Impoverished auditory cues fail to engage brain networks controlling spatial selective attention

Yuqi Deng[1], Inyong Choi[2], Barbara Shinn-Cunningham[1, 3], Robert Baumgartner[1, 4, *]

[1] Biomedical Engineering, Boston University, Boston, MA 02215, USA

[2] Communication Sciences & Disorders, University of Iowa, Iowa City, IA 52242, USA

[3] Neuroscience Institute, Carnegie Mellon University, Pittsburgh, PA 15213 USA

[4] Acoustics Research Institute, Austrian Academy of Sciences, Vienna, Austria

[*] Corresponding Author: robert.baumgartner@oeaw.ac.at

Author contributions: R.B., B.S.-C., Y.D., and I.C. designed research; Y.D. and R.B. performed research; Y.D. and R.B. analyzed data; and Y.D., R.B., B.S.-C., and I.C. wrote the paper.

Abstract

Spatial selective attention enables listeners to process a signal of interest in natural settings. However, most past studies on auditory spatial attention used impoverished spatial cues: presenting competing sounds to different ears, using only interaural differences in time (ITDs) and/or intensity (IIDs), or using non-individualized head-related transfer functions (HRTFs). Here we tested the hypothesis that impoverished spatial cues impair spatial auditory attention by only weakly engaging relevant cortical networks. Eighteen normal-hearing listeners reported the content of one of two competing syllable streams simulated at roughly +30 ˚ and -30˚ azimuth. The competing streams consisted of syllables from two different-sex talkers. Spatialization was based on natural spatial cues (individualized HRTFs), individualized IIDs, or generic ITDs. We measured behavioral performance as well as electroencephalographic markers of selective attention. Behaviorally, subjects recalled target streams most accurately with natural cues. Neurally, spatial attention significantly modulated early evoked sensory response magnitudes only for natural cues, not in conditions using only ITDs or IIDs. Consistent with this, parietal oscillatory power in the alpha band (8-14 Hz; associated with filtering out distracting events from unattended directions) showed significantly less attentional modulation with isolated spatial cues than with natural cues. Our findings support the hypothesis that spatial selective attention networks are only partially engaged by impoverished spatial auditory cues. These results not only suggest that studies using unnatural spatial cues underestimate the neural effects of spatial auditory attention, they also illustrate the importance of preserving natural spatial cues in assistive listening devices to support robust attentional control.

Keywords: auditory spatial selective attention, head-related transfer functions, electroencephalography

Introduction

Spatial hearing is crucial to selectively attend to sounds of interest in everyday social settings. The remarkable ability of normal-hearing listeners to focus on a sound source within a complex acoustic scene is often referred to as "the cocktail party phenomenon," and has a rich history (Cherry, 1953). Nevertheless, the mechanisms controlling spatial selective attention are still poorly understood. Acoustically, in everyday situations, the two ears provide the listener with a listener-specific combination of spatial cues that include interaural time and intensity differences (ITDs and IIDs, respectively), as well as spectral cues caused by acoustical filtering of the pinnae (Blauert, 1997). Together, these cues, captured by individualized head-related transfer functions (HRTFs), allow the brain to create a clear, punctate internal representation of the location of sound sources in the environment (Majdak et al., 2019; Middlebrooks, 2015).

When only isolated or impoverished spatial cues are present, auditory localization performance degrades and the natural perception of external auditory objects may even collapse into the listener's head (e.g., Baumgartner et al., 2017; Callan et al., 2013a; Hartmann and Wittenberg, 1996). Nevertheless, isolated ITDs and IIDs still create a strong sense of lateralization within the head; moreover, even such highly impoverished spatial cues can be used to achieve spatial release from speech-on-speech masking, behaviorally (Culling et al., 2004; Ellinger et al., 2017; Glyde et al., 2013; Kidd et al., 2010b; Loiselle et al., 2016). The relative importance of ITDs and IIDs in spatial release from masking remains unclear, with past studies reporting conflicting results when directly comparing different binaural conditions (Ellinger et al., 2017; Glyde et al., 2013; Higgins et al., 2017; Shinn-Cunningham et al., 2005).

Previous electroencephalography (EEG) and magnetoencephalography (MEG) studies have demonstrated that rich spatial cues in sound stimuli lead to different cortical activity compared to using isolated cues during sound localization (Callan et al., 2013a; Leino et al., 2007; Palomäki et al., 2005) and auditory motion processing (Getzmann and Lewald, 2010). However, the apparently minor behavioral consequences of using unnatural, non-individualized spatial cues on spatial release from masking, combined with the ease of implementing studies with simple, non-individualized spatial cues, led to their wide usage in auditory neuroscience studies (Cusack et al., 2001; Dahmen et al., 2010; Dai et al., 2018; Itoh et al., 2000; Kong et al., 2014; Sach et al., 2000). Indeed, in the auditory neuroscience literature, many studies did not even present true binaural signals, but instead studied "spatial"

attention by using dichotic signals, with one sound presented monaurally to one ear, and a competing sound presented monaurally to the other ear (e.g., Ahveninen et al., 2011; Alho et al., 1999; Wöstmann et al., 2016). These studies implicitly assumed that because listeners were able to use impoverished spatial cues to listen to one sound from a particular (relative) direction, the cognitive networks responsible for controlling spatial attention must be engaged just as they are when listening to rich, natural spatial cues. Nonetheless, it is unclear whether and how engagement of higher-order cognitive processes such as deployment of selective attention is affected by the use of unnatural or impoverished spatial cues.

Modulation of neural signatures, such as event-related potentials (ERPs) and induced oscillatory activity, is often taken as evidence of effective attentional control (Herrmann and Knight, 2001; Siegel et al., 2012). In particular, auditory spatial attention is known to modulate early sensory ERPs in the N1 time range (processing latencies of 100 to 150 ms) (Choi et al., 2013; Röder et al., 1999), whereas modulation of P1 ERPs (50 to 100 ms) has only recently been demonstrated in a free field experiment (Giuliano et al., 2014). Induced alpha oscillation (8 to 14 Hz) has been hypothesized to function as an information gating mechanism (Klimesch et al., 2007). During auditory spatial attention, parietal alpha power often decreases in the contralateral hemisphere of attended stimuli and/or increases in the ipsilateral hemisphere (Banerjee et al., 2011; Lim et al., 2015; Wöstmann et al., 2016). These neural modulations constitute objective metrics of the efficacy of attentional control.

Here, we test listeners in a selective attention paradigm with simultaneous, spatially separated talkers. We use the aforementioned EEG measures to compare both perceptual ability and the neural signatures of attentional control for simulations with impoverished vs. natural spatial cues. Eighteen subjects performed an auditory spatial attention task with two competing streams located at roughly +30 ˚ and -30˚ azimuth (Figure 1). On every trial, listeners were cued by an auditory cue to attend to either the left or right stream and report the content of the cued stream. The competing streams consisted of syllables (/ba/, /da/ or /ga/) from two different-sex talkers. Sound stimuli (including the cuing sound) were spatialized using three different levels of naturalness and richness: 1) generic ITDs only, 2) individualized IIDs, or 3) individualized HRTFs containing all of the naturally occurring spatial cues a listener experiences in the everyday world. We show that behavioral performance is better when listeners hear natural, individualized spatial cues than when they hear impoverished cues. Importantly, only natural spatial cues yield significant attentional modulation of P1 amplitudes. Moreover, induced alpha activity is less robust and poorly lateralized with isolated spatial cues compared to rich, natural spatial cues.

