Spatial Release From Informational Masking: Evidence From Functional Near Infrared Spectroscopy

Min Zhang[1,2], Yu-Lan Mary Ying[3], and Antje Ihlefeld[1]

Abstract

Informational masking (IM) can greatly reduce speech intelligibility, but the neural mechanisms underlying IM are not understood. Binaural differences between target and masker can improve speech perception. In general, improvement in masked speech intelligibility due to provision of spatial cues is called spatial release from masking. Here, we focused on an aspect of spatial release from masking, specifically, the role of spatial attention. We hypothesized that in a situation with IM background sound (a) attention to speech recruits lateral frontal cortex (LFCx) and (b) LFCx activity varies with direction of spatial attention. Using functional near infrared spectroscopy, we assessed LFCx activity bilaterally in normal-hearing listeners. In Experiment 1, two talkers were simultaneously presented. Listeners either attended to the target talker (speech task) or they listened passively to an unintelligible, scrambled version of the acoustic mixture (control task). Target and masker differed in pitch and interaural time difference (ITD). Relative to the passive control, LFCx activity increased during attentive listening. Experiment 2 measured how LFCx activity varied with ITD, by testing listeners on the speech task in Experiment 1, except that talkers either were spatially separated by ITD or colocated. Results show that directing of auditory attention activates LFCx bilaterally. Moreover, right LFCx is recruited more strongly in the spatially separated as compared with colocated configurations. Findings hint that LFCx function contributes to spatial release from masking in situations with IM.

Keywords

auditory attention, informational masking, functional infrared spectroscopy, lateral frontal cortex, spatial release from masking

Date received: 28 June 2018; revised: 31 October 2018; accepted: 13 November 2018

Introduction

In everyday life, background speech often interferes with recognition of target speech. At least two forms of masking contribute to this reduced intelligibility, referred to as *energetic* and *informational* masking (EM and IM, Brungart, 2001; Freyman, Balakrishnan, & Helfer 2001; Jones & Litovsky, 2011; Mattys, Brooks, & Cooke, 2009). EM occurs when sound sources have energy at the same time and frequency (e.g., Brungart, Chang, Simpson, & Wang, 2006). IM broadly characterizes situations when target and background sources are perceptually similar to each other or when the listener is uncertain about what target features to listen for in an acoustic mixture (for a recent review, see Kidd & Colburn, 2017). IM is thought to be a major factor limiting performance of hearing aid and cochlear implant devices (Marrone, Mason, & Kidd, 2008; Shinn-Cunningham & Best, 2008; Xia, Kalluri,

Micheyl, & Hafter, 2017). However, the neural mechanisms underlying IM are not understood. The current study explores cortical processing of speech detection and identification in IM.

In EM-dominated tasks, computational models based on the output of the auditory nerve can closely capture speech identification performance (review: Goldsworthy

[1]Department of Biomedical Engineering, New Jersey Institute of Technology, Newark, NJ, USA
[2]Graduate School of Biomedical Sciences, Rutgers University, Newark, NJ, USA
[3]Department of Otolaryngology-Head and Neck Surgery, Rutgers New Jersey Medical School, Newark, NJ, USA

Corresponding author:
Antje Ihlefeld, Department of Biomedical Engineering, New Jersey Institute of Technology, Newark, NJ 07102, USA.
Email: antje.ihlefeld@njit.edu

& Greenberg, 2004). Consistent with this interpretation, subcortical responses encode the fidelity by which a listener processes speech in EM noise (Anderson & Kraus, 2010). However, peripheral models fail to account for speech intelligibility in IM-dominated tasks (e.g., Cooke, Garcia Lecumberri, & Barker, 2008), suggesting that performance in IM is mediated at least partially by mechanisms of the central nervous system.

In IM-dominated tasks, previous behavioral studies are consistent with the idea that in order to understand a masked target voice, listeners need to segregate short-term speech segments from the acoustic mixture, stream these brief segments across time to form a perceptual object, and selectively attend to those perceptual features of the target object that distinguish the target talker from competing sound (Cusack, Decks, Aikman, & Carlyon, 2004; Ihlefeld & Shinn-Cunningham, 2008a; Jones, Alford, Bridges, Tremblay, & Macken, 1999). Previous work suggests that common onsets and harmonicity determine how short-term segments form (Darwin & Hukin, 1998; Micheyl, Hunter, & Oxenham, 2010). Differences in higher order perceptual features, including spatial direction and pitch, then allow listeners to link these short-term segments across time to form auditory objects (Brungart & Simpson, 2002; Darwin, Brungart, & Simpson, 2003; Darwin & Hukin, 2000), enabling the listener to selectively attend to a target speaker and ignore the masker (Carlyon, 2004; Ihlefeld & Shinn-Cunningham, 2008b; Shinn-Cunningham, 2008).

