation in the hemisphere contralateral to the ear receiving the more intense signal (Higgins et al., 2017). A direct consequence of this cortical lateralization is that the neural representations of competing sounds with opposing ILDs show inherently less overlap than representations of sources containing only ITDs. This lateralization may support perceptual source segregation even in the absence of explicit spatial effects. These differences suggest that ITDs and ILDs may support SRM through different mechanisms.

The extent to which ILDs support SRM will also depend on the spectrotemporal properties of competing sounds. Natural ILDs are substantial only in high frequencies (Yost & Dye, 1988). Because high-frequency components of speech group with their corresponding low-frequency components (Best et al., 2007; Bregman, 1990; Darwin, 1997), natural ILDs may nevertheless help with perceptual segregation of competing speech. Yet when paired with uninformative or zero ITDs, the perceived location of speech containing only natural ILDs and no usable ITDs will be nearer midline, since ITDs at low and mid frequencies dominate lateralization (Best et al., 2007; Ellinger et al., 2017; Glyde, Buchholz, Dillon, Cameron, & Hickson, 2013; Wightman & Kistler, 1992). As a result, competing sources with only natural, opposing ILDs may be less likely to be perceived as coming from distinct locations, making source selection difficult. In contrast, broadband ILDs—applied uniformly across frequencies—can provide strong grouping and separation cues.

Magnifying ILDs to larger-than-natural magnitudes can enhance SRM in both normal-hearing listeners and cochlear implant (CI) users (Brown, 2014; Richardson et al., 2025). Without ILD magnification, CI users typically do not show significant SRM, likely because cochlear implants do not preserve ITDs and compress ILDs (Brown, 2018; Dorman et al., 2014; Gray et al., 2021). The current study seeks to clarify how ILDs contribute to SRM in normal hearing listeners in order to begin to identify how enhanced ILDs benefit CI users.

Functional neuroimaging studies reveal that spatial attention engages a visuo-spatial control network that includes prefrontal cortex (PFC; Kong et al., 2014; Michalka et al., 2015; Noyce et al., 2022) and enhances contralateral bias in the representation of sound in auditory cortex (Alho et al., 1998, 1999, 2003; Ciaramitaro et al., 2007; Jäncke & Shah, 2002; Yang & Mayer, 2014). Furthermore, previous functional near-infrared spectroscopy (fNIRS) studies show that PFC activity varies with cognitive demands, reflecting individual differences in speech intelligibility (Lawrence et al., 2018; Zhou, Burg, et al., 2022, Zhou, Burg, et al., 2022), perceived task difficulty (Zhou, Burg, et al., 2022; Zhou, Burg, et al., 2022), and spatial separation between target and masker sounds (Zhang et al., 2021a). Measures of effort, including hemodynamic activity in PFC, show a non-monotic relationship with task difficulty: Activity is larger when auditory attention is actively deployed compared to passive listening (Zhang et al., 2018), and reduced responses can reflect either very low or very high difficulty (Lawrence et al., 2018). In a similar spatial selective attention task to that used in the present study, PFC showed less activity when target and masker were dichotic than when lateralized with relatively small ITDs (Zhang et al., 2021a). This result could

reflect reduced demands on spatial attention when target and masker locations are more perceptually distinct (infinite ILDs compared to small ITDs); however, it could also reflect reduced effort expended to segregate sources lateralized with ILDs than with ITDs. Specifically, because neural representations are automatically lateralized for sources with ILDs, computational demands of segregation and selection may be reduced, resulting in decreased activation in PFC.

Previous investigations form a general picture of how ITDs and ILDs contribute to segregation and spatial selection. ITDs tend to dominate lateralization, and ILDs as they naturally occur are small at frequencies containing speech energy. ILDs do, however, result in the preferential representation of a sound in the contralateral auditory cortex. This separation of neural populations that represent competing sounds, along with top-down modulation from PFC, is required for successful selective attention. This has been explored using fNIRS imaging in both PFC and STG, but often separating competing sounds diotically (Zhou, Sobczak, et al., 2022) or using only ITDs (Zhang et al., 2018; 2021a). Here, we address this gap by comparing behavioral responses and neural activity in PFC and STG with either ITDs or ILDs during a spatial selective attention task.

In the current study, competing sounds were separated with small ITDs, large ITDs, natural ILDs, or broadband ILDs under conditions where the masker was either competing speech or white noise. Participants responded to target color words in one stream. We had several hypotheses about how the masker type (speech, noise) and spatial cues would interact to affect segregation and spatial selection in behavioral performance. First, we expected that a speech masker would more readily mask the target speech sound than a noise masker. Because of a lack of segregation cues in this case, participants would more heavily rely on spatial cues (Culling & Summerfield, 1995; Darwin, 2007; Kubovy & Van Valkenburg, 2001; Noyce et al., 2023): Larger spatial separation would engender better performance as indexed by hit rate and false alarm rate. Second, we hypothesized that conditions that provided small spatial cues in low frequencies would result in greater listening difficulty, reflected in poorer performance. That is, we expected worse behavioral performance with small ITD cues compared to large ITD cues, and worse performance for natural ILDs compared to broadband ILDs. Of particular interest is how these performance improvements might manifest with respect to hit rates and false alarms. For instance, if a change in spatial cues were to induce an increase in hit rates but also in false alarms, we would infer that listeners are able to accurately segregate the two streams (leading to enhanced hit rates) but that they struggle with selection (leading to them falsely respond to color words in the non-target stream). Accordingly, we view hit rates and false alarm rates in this task as indices of segregation and selection, respectively.

We further hypothesized that greater listening effort, as reflected in performance, would also manifest in greater activity in PFC. Specifically, we expected that small ITDs would drive greater PFC engagement relative to large ITDs, and natural ILDs than broadband ILDs. Finally, we predicted that STG activity would lateralize contralateral to the attended direction, consistent with previous reports (Alho et al., 1999, 2003; Ciaramitaro et al., 2007; Jäncke & Shah, 2002; Yang & Mayer, 2014).

Materials & Methods

Participants

Thirty native English speaking listeners participated in this experiment (27 female, 3 male, age 22.6 $\pm$ 4.45 S.D.). Although our sample was predominantly female and sampled from university students, we did not anticipate that there are any gender differences in performance on this psychometric task. All statistical analyses were conducted within-subject. All participants had normal hearing, defined as having pure-tone audiometric thresholds of $\leq$ 20 dB HL at octave spaced frequencies from 250 Hz to 8 kHz. All listeners gave written informed consent prior to participating in the study. All testing was administered according to the guidelines of the Institutional Review Board of the University of Pittsburgh.

Overview of the Task

Subjects completed a color-word detection task in a selective attention paradigm with two competing, spatially separated sound streams. Figure 1 shows a schematic of target and masker streams. The target stream consisted of temporally jittered words spoken by a single male talker. The competing stream was either speech from the same talker or white noise. Within an experimental block, target and masker sequences were spatialized such that they were perceived to originate from opposing hemifields, one on the left and the other on the right, assigned randomly on each block.

For the speech masker condition, target and masker sequences each consisted of 18 word tokens, which were presented in pairs (one target token and one masker token, see *Stimuli*). A word token pair occurred every 600 ms. Within each token pair, the two 300-ms-long tokens partially overlapped in time, with the second one starting 150 ms after the first. To ensure that listeners relied on spatial attention rather than temporal regularity, which token in a pair occurred first was randomly selected,

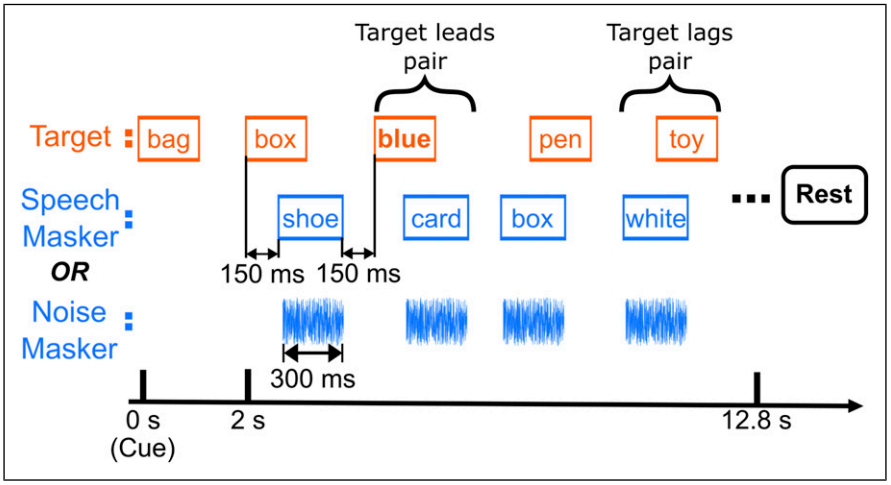


Figure 1. Schematic of a single block. Listeners first heard a cue word "bag" from the target stream. Then, they responded by button press to color words in the target stream (top row). The target stream was presented against either a speech or noise masker at the same intensity (i.e., with a 0-dB target-to-masker ratio). Each block was followed by 13-18 second rest

independently, for each pair in a block. A 150-ms-long silent gap followed the second token before the next pair began. For the noise masker condition, masker word tokens were replaced by white noise bursts.

Listeners were instructed to press a key on a handheld keypad whenever they detected a color word within the target stream. Data collection was divided into blocks, each of which presented one of the two different masker streams (speech or noise; see Figure 1) and one of four different spatialization conditions, described in detail below (see Stimuli). At the start of each block, a single presentation of the word "bag" was presented from the side of the target stream to cue the participant as to which direction to attend. Two seconds after the start of the cue word, the target and masker streams began playing. Thus, each block consisted of the 2-sec cue interval, followed by 18 token pairs, each of which was 600 ms long, for a total block duration of 12.8 s (see Figure 1).

Each unique combination of masker type and spatialization condition was presented in 8 different blocks, yielding a total of 64 blocks (8 blocks for each of 2 masker types x 4 spatializations). In addition, for each masker and spatialization condition, 4 blocks presented the target stream to the left (masker to the right), and 4 had the target to the right (masker to the left). The order of these 64 blocks was randomized separately for each participant. A silent period followed each block, with the duration selected randomly from a uniform distribution from 13-18 seconds. This silent period allowed hemodynamic responses to return to baseline (see *fNIRS Data Acquisition and Analysis*). Subjects were given the option to pause after each block and were prompted to take a break at the halfway point of the experiment.