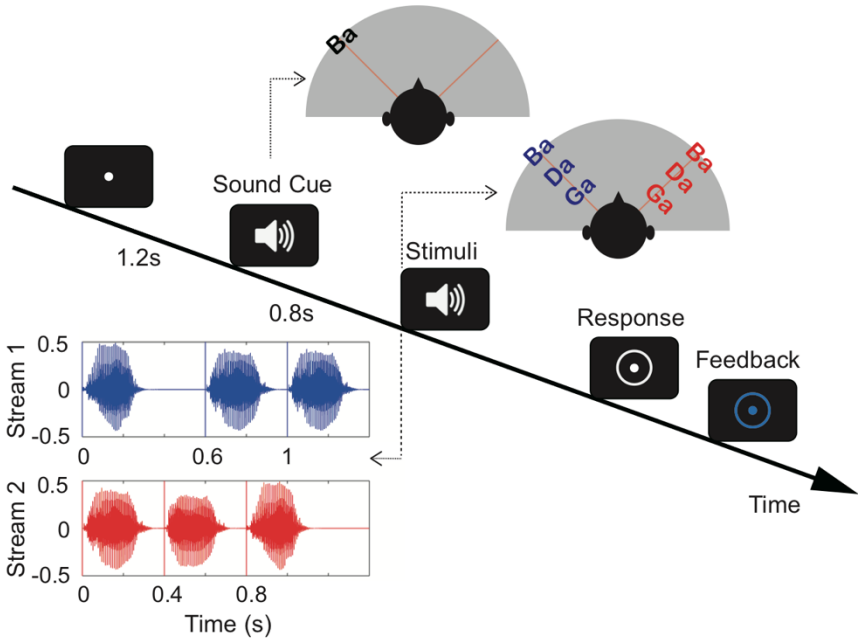

Fig. 1. *Auditory spatial attention task with two competing streams was used to assess the consequence of impoverished auditory spatial cues on neural proxies of attention control. An auditory cue was presented first from the location of the upcoming target stream, processed by the same spatialization scheme as the upcoming mixture. Following the cue, the competing streams began, one from around +30° the other from around -30° azimuth. Listeners were asked to recall the syllable sequence presented from the cued side. The first syllables of both streams were temporally aligned; however, the latter two syllables in the competing streams were staggered, enabling us to isolate neural responses to each. Feedback was provided after every trial.*

Results

Natural spatial cues facilitate behavioral performance. We compared the percentage of correctly recalling each syllable of the target stream across the three spatialization conditions (Figure 2). Paired t-tests were performed on the logit transformed percentages. For the first syllable, where the target sound and the distractor sound overlapped in time, subjects were least accurate in discriminating the target sound in the ITD condition compared to the IID condition ($t_{(17)} = 5.15$, $P < 0.001$) and HRTF condition ($t_{(17)} = 7.09$, $P < 0.001$). However, no statistically significant difference was observed between IID and HRTF conditions for the first syllable ($t_{(17)} = 1.4$, $P = 0.18$). For the second and the third syllable, where target and distractor streams occurred staggered in time, subjects performed significantly better in the HRTF condition than in both the ITD condition (2nd syllable: $t_{(17)} = 2.78$, $P = 0.013$; 3rd syllable:

$t_{(17)} = 2.68$, $P = 0.016$) and the IID condition (2nd syllable: $t_{(17)} = 3.15$, $P < 0.01$; 3rd syllable: $t_{(17)} = 2.15$, $P = 0.046$). There was no significant difference between the ITD and IID conditions for the two staggered syllables (2nd syllable: $t_{(17)} = 0.49$, $P = 0.63$; 3rd syllable: $t_{(17)} = 1.74$, $P = 0.099$).

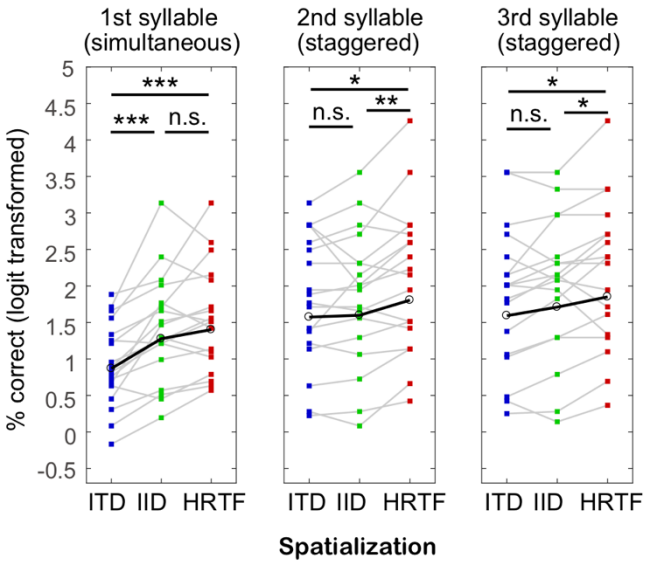

Fig. 2. *Listeners' (N = 18) recall performance was evaluated for every syllable and different spatialization conditions. Sounds were spatialized either based on generic ITDs, individualized IIDs, or the natural combination of ITDs, IIDs, and spectral cues in individualized HRTFs. Behavioral advantages of having more consistent spatial information were significant but small. * P < .05; ** P < .001; *** P < .0001*

Impoverished spatial cues affect attentional modulation of ERPs. Figure 3A shows the ERPs evoked by the onset of the second syllable of the attended target sound and the unattended distractor sound, aligning the onsets of the target and distractor syllables to 0 s to allow direct comparison. Stimulus onsets elicited a fronto-central positivity (P1) between 50 to 100 ms followed by a negativity (N1) between 100 to 180 ms (Figure 3A-B). The amplitudes of these two components were extracted and the difference between attended stimuli (target sound) and unattended stimuli (distractor sound) was calculated in order to quantify attentional modulation for both the P1 and N1 components (Figure 3C).

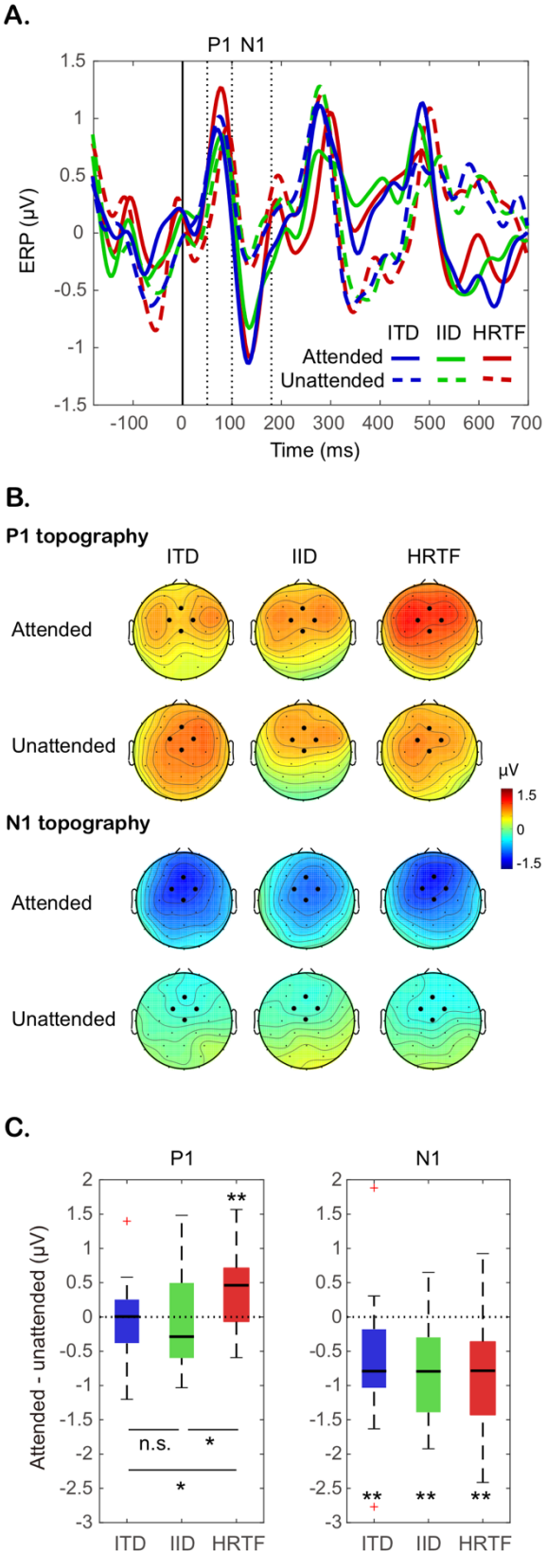
A.
P1 N1
ERP (μV)
Time (ms)
ITD IID HRTF
Attended
Unattended
B.
P1 topography
ITD
IID
HRTF
Attended
Unattended
μV
1.5
0
-1.5
N1 topography
Attended
Unattended
C.
P1
N1
Attended - unattended (μV)
n.s.
*
**
ITD IID HRTF