Rejection of competing auditory streams correlates with behavioral measures of short-term working memory, where a person's ability to suppress unwanted sound decreases with decreasing working memory capacity (Conway, Cowan, & Bunting, 2001). This raises the possibility that central regions linked to auditory short-term memory tasks are recruited in situations with IM. To test this prediction, here, we conducted two experiments to characterize oxy-hemoglobin (HbO) correlates of cortical responses, while normal hearing (NH) subjects listened, either actively or passively, to speech in IM background sound. Recent work in NH listeners demonstrates that auditory short-term memory tasks can alter blood oxygenation level-dependent signals bilaterally in two areas of lateral frontal cortex (LFCx): (a) the transverse gyrus intersecting precentral sulcus (tgPCS) and (b) the caudal inferior frontal sulcus (cIFS; Michalka, Kong, Rosen, Shinn-Cunningham, & Somers, 2015; Noyce, Cestero, Michalka, Shinn-Cunningham, & Somers, 2017). This suggests that LFCx should engage when listeners are actively trying to reject unwanted sound but be less active when listeners are passively hearing the same sound. Using functional near infrared spectroscopy (fNIRS) to record HbO signals at the tgPCS and cIFS

bilaterally, we here examined how LFCx engages when a listeners tries to filter out IM.

In two experiments, we tested rapid-serial auditory presentation stimuli adapted from previous work by Michalka and coworkers (2015). Our goal was to examine how direction of auditory attention alters the HbO responses in LFCx in a situation with IM, as assessed with fNIRS. In Experiment 1, NH listeners were asked to detect keywords in a target message on the left side, while a background talker producing IM was simultaneously presented on the right. In a control condition, participants listened passively to an unintelligible, acoustically scrambled version of the same stimuli. We hypothesized that unlike in passive listening, when listeners actively tried to hear out speech in IM background sound, this would recruit LFCx.

We further hypothesized that interactions between spatially directed auditory attention and LFCx activity would arise. An extensive literature documents that speech intelligibility improves and IM is released when competing talkers are spatially separated as opposed to being colocated, a phenomenon referred to as spatial release from masking (e.g., Carhart, Tillman, & Johnson, 1967; Darwin & Hukin, 1997; Glyde, Buchholz, Dillon, Cameron, & Hickson, 2013; Kidd, Mason, Best, & Marrone, 2010). Using similar speech stimuli as in Experiment 1, we looked whether the mechanisms underlying spatial release from IM recruit LFCx, by comparing LFCx HbO responses in the spatially separated configuration from Experiment 1 versus a colocated configuration of the same stimuli. We reasoned that a stronger HbO response in the spatially separated versus colocated configurations would support the view that spatial attention under IM activates LFCx. In contrast, a stronger LFCx response in the colocated configuration would suggest that LFCx does not encode the direction of spatial auditory attention.

Participants

A total of 29 listeners (age 19 to 25 years, 9 women) participated in the study and were paid for their time, with 14 participants in Experiment 1 and 15 participants in Experiment 2. All listeners were native speakers of English, right handed, and had normal audiometric pure-tone detection thresholds as assessed through standard audiometric testing at all octave frequencies from 250 Hz to 8 kHz. At each tested frequency, tone detection thresholds did not differ by more than 10 dB across ears, and all thresholds were 20 dB HL or better. All listeners gave written informed consent to participate in the study. All testing was administered according to the guidelines of the institutional review board of the New Jersey Institute of Technology.

Methods

Recording Setup

Each listener completed one session of behavioral testing, while we simultaneously recorded bilateral hemodynamic responses over the listener's left and right dorsal and ventral LFCx. The listener was seated approximately 0.8 m away from a computer screen with test instructions (Lenovo ThinkPad T440P), inside a testing suite with a moderately quiet background sound level of less than 44 dBA. The listener held a wireless response interface in the lap (Microsoft Xbox 360 Wireless Controller) and wore insert earphones (Etymotic Research ER-2) for delivery of sound stimuli. The setup is shown in Figure 1(a).

A camera-based three-dimensional location tracking and pointer tool system (Brainsight 2.0 software and hardware by Rogue Research Inc., Canada) allowed the experimenter to record four coordinates on the listener's head: nasion, inion, and bilateral preauricular points. Following the standard Montreal Neurological Institute ICBM-152 brain atlas (Talairach, Rayport, & Tournoux, 1988), these four landmark coordinates were then used as reference for locating the four regions of interest (ROIs, locations illustrated in Figure 1(b)). Infrared optodes were placed on the listener's head directly above the four ROIs, specifically, the left tgPCS, left cIFS, right tgPCS, and right cIFS. A custom-built head cap, fitted to the listener's head via adjustable straps, embedded the optodes, and held them in place.