Stimuli

All stimulus processing was completed in the Python programming language (Van Rossum & Drake, 2009), with signal processing functions that were either from the *scipy* package (Virtanen et al., 2020) or custom written.

Target tokens were American English words. These speech stimuli were constructed from a list of 16 words, recorded in isolation by a single male talker (Kidd et al., 2008). The words were selected from a set of 12 object words: <*hat*, *bag*, *card*, *chair*, *desk*, *glove*, *pen*, *shoe*, *sock*, *spoon*, *table*, *toy*> and a set of four color words: <*red*, *white*, *blue*, *green*>. All word tokens were edited to have a duration of 300 ms by using a waveform editor to isolate each token and the soundstretch synchronous-overlap-and-add algorithm to adjust duration (SoundStretch Audio Processing Utility, 2017). Finally, tokens were equated in root-mean-square amplitude.

Target and masker word sequences were generated by randomly choosing 18 words, with replacement, from the object and color word sets. Each target stream included a minimum of 3 and a maximum of 5 color words; the other words were all from the object set. The number of color words in the masker stream was also drawn randomly between 3 and 5 words. The temporal position of color words was restricted such that color words could not occur within 2 indices of each other in either stream.

Two types of maskers were used: speech (second row, Figure 1) and noise (third row, Figure 1). Speech masker sequences were generated identically to target sequences, but with the additional constraint that within each target-masker pair, the masker word differed from the target word. Note that color words could occur in the speech masker sequence, but the listener

was asked to ignore these "foil" color words. The speech masker contained words drawn from the same set (i.e., they were spoken by the same talker). This choice and the lead-lag jittered timing resulted in stimuli in which target and masker could only be differentiated based on their perceived directions: Participants did not have access to pitch differences, word-level timing cues for the task, or other features that could be used to determine whether a word was from the target or the masker.

Noise masker streams comprised independent 300-ms white noise bursts. The noise tokens were each ramped with 20 ms raised cosine onset and offset ramps. Each noise burst was flat spectrum broadband white noise and sampled independently, and so covered frequencies 0-22 kHz. Noise bursts were subjected to the same timing constraints as speech masker word tokens (Figure 1). Color words across both target and speech masker streams were constrained to ensure that color words could not occur in successive word token pairs. The presentation level of each stream was set to 65 dB SPL resulting in a 0 dB target-to-masker ratio prior to applying spatial cues.

The four different spatialization conditions are illustrated in Figure 2. To each stream (target and masker, spatialized to opposite lateral locations), we either applied ± 50 μs ITD ("Small ITD"), ± 500 μs ITD ("Large ITD"), a naturally occurring ILD corresponding to 70 degrees azimuth ("Natural ILD"), or a 10 dB broadband ILD ("Broadband ILD"). To apply ITDs, the signal in the ear contralateral to the apparent source location was delayed in time by the appropriate number of samples (corresponding to 50 μs or 500 μs). In the natural ILD condition, the magnitude spectrum from a head-related transfer function (HRTF) for 70° azimuth, measured on a KEMAR manikin, was applied to spatialize sources to the right (Gardner & Martin, 1995). For sources to the left, the left and right magnitude spectra from this HRTF were swapped. The 70° location was chosen because it yielded the greatest average broadband ILD magnitude, equal to about 16 dB. In the broadband ILD condition, the signal in the contralateral ear was attenuated by 10 dB (equally across all frequencies). After spatialization, target and masker streams, one from the left and one from the right, were summed to create left and right channels.

Experimental Procedure

Participants were seated in an anechoic chamber, fitted with Etymotic ER-3a insert phones, and provided with a numeric keypad. Before the start of the experiment, subjects performed practice runs to familiarize themselves with the stimuli and task. Practice runs presented only a diotic target sequence with no masker and no spatial cues applied. Participants practiced pressing the keypad "enter" button when they heard a color word from the talker. Participants had the opportunity to discuss the task with the experimenter between practice runs. At least three practice runs were presented to each participant; however, each participant could perform as many practice runs as they needed to feel comfortable with the task. Participants were then instructed that there would be two streams during testing, each coming from a different location, and that the cue word 'bag' would be presented at the start of each block to indicate the location of the target sequence in that block.

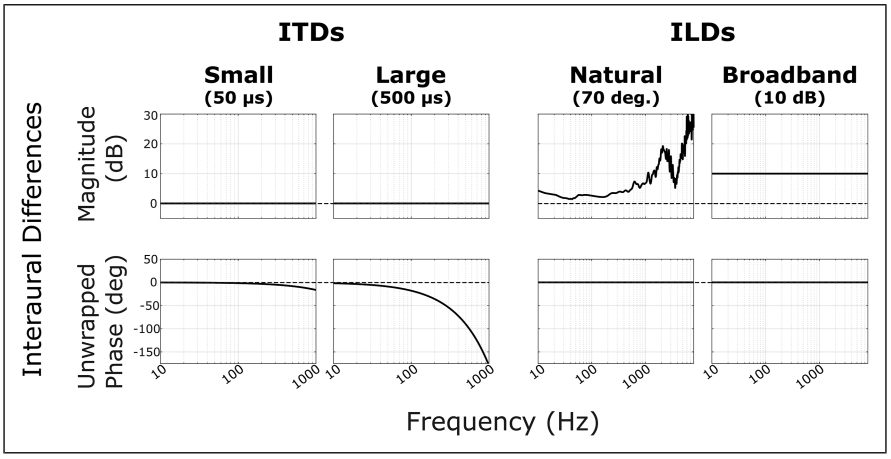


Figure 2. Spatialization conditions. The left (L) and right (R) ear signals for a given stream were spatialized with a 50 μs ITD ("Small ITD"), 500 μs ITD ("Large ITD"), naturally occurring ILD (HRTF at 70° azimuth, "Natural ILD"), or a 10 dB attenuation ("Broadband ILD"). Each subplot shows the magnitude (top row) and unwrapped phase (bottom row) properties of the interaural differences used. ITDs were implemented by pure time delay of the contralateral ear. Note that for the ITD plots, the unwrapped phase plots show a characteristic falloff with frequency proportional to 1/f, which correspond to group delays of 50 μs and 500 μs for small and large ITDs, respectively

Behavioral Performance Estimation

In each block, we recorded the time of each button press and compared it to the onset of word tokens to determine rates of "hits" (correct responses to a target color word) and, for cases with the speech masker, "false alarms" (incorrect responses following a color word in the distractor stream). For each of these two response types, we defined windows for scoring response rates. "Hit windows" encompassed 300-1400 ms following the onset of each color word in the target stream; "false alarm windows" spanned 300-1400 ms after each color in the masker stream. To resolve ambiguity due to overlap of these potential windows, if a potential hit window overlapped a potential false alarm window, both the hit and false alarm windows were removed. Additionally, if two potential time windows of the same type overlapped, the first of the two was kept and the subsequent window thrown out. Using this approach, all response windows were of equal length in time and did not overlap with one another. Since these windows were defined independent of when button presses occurred, they allowed us to operationally define hit rates and false alarm rates in a fair and unbiased manner. Finally, any button-press responses not associated with a color word window (whether or not that window was thrown out) were termed "random responses."

We computed hit rates and random response counts for both masker types; we computed hit rates separately for lead and lag color word positions. Overall, few responses were removed due to overlapping hit and FA windows, although more were removed for a speech masker than a noise masker (mean 3.21 ± 0.20 S.D. clicks for speech masker; mean 0.00 ± 0.00 S.D. clicks for noise masker). Very few random responses (i.e., those that fell completely outside any color word response window, including deleted ones) occurred in any condition (mean 0.317 ± 0.053 S.D. clicks for speech masker; mean 0.283 ± 0.050 S.D. clicks for noise masker).

For the speech masker, we also calculated false alarm (FA) rates. We opted to analyze hit and FA rates because false alarms did not exist in the noise masker case, as there were no color words in the noise masker stream. In analyzing hit rates and false alarm rates separately, we also tap into two different aspects of task performance. Hit rate reflects the ability to segregate target from masker; a failure to segregate the competing streams will lead to low hit rates, because none of the words will be intelligible. In contrast, false alarm rates reflect failures to differentiate target and masker word locations. If target and masker streams are perceptually segregated and sufficiently distinct in their perceived locations, listeners should be able to focus attention on the target stream and ignore the masker stream (not responding to color words from the wrong direction).

Hit and false alarm response rates were calculated as the total sum of button presses in each of the corresponding remaining time windows divided by the total number of those windows. Rates were computed individually for each subject in each condition.

fNIRS Data Acquisition and Analysis

Functional near-infrared spectroscopy (fNIRS) was used to record changes in concentration of oxygenated and deoxygenated hemoglobin concentration during the task. We used a NIRx NIRSport 2 (NIRx, Berlin) system and recorded at 760 and 850 nm wavelengths with a source-detector distance of 3 cm. The source-detector montage and sensitivity map are shown in Figure 3. Subjects wore an fNIRS head-cap fitted with light sources and detectors covering the prefrontal cortex (PFC) and superior temporal gyrus (STG) bilaterally, resulting in 14 source-detector pairs. Source and detector locations were chosen by selecting "dorsolateral prefrontal cortex" and "superior temporal gyrus" in the fNIRS Optodes' Location Decider (fOLD) software (Zimeo Morais et al., 2018). Short channel detectors were attached to each source in order to record systemic physiological signals from the scalp.

fNIRS data processing was performed in Python using the MNE software package (Gramfort et al., 2013; Luke et al., 2021-4). Raw intensity values were first converted to optical density. To exclude source-detector pairs (channels) with poor data quality, we calculated the scalp coupling index (SCI) for each channel, a measure of the presence of the heart rate signal in the data (Pollonini et al., 2014). Channels with an SCI below 0.8 were removed from further analysis. Out of 14 long-distance channels, across the different participants an average of 1.42 ± 1.47 S.D. channels were removed due to low SCI. Next, high-frequency motion artifacts were removed using temporal derivative distribution repair (Fishburn et al., 2019). Time traces were then bandpass filtered between 0.01 and 0.3 Hz using a fourth order, zero-phase Butterworth filter to remove heartbeat and respiration signals. In order to further remove physiological artifacts, short channel information was regressed out of the optical density signal using the short_channel_regression function in MNE (Fabbri et al., 2004; Saager & Berger, 2005; Scholkmann et al., 2014). Finally, we applied the modified Beer-Lambert law (Kocsis et al., 2006) to the time traces of optical density to compute changes in oxygenated hemoglobin (ΔHbO) and deoxygenated hemoglobin (ΔHbR) concentrations.