Fig. 3. *P1 amplitudes were only modulated by the direction of attention in the HRTF condition, whereas N1 amplitudes were modulated equivalently across spatialization conditions (N = 18).* ***A.*** *ERP waveforms at fronto-central electrodes were compared between the attended target stream and the unattended distractor stream for every spatialization condition. The P1 time range was defined between 50 ms and 100 ms, and the N1 time range was defined between 100 ms and 180 ms.* ***B.*** *Most topographies of both ERP components show maxima at the fronto-central sites (black dots) used for evaluation.* ***C.*** *The modulation strength of ERP components was assessed by the amplitude differences between attended and unattended streams.*

P1 responses to attended stimuli were significantly larger than to unattended stimuli in the HRTF condition ($t_{(17)} = 3.12$, $P < 0.01$). However, no significant attentional modulation was found in the ITD ($t_{(17)} = 0.50$, $P = 0.62$) and IID conditions ($t_{(17)} = 0.06$, $P = 0.95$). Consistent with this, a direct comparison of the amount of modulation (difference in ERP magnitudes for attended versus unattended sounds) found that attentional modulation was significantly larger in the HRTF condition than in the ITD ($t_{(17)} = 2.48$, $P = 0.024$) and IID conditions ($t_{(17)} = 2.29$, $P = 0.035$); however, modulation did not differ significantly between the ITD and IID conditions ($t_{(17)} = 0.27$, $P = 0.79$) (Figure 3C).

In all three spatialization conditions, the N1 amplitude was modulated by spatial attention. An attended sound elicited larger N1 amplitudes than when a sound was ignored in all three conditions (ITD: $t_{(17)} = 3.01$, $P < 0.01$; IID: $t_{(17)} = 4.12$, $P < 0.001$; HRTF: $t_{(17)} = 3.56$, $P < 0.01$). The magnitude of this N1 modulation did not differ significantly across the three spatialization conditions (ITD vs IID: $t_{(17)} = 0.32$, $P = 0.76$; ITD vs HRTF: $t_{(17)} = 0.21$, $P = 0.84$; HRTF vs IID: $t_{(17)} = 0.13$, $P = 0.90$) (Figure 3C).

Alpha oscillation power shows less attentional modulation with impoverished spatial cues. To investigate the effect of spatialization on attentional control, we analyzed the power in alpha oscillations during the attentional preparation period (-800 ms to 0 ms), a time period in which listeners know where to orient spatial attention based on the preceding acoustic cue, but before the sound mixture of competing streams begins. We averaged the power in alpha across all trials for each spatialization condition, regardless of where spatial attention was focused, to get a measure of the total engagement of alpha activity. We then compared relative power for different attentional directions. Averaged across direction of attentional focus, we calculated the time courses of alpha global field power (GFP, Figure 4A) and compared within-subject differences of the temporal average within the preparatory time period across spatialization conditions (Figure 4B). We found significantly larger alpha GFP differences in

the HRTF condition than in the other two conditions (HRTF vs ITD: $t_{(17)} = 2.75$, $P = 0.014$; HRTF vs IID: $t_{(17)} = 2.63$, $P = 0.017$). No significant difference was found between the ITD and IID conditions ($t_{(17)} = 0.021$, $P = 0.98$).

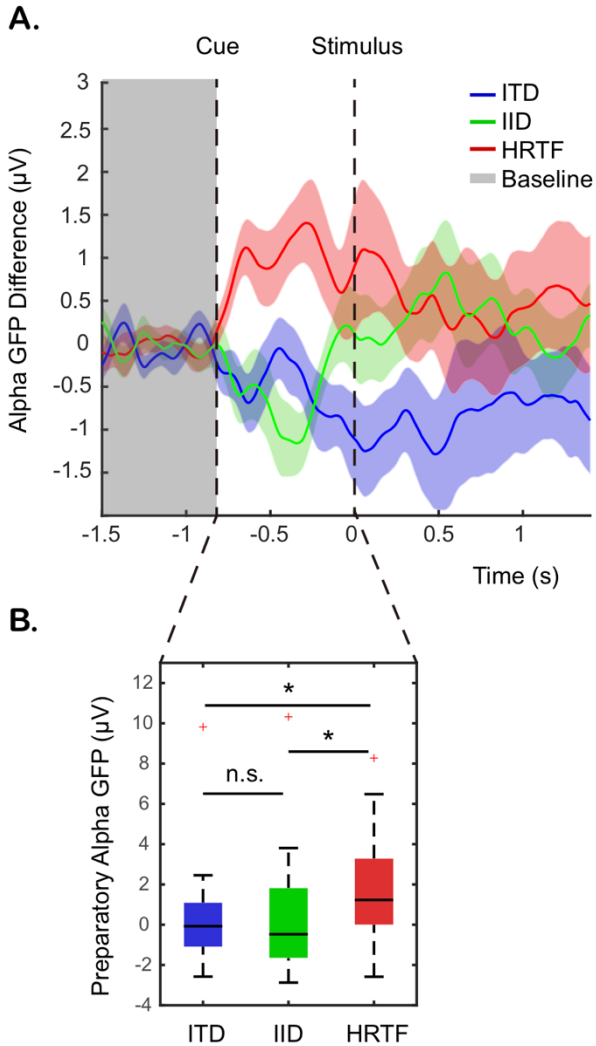

Fig. 4. *Within-subject differences in alpha-band GFP are larger in the HRTF condition, especially during the preparatory time window (after the sound cue but before the first syllables of the competing streams).* ***A.*** *Waveforms of the average (± SEM) GFP differences are shown during the baseline period, preparatory phase, and stimulus phase with stream competition.* ***B.*** *The temporal average of the preparatory alpha GFP difference is larger for the HRTF condition.*

We next explored whether the lateralization of alpha with the spatial focus of attention was greater with more natural spatial cues than with impoverished cues. To this end, for each spatialization condition we calculated the alpha power "attentional modulation index" (AMI), defined as the alpha power difference between attend left and attend right trials divided by the overall alpha power (Wöstmann et al., 2016). Calculated in this way, the AMI is expected to be positive in left and negative in right parietal channels. Scalp topographies of the resulting AMI are shown in Figure 5. The alpha AMI was then averaged within the left and the right parietal channels (see Methods). Statistical comparisons between left- and right-hemisphere AMIs were made within conditions to quantify lateralization of alpha power. Statistically significant hemispheric differences were found in the HRTF condition only ($t_{(17)} = 3.09$, $P < 0.01$), not in the ITD ($t_{(17)} = 1.29$, $P = 0.21$) or IID conditions ($t_{(17)} = 0.15$, $P = 0.88$). To compare alpha lateralization directly across conditions, the hemispheric difference in alpha AMI was calculated by subtracting average right-hemisphere AMI from the average left-hemisphere AMI. Direct comparison of conditions shows that alpha lateralization was significantly greater in the HRTF condition than in the ITD condition ($t_{(17)} = 2.18$, $P = 0.043$); alpha lateralization in the IID condition did not differ significantly from that of either the HRTF condition ($t_{(17)} = 1.82$, $P = 0.086$) or the ITD condition ($t_{(17)} = 0.72$, $P = 0.48$).