Acoustic stimuli were generated in MATLAB (Release R2016a, The Mathworks, Inc., Natick, MA, USA), digital-to-analog converter with a sound card (Emotiva Stealth DC-1; 16-bit resolution, 44.1 kHz sampling frequency) and presented over the insert earphones. This acoustic setup was calibrated with a 2-cc coupler, $1/2''$ pressure-field microphone and a sound level meter (Bruel & Kjaer 2250-G4).

Using a total of 4 source optodes and 16 detector optodes, a continuous-wave diffuse optical NIRS system (CW6; TechEn Inc., Milford, MA) simultaneously recorded light absorption at two different wavelengths, 690 nm and 830 nm, with a sampling frequency of 50 Hz. Sound delivery and optical recordings were synchronized via trigger pulse with a precision of 20 ms. Using a time-multiplexing algorithm developed by Huppert, Diamond, Franceschini, and Boas (2009), multiple source optodes were paired with multiple detector optodes. A subset of all potential combinations of
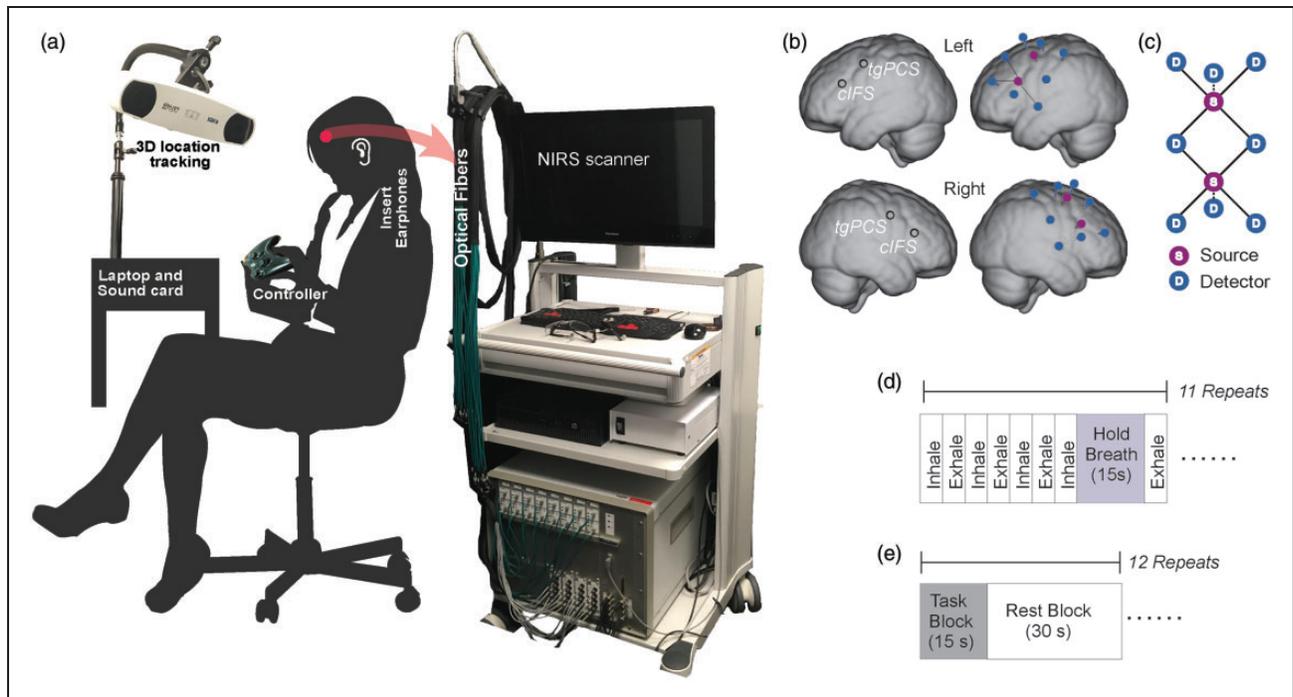
Figure 1. (a) Experimental apparatus and setup. (b) ROIs and optode placement for a representative listener. Blue circles show placements of detector optodes and red circles of source optodes. (c) fNIRS optical probes design with deep neurovascular (solid line) and shallow nuisance (dotted line) channels. (d) Block design, controlled breathing task, and (e) Block design, auditory task. S = source; D = detector.