The 13-18 second gap after each block ensured that hemoglobin levels returned to baseline before the next block began. The length of the silent period was chosen randomly from that range on each block in order to avoid the presence of a coherent signal in the fNIRS data at the rate of stimulus presentation. Raw fNIRS data for each block were epoched from -5 to 20 seconds from the onset of the sound cue. The first 5 seconds of each epoch (before and up to the cue) were used to set the

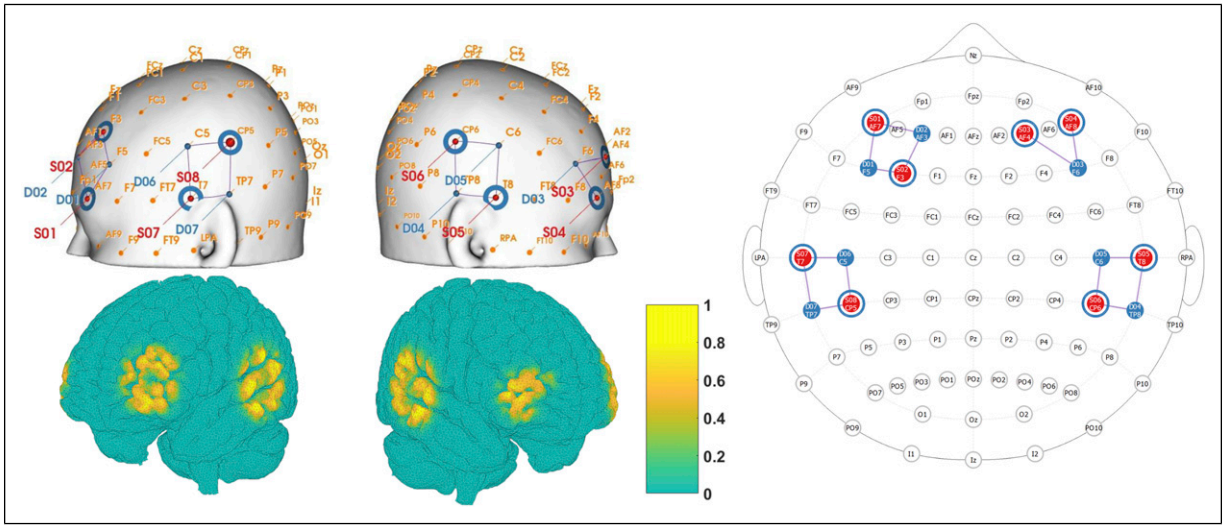


Figure 3. fNIRS montage and sensitivity map. Sensors covered the prefrontal cortex and superior temporal gyrus, bilaterally. Left: The top row shows placement of near-infrared light sources and detectors (end points of white lines), and the location of data channels (between each source and detector, yellow dots). The bottom row shows the normalized sensitivity map projected onto the ICBM 2009c nonlinear asymmetric brain model (Fonov et al., 2009). Right: 2D representation of source detector placement generated using the NIRSite montage creation software (NIRx)

baseline of ΔHbO and ΔHbR by subtracting the mean value over this period from the raw measure. Baselining was performed separately for ΔHbO and ΔHbR data. To quantify the strength of the hemodynamic response in each block, we calculated the mean ΔHbO value during sound stimulation (from stimulus onset to 12.8 seconds later) in each channel for each block.

Statistical Analysis

We used linear mixed effects regression to analyze task performance and by-channel hemodynamic response magnitudes. All models were implemented in R version 4.2.1 (R Core Team, 2024) and fit using the "lmer" function in the lme4 library (Bates et al., 2015). To estimate main effects and interactions, we used the "mixed" function in the afex library (Singmann, H., Bolker, B., Westfall, J., & Aust, F., 2018), which uses likelihood ratio tests to assess the fit of a model with the effect of interest versus that of a simplified model without that effect. Thus, main effects and interactions are reported with chi-square values that index whether inclusion of the effect leads to a significant improvement in model fit. For all measures, we followed up on significant effects with Bonferroni-corrected pairwise t tests.

To analyze task performance, we fit separate models for hit rates and false alarm rates. For both analyses, we included fixed effects of spatialization (small ITD, large ITD, natural ILD, broadband ILD) and color word position (lead, lag), as well as random intercepts for each participant. We analyzed all spatialization conditions as levels of a single factor (rather than ITD/ILD as one factor and small/large as another) because 'small' and 'large' conditions within each cue type were not perceptually or physically equivalent; accordingly, potential interactions between cue type and cue magnitude would be misleading. For the hit rate model only, we included a fixed effect of masker type (speech, noise). We did not include this effect for false alarms, since false alarms were not possible in the noise masker condition.

To compare hemodynamic response magnitudes, we fit separate statistical models for mean ΔHbO in PFC, mean ΔHbR in PFC, mean ΔHbO in STG, and mean ΔHbR in STG. We fit separate models for each masker type, given our a priori hypothesis that separate psychoacoustic mechanisms underlay performance with each masker type. For all fNIRS models, we included a fixed effect of spatialization, random intercepts for each subject, and random intercepts for each channel, allowing the model to account for baseline differences in activation across channels. Finally, for the STG models only, we also included a fixed effect of cortical hemisphere (contralateral, ipsilateral), driven by our a priori hypothesis that spatial selective attention will result in an increase in neural activity in the auditory cortex contralateral to attention (Alho et al., 2003; Ciaramitaro et al., 2007). We did not include a fixed effect of cortical hemisphere for the PFC data because the sensor montage was not symmetrical and therefore did not image identical populations in left and right hemispheres.

For each behavioral and neural analysis, the omnibus model used sum-coded contrasts for all factors (e.g., the fixed effect of cortical hemisphere might be coded with a [1,-1] contrast for [contralateral, ipsilateral]). To perform pairwise comparisons between different levels of spatialization (e.g., small ITD vs. large ITD), we re-fit the omnibus model using treatment coding.

With a treatment coding scheme, one level (e.g., small ITD) was set as a reference level, and beta estimates reflect pairwise comparisons between the reference level and each other level. To obtain all pairwise estimates, we ran four models, rotating which level was the reference level. Critically, these follow-up models do not differ in their fit to the data; all that differs across the models is the comparison captured by each beta coefficient. For all pairwise comparisons, we therefore report beta coefficients that capture the difference between those conditions in the same units as the dependent measure (e.g., hit rate).

Results

Hit Rates Varied Between Spatialization Conditions

Hit rates across spatialization, masker type, and color word position are shown in the first column of Figure 4. The hit rates for the small ITD condition (leftmost data points) tended to be lower than for any other condition. Hit rates were slightly higher for the noise masker than the speech masker (compare bottom left and top left panels). Additionally, hit rates tended to be higher for a color word that occurred in the leading position within a word token pair than for a lagging position across all spatialization conditions (compare blue circles to corresponding yellow triangles).

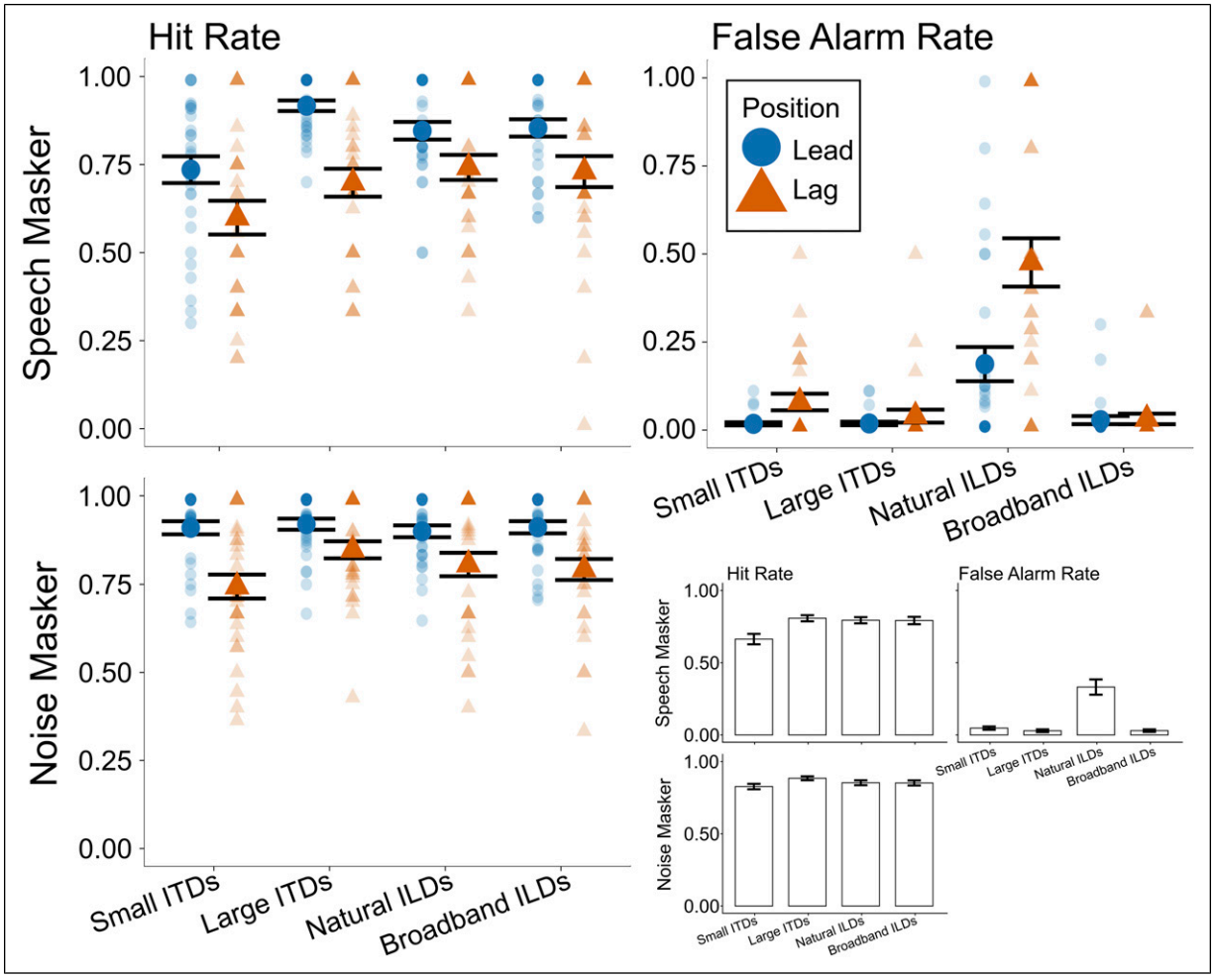


Figure 4. Hit and False Alarm Rates. First column (Hit Rates): The percentage of responses to color words in the target stream when the masker was speech (top row) or noise (bottom row). Second column, Top (False Alarm Rates): The percentage of responses to color words in the masker stream only when the masker was speech. For all plots, blue circles show results when a color word occurred in the leading position, and orange triangles show results for a color word in the lagging position. Small markers show individual participant data, and large markers show the mean across participants. Error bars indicate the mean and SEM across subjects in each condition. Second column bottom: The data represented in the other three subplots is collapsed across lead and lag word positions for ease of comparison. Bar height represents the mean proportion of responses in each category (hit rate, false alarm rate), and error bars show SEM across subjects in each condition