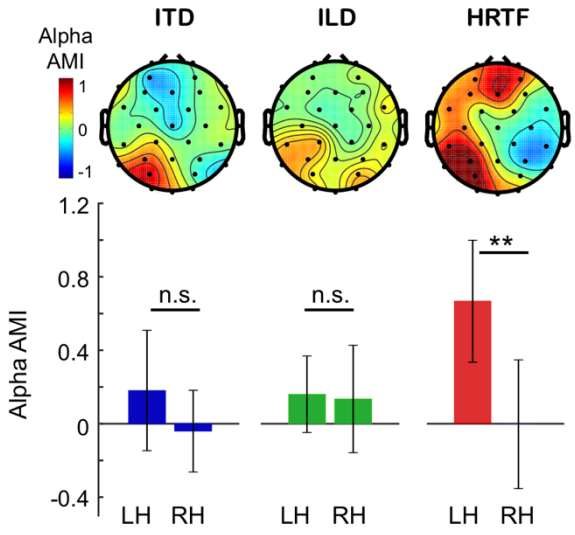

Fig. 5. *Attentional modulation of alpha activity was lateralized to the hemisphere ipsilateral to the target stream only in the HRTF condition. AMI topographies and hemispheric averages are shown for every spatialization condition (N = 18). ** P < .001*

Impoverished spatial cues lead to worse behavioral performance, smaller P1 modulation, reduced modulation of preparatory alpha GFP, and reduced lateralization of alpha power with attentional focus, confirming our hypothesis that impoverished spatial cues impaired engagement of spatial attention. We then explored relationships between these metrics to see if there were ordered relationships in the individual measures of performance and neural signatures of attentional control. To investigate the relationship between evoked response modulation and alpha oscillatory activities, we first calculated the regression slope relating P1 amplitude to preparatory alpha GFP for each subject, and then performed a paired t-test on the coefficients obtained. No consistent relationship between alpha GFP and P1 amplitudes was observed ($t_{(17)} = 0.90$, $P = 0.38$). Correlation analysis was also conducted comparing behavioral accuracy to P1 modulation, defined as the attended P1 amplitude minus unattended P1 amplitude. No consistent relationships between P1 modulation and behavioral performance were observed for any syllable (1st syllable: $t_{(17)} = 0.54$, $P = 0.59$; 2nd syllable: $t_{(17)} = 0.31$, $P = 0.76$; 3rd syllable: $t_{(17)} = 0.69$, $P = 0.50$). Similarly, we did not observe consistent relationships between alpha AMI lateralization and response accuracy for any syllable (1st syllable: $t_{(17)} = 0.19$, $P = 0.85$; 2nd syllable: $t_{(17)} = 1.39$, $P = 0.18$; 3rd syllable: $t_{(17)} = 0.11$, $P = 0.91$). Thus, although there were significant differences in engagement of attention across spatial conditions as measured both behaviorally and neurally, the individual subject differences in these metrics were not closely related.

Discussion

Behaviorally, we found that impoverished spatial cues impair performance on an auditory spatial attention task in a multi-talker scene. We used objective electrophysiological measures to assess whether the naturalness and richness of spatial cues also impacts how strongly auditory spatial attention modulates brain responses. We found that impoverished spatial cues reduce the strength of the evoked and induced neural signatures of attentional control. Specifically, evoked P1 amplitudes and induced alpha oscillatory power showed less attentional modulation for sound stimuli with impoverished spatial cues compared to when spatial cues were tailored to recreate the natural, rich experience of individual listeners.

Impoverished spatial cues result in less neural modulation during selective attention. We investigated attentional modulation of four established neural signatures of selective attention: evoked P1 and N1 amplitudes and induced power and lateralization of alpha oscillation. While attentional modulation of N1 amplitude was observed in all conditions, attentional modulation of the earlier P1 amplitude was not

214 observed or was significantly weaker in the impoverished cue conditions compared to the natural cue
215 condition. Similarly, we found less preparatory alpha power activity in the impoverished spatial cue
216 conditions than in the natural cue condition, reflected by two indexes quantifying the amount of spatial
217 variability of alpha power: alpha GFP (Figure 4) and AMI (Figure 5). In the ITD and IID conditions,
218 although there was a hint of preparatory alpha lateralization over parietal sensors, the amount of
219 lateralization was significantly smaller than in the HRTF condition and did not reach statistical
220 significance. Preparatory alpha activity during spatial attention tasks has been well documented to form
221 a specific lateralization pattern in both vision and audition (Banerjee et al., 2011; Kelly, 2006; Sauseng et
222 al., 2005; Worden et al., 2000), which is thought to be evidence of a preparatory information-gating
223 mechanism (Foxe and Snyder, 2011; Jensen and Mazaheri, 2010; Klimesch, 2012; Klimesch et al., 2007).
224 In vision, alpha lateralization has been observed to increase with the laterality of attention focus (Rihs et
225 al., 2007a; Samaha et al., 2016), reflecting an inhibition pattern topographically specific to attention
226 focus. Moreover, evidence for active top-down control of the phase of alpha oscillation during visual
227 spatial attention suggests that alpha oscillatory activity represents active engagement and
228 disengagement of the attentional network (Samaha et al., 2015). In addition, a previous somatosensory
229 study has revealed that the alpha lateralization is positively correlated to pre-stimulus cue reliability,
230 further suggesting that alpha lateralization reflects the top-down control in order to optimize the
231 processing of upcoming stimuli (Haegens et al., 2011). Although relatively few studies have investigated
232 alpha activity in audition, studies suggest that alpha control mechanisms are supra-modal rather than
233 sensory specific (33). In the current experiment, a pre-stimulus auditory cue directed listeners where to
234 focus attention in an upcoming sound mixture. The cue was spatialized using the same auditory features
235 used to spatialize the stream mixture. Our results thus suggest that compared to stimuli with natural
236 spatial cues, stimuli featuring only ITDs or only IIDs are less reliable in directing attentional focus,
237 producing weaker engagement of spatial attention and reduced attentional modulation of neural
238 responses.

239 Consistent with the idea that impoverished spatial cues lead to weaker engagement of spatial attention,
240 we found that the P1 ERP component was modulated by attention only with natural spatial cues, not
241 with impoverished cues; this result is consistent with a weak spatial representation failing to engage
242 attentional modulation of early sensory responses (Figure 3). Our finding that attentional focus leads to
243 a modulation of P1 amplitude for natural spatial cues is consistent with previous reports of effects of
244 attention on the P1 amplitude observed in previous spatial attention studies across sensory modalities

[auditory: (Giuliano et al., 2014); visual: (Hillyard and Anllo-Vento, 1998; Hopfinger et al., 2004)]. Past studies agree that P1 modulation reflects an early sensory inhibition mechanism related to suppression of task-irrelevant stimuli. Although debates remain as to whether P1 modulation results from bottom-up sensory gain control (Hillyard and Anllo-Vento, 1998; Luck, 1995; Slagter et al., 2016) or for some top-down inhibitory process (Freunberger et al., 2008; Klimesch, 2011a), it is generally accepted in visual spatial studies that greater P1 amplitude modulation is associated with greater inhibition of to-be-ignored stimuli (Couperus and Mangun, 2010; Hillyard and Anllo-Vento, 1998; Klimesch, 2012). Interestingly, attentional modulation of auditory P1 has been found to be positively correlated to visual working memory capacity, a result that was used to suggest that stronger P1 modulation reflects better attentional control of the flow of sensory information into working memory (Fukuda and Vogel, 2009; Giuliano et al., 2014). Our result is consistent with the hypothesis that P1 modulation directly reflects attentional control. Specifically, impoverished spatial cues likely produce a “muddy” representation of auditory space that supports only imprecise, poorly focused top-down spatial attention. The resulting lack of control and specificity of spatial auditory attention results in early P1 responses that are unmodulated by attentional focus.

N1 modulation is well documented as a neural index of attentional control (Choi et al., 2013; Hillyard and Anllo-Vento, 1998; Stevens et al., 2008; Wyart et al., 2012). The attentional modulation of N1 is thought to reflect attentional facilitation rather than inhibition (Couperus and Mangun, 2010; Marzecová et al., 2018; Slagter et al., 2016). In contrast to preparatory alpha and P1, we found that the later N1 evoked response was modulated similarly, regardless of the richness and naturalness of spatial cues.

Due to the robustness and relatively large amount of modulation, changes in auditory N1 amplitude have been used as a biomarker and a primary feature for classification of attentional focus ((Blankertz et al., 2011; Schreuder et al., 2011); see also recent work on decoding attentional focus for running speech using the correlation between neural responses and the power envelope of the speech streams: (Chait et al., 2010; Mesgarani and Chang, 2012; Rimmele et al., 2015)). However, there is little known about how N1 amplitudes reflects the processing of different spatial cues during auditory spatial attention. Previous studies have revealed different N1 topographies during ITD and IID processing, leading to the conclusion that ITD and IID are processed by different neural populations in the auditory cortex (Johnson and Hautus, 2010; Tardif et al., 2006; Ungan et al., 2001). However, debates remain about whether this difference in topography depends on perceived laterality, instead of different neural

populations specialized for processing different spatial cues. Results from a more recent study show that auditory N1 modulation does not differ across spatial cue conditions, indicating integrated processing of sound locations in auditory cortex regardless of cues (Salminen et al., 2015). In the current study, N1 modulation did not differ across the three spatialization conditions. Thus, our results support the idea that the same cortical neural population is responsible for processing different binaural spatial cues.