optode-detector pairs was interpreted as response channels and further analyzed. Specifically, on both sides of the head, we combined one optical source and four detectors into one probe set according to the channel geometry shown in Figure 1(b). On each side of the head, we had two probe sets placed directly above cIFS and tgPCS on the scalp. Within each source-detector channel, the distance between source and detector determined the depth of the light path relative to the surface of the skull (review: Ferrari & Quaresima, 2012). To enable us to partial out the combined effects of nuisance signals such as cardiac rhythm, respiratory induced change, and blood pressure variations from the desired hemodynamic response driven neural events in cortex, we used two recording depths. Deep channels, used to estimate the neurovascular response of cortical tissue between 0.5 and 1 cm below the surface of the skull, had a 3-cm source-detector distance (solid lines in Figure 1(c)), whereas shallow channels, used to estimate physiological noise, had a source-detector distance of 1.5 cm (dotted line in Figure 1(c)). At each of the four ROIs, we recorded with four concentrically arranged deep channels and one shallow channel and averaged the traces of the four deep channels, to improve the noise floor. As a result, for each ROI, we obtained one deep trace, which we interpreted as neurovascular activity, and one shallow trace, which we interpreted as nuisance activity.

Controlled Breathing Task

Variability in skull thickness, skin pigmentation, and other idiosyncratic factors can adversely affect recording quality with fNIRS (Bickler, Feiner, & Rollins, 2013; Yoshitani et al., 2007). As a control for reducing group variance and to monitor recording quality, listeners initially performed a nonauditory task, illustrated in Figure 1(d). This nonauditory task consisted of 11 blocks of controlled breathing (Thomason, Foland, & Glover, 2006).

During each of these blocks, visuals on the screen instructed listeners to (a) *inhale* via a gradually expanding green circle, or (b) *exhale* via a shrinking green circle, or (c) *hold breath* via a countdown on the screen. Using this controlled breathing method, listeners were instructed to follow a sequence of inhaling for 5 s, followed by exhaling for 5 s, for a total of 30 s. At the end of this sequence, listeners were instructed to inhale for 5 s and then hold their breath for 15 s. Our criterion for robust recording quality was that for each listener, breath holding needed to induce a significant change in the hemodynamic response at all ROIs (analysis technique and statistical tests described later), otherwise that listener's data would have been excluded from further analysis. Moreover, we used the overall activation

strength of the hemodynamic response during breath holding for normalizing the performance in the auditory tasks (details described later).

Auditory Tasks

Following the controlled breathing task, listeners performed Experiment 1, consisting of 24 blocks of behavioral testing with their eyes closed. Each listener completed 12 consecutive blocks of an active and 12 consecutive blocks of a passive listening task, with task order (active vs. passive) counter-balanced across listeners. In each block, two competing auditory streams of 15 s duration each were presented simultaneously. In the active listening task, we presented intelligible speech utterances, whereas in the passive listening task, we presented unintelligible scrambled speech. Figure 2 shows a schematic of the paradigm (a) and spectrograms for two representative stimuli (b).

In Experiment 1, the target stream was always presented with a left-leading interaural time difference (ITD) of 500 μs, while the concurrent masker stream was presented with a right-leading ITD of 500 μs (spatially separated configuration). In Experiment 2, we also tested a spatially colocated configuration, where both the target and the masker had 0 μs ITD. In Experiment 1, the broadband root means square values of the stimuli were equated at 59 dBA, then randomly roved from 53 to 65 dBA, resulting in broadband signal-to-noise ratios from −6 to 6 dB, so that listeners could not rely on
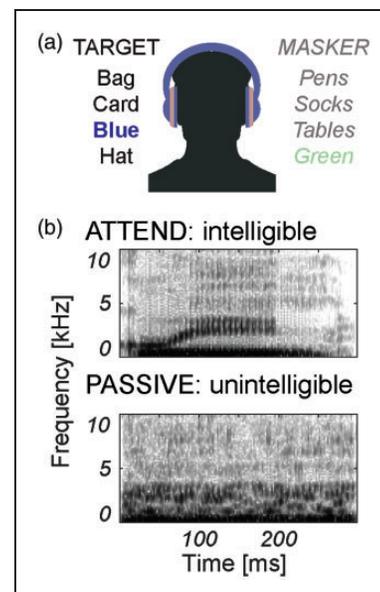
Figure 2. (a) Speech paradigm. (b) Spectrograms of the word *green*. Unprocessed speech in the ATTEND condition (top) and scrambled speech in the PASSIVE condition (bottom).

level cues to detect the target. To remove level cues entirely, giving spatial cues even more potential strength for helping the listener attend to the target, in Experiment 2, we made the target and masker equally loud. In Experiment 2, both target and masker were presented at 59 dBA.

Unfortunately, due to a programming error, listeners' responses were inaccurately recorded during the auditory tasks of Experiments 1 and 2 and are thus not reported here. During pilot testing with the tested stimulus parameters (not shown here), speech detection performance was 90% correct or better across all conditions.