Statistical analyses supported these observations. The linear mixed effects model revealed a significant main effect of masker type ($\chi^2(1) = 37.17$, $p < 0.001$), driven by higher hit rates with the noise masker (mean: 0.854) compared to with the speech masker (mean: 0.765). We also observed a significant main effect of color word position ($\chi^2(1) = 75.26$, $p < 0.001$), driven by higher hit rates for color words occurring in the leading position (mean: 0.875) compared to in the lagging position (mean: 0.745). The model also revealed a significant main effect of spatialization ($\chi^2(3) = 26.40$, $p < 0.001$) and a significant interaction between spatialization and masker type ($\chi^2(3) = 8.35$, $p = 0.039$). The three-way interaction between spatialization, masker, and color word position was not significant ($\chi^2(3) = 5.45$, $p = 0.142$).

To interpret the significant interaction between spatialization and masker type, we conducted post hoc separate analyses of the effect of spatialization within each masker type. Within the speech masker data, the hit rates with small ITDs (mean: 0.667) were significantly lower than with large ITDs (mean: 0.808; $\beta = 0.140$, $p = 0.0001$), natural ILDs (mean: 0.794; $\beta = 0.127$, $p = 0.0005$), and broadband ILDs (mean: 0.792; $\beta = 0.125$, $p = 0.0006$), but the large ITD, natural ILD, and broadband ILD hit rates did not differ significantly from one another. For the noise masker, the hit rates with small ITDs (mean: 0.827) were significantly lower than with large ITDs (mean: 0.884; $\beta = 0.057$, $p = 0.034$). Hit rates for all other pairs of conditions were not significantly different when the masker was noise.

False Alarm Rates Were Larger With Natural ILDs

As noted above, false alarm rates (reflecting responses to a color word in the distracting stream) could only be computed for the speech masker conditions (top row of Figure 4). As seen in the second column of Figure 4, the false alarm rates were near zero for small ITD, large ITD, and broadband ILD conditions; however, they were higher in the natural ILD condition, especially for lagging color words.

Statistical analyses confirmed these observations: The mixed effects model showed a significant main effect of spatialization ($\chi^2(3) = 105.78$, $p < 0.001$), a significant main effect of color word position ($\chi^2(1) = 17.45$, $p < 0.001$), and a significant interaction between the two ($\chi^2(3) = 25.75$, $p < 0.001$). To interpret the interaction, we conducted post hoc analyses across spatialization condition within each color word position. For the leading color word, false alarm rates were significantly larger in the natural ILD condition (mean: 0.187) than in the small ITD (mean: 0.018; $\beta = -0.170$, $p < 0.0001$), large ITD (mean: 0.019; $\beta = -0.169$, $p < 0.0001$), and broadband ILD (mean: 0.028; $\beta = -0.159$, $p < 0.0001$) conditions. An identical, but statistically stronger pattern arose for the lagging color word: False alarm rates were significantly larger in the natural ILD condition (mean: 0.476) than in the small ITD (mean: 0.079; $\beta = -0.396$, $p < 0.0001$), large ITD (mean: 0.040; $\beta = -0.436$, $p < 0.0001$), and broadband ILD (mean: 0.032; $\beta = -0.444$, $p < 0.0001$) conditions. None of the false alarm rates for leading and lagging color words in the small ITD, large ITD, or broadband ILD conditions differed significantly from one another.

fNIRS Results

Figure 5 shows the block averaged ΔHbO (red lines) and ΔHbR (blue lines) mean time traces and standard errors across subjects (colored ribbons), averaged within each ROI and masker type. The left plot in each panel compares the small (solid) and large (dashed) ITD spatialization conditions, and the right plot compares natural (solid) and broadband (dashed) ILD spatializations. Hemodynamic responses in all regions and conditions show a canonical shape, with ΔHbO increasing with respect to baseline during sound stimulation (between 0 and 12.8 seconds) and decreasing thereafter (Luke et al., 2021-4; Quaresima & Ferrari, 2019).

Figure 6 summarizes the differences in hemodynamic responses by plotting the mean ΔHbO and ΔHbR averaged within prefrontal cortex during the block (0-12.8 seconds from cue onset). Mean ΔHbO for the small ITD condition was larger than for any other condition when the masker was speech, but mean ΔHbO did not vary strongly with spatialization when the masker was noise.

Consistent with this summary, our statistical models only found a significant effect of spatialization in mean ΔHbO in PFC ($\chi^2(3) = 27.15$, $p < 0.001$). Post hoc tests showed this main effect was driven by significantly larger mean ΔHbO in the small ITD condition (mean: 0.086 μM) compared to the large ITD (mean: 0.039 μM; $\beta = -0.047$, $p < 0.001$), natural ILD (mean: 0.037 μM; $\beta = -0.049$, $p < 0.001$), and broadband ILD (mean: 0.044 μM; $\beta = -0.042$, $p < 0.001$) conditions, with no significant differences between the other three spatialization conditions. When the masker was noise, the effect of spatialization on mean ΔHbO in PFC was not significant ($\chi^2(3) = 3.82$, $p = 0.282$). Mean ΔHbR in PFC did not significantly vary in either the speech masker case ($\chi^2(3) = 4.94$, $p = 0.176$) or the noise masker case ($\chi^2(3) = 4.48$, $p = 0.214$).

Figure 7 shows the across-channel mean ΔHbO and mean ΔHbR in STG for each hemisphere (contralateral, ipsilateral to the attended direction). For the speech masker, hemodynamic responses in STG were similar across hemispheres and across spatialization cues for the large ITD, natural ILD, and broadband ILD conditions. However, in the small ITD condition, the

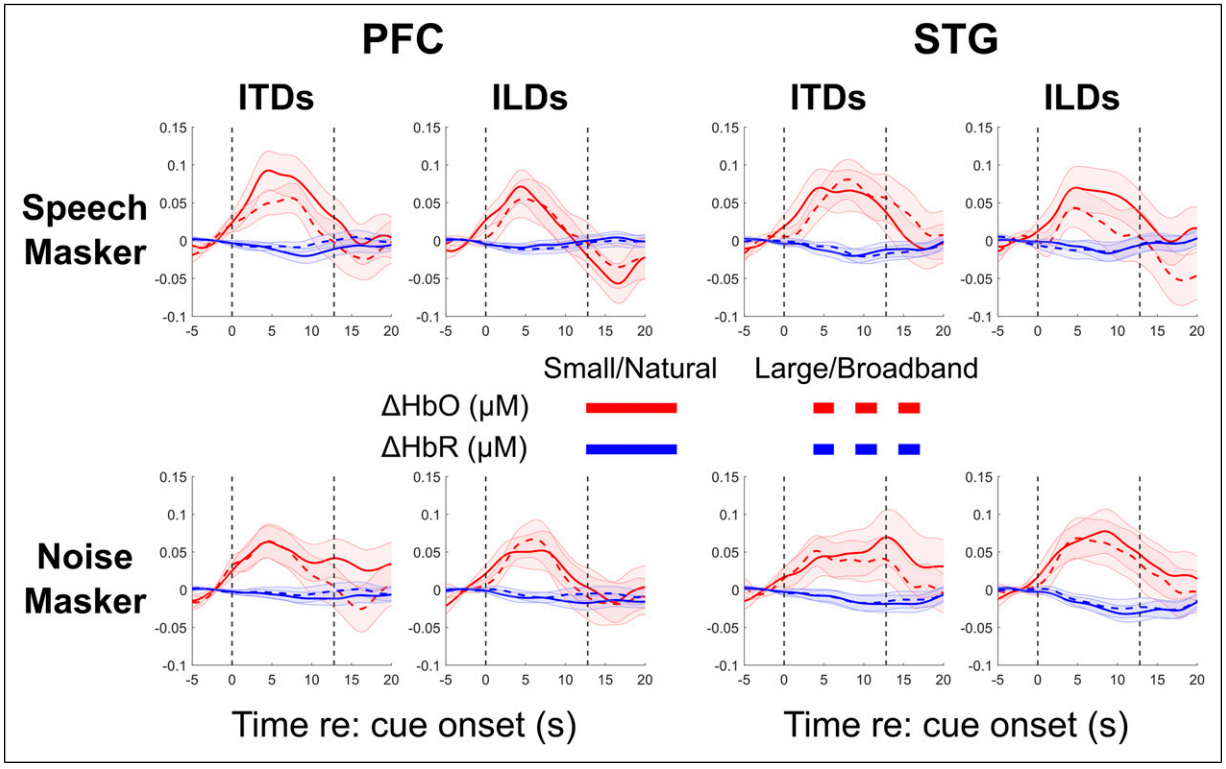


Figure 5. Block average hemodynamic responses. Each individual subplot shows the ΔHbO (red) and ΔHbR (blue) curves for each masker type and in each region. Shaded error bars show ± 1 S.E.M. Top left panel: Hemodynamic responses in PFC when the masker was speech (small vs. large ITD, left subpanel; Natural vs. Broadband ILD, right subpanel). Top right panel: as before, in STG when the masker was speech. Bottom left panel: as before, in PFC when the masker was noise. Bottom right panel: as before, in STG when the masker was noise

hemodynamic response was stronger in the cortical hemisphere contralateral to the direction of attention. For the noise masker, hemodynamic responses varied more with spatial cue condition but did not differ between hemispheres for large ITD or natural ILD conditions; however, for both small ITD and broadband ILD conditions, hemodynamic responses were larger in the hemisphere contralateral to the direction of attention.