Behavioral disadvantages associated with impoverished spatial cues are modest and depend on sound stimulus characteristics. Despite the influence of spatial cue richness on neural metrics, our behavioral results showed only small (albeit significant) behavioral differences between impoverished spatial cues and natural, individualized spatial cues (Figure 2). In line with previous studies that observed greater spatial release from masking with combined spatial cues compared to with isolated cues (Culling et al., 2004; Ellinger et al., 2017), accuracy was best in the HRTF condition. The small accuracy improvement over using impoverished cues is seen consistently across subjects. In the first syllable where the target and distractor streams overlap in time, the HRTF condition yielded an increase of 13% points in accuracy over the ITD condition, but is comparable to performance in the IID condition. In the two staggered syllables, accuracy in the HRTF condition is greater than in the ITD and IID conditions by only about 6% and 1% point, respectively. These differences in behavioral performance across syllables suggest that the characteristics of sound stimuli influence the difficulty of the task and may influence the behavioral advantages of having richer, more robust spatial cues (Kidd et al., 2010a). Concordantly, a previous study with complex tone stimuli has shown much larger differences in behavioral performance of up to 20% points (Schröger, 1996), whereas studies presenting speech stimuli in a multi-talker environment found no behavioral advantage of having combined cues compared to impoverished cues (Glyde et al., 2013). These behavioral discrepancies, in combination with our neural findings, indicate that behavioral performance alone is not a sensitive metric for determining whether cortical networks controlling spatial selective attention are fully engaged.

Summary

Our results indicate that although impoverished spatial cues can support spatial segregation of speech in a multi-talker environment, they do not fully engage the brain networks controlling spatial attention and lead to weak attentional control. Previous auditory studies have provided evidence that impoverished spatial cues do not evoke the same neural processing mechanisms as natural cue combinations during localization tasks with single sounds (Callan et al., 2013b; Getzmann and Lewald, 2010; Leino et al.,

2007; Palomäki et al., 2005). The current study extends these findings, demonstrating that the efficacy of higher-level cognitive processing, such as deployment of auditory selective attention, also depends on the naturalness of spatial cues. Poor attentional control was reflected in limited modulation of neural biomarkers of attentional processes. These findings suggest that the many past auditory attention studies using impoverished spatial cues may have underestimated the robust changes in cortical activity associated with deployment of spatial auditory attention in natural settings. Although impoverished auditory spatial cues can allow listeners to deploy spatial attention effectively enough to perform well in simple acoustic scenes, noisy, complex listening environments like those encountered in everyday environments pose greater challenges to attentional processing. In natural settings, spatial attention may fail unless attentional control networks are fully engaged. Thus, these results demonstrate the importance of preserving rich, natural spatial cues in hearing aids and other assistive listening devices.

Materials and Methods

Subjects. Twenty-one paid volunteers and one author within the age of 18-42 years (M = 22.9, SD = 5.5; 12 females, 10 males) participated in this study. None of the subjects had audiometric thresholds greater than 20 dB for frequencies from 250 Hz to 8 kHz. All participants gave informed consent as approved by the Boston University Institutional Review Board. Two subjects were withdrawn from the study due to the inability to perform the task (percentage of correct response less than 30% after training), and two subjects were removed during EEG data preprocessing due to excessive artifacts. Therefore 18 subjects remained for further analysis (N = 18).

Stimuli and Procedure. The sound stimuli consisted of consonant-vowel syllables (/ba/, /da/, & /ga/), each 0.4 s in duration. These syllables were recorded from three talkers that naturally differed in fundamental frequency (F0). The auditory cue was a single syllable /ba/ spoken by a low-pitch male voice (F0 = 91 Hz, estimated by Praat software) (Boersma, 2001). The following target and distractor streams both consisted of three syllables randomly chosen out of the set of three syllables (with replacement). The target stream was spoken by either a female (F0 = 189 Hz) or a high-pitch male talker (F0 = 125 Hz), and the distractor stream was spoken by the other talker different than the target stream. The first syllable of the target and distractor sound overlapped in time, while the latter two syllables were separated by 200 ms, onset to onset (Figure 1). To avoid engagement of temporal attention rather than spatial attention, the assignment of the target stream being leading or lagging was equally distributed across trials. In the leading stream, the onsets of all three syllables were separated by 400 ms; in the lagging stream, the onsets of the first and the second syllable were separated by 600 ms, whereas those of the second and the third syllable were separated by 400 ms. All sound stimuli were presented at a sound pressure level of approximately 75 dB.

Cue and stimuli were presented via earphones (ER-2, Etymotic Research, Inc.) and spatialized to approximately ±30° azimuth (0° elevation). Three different spatialization conditions were used: HRTF, IID, and ITD. In the HRTF condition, individualized HRTFs, providing natural combinations of ITDs, IIDs, and spectral cues, were used. In the IID condition, ITDs were removed from the individualized HRTFs by computing minimum-phase representations of the filters via removing the non-causal part of the cepstrum (Oppenheim and Schafer, 1975). Hence, the IID and HRTF conditions provided the same monaural magnitude spectra and thus the same energetic advantage of the ear ipsilateral to the target.

346 In the ITD condition, spatialization was based on simply delaying the signal presented to the
347 contralateral ear by 300 µs, thus providing no energetic advantage to the ipsilateral ear.

348 The measurement procedure for individualized HRTFs was identical to that described in (Baumgartner et
349 al., 2017). In short, acoustical transfer characteristics between loudspeakers positioned at the target
350 angles 1.5 m away from the subject's head and miniature microphones inside the subject's blocked ear
351 canals were measured inside a semi-anechoic chamber. Room reflections were removed from the
352 resulting impulse responses via temporal windowing (0.5-ms cosine ramps) to a duration of 3 ms.

353 Subjects performed a spatial attention task in a Posner paradigm (Figure 1) (Posner, 1980). Sound
354 spatialization was realized by one of the three spatialization conditions fixed within trials but pseudo-
355 randomized across trials. Subjects were instructed to fixate their gaze on a dot at the center of the
356 screen at the beginning of each trial. The fixation dot lasted for 1.2 s before an auditory cue was
357 presented. The auditory cue came from either left or right indicating where the target sound would
358 come from. A target sound started 0.8 s later from the cued location. At the same time a distractor
359 sound started from the opposite location of the target sound. Subjects were asked to report the syllable
360 sequence of the target sound by pressing a keyboard after the sounds finished and a response cue was
361 shown. Feedback about whether or not they correctly reported the syllables was given at the end of
362 every trial.

363 Each subject performed 450 randomized trials of this task, divided into 9 blocks each consisting of 50
364 trials. In total, every subject performed 150 trials for each of the three sound spatialization conditions
365 (75 trials attending left and 75 trials attending right). Prior to the test sessions, all participants received a
366 practice session to get familiarized with the task. Participants with a percentage of correct response
367 lower than 30% after 3 blocks of training (50 trials per block) were excluded from the study.

368 **EEG Acquisition and Preprocessing.** 32-channel scalp EEG data was recorded (Activetwo system with
369 Activeview acquisition software, Biosemi B.V.) in a sound proof booth (Eckel Industries, Inc.) while
370 subjects were performing the task. Two additional reference electrodes were placed on the earlobes.
371 Horizontal eye movements were recorded by two electrooculography (EOG) electrodes placed on the
372 outer canthi of each eye. Vertical eye movement was recorded by one EOG electrode placed below the
373 right eye. The timing of stimulus was controlled by Matlab (Mathworks) with Psychtoolbox (extension 3)
374 (Brainard, 1997; Delorme and Makeig, 2004). EEG preprocessing was conducted with Eeglab toolbox in
375 Matlab.