In the active task, stimuli consisted of two concurrent rapid serial streams of spoken words. Speech utterances were chosen from a closed-set corpus (Kidd, Best, & Mason, 2008). There were 16 possible words, consisting of the colors <red, white, blue, and green> and the objects <hats, bags, card, chairs, desks, gloves, pens, shoes, socks, spoons, tables, and toys>. Those words were recorded from two male talkers, spoken in isolation. The target talker had an average pitch of 115 Hz versus 144 Hz for the masker talker. Using synchronized overlap-add with fixed synthesis (Hejna & Musicus, 1991), all original utterances were time-scaled to make each word last 300 ms. Words from both the target and masker talkers were simultaneously presented, in random order with replacement. Specifically, target and masker streams each consisted of 25 words with 300 ms of silence between consecutive words (total duration 15 s).

To familiarize the listener with the target voice, at the beginning of each active block, we presented the target voice speaking the sentence "Bob found five small cards" at 59 dBA and instructed the listeners to remember this voice.

Listeners were further instructed to press the right trigger button on the handheld response interface each time the target talker to their left side uttered any of the four *color* words, while ignoring all other words from both the target and the masker. A random number (between three and five) of color words in the target voice would appear during each block. No response feedback was provided to the listener.

In the passive task, we simultaneously presented two streams of concatenated scrambled speech tokens that were processed to be unintelligible. Stimuli in the passive task were derived from the stimuli in the active task. Specifically, using an algorithm by Ellis (2010), unprocessed speech tokens were time-windowed into snippets of 25 ms duration, with 50% temporal overlap between consecutive time-steps. Using a bank of 64 GammaTone filters with center frequencies that were spaced linearly along the human equivalent rectangular bandwidth scale (Patterson & Holdsworth, 1996) and that had bandwidths of 1.5 equivalent rectangular bandwidth, the time-windowed snippets were bandpass filtered. Within each of the 64 frequency bands, the bandpass-filtered time-windowed snippets were permutated with a Gaussian probability distribution over a radius of 250 ms, and added back together, constructing scrambled tokens of speech.

Thus, the scrambled speech tokens had similar magnitude spectra and similar temporal-fine structure characteristics as the original speech utterances, giving them speech-like perceptual qualities. However, because the sequence of the acoustic snippets was shuffled, the scrambled speech was unintelligible.

Furthermore, the passive differed from the active task in that the handheld response vibrated randomly between 3 and 5 times during each block. Listeners were instructed to passively listen to the sounds and press the right trigger button on the handheld response interface each time the interface vibrated, ensuring that the listener stayed engaged in this task. Listeners need to correctly detect at least two out of three vibrations, otherwise they were excluded from the study.

In the active task of Experiment 1, target and masker differed in both voice pitch and perceived spatial direction, and listeners could use either cue to direct their attention to the target voice. Experiment 2 further assessed the role of spatial attention in two active tasks. The first task (*spatial cues*) was identical to the active condition of Experiment 1. The second task (*no spatial cues*) used similar stimuli as the active task in Experiment 1, except that both sources had 0 µs ITD. Thus, in Experiment 2, each listener completed six blocks of an active listening task that was identical to the active task in Experiment 1 and six blocks of another active listening task that was similar to the active task in Experiment 1, except that the spatial cues were removed. Blocks were randomly interleaved. Listeners indicated when they detected the target talker uttering one of the four color words, by pressing the right trigger on the handheld response interface.

Signal Processing of the fNIRS Traces

We used HOMER2 (Huppert et al., 2009), a set of MATLAB-based scripts, to analyze the raw recordings of the deep and shallow fNIRS channels at each of the four ROIs. First, the raw recordings were bandpass filtered between 0.01 and 0.3 Hz, using a fifth order zero-phase Butterworth filter. Next, we removed slow temporal drifts in the bandpass filtered traces by de-trending each trace with a 20th-degree polynomial (Pei et al., 2007). To remove artefacts due to sudden head movement during the recording, the detrended traces were then wavelet transformed using Daubechies 2 (db2) base functions. We removed wavelet coefficients that were outside of one interquartile range (Molavi & Dumont, 2012).

We applied the modified Beer–Lambert law (Cope & Delpy, 1988; Kocsis, Herman, & Eke, 2006) to these processed traces and obtained the estimated HbO concentrations for the deep and shallow channels at each ROI. To partial out physiological nuisance signals, thus reducing across-listener variability, we then normalized all HbO traces from the task conditions by dividing each trace by the maximal HbO concentration change in that source-detector pair during controlled breathing.