These observations were supported by our statistical analyses. The mean ΔHbO models showed a significant interaction between spatialization and hemisphere both when the masker was speech ($\chi^2(3) = 7.99$, $p = 0.046$) and when the masker was noise ($\chi^2(3) = 11.02$, $p = 0.012$). To follow up on this interaction, we conducted post hoc tests comparing across hemispheres within each spatialization condition. For the speech masker, the interaction effect was driven by an effect of hemisphere in the small ITD condition only, with larger mean ΔHbO in contralateral channels (mean: 0.078 μM) compared to the ipsilateral channels (mean: 0.045 μM; $\beta = -0.033$, $p = 0.0417$). A similar pattern emerged in the noise masker data, with larger responses in contralateral channels (mean: 0.0645 μM) than in ipsilateral channels (mean: 0.0285 μM; $\beta = -0.036$, $p = 0.017$) for the small ITD condition only; the effect of hemisphere approached significance in the broadband ILD condition (contralateral mean: 0.0673, ipsilateral mean: 0.0406; $\beta = -0.027$, $p = 0.077$).

Discussion

We examined how ITDs and ILDs shaped both perceptual performance and cortical activity in a color word detection task; to force listeners to rely on spatial cues, the two streams were spoken by the same talker and the timing of stimuli within each stream was irregular. Accordingly, participants were forced to segregate and select based on spatial cues. Differences in spatial cues across frequency interact with this process (Best et al., 2007), leading to patterns of behavior and neural activation.

Performance was generally better for a noise masker than a speech masker, for leading color words than lagging color words, and for stronger spatial cues (large ITDs, broadband ILDs) versus less robust spatial differences (small ITDs and natural ILDs). When listening with a speech distractor with small ITDs, hemodynamic activity in PFC was stronger than for

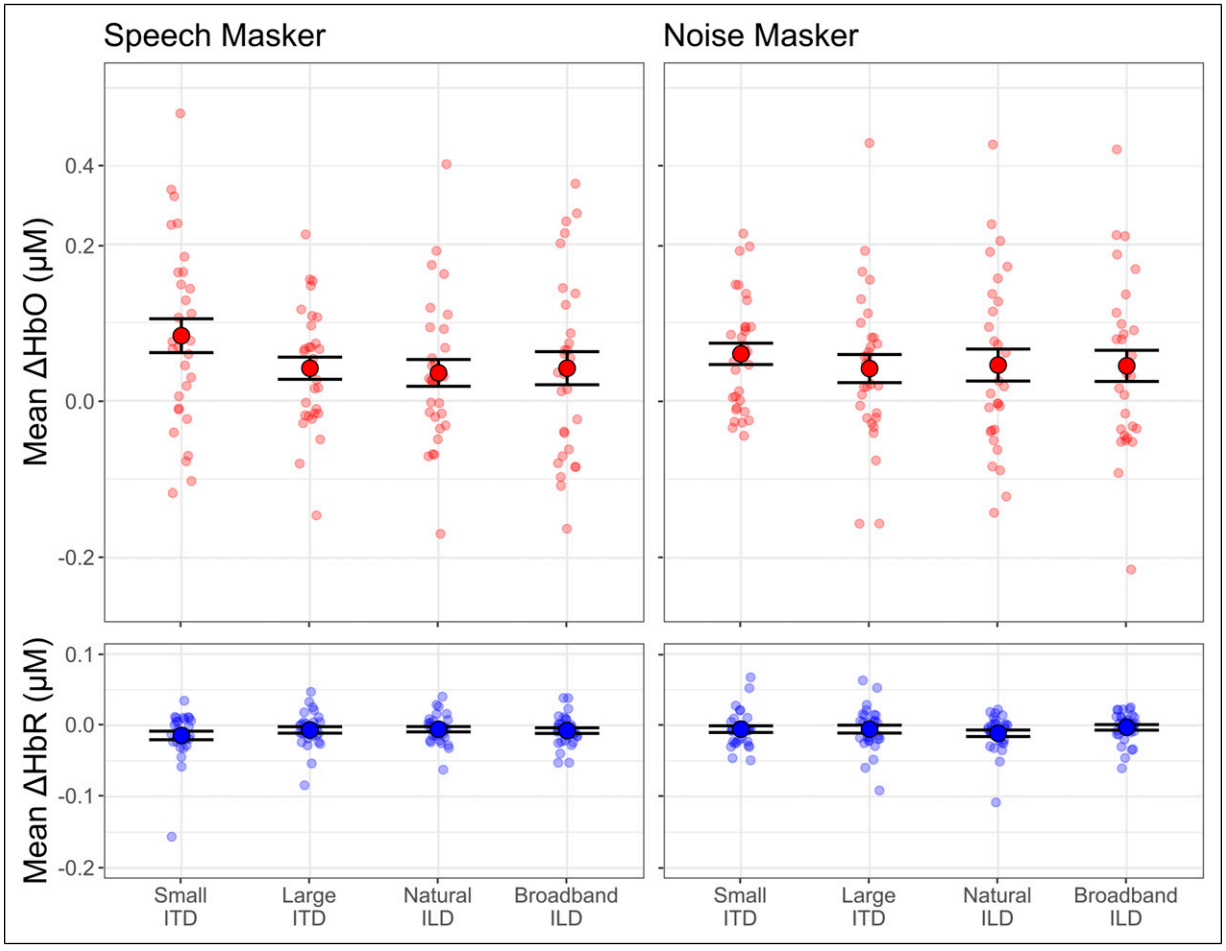


Figure 6. Mean hemodynamic response magnitudes in prefrontal cortex. Top row: Mean changes in oxygenated hemoglobin concentration (ΔHbO) when the masker was speech (left panel) and noise (right panel). Bottom row: Mean changes in deoxygenated hemoglobin concentration (ΔHbR) when the masker was speech (left panel) and noise (right panel). Small markers show individual participant data, and large markers show the mean across participants. Error bars indicate the mean and SEM across subjects in each condition

other conditions. For both speech and noise maskers, activity in STG was lateralized contralateral to the direction of attention when sources were lateralized with small ITDs, but not for any other spatial conditions. Together, these results provide new insights into how different spatial cues contribute to SRM and speech understanding in noisy settings.

Energetic Masking Accounts for Noise Masker Results

In all spatial conditions, the hit rates were higher on average when the masker consisted of white noise bursts rather than speech. Hit rates were near ceiling, indicating that the task was relatively easy; however, rates still varied with masker type and spatialization. This supports the idea that participants were better able to hear out the target words when the background was spectrotemporally distinct from the target speech. This effect is well documented: Listeners perform better on SRM tasks when the masker is noise than when it is speech, even at negative signal to noise ratios (Best et al., 2012; Johnstone & Litovsky, 2006; Moore, 2008; Zhang et al., 2021a). Noise maskers do not contain any structured spectrotemporal features or semantic or linguistic content, and thus are not confused with words in the target stream. With a noise masker, both target segregation and selection are relatively trivial, making spatial information redundant. It is also the case that, for these stimuli, the flat-spectrum white noise bursts led to less masking in lower frequencies than did the word token maskers of equal root-mean-square intensity. The low frequencies contained both speech information and ITD cues and, in the Broadband ILD condition, substantial ILD information. Compared to the speech masker, the noise masker thus produced less energetic masking in addition to causing less interference with target speech segregation and selection.

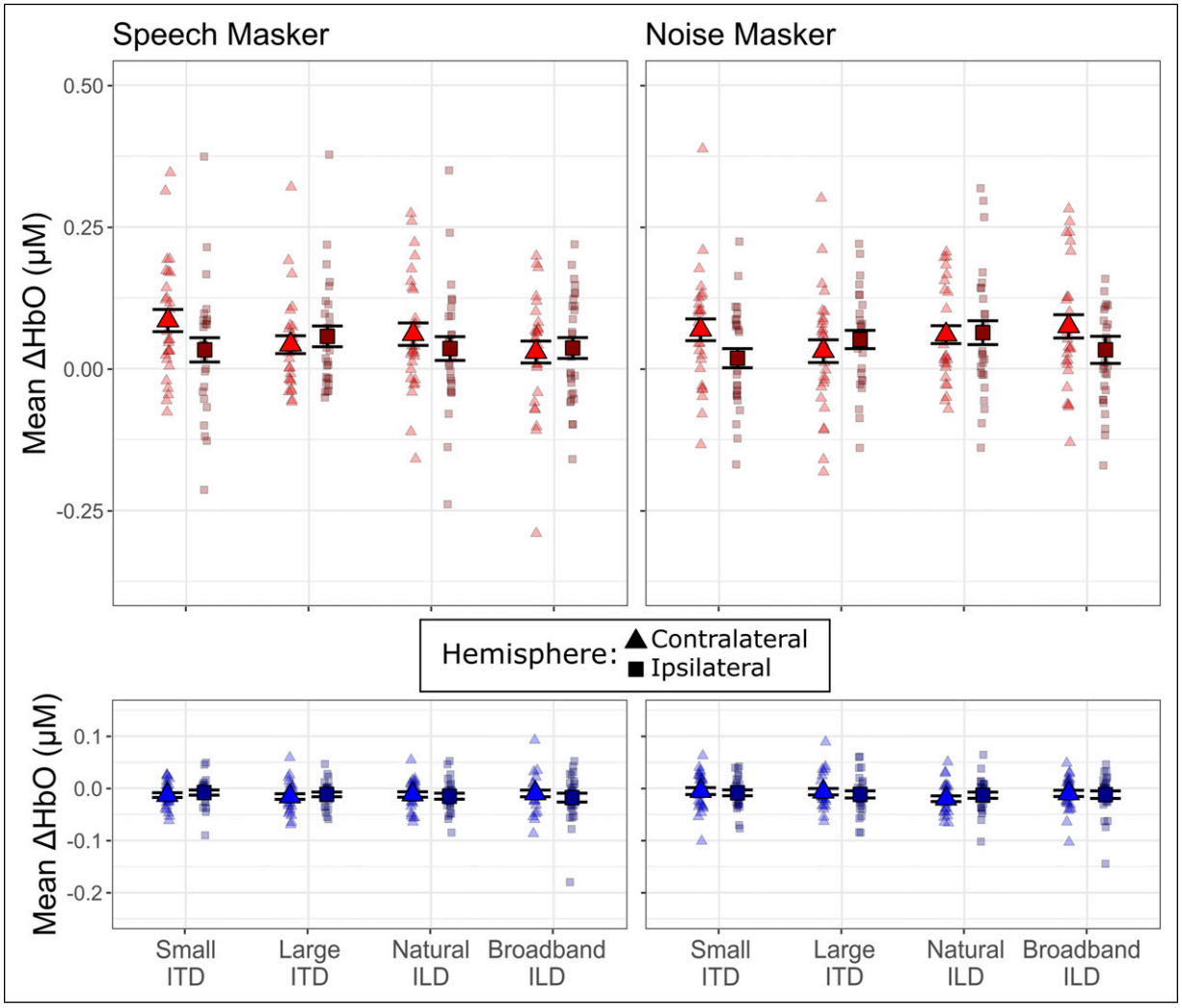


Figure 7. Mean hemodynamic response magnitudes in superior temporal gyrus. Top row: Mean ΔHbO when the masker was speech (left panel) and noise (right panel). Data within each spatialization condition are shown for sensors contralateral (lighter triangles) and ipsilateral (darker squares) to the direction of the target sound stream. Bottom row: As in the top row, but for mean ΔHbR. Small markers show individual participant data, and large markers show the mean across participants. Error bars indicate the mean and SEM across subjects in each condition

When segregation and selection are easy (i.e., when the masker is noise), energetic masking is likely the limiting factor on performance. In these conditions, hit rates were larger for leading color words than for lagging color words. This asymmetry likely reflects the fact that the lagging token (word or noise) onset was energetically masked by the end of the leading token in each pair. Specifically, syllable onsets convey more lexical information than syllable codas (Pimentel et al., 2021; Sun & Poeppel, 2023); therefore, energetic masking of the onset of a lagging word has a greater impact on performance than masking of the coda of a leading word. For the noise masker, hit rates were lower with small ITDs than with large ITDs, which might be explained by the influence that robust ITD differences can have in combatting energetic masking (Culling & Lavandier, 2021; Durlach, 1963).