EEG data were corrected against the average of the two reference channels. Bad channels were marked by manual selection during recording and automatically detected based on joint probability measures of Eeglab (Delorme et al., 2007). EEG signals were then down-sampled to 256 Hz and epochs containing responses to individual trials were extracted. Each epoch was baseline corrected against 100 ms prior to the cue onset by removing the mean of the baseline period from the whole trial. ICA artifact rejection was performed with Eeglab to remove components of eye movements, blinks, and muscle artifacts (Delorme and Makeig, 2004). The maximum number of independent components rejected for each subject was five. After ICA rejection, bad channels were removed and interpolated. Trials with a maximum absolute value over 80 µV were rejected (Delorme et al., 2007). Two subjects with excessive artifacts were removed from further EEG analysis because less than 50% of trials remained after thresholding. For the rest of the 18 subjects, at least about two thirds of the trials (minimum was 48 out of 75 trials) remained for each condition after artifact rejection. Trial numbers were equalized within and across subjects by randomly selecting the minimum number of available trials (N = 48) for each condition across the whole recording session.

Data analysis. Behavioral performance was quantified by the percentage of correct responses for every one of the three syllables in the target stream and each spatialization condition. Behavioral results were collapsed across the attend-left and attend-right trials. The percentages of correct response were then normalized by logit transformation before parametric statistical testing was performed on the resulting data.

ERP responses were evaluated for the second syllable of the target sound and distractor sound, respectively. The reason we looked at the second syllable only is that 1) the first syllable of the target and distractor aligned in time and therefore the ERPs were inseparable, and 2) the ERP amplitude in response to the third syllable was small, and therefore more contaminated by noise. The preprocessed data (details see EEG preprocessing) was bandpass filtered from 0.5 to 20 Hz by a finite impulse response filter with Kaiser window design ($\beta = 7.2$, $n = 1178$). Data from four fronto-central channels (Cz, Fz, FC1, and FC2) were averaged to get the auditory ERP response. We picked these four channels a priori because auditory ERP responses in sensor space are largest in the fronto-central area of the scalp. To quantify the amplitudes of ERP components, the maximum value within the window of 50 to 100 ms after the second syllable onset was taken to be the P1 amplitude; the minimum value within the window of 100 to 180 ms after the second syllable onset was calculated to be the N1 amplitude. The values extracted from the selected windows were calculated for each channel and plotted onto a 2D scalp map

to generate topography plots. The values of the ERP components from the four selected channels were then averaged and compared across different spatialization conditions.

To get the amplitude of alpha oscillation, the preprocessed EEG data was bandpass filtered to the alpha range (8 to 14 Hz) before a Hilbert transform was applied. The magnitude of the resulting data was taken as the extracted alpha power envelope. To get induced alpha power, the alpha power was calculated for single trials first and then averaged across trials (Snyder and Large, 2005). The time course of alpha power was baseline corrected against 700 ms before the auditory cue onset. GFP (Murray et al., 2008; Skrandies, 1990) constitutes the spatial standard deviation across all scalp electrodes; it has been used as a measurement to quantify the amount of alpha variation across the scalp (Lim et al., 2015). We calculated the time courses of alpha GFP by taking the standard deviation of all electrodes. To quantify the degree of alpha lateralization based on direction of attention, we calculated the Attentional Modulation Index (AMI) of alpha power, defined as the alpha power difference between attended left and attended right trials divided by the overall alpha power (Wöstmann et al., 2016). The AMI of alpha was calculated for each time point, yielding the time course of AMI for each spatialization condition. We then averaged the alpha AMI of each spatialization condition over the 800 ms immediately before stimulus onset (-800 ms to 0 ms, re: onset). This is the period where subjects have been cued to orient their spatial attention in preparation for the target sound, but before the speech streams begin. Scalp topographies of the preparatory alpha AMI were plotted for each condition. Hemispheric lateralization of alpha AMI was further compared across spatialization conditions and evaluated as the difference between the left hemisphere and the right hemisphere.

For hypothesis testing, we compared group level differences between conditions with paired t-tests. Due to the relatively small number of subjects (N=18), the Lilliefors test was performed prior to statistical testing to check normality of the data. Data was considered normally distributed at $P>0.05$.

Data availability. All data and code used to conduct this study and analyze the results will be provided via Zenodo immediately after acceptance of this manuscript for publication.

Acknowledgements

We thank Ashvini Meltoke for collecting part of the data, Gerald Kidd Jr. for providing the facilities for individualized HRTF measurements, and Virginia Best for fruitful discussions. This work was supported by the Austrian Science Fund (grant J3803-N30 to R.B.) and NICD (grant DC013825 to B.S.-C.).

References

Ahveninen, J., Hämäläinen, M., Jääskeläinen, I. P., Ahlfors, S. P., Huang, S., Lin, F.-H., Raij, T., et al. (**2011**). "Attention-driven auditory cortex short-term plasticity helps segregate relevant sounds from noise," *Proc. Natl. Acad. Sci.*, **108**, 4182–4187, doi: 10.1073/pnas.1016134108.

Alho, K., Medvedev, S. V., Pakhomov, S. V., Roudas, M. S., Tervaniemi, M., Reinikainen, K., Zeffiro, T., et al. (**1999**). "Selective tuning of the left and right auditory cortices during spatially directed attention," *Cogn. Brain Res.*, **7**, 335–341, doi: 10.1016/S0926-6410(98)00036-6.

Banerjee, S., Snyder, A. C., Molholm, S., and 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**, 9923–9932, doi: 10.1523/JNEUROSCI.4660-10.2011.

Baumgartner, R., Reed, D. K., Tóth, B., Best, V., Majdak, P., Colburn, H. S., and Shinn-Cunningham, B. (**2017**). "Asymmetries in behavioral and neural responses to spectral cues demonstrate the generality of auditory looming bias," *Proc. Natl. Acad. Sci.*, **114**, 9743–9748, doi: 10.1073/pnas.1703247114.

Blankertz, B., Lemm, S., Treder, M., Haufe, S., and Müller, K.-R. (**2011**). "Single-trial analysis and classification of ERP components — A tutorial," *NeuroImage*, **56**, 814–825, doi: 10.1016/J.NEUROIMAGE.2010.06.048.

Blauert, J. (**1997**). *Spatial hearing. The Psychophysics of Human Sound Localization*, (MIT-Press, Cambridge, MA), pp. 497.

Boersma, P. (**2001**). "Praat, a system for doing phonetics by computer," *Glot Int.*, **5**, 341–345.

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

Callan, A., Callan, D. E., and Ando, H. (**2013**). "Neural correlates of sound externalization," *NeuroImage*, **66**, 22–27, doi: 10.1016/J.NEUROIMAGE.2012.10.057.

Callan, A., Callan, D. E., and Ando, H. (**2013**). "Neural correlates of sound externalization," *NeuroImage*, **66**, 22–27, doi: 10.1016/j.neuroimage.2012.10.057.

Chait, M., de Cheveigné, A., Poeppel, D., and Simon, J. Z. (**2010**). "Neural dynamics of attending and ignoring in human auditory cortex," *Neuropsychologia*, **48**, 3262–3271, doi: 10.1016/J.NEUROPSYCHOLOGIA.2010.07.007.