Calculation of Activation levels

For each of the auditory task conditions and ROIs, we wished to determine what portion of each hemodynamic response could be attributed to the behavioral task. Therefore, HbO traces were fitted by four general linear models (GLM), one GLM for each ROI. Each GLM was of the form:

$$y(t) = x_{\text{task1}}(t)\beta 1 + x_{\text{task2}}(t)\beta 2 + x_{\text{nuisance}}(t)\beta 3 + \varepsilon(t)$$

where y is the HbO trace, t is time, and the β_i values indicate the activation levels of each of the regressors. We calculated the β_i values for each listener and ROI. Specifically, $x_{\text{task }i}(t)$ was the regressor of the hemodynamic change attributed to behavioral task i. $x_{\text{nuisance}}(t)$ was the HbO concentration in the shallow channel (Brigadoi & Cooper, 2015), and $\varepsilon(t)$ was the residual error of the GLM.

The task regressors $x_{\text{task }i}$ in the GLM design matrix then contained reference functions for the corresponding task, each convolved with a canonical hemodynamic response function (Lindquist, Loh, Atlas, & Wager, 2009):

$$HRF(t) = \frac{1}{\Gamma(6)}\, t^5 e^{-1} - \frac{1}{6\Gamma(16)}\, t^{15} e^{-t}$$

where Γ was the gamma function.

Task reference functions were built from unit step functions as follows. In the controlled breathing task, the reference function equaled 1 during the breath holding time intervals and 0 otherwise. Only one task regressor was used to model the controlled breathing task. In the auditory tasks, two reference functions were built, one for each task, and set to 1 for stimulus present, and 0 for stimulus absent.

In general, fNIRS allows for calculation of both HbO and deoxy-hemoglobin (HbR) levels. Neurovascular activity couples HbO and HbR, such that both measures are anticorrelated. In contrast, systemic changes in oxygen level couples HbO and HbR such that the two are correlated. To date, no standardized method exists for estimating brain activity from HbO and HbR (e.g., Knauth, Heldmann, Münte, & Royl, 2017). During pilot

testing, we here analyzed both HbO and HbR and found that both measures lead to highly consistent interpretations for the current task. However, HbR was generally at much reduced amplitude compared with HbO, thus resulting in recordings that were often close to the noise floor. For clarity, the analysis in this manuscript is based on HbO, the cleaner signal.

Statistical Analysis

To assess whether the HbO activation levels at each ROI differed from 0, we applied two-sided Student's t tests. Furthermore, to determine whether HbO activation levels differed from each other across the two task conditions of each experiment, left or right hemispheres and dorsal (tgPCS) or ventral (cIFS) sites, $2 \times 2 \times 2$ repeated-measures analyses of variance (rANOVA) were applied to the β_i values, at the .05 alpha level for significance. To correct for multiple comparisons, all reported p values were Bonferroni-corrected.

Results

Controlled Breathing Task

Figure 3 shows the HbO traces during the controlled breathing task for both Experiments 1 and 2, at each of the four ROIs. Two-sided Student's t test on the β-values of the GLM fit on HbO concentration changes revealed that at each ROI, the mean activation levels during breath holding differed significantly from 0 ($t(13) = 7.6$, $p < .001$ at left tgPCS; $t(13) = -6.8$, $p < .001$ at right tgPCS; $t(13) = -6.5$, $p < .001$ at left cIFS; $t(13) = -7.5$, $p < .001$ at right cIFS, after Bonferroni corrections). Two-sided Student's t test on
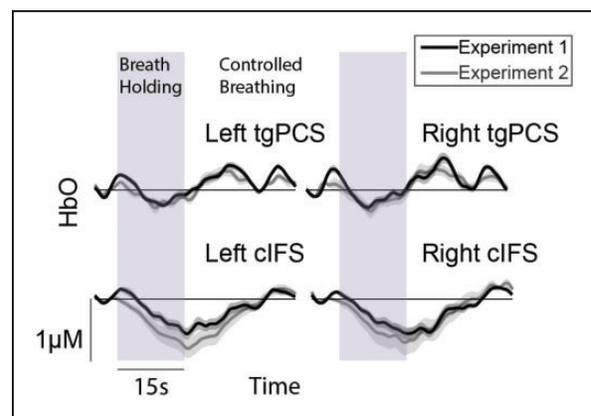
Figure 3. HbO concentration change during controlled breathing in Experiments 1 and 2.
HbO = oxy-hemoglobin; tgPCS = transverse gyrus intersecting precentral sulcus; cIFS = caudal inferior frontal sulcus.

the β-values of the GLM fit on HbR concentration changes revealed that only at left cIFS and right cIFS, the mean activation levels during breath holding differed significantly from 0 ($t(13) = 3.1$, $p = .03$ at left cIFS; $t(13) = 3.4$, $p = .02$ at right cIFS, after Bonferroni corrections).