For the speech masker, hit rates were higher for leading than for lagging color words. This difference mirrors the effect for noise maskers, suggesting that energetic masking of the onsets of lagging target words degrades performance for speech maskers like it does for noise maskers. However, for speech maskers, hit rates were generally lower than for the noise masker and also varied with spatial cues, hinting that other factors affect performance, as well. These other factors are considered in the following sections.

ITDs and ILDs yield different modes of failure for speech maskers

We hypothesized that both the small ITD and natural ILD conditions would produce relatively weak spatial cues, which could not only make it difficult to segregate target from masker but also to differentiate target words from masker words. In the small ITD condition, listeners had relatively low hit rates but also low false alarm rates: Listeners struggled to identify color words in either stream, suggesting that they struggled to group spectrotemporal features into intelligible words (i.e., a failure of segregation). This is schematized in the first row of Figure 8. Without access to strong spatial cues (space is conveyed with color in Figure 8), a listener would struggle to separate "shoe" from "blue;" instead, they may have sometimes heard a mixture of both words that was difficult to interpret ("shboleue"). This also helps explain why the small ITD condition yielded a larger hemodynamic response in PFC than the other spatialization conditions; listeners expended greater cognitive effort in their attempt to segregate the competing words (Lawrence et al., 2018; Zhang et al., 2021b; Zhou, Burg, et al., 2022, Zhou, Burg, et al., 2022). With small ITDs, activity in STG also lateralized, increasing in the hemisphere contralateral to the direction of attention — as is often observed when spatial auditory attention is strongly deployed (Alho et al., 1998, 1999, 2003; Ciaramitaro et al., 2007; Jäncke & Shah, 2002; Yang & Mayer, 2014).

In contrast, natural ILDs elicited both high hit rates and high false alarm rates. This pattern suggests that listeners could hear out individual word tokens, but could not differentiate whether color words belonged to the target or masker stream — a failure of selection, rather than segregation. Unlike in the small ITD case, with natural ILDs the two streams were discriminable from one another and listeners could hear out individual color words. However, they failed to identify whether words were from the target or the masker stream. As illustrated in the middle panel of Figure 8, ILDs in the high frequencies may help support segregation by binding with comodulated lower frequency sound components; however, the perceived lateralities of the resulting words are dominated by the lower-frequency energy, which carried zero-μs ITDs and only modest

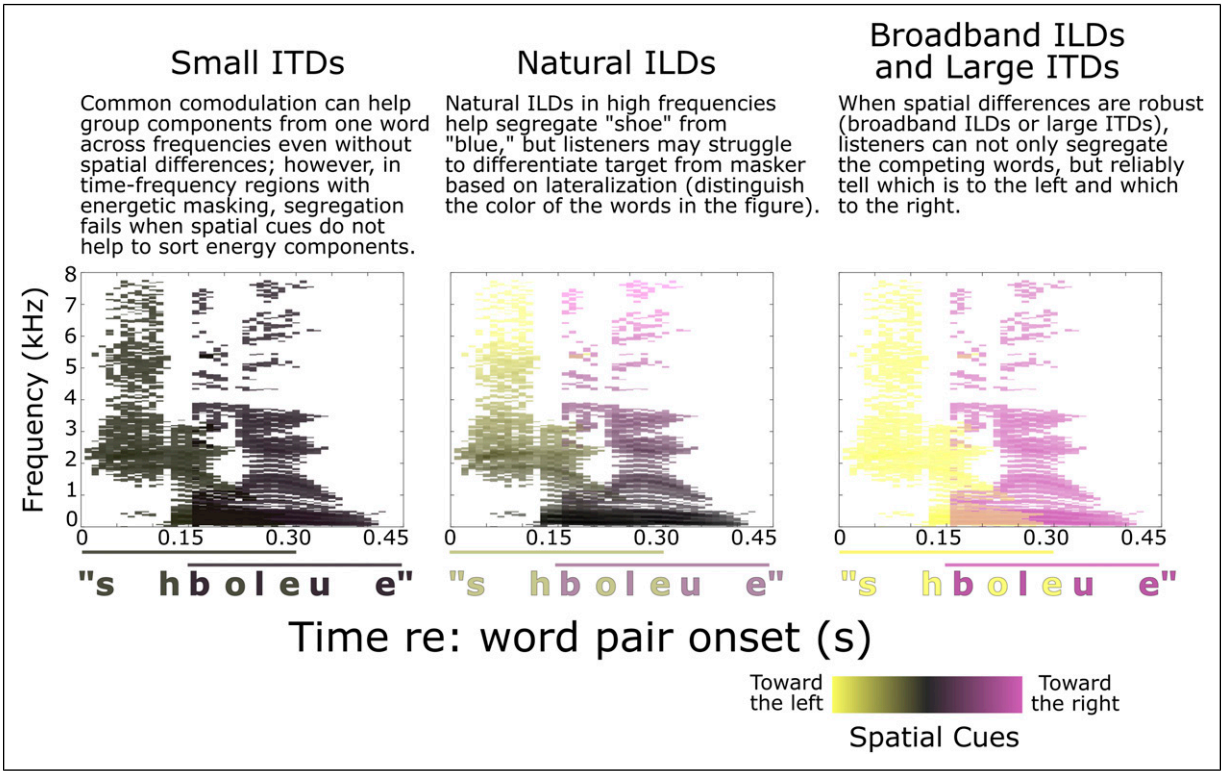


Figure 8. Schematic representation of how different spatial cues impact segregation and spatial perception. Each panel shows a spectrogram representation of the word pair "shoe" (leading) and "blue" (lagging). In each time-frequency pixel, the dominant spatial cues determine the color (yellow towards the left, magenta to the right). Time-frequency pixels where there is significant overlap between words are colored by the mean of the two color schemes. The horizontal lines at the bottom of each panel denote the time extent of the lead and lag words; the color of the lines and the corresponding word labels denote the perceived laterality of the overall word, integrating spatial information across the grouped spectrum. Small ITDs may fail to support segregation, leading listeners to hear an unintelligible mixture of both words (orthographically, "shboleue," left panel). Natural ILDs may support segregation, but the perceived words may be perceived near midline, making it hard to select the target and ignore the masker. Robust spatial cues (broadband ILDs and large ITDs) enable both segregation and spatial selection

ILDs in the natural ILD condition. This is consistent with previous investigations: A low frequency interferer that is perceptually grouped with a high frequency target dominates the perceived lateral location of the composite perceptual object (Best et al., 2007). Here, all of the competing words were perceived near midline due to the dominant zero-μs ITDs, making it hard to discriminate target from masker.

Both target and masker words are thus likely to be heard near midline, leading to increased confusions about which word was from the target and which from the masker. Notably, the false alarm rate is higher for lagging than leading masker color words, suggesting that the lateralization of the lagging word is especially weak. A number of studies have shown that interaural differences at sound onsets strongly influence lateralization (Freyman & Zurek, 2017; Klein-Hennig et al., 2011). Given that the coda of the leading word energetically masks the onset of the lagging word, it is unsurprising that listeners may be more unsure of the lateralization of a lagging color masker word than a leading color masker word, leading to an even larger false alarm rate.

Although performance was relatively poor with natural ILDs, PFC activity was not stronger in this condition than with other spatializations. In the conditions in which segregation was easier (natural ILD, large ITD, and broadband ILD), PFC was less strongly engaged and STG activity less lateralized. Together, these results hint that PFC is more strongly engaged when listeners are struggling to segregate sources but not when they are having difficulty discriminating differences in competing stream locations (for spatial selection). For small ITDs, strong engagement of spatial attention (seen in PFC) likely increases sensory responses in contralateral STG to enhance the target stream representation, albeit with limited success.

Broadband ILDs and large ITDs supported both stream segregation and spatial selection: Hit rates were high and false alarm rates were low. With large perceived spatial separation (distinct yellow and magenta in row 3 of Figure 8), listeners both properly segregate constituent words and select the target.

These results highlight the importance of separately considering failures of segregation and spatial selection when trying to understand the processes allowing listeners to make sense of competing sounds in a mixture. Specifically, listeners must group sounds into objects based on spectrotemporal information (comodulation of different frequency components, harmonic structure, etc.). Natural ILDs alone can help support object formation of overlapping sounds by helping the brain sort out which components belong to which source. The perceived location of each object can provide a powerful cue for guiding selective attention when sources are perceived at different locations (Allen et al., 2008; Bregman, 1990; Darwin, 2007; Kidd et al., 2010; Shinn-Cunningham, 2008b). However, even spatial cues that are sufficient to aid source segregation (like natural ILDs) may not produce robust differences in perceived location, leading to confusions between properly segregated sources.

Summary, Limitations, and Future Directions

Here, we compared behavioral performance and hemodynamic responses in ILD and ITD spatialization conditions. We found that weaker ITDs and ILDs (small and natural, respectively) led to poorer performance than stronger ITDs and ILDs (large and broadband, respectively). Moreover, differences in the response patterns suggest that small ITDs led to failures of segregation, while natural ILDs, which are modest in the frequency range where most speech energy resides, produced more selection errors. It is tempting to conclude that ITDs and ILDs thus contribute differently to SRM performance, in general. However, we did not match these "weak" ITD and ILD conditions to produce perceptually similar lateralization of our stimuli, limiting our ability to directly compare across conditions.