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

467 Choi, I., Rajaram, S., Varghese, L. A., and Shinn-Cunningham, B. G. (**2013**). "Quantifying attentional
468 modulation of auditory-evoked cortical responses from single-trial electroencephalography,"
469 *Front. Hum. Neurosci.*, **7**, 115, doi: 10.3389/fnhum.2013.00115.
470 Couperus, J. W. and Mangun, G. R. (**2010**). "Signal enhancement and suppression during visual–spatial
471 selective attention," *Brain Res.*, **1359**, 155–177, doi: 10.1016/J.BRAINRES.2010.08.076.
472 Culling, J. F., Hawley, M. L., and Litovsky, R. Y. (**2004**). "The role of head-induced interaural time and
473 level differences in the speech reception threshold for multiple interfering sound sources," *J.*
474 *Acoust. Soc. Am.*, **116**, 1057–1065, doi: 10.1121/1.1772396.
475 Cusack, R., Carlyon, R. P., and Robertson, I. H. (**2001**). "Auditory Midline and Spatial Discrimination in
476 Patients with Unilateral Neglect," *Cortex*, **37**, 706–709, doi: 10.1016/S0010-9452(08)70620-8.
477 Dahmen, J. C., Keating, P., Nodal, F. R., Schulz, A. L., and King, A. J. (**2010**). "Adaptation to Stimulus
478 Statistics in the Perception and Neural Representation of Auditory Space," *Neuron*, **66**, 937–948,
479 doi: 10.1016/J.NEURON.2010.05.018.
480 Dai, L., Best, V., and Shinn-Cunningham, B. G. (**2018**). "Sensorineural hearing loss degrades behavioral
481 and physiological measures of human spatial selective auditory attention," *Proc. Natl. Acad. Sci.*,
482 **115**, E3286--E3295, doi: 10.1073/pnas.1721226115.
483 Delorme, A. and Makeig, S. (**2004**). "EEGLAB: an open source toolbox for analysis of single-trial EEG
484 dynamics including independent component analysis," *J. Neurosci. Methods*, **134**, 9–21, doi:
485 10.1016/j.jneumeth.2003.10.009.
486 Delorme, A., Sejnowski, T., and Makeig, S. (**2007**). "Enhanced detection of artifacts in EEG data using
487 higher-order statistics and independent component analysis," *NeuroImage*, **34**, 1443–1449, doi:
488 10.1016/j.neuroimage.2006.11.004.
489 Ellinger, R. L., Jakien, K. M., and Gallun, F. J. (**2017**). "The role of interaural differences on speech
490 intelligibility in complex multi-talker environments," *J. Acoust. Soc. Am.*, **141**, EL170–EL176, doi:
491 10.1121/1.4976113.
492 Foxe, J. J. and Snyder, A. C. (**2011**). "The role of alpha-band brain oscillations as a sensory suppression
493 mechanism during selective attention," *Front. Psychol.*, **2**, 154, doi: 10.3389/fpsyg.2011.00154.
494 Freunberger, R., Höller, Y., Griesmayr, B., Gruber, W., Sauseng, P., and Klimesch, W. (**2008**). "Functional
495 similarities between the P1 component and alpha oscillations," *Eur. J. Neurosci.*, **27**, 2330–2340,
496 doi: 10.1111/j.1460-9568.2008.06190.x.
497 Fukuda, K. and Vogel, E. K. (**2009**). "Human Variation in Overriding Attentional Capture," *J. Neurosci.*, **29**,
498 8726–8733, doi: 10.1523/JNEUROSCI.2145-09.2009.

499 Getzmann, S. and Lewald, J. (**2010**). "Effects of natural versus artificial spatial cues on
500 electrophysiological correlates of auditory motion," *Hear. Res.*, **259**, 44–54, doi:
501 10.1016/j.heares.2009.09.021.
502 Giuliano, R. J., Karns, C. M., Neville, H. J., and Hillyard, S. A. (**2014**). "Early Auditory Evoked Potential Is
503 Modulated by Selective Attention and Related to Individual Differences in Visual Working
504 Memory Capacity," *J. Cogn. Neurosci.*, doi: 10.1162/jocn_a_00684.
505 Glyde, H., Buchholz, J. M., Dillon, H., Cameron, S., and Hickson, L. (**2013**). "The importance of interaural
506 time differences and level differences in spatial release from masking," *J. Acoust. Soc. Am.*, **134**,
507 EL147–EL152, doi: 10.1121/1.4812441.
508 Haegens, S., Luther, L., and Jensen, O. (**2011**). "Somatosensory Anticipatory Alpha Activity Increases to
509 Suppress Distracting Input," *J. Cogn. Neurosci.*, **24**, 677–685, doi: 10.1162/jocn_a_00164.
510 Hartmann, W. M. and Wittenberg, A. (**1996**). "On the externalization of sound images," *J Acoust Soc Am*,
511 **99**, 3678–3688.
512 Herrmann, C. S. and Knight, R. T. (**2001**). "Mechanisms of human attention: event-related potentials and
513 oscillations," *Neurosci. Biobehav. Rev.*, **25**, 465–476, doi: 10.1016/S0149-7634(01)00027-6.
514 Higgins, N. C., McLaughlin, S. A., Rinne, T., and Stecker, G. C. (**2017**). "Evidence for cue-independent
515 spatial representation in the human auditory cortex during active listening," *Proc. Natl. Acad.*
516 *Sci.*, **114**, E7602--E7611, doi: 10.1073/pnas.1707522114.
517 Hillyard, S. A. and Anllo-Vento, L. (**1998**). "Event-related brain potentials in the study of visual selective
518 attention," *Proc. Natl. Acad. Sci.*, **95**, 781–787, doi: 10.1073/pnas.95.3.781.
519 Hopfinger, J. B., Luck, S. J., and Hillyard, S. A. (**2004**). "Selective Attention: Electrophysiological and
520 Neuromagnetic Studies," in *The cognitive neurosciences, 3rd ed*, (MIT Press, Cambridge, MA,
521 US), pp. 561–574.
522 Itoh, K., Yumoto, M., Uno, A., Kurauchi, T., and Kaga, K. (**2000**). "Temporal stream of cortical
523 representation for auditory spatial localization in human hemispheres," *Neurosci. Lett.*, **292**,
524 215–219, doi: 10.1016/S0304-3940(00)01465-8.
525 Jensen, O. and Mazaheri, A. (**2010**). "Shaping Functional Architecture by Oscillatory Alpha Activity:
526 Gating by Inhibition," *Front. Hum. Neurosci.*, **4**, 186, doi: 10.3389/fnhum.2010.00186.
527 Johnson, B. W. and Hautus, M. J. (**2010**). "Processing of binaural spatial information in human auditory
528 cortex: Neuromagnetic responses to interaural timing and level differences," *Neuropsychologia*,
529 **48**, 2610–2619, doi: 10.1016/J.NEUROPSYCHOLOGIA.2010.05.008.

530 Kelly, S. P. (**2006**). "Increases in Alpha Oscillatory Power Reflect an Active Retinotopic Mechanism for
531 Distracter Suppression During Sustained Visuospatial Attention," *J. Neurophysiol.*, **95**, 3844–
532 3851, doi: 10.1152/jn.01234.2005.
533 Kidd, G., Mason, C. R., Best, V., and Marrone, N. (**2010**). "Stimulus factors influencing spatial release
534 from speech-on-speech masking," *J. Acoust. Soc. Am.*, **128**, 1965–1978, doi: 10.1121/1.3478781.
535 Klimesch, W. (**2011**). "Evoked alpha and early access to the knowledge system: The P1 inhibition timing
536 hypothesis," *Brain Res.*, **1408**, 52–71, doi: 10.1016/j.brainres.2011.06.003.
537 Klimesch, W. (**2012**). "Alpha-band oscillations, attention, and controlled access to stored information,"
538 *Trends Cogn. Sci.*, **16**, 606–617, doi: 10.1016/J.TICS.2012.10.007.
539 Klimesch, W., Sauseng, P., and Hanslmayr, S. (**2007**). "EEG alpha oscillations: The inhibition-timing
540 hypothesis," *Brain Res. Rev.*, **53**, 63–88, doi: 10.1016/j.brainresrev.2006.06.003.
541 Kong, L., Michalka, S. W., Rosen, M. L., Sheremata, S. L., Swisher, J. D., Shinn-Cunningham, B. G., and
542 Somers, D. C. (**2014**). "Auditory Spatial Attention Representations in the Human Cerebral
543 Cortex," *Cereb. Cortex*, **24**, 773–784, doi: 10.1093/cercor/bhs359.
544 Leino, S., May, P. J. C., Alku, P., Liikkanen, L. A., and Tiitinen, H. (**2007**). "The contribution of high
545 frequencies to human brain activity underlying horizontal localization of natural spatial sounds,"
546 *BMC Neurosci.*, **8**, 78, doi: 10.1186/1471-2202-8-78.
547 Lim, S.-J., Wöstmann, M., and Obleser, J. (**2015**). "Selective Attention to Auditory Memory Neurally
548 Enhances Perceptual Precision," *J. Neurosci.*, **35**, 16094–16104, doi: 10.1523/JNEUROSCI.2674-
549 15.2015.
550 Loiselle, L. H., Dorman, M. F., Yost, W. A., Cook, S. J., and Gifford, R. H. (**2016**). "Using ILD or ITD Cues for
551 Sound Source Localization and Speech Understanding in a Complex Listening Environment by
552 Listeners With Bilateral and With Hearing-Preservation Cochlear Implants," *J. Speech Lang. Hear.*
553 *Res. JSLHR*, **59**, 810–818, doi: 10.1044/2015_JSLHR-H-14-0355.
554 Luck, S. J. (**1995**). "Multiple mechanisms of visual-spatial attention: recent evidence from human
555 electrophysiology," *Behav. Brain Res.*, **71**, 113–123, doi: 10.1016/0166-4328(95)00041-0.
556 Majdak, P., Baumgartner, R., and Jenny, C. (**2019**). "Formation of three-dimensional auditory space,"
557 *ArXiv:*1901.03990 Q-Bio (13 January 2019).
558 Marzecová, A., Schettino, A., Widmann, A., SanMiguel, I., Kotz, S. A., and Schröger, E. (**2018**).
559 "Attentional gain is modulated by probabilistic feature expectations in a spatial cueing task: ERP
560 evidence," *Sci. Rep.*, **8**, doi: 10.1038/s41598-017-18347-1.