Two-sided Student's t test confirmed that also in Experiment 2, HbO activation levels during breath holding significantly differed from 0 ($t(13) = -5.6$, $p < .001$ at left tgPCS; $t(13) = -3.4$, $p < 0.001$ at right tgPCS; $t(13) = -4$, $p < .001$ at left cIFS; $t(13) = -3.7$, $p = 0.006$ at right cIFS). Thus, breath holding induced a significant change in the HbO response at all four ROIs, confirming feasibility of the recording setup and providing a baseline reference for normalizing the task-evoked HbO traces of Experiments 1 and 2.

Experiment 1

Figure 4(a) shows the HbO traces during active versus passive listening, at each of the four ROIs. Solid lines denote the auditory attention condition, dotted lines passive listening. The ribbons around each trace show one standard error of the mean across listeners. Figure 4(b) shows HbO activation levels β, averaged across listeners, during the auditory attention (solid fill) and the passive listening tasks (hatched fill). Error bars show one standard error of the mean. All listeners reached criterion performance during behavioral testing and were included in the group analysis. rANOVA revealed significant main effects of task, $F(1, 13) = 6.5$, $p = .024$, and dorsal (tgPCS) or ventral (cIFS) site, $F(1, 13) = 6.1$, $p = .028$. The effect of hemisphere was not significant, $F(1, 13) = 0.015$, $p = .9$. In Experiment 1, listeners were tested over 12 blocks, a number we initially chose conservatively.

To investigate the minimum number of blocks needed to see a robust difference between active and passive listening conditions, we applied a power analysis. Using bootstrapping of sampling without replacements, we calculated activation levels β during active versus passive listening in 100 repetitions and found that a minimum of six blocks suffices to show a robust effect. Therefore, in Experiment 2, listeners were tested using six blocks per condition.

Experiment 2

Figure 5(a) and (b) display the HbO traces (red lines denote spatially separated, blue lines colocated configurations) and the across-listener average in HbO activation β-levels for the spatially separated (red fill) versus colocated configurations (blue fill), at each of the four ROIs; 14 listeners reached criterion performance during behavioral testing and were included in the group analysis.
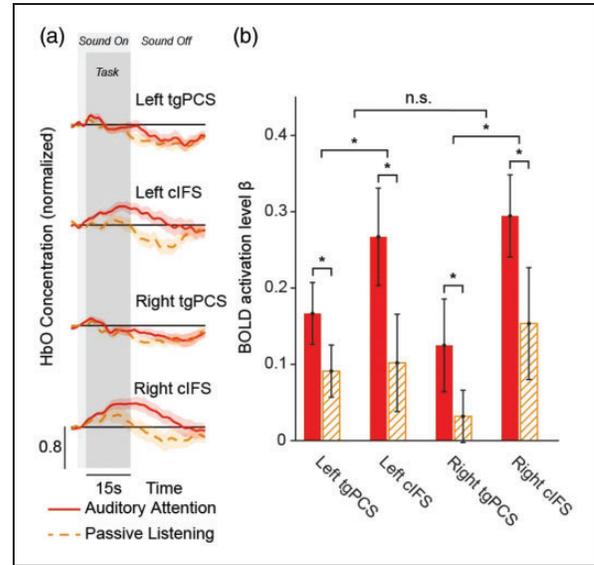
Figure 4. Results from Experiment 1. (a) Normalized HbO traces during the direction of auditory attention versus passive listening, at each of the four ROIs in Experiment 1. The ribbons around each trace show one standard error of the mean across listeners. (b) Normalized HbO traces during pitch and spatial cues condition versus pitch cue only condition, at each of the four ROIs in Experiment 2. The ribbons around each trace show one standard error of the mean across listeners. HbO activation levels β, error bars show one standard error of the mean.
HbO = oxy-hemoglobin; tgPCS = transverse gyrus intersecting precentral sulcus; cIFS = caudal inferior frontal sulcus.
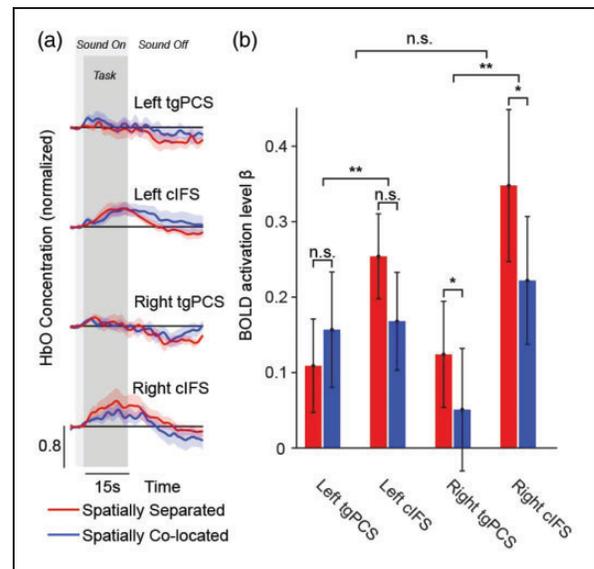
Figure 5. Results from Experiment 2, formatting similar to Figure 4.
HbO = oxy-hemoglobin; tgPCS = transverse gyrus intersecting precentral sulcus; cIFS = caudal inferior frontal sulcus.