In the small ITD condition, we observed greater hemodynamic responses in PFC and a lateralization of hemodynamic responses in STG, with greater activation in the hemisphere contralateral to the direction of the target. Considered alongside our behavioral results, this pattern suggests that PFC was more strongly engaged in the one condition where segregation failures limited performance. Although these findings are intriguing, they are not definitive. The measures of hemodynamic activity obtained with this fNIRS montage were relatively coarse with four channels of data per hemisphere of STG and seven channels of data in PFC. We also gathered only eight blocks of data per condition. Further investigations employing denser fNIRS montages, presenting more blocks in each spatialization condition, and co-registering results with MRI structural scans would provide more definitive results (Yücel et al., 2024). Such studies would provide better ground-truth estimates of the neural locus of activity with better signal-to-noise ratio.

Future experiments should investigate psychophysically matched ITDs and ILDs or parametrically vary ITD and ILD magnitudes to support direct comparisons between ITDs and ILDs that produce similar spatial percepts. Clinically focused experiments could also explore whether improved spatial hearing outcomes in bilateral cochlear implant users with magnified ILDs (Brown, 2014, 2018; Richardson et al., 2025) are driven by the same mechanisms as in normal hearing listeners.

ORCID iD

Benjamin N. Richardson https://orcid.org/0000-0003-0454-6337

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by National Institutes of Health (F31-DC022481), (R01-DC015988), (R21-DC018408).

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

Alho, K., Connolly, J. F., Cheour, M., Lehtokoski, A., Huotilainen, M., Virtanen, J., Aulanko, R., & Ilmoniemi, R. J. (1998). Hemispheric lateralization in preattentive processing of speech sounds. *Neurosci. Lett*, *258*(1), 9–12. https://doi.org/10.1016/s0304-3940(98)00836-2

Alho, K., Medvedev, S. V., Pakhomov, S. V., Roudas, M. S., Tervaniemi, M., Reinikainen, K., Zeffiro, T., & Näätänen, R. (1999). Selective tuning of the left and right auditory cortices during spatially directed attention. *Brain Res. Cogn. Brain Res*, *7*(3), 335–341. https://doi.org/10.1016/s0926-6410(98)00036-6

Alho, K., Vorobyev, V. A., Medvedev, S. V., Pakhomov, S. V., Roudas, M. S., Tervaniemi, M., van Zuijen, T., & Näätänen, R. (2003). Hemispheric lateralization of cerebral blood-flow changes during selective listening to dichotically presented continuous speech. *Brain Res. Cogn. Brain Res*, *17*(2), 201–211. https://doi.org/10.1016/s0926-6410(03)00091-0

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

Arbogast, T. L., Mason, C. R., & Kidd, G., Jr (2002). The effect of spatial separation on informational and energetic masking of speech. *J. Acoust. Soc. Am*, *112*(5 Pt 1), 2086–2098. https://doi.org/10.1121/1.1510141

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*. *67*(1). https://doi.org/10.18637/jss.v067.i01

Best, V., Gallun, F. J., Carlile, S., & Shinn-Cunningham, B. G. (2007). Binaural interference and auditory grouping. *J. Acoust. Soc. Am*, *121*(2), 1070–1076. https://doi.org/10.1121/1.2407738

Best, V., Marrone, N., Mason, C. R., & Kidd, G., Jr (2012). The influence of non-spatial factors on measures of spatial release from masking. *J. Acoust. Soc. Am*, *131*(4), 3103–3110. https://doi.org/10.1121/1.3693656

Bregman, A. S. (1990). *Auditory Scene Analysis: The perceptual organization of sound*. The MIT Press. 10.7551/mitpress/1486.001.0001.

Bronkhorst, A. W., & Plomp, R. (1988). The effect of head-induced interaural time and level differences on speech intelligibility in noise. *J. Acoust. Soc. Am*, *83*(4), 1508–1516. https://doi.org/10.1121/1.395906

Brown, C. A. (2014). Binaural enhancement for bilateral cochlear implant users. *Ear Hear*, *35*(5), 580–584. https://doi.org/10.1097/AUD.0000000000000044

Brown, C. A. (2018). Corrective binaural processing for bilateral cochlear implant patients. *PLoS One*, *13*(1), Article e0187965. https://doi.org/10.1371/journal.pone.0187965

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

Ciaramitaro, V. M., Buracas, G. T., & Boynton, G. M. (2007). Spatial and cross-modal attention alter responses to unattended sensory information in early visual and auditory human cortex. *J. Neurophysiol*, *98*(4), 2399–2413. https://doi.org/10.1152/jn.00580.2007

Culling, J. F., & Lavandier, M. (2021). Binaural unmasking and spatial release from masking. *Springer Handbook of Auditory Research, Springer Handbook of Auditory Research* (pp. 209–241): Springer International Publishing. 10.1007/978-3-030-57100-9_8.

Culling, J. F., & Summerfield, Q. (1995). Perceptual separation of concurrent speech sounds: absence of across-frequency grouping by common interaural delay. *J. Acoust. Soc. Am*, *98*(2 Pt 1), 785–797. https://doi.org/10.1121/1.413571

Darwin, C. J. (1997). Auditory grouping. *Trends Cogn. Sci*, *1*(9), 327–333. https://doi.org/10.1016/S1364-6613(97)01097-8

Darwin, C. J. (2007). Spatial Hearing and Perceiving Sources. *Auditory Perception of Sound Sources* (pp. 215–232): Springer US. 10.1007/978-0-387-71305-2_8.

Dorman, M. F., Loiselle, L., Stohl, J., Yost, W. A., Spahr, A., Brown, C., & Cook, S. (2014). Interaural level differences and sound source localization for bilateral cochlear implant patients. *Ear Hear*, *35*(6), 633–640. https://doi.org/10.1097/AUD.0000000000000057

Durlach, N. I. (1963). Equalization and cancellation theory of binaural masking-level differences. *J. Acoust. Soc. Am*, *35*(8), 1206–1218. https://doi.org/10.1121/1.1918675

Ellinger, R. L., Jakien, K. M., & Gallun, F. J. (2017). The role of interaural differences on speech intelligibility in complex multi-talker environments. *J. Acoust. Soc. Am*, *141*(2), EL170. https://doi.org/10.1121/1.4976113

Fabbri, F., Sassaroli, A., Henry, M. E., & Fantini, S. (2004). Optical measurements of absorption changes in two-layered diffusive media. *Phys. Med. Biol*, *49*(7), 1183–1201. https://doi.org/10.1088/0031-9155/49/7/007

Fishburn, F. A., Ludlum, R. S., Vaidya, C. J., & Medvedev, A. V. (2019). Temporal Derivative Distribution Repair (TDDR): A motion correction method for fNIRS. *Neuroimage*, *184*, 171–179. https://doi.org/10.1016/j.neuroimage.2018.09.025

Fonov, V. S., Evans, A. C., McKinstry, R. C., Almli, C. R., & Collins, D. L. (2009). Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *Neuroimage*, *47*, S102. https://doi.org/10.1016/s1053-8119(09)70884-5

Freyman, R. L., & Zurek, P. M. (2017). Strength of onset and ongoing cues in judgments of lateral position. *J. Acoust. Soc. Am*, *142*(1), 206. https://doi.org/10.1121/1.4990020

Freyman, R. L., Helfer, K. S., McCall, D. D., & Clifton, R. K. (1999). The role of perceived spatial separation in the unmasking of speech. *J. Acoust. Soc. Am*, *106*(6), 3578–3588. https://doi.org/10.1121/1.428211

Gardner, W. G., & Martin, K. D. (1995). HRTF measurements of a KEMAR. *J. Acoust. Soc. Am*, *97*(6), 3907–3908. https://doi.org/10.1121/1.412407

Glyde, H., Buchholz, J., Dillon, H., Best, V., Hickson, L., & Cameron, S. (2013a). The effect of better-ear glimpsing on spatial release from masking. *J. Acoust. Soc. Am*, *134*(4), 2937–2945. https://doi.org/10.1121/1.4817930

Glyde, H., Buchholz, J. M., Dillon, H., Cameron, S., & Hickson, L. (2013b). The importance of interaural time differences and level differences in spatial release from masking. *J. Acoust. Soc. Am*, *134*(2), EL147–EL152. https://doi.org/10.1121/1.4812441

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., Goj, R., et al. (2013). "MEG and EEG data analysis with MNE-Python," *Front. Neurosci*. https://doi.org/10.3389/fnins.2013.00267. Retrieved from.

Gray, W. O., Mayo, P. G., Goupell, M. J., & Brown, A. D. (2021). Transmission of Binaural Cues by Bilateral Cochlear Implants: Examining the Impacts of Bilaterally Independent Spectral Peak-Picking, Pulse Timing, and Compression. *Trends in Hearing*, *25*, Article 23312165211030411. https://doi.org/10.1177/23312165211030411

Higgins, N. C., McLaughlin, S. A., Rinne, T., & Stecker, G. C. (2017). Evidence for cue-independent spatial representation in the human auditory cortex during active listening. *Proceedings of the National Academy of Sciences*, *114*(36), E7602–E7611. https://doi.org/10.1073/pnas.1707522114

Ihlefeld, A., & Shinn-Cunningham, B. (2008). Disentangling the effects of spatial cues on selection and formation of auditory objectsa). *J. Acoust. Soc. Am*, *124*(4), 2224–2235. https://doi.org/10.1121/1.2973185

Jäncke, L., & Shah, N. J. (2002). Does dichotic listening probe temporal lobe functions? *Neurology*, *58*(5), 736–743. https://doi.org/10.1212/wnl.58.5.736

Johnstone, P. M., & Litovsky, R. Y. (2006). Effect of masker type and age on speech intelligibility and spatial release from masking in children and adults. *J. Acoust. Soc. Am*, *120*(4), 2177–2189. https://doi.org/10.1121/1.2225416

Kidd, G., Jr, Best, V., & Mason, C. R. (2008). Listening to every other word: examining the strength of linkage variables in forming streams of speech. *J. Acoust. Soc. Am*, *124*(6), 3793–3802. https://doi.org/10.1121/1.2998980