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

Middlebrooks, J. C. (**2015**). "Sound localization," *Handb. Clin. Neurol.*, *The Human Auditory System*, **129**, 99–116, doi: 10.1016/B978-0-444-62630-1.00006-8.

Murray, M. M., Brunet, D., and Michel, C. M. (**2008**). "Topographic ERP Analyses: A Step-by-Step Tutorial Review," *Brain Topogr.*, **20**, 249–264, doi: 10.1007/s10548-008-0054-5.

Oppenheim, A. V. and Schafer, R. W. (**1975**). *Digital Signal Processing*, (Prentice-Hall, Englewood Cliffs, NJ).

Palomäki, K. J., Tiitinen, H., Mäkinen, V., May, P. J. C., and Alku, P. (**2005**). "Spatial processing in human auditory cortex: The effects of 3D, ITD, and ILD stimulation techniques," *Cogn. Brain Res.*, **24**, 364–379, doi: 10.1016/j.cogbrainres.2005.02.013.

Posner, M. I. (**1980**). "Orienting of attention," *Q. J. Exp. Psychol.*, **32**, 3–25, doi: 10.1080/00335558008248231.

Rihs, T. A., Michel, C. M., and Thut, G. (**2007**). "Mechanisms of selective inhibition in visual spatial attention are indexed by α-band EEG synchronization," *Eur. J. Neurosci.*, **25**, 603–610, doi: 10.1111/j.1460-9568.2007.05278.x.

Rihs, T. A., Michel, C. M., and Thut, G. (**2009**). "A bias for posterior α-band power suppression versus enhancement during shifting versus maintenance of spatial attention," *NeuroImage*, **44**, 190–199, doi: 10.1016/j.neuroimage.2008.08.022.

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

Röder, B., Teder-Sälejärvi, W., Sterr, A., Rösler, F., Hillyard, S. A., and Neville, H. J. (**1999**). "Improved auditory spatial tuning in blind humans," *Nature*, **400**, 162–166, doi: 10.1038/22106.

Sach, A. J., Hill, N. I., and Bailey, P. J. (**2000**). "Auditory spatial attention using interaural time differences," *J. Exp. Psychol. Hum. Percept. Perform.*, **26**, 717–729, doi: 10.1037//0096-1523.26.2.717.

Salminen, N. H., Takanen, M., Santala, O., Lamminsalo, J., Altoè, A., and Pulkki, V. (**2015**). "Integrated processing of spatial cues in human auditory cortex," *Hear. Res.*, **327**, 143–152, doi: 10.1016/J.HEARES.2015.06.006.

591 Samaha, J., Bauer, P., Cimaroli, S., and Postle, B. R. (**2015**). "Top-down control of the phase of alpha-
592 band oscillations as a mechanism for temporal prediction," *Proc. Natl. Acad. Sci. U. S. A.*, **112**,
593 8439–8444, doi: 10.1073/pnas.1503686112.
594 Samaha, J., Sprague, T. C., and Postle, B. R. (**2016**). "Decoding and Reconstructing the Focus of Spatial
595 Attention from the Topography of Alpha-band Oscillations," *J. Cogn. Neurosci.*, 1–8, doi:
596 10.1162/jocn_a_00955.
597 Sauseng, P., Klimesch, W., Stadler, W., Schabus, M., Doppelmayr, M., Hanslmayr, S., Gruber, W. R., et al.
598 (**2005**). "A shift of visual spatial attention is selectively associated with human EEG alpha
599 activity," *Eur. J. Neurosci.*, **22**, 2917–2926, doi: 10.1111/j.1460-9568.2005.04482.x.
600 Schreuder, M., Rost, T., and Tangermann, M. (**2011**). "Listen, You are Writing! Speeding up Online
601 Spelling with a Dynamic Auditory BCI," *Front. Neurosci.*, **5**, 112, doi: 10.3389/fnins.2011.00112.
602 Schröger, E. (**1996**). "Interaural time and level differences: Integrated or separated processing?," *Hear.*
603 *Res.*, **96**, 191–198, doi: 10.1016/0378-5955(96)00066-4.
604 Shinn-Cunningham, B. G., Kopco, N., and Martin, T. J. (**2005**). "Localizing nearby sound sources in a
605 classroom: Binaural room impulse responses," *J. Acoust. Soc. Am.*, **117**, 3100–3115, doi:
606 10.1121/1.1872572.
607 Siegel, M., Donner, T. H., and Engel, A. K. (**2012**). "Spectral fingerprints of large-scale neuronal
608 interactions," *Nat. Rev. Neurosci.*, **13**, 121–134, doi: 10.1038/nrn3137.
609 Skrandies, W. (**1990**). "Global field power and topographic similarity," *Brain Topogr.*, **3**, 137–141.
610 Slagter, H. A., Prinssen, S., Reteig, L. C., and Mazaheri, A. (**2016**). "Facilitation and inhibition in attention:
611 Functional dissociation of pre-stimulus alpha activity, P1, and N1 components," *NeuroImage*,
612 **125**, 25–35, doi: 10.1016/j.neuroimage.2015.09.058.
613 Snyder, J. S. and Large, E. W. (**2005**). "Gamma-band activity reflects the metric structure of rhythmic
614 tone sequences," *Brain Res. Cogn. Brain Res.*, **24**, 117–126, doi:
615 10.1016/j.cogbrainres.2004.12.014.
616 Stevens, C., Fanning, J., Coch, D., Sanders, L., and Neville, H. (**2008**). "Neural mechanisms of selective
617 auditory attention are enhanced by computerized training: Electrophysiological evidence from
618 language-impaired and typically developing children," *Brain Res.*, **1205**, 55–69, doi:
619 10.1016/J.BRAINRES.2007.10.108.
620 Tardif, E., Murray, M. M., Meylan, R., Spierer, L., and Clarke, S. (**2006**). "The spatio-temporal brain
621 dynamics of processing and integrating sound localization cues in humans," *Brain Res.*, **1092**,
622 161–176, doi: 10.1016/J.BRAINRES.2006.03.095.

623 Ungan, P., Yagcioglu, S., and Goksoy, C. (**2001**). "Differences between the N1 waves of the responses to
624 interaural time and intensity disparities: scalp topography and dipole sources," *Clin.*
625 *Neurophysiol.*, **112**, 485–498, doi: 10.1016/S1388-2457(00)00550-2.
626 Worden, M. S., Foxe, J. J., Wang, N., and Simpson, G. V. (**2000**). "Anticipatory biasing of visuospatial
627 attention indexed by retinotopically specific alpha-band electroencephalography increases over
628 occipital cortex," *J. Neurosci.*, **20**, RC63, doi: 10.1523/JNEUROSCI.20-06-j0002.2000.
629 Wöstmann, M., Herrmann, B., Maess, B., and Obleser, J. (**2016**). "Spatiotemporal dynamics of auditory
630 attention synchronize with speech," *Proc. Natl. Acad. Sci.*, **113**, 3873–3878, doi:
631 10.1073/pnas.1523357113.
632 Wyart, V., Dehaene, S., and Tallon-Baudry, C. (**2012**). "Early dissociation between neural signatures of
633 endogenous spatial attention and perceptual awareness during visual masking," *Front. Hum.*
634 *Neurosci.*, **6**, 16, doi: 10.3389/fnhum.2012.00016.
635