Kidd, G., Jr, Mason, C. R., Best, V., & Marrone, N. (2010). Stimulus factors influencing spatial release from speech-on-speech masking. *J. Acoust. Soc. Am*, *128*(4), 1965–1978. https://doi.org/10.1121/1.3478781

Klein-Hennig, M., Dietz, M., Hohmann, V., & Ewert, S. D. (2011). The influence of different segments of the ongoing envelope on sensitivity to interaural time delays. *J. Acoust. Soc. Am*, *129*(6), 3856–3872. https://doi.org/10.1121/1.3585847

Kocsis, L., Herman, P., & Eke, A. (2006). The modified Beer-Lambert law revisited. *Phys. Med. Biol*, *51*(5), N91–N98. https://doi.org/10.1088/0031-9155/51/5/N02

Kong, L., Michalka, S. W., Rosen, M. L., Sheremata, S. L., Swisher, J. D., Shinn-Cunningham, B. G., & Somers, D. C. (2014). Auditory spatial attention representations in the human cerebral cortex. *Cereb. Cortex*, *24*(3), 773–784. https://doi.org/10.1093/cercor/bhs359

Kubovy, M. (1988). Should we resist the seductiveness of the space:time::vision:audition analogy? *J. Exp. Psychol. Hum. Percept. Perform*, *14*(2), 318–320. https://doi.org/10.1037/0096-1523.14.2.318

Kubovy, M., & Van Valkenburg, D. (2001). Auditory and visual objects. *Cognition*, *80*(1-2), 97–126. https://doi.org/10.1016/s0010-0277(00)00155-4

Lawrence, R. J., Wiggins, I. M., Anderson, C. A., Davies-Thompson, J., & Hartley, D. E. H. (2018). Cortical correlates of speech intelligibility measured using functional near-infrared spectroscopy (fNIRS). *Hear. Res*, *370*, 53–64. https://doi.org/10.1016/j.heares.2018.09.005

Luke, R., Larson, E., Shader, M. J., Innes-Brown, H., Van Yper, L., Lee, A. K. C., Sowman, P. F., & McAlpine, D. (2021). Analysis methods for measuring passive auditory fNIRS responses generated by a block-design paradigm. *Neurophotonics*, *8*(2), Article 025008. https://doi.org/10.1117/1.NPh.8.2.025008

Michalka, S. W., Kong, L., Rosen, M. L., Shinn-Cunningham, B. G., & Somers, D. C. (2015). Short-Term Memory for Space and Time Flexibly Recruit Complementary Sensory-Biased Frontal Lobe Attention Networks. *Neuron*, *87*(4), 882–892. https://doi.org/10.1016/j.neuron.2015.07.028

Middlebrooks, J. C. (2017). Spatial Stream Segregation. In J. C. Middlebrooks, J. Z. Simon, A. N. Popper, & R. R. Fay (Eds.), *The Auditory System at the Cocktail Party* (pp. 137–168): Springer International Publishing. 10.1007/978-3-319-51662-2_6.

Middlebrooks, J. C., & Waters, M. F. (2020). Spatial Mechanisms for Segregation of Competing Sounds, and a Breakdown in Spatial Hearing. *Front. Neurosci*, *14*, Article 571095. https://doi.org/10.3389/fnins.2020.571095

Moore, B. C. J. (2008). Basic auditory processes involved in the analysis of speech sounds. *Philos. Trans. R. Soc. Lond. B Biol. Sci*, *363*(1493), 947–963. https://doi.org/10.1098/rstb.2007.2152

Moore, B. C. J., & Gockel, H. (2002). Factors Influencing Sequential Stream Segregation. *Acta Acustica united with Acustica*, *88*, 320–333. Retrieved from. https://www.ingentaconnect.com/content/dav/aaua/2002/00000088/00000003/art00004

Noyce, A. L., Kwasa, J. A. C., & Shinn-Cunningham, B. G. (2023). Defining attention from an auditory perspective. *Wiley Interdiscip. Rev. Cogn. Sci*, *14*(1), Article e1610. https://doi.org/10.1002/wcs.1610

Noyce, A. L., Lefco, R. W., Brissenden, J. A., Tobyne, S. M., Shinn-Cunningham, B. G., & Somers, D. C. (2022). Extended Frontal Networks for Visual and Auditory Working Memory. *Cereb. Cortex*, *32*(4), 855–869. https://doi.org/10.1093/cercor/bhab249

Pimentel, T., Cotterell, R., & Roark, B. (2021). *Disambiguatory Signals are Stronger in Word-initial Positions arXiv [cs.CL]*. 10.48550/arXiv.2102.02183.

Pollonini, L., Olds, C., Abaya, H., Bortfeld, H., Beauchamp, M. S., & Oghalai, J. S. (2014). Auditory cortex activation to natural speech and simulated cochlear implant speech measured with functional near-infrared spectroscopy. *Hear. Res*, *309*, 84–93. https://doi.org/10.1016/j.heares.2013.11.007

Quaresima, V., & Ferrari, M. (2019). Functional Near-Infrared Spectroscopy (fNIRS) for Assessing Cerebral Cortex Function During Human Behavior in Natural/Social Situations: A Concise Review. *Organizational Research Methods*, *22*(1), 46–68. https://doi.org/10.1177/1094428116658959

R Core Team. (2024). R: A Language and Environment for Statistical Computing. *MSOR connections*, *1*, 23–25. Retrieved from. https://www.R-project.org/

Richardson, B. N., Kainerstorfer, J. M., Shinn-Cunningham, B. G., & Brown, C. A. (2025). Magnified interaural level differences enhance binaural unmasking in bilateral cochlear implant users. *J. Acoust. Soc. Am*, *157*(2), 1045–1056. https://doi.org/10.1121/10.0034869

Saager, R. B., & Berger, A. J. (2005). Direct characterization and removal of interfering absorption trends in two-layer turbid media. *J. Opt. Soc. Am. A Opt. Image Sci. Vis*, *22*(9), 1874–1882. https://doi.org/10.1364/josaa.22.001874

Scholkmann, F., Metz, A. J., & Wolf, M. (2014). Measuring tissue hemodynamics and oxygenation by continuous-wave functional near-infrared spectroscopy--how robust are the different calculation methods against movement artifacts? *Physiol. Meas*, *35*(4), 717–734. https://doi.org/10.1088/0967-3334/35/4/717

Shinn-Cunningham, B. G. (2008a). Object-based auditory and visual attention. *Trends Cogn. Sci*, *12*(5), 182–186. https://doi.org/10.1016/j.tics.2008.02.003

Shinn-Cunningham, B. G. (2008b). Object-based auditory and visual attention. *Trends Cogn. Sci*, *12*(5), 182–186. https://doi.org/10.1016/j.tics.2008.02.003

Singmann, H., Bolker, B., Westfall, J., & Aust, F. (2018). "afex: Analysis of Factorial Experiments. *R package version 0.21-2*. https://CRAN.R-project.org/package=afex. Retrieved from. https://CRAN.R-project.org/package=afex

SoundStretch Audio Processing Utility. (2017). SoundStretch Audio Processing Utility. (date last viewed: 21-Mar-25). https://www.surina.net/soundtouch/soundstretch.html

Sun, Y., & Poeppel, D. (2023). Syllables and their beginnings have a special role in the mental lexicon. *Proc. Natl. Acad. Sci. U. S. A*, *120*(36), Article e2215710120. https://doi.org/10.1073/pnas.2215710120

Van Rossum, G., & Drake, F. L., Jr (2009). Python 3 Reference Manual: (Python Documentation Manual Part 2). *Createspace*, 244. Retrieved from. https://books.google.com/books/about/Python_3_Reference_Manual.html?id=KIybQQAACAAJ

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., & SciPy 1.0 Contributors. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods*, *17*(3), 261–272. https://doi.org/10.1038/s41592-019-0686-2

Wightman, F. L., & Kistler, D. J. (1992). The dominant role of low-frequency interaural time differences in sound localization. *J. Acoust. Soc. Am*, *91*(3), 1648–1661. https://doi.org/10.1121/1.402445

Yang, Z., & Mayer, A. R. (2014). An event-related FMRI study of exogenous orienting across vision and audition: Cross-Modal Orienting. *Hum. Brain Mapp*, *35*(3), 964–974. https://doi.org/10.1002/hbm.22227

Yücel, M. A., Luke, R., Mesquita, R. C., von Lühmann, A., Mehler, D. M. A., Lührs, M., Gemignani, J., et al. (2024). *The fNIRS reproducibility study hub (FRESH): Exploring variability and enhancing transparency in fNIRS neuroimaging research BITSS*. 10.31222/osf.io/pc6x8.

Yost, W. A., & Dye, R. H., Jr (1988). Discrimination of interaural differences of level as a function of frequency. *J. Acoust. Soc. Am*, *83*(5), 1846–1851. https://doi.org/10.1121/1.396520

Zhang, M., Alamatsaz, N., & Ihlefeld, A. (2021a). Hemodynamic Responses Link Individual Differences in Informational Masking to the Vicinity of Superior Temporal Gyrus. *Front. Neurosci*, *15*, Article 675326. https://doi.org/10.3389/fnins.2021.675326

Zhang, M., Alamatsaz, N., & Ihlefeld, A. (2021b). *Hemodynamic responses link individual differences in informational masking to the vicinity of superior temporal gyrus* p. 2020.08.21.261222. Retrieved from. https://doi.org/10.1101/2020.08.21.261222v2

Zhang, M., Mary Ying, Y.-L., & Ihlefeld, A. (2018). Spatial Release From Informational Masking: Evidence From Functional Near Infrared Spectroscopy. *Trends Hear*, *22*, Article 2331216518817464. https://doi.org/10.1177/2331216518817464

Zhou, X., Burg, E., Kan, A., & Litovsky, R. Y. (2022a). Investigating effortful speech perception using fNIRS and pupillometry measures. *Current Research in Neurobiology*, *3*, Article 100052. https://doi.org/10.1016/j.crneur.2022.100052

Zhou, X., Sobczak, G. S., McKay, C. M., & Litovsky, R. Y. (2022b). Effects of degraded speech processing and binaural unmasking investigated using functional near-infrared spectroscopy (fNIRS). *PLoS One*, *17*(4), Article e0267588. https://doi.org/10.1371/journal.pone.0267588

Zimeo Morais, G. A., Balardin, J. B., & Sato, J. R. (2018). fNIRS Optodes' Location Decider (fOLD): a toolbox for probe arrangement guided by brain regions-of-interest. *Sci. Rep*, *8*(1), 3341. https://doi.org/10.1038/s41598-018-21716-z