One listener's data had to be excluded, because the participant had fallen asleep during testing. An rANOVA on the activation levels found a significant main effect of dorsal or ventral site, $F(1, 13) = 10.3$, $p = .007$. Main effects of spatial configuration and left or right hemisphere were not significant, $F(1, 13) = 1.6$, $p = .212$ for effect of task; $F(1, 13) = 0.153$, $p = .702$ for effect of hemisphere. In addition, the interaction between task and left or right hemisphere was significant, $F(1, 13) = 7.2$, $p = .019$, confirming an overall stronger activation in the right hemisphere in the spatially separated as compared with the colocated configuration. No difference between spatial configurations was discovered in the HbO concentration changes in the left hemisphere.

Discussion

Physiological Correlates of Active Listening Exist in LFCx

In Experiment 1, we presented two competing streams of rapidly changing words. All target and masker words were drawn from an identical corpus of possible words, uttered by two male talkers and played synchronously. As a result, both EM and IM interfered with performance. When the sounds were unintelligible scrambled speech and the participants listened passively, across all ROIs, the LFCx responses were smaller as compared with the active auditory attention task.

Thus, direction of auditory attention increased bilateral HbO responses in LFCx. These results support and extend previous finding on the role of LFCx. Using rapid serial presentation task with two simultaneous talkers, where listeners monitored a target stream in search for targets and were tasked to detect-and-identify target digits, prior work had revealed an auditory bias of LFCx regions (Michalka et al., 2015). Here, we found that even when listeners were performing a detection-only task under conditions of IM, this resulted in robust recruitment of LFCx. Moreover, the current results show that attentive listening in a situation with IM recruits LFCx, whereas passive listening does not.

Right LFCx Activation Associated With Spatial Release from Masking

We wished to disentangle the role of spatial attention on the LFCx HbO response. In Experiment 1, spatial differences between target and masker were available. However, the target voice also had a slightly lower pitch than the masker voice, and listeners could utilize either or both cues to attend to the target (Ihlefeld & Shinn-Cunningham, 2008b). Therefore, we presented two different spatial configurations in Experiment 2—a spatially separated configuration, where spatial attention could help performance, and a spatially colocated configuration, where spatial attention cues were not available. Contrasting active listening across these two spatial configurations, Experiment 2 revealed that right LFCx was more strongly recruited in the spatially separated as compared with the colocated configuration. In contrast, in left LFCx, no difference in HbO signals was observed across the two spatial configurations. Therefore, these findings are consistent with the interpretation that right LFCx HbO activation contained significant information about the direction of spatial attention. Indeed, previous work finds asymmetrical recruitment with stronger activation in the hemisphere that is contralateral to sound location, at least for ITDs within the physiologically plausible range of naturally occurring sound (Undurraga, Haywood, Marquardt, & McAlpine, 2016; von Kriegstein, Griffiths, Thompson, & McAlpine, 2008).

In general, spatial release from masking is thought to arise from three different mechanisms (e.g., Shinn-Cunningham, Ihlefeld, Satyavarta, & Larson, 2005), monaural head shadow, assumed to be a purely acoustic phenomenon, binaural decorrelation processing, and spatial attention. The current stimuli did not provide head shadow. Therefore, in the current paradigm, spatial cues could have contributed to spatial release from masking through two mechanisms, binaural decorrelation, presumably arising at or downstream from the brainstem (Dajani & Picton, 2006; Wack et al., 2012; Wong & Stapells, 2004) and spatial attention, assumed to arise at cortical processing levels (Ahveninen et al., 2006; Larson & Lee, 2014; Shomstein & Yantis, 2006; Wu, Weissman, Roberts, & Woldorff, 2007; Zatorre, Mondor, & Evans, 1999).

Alternatively, or in addition, a stronger HbO response in the spatially separated versus colocated configurations could also be interpreted in support of the notion that right LFCx HbO activity correlates with overall higher speech intelligibility in the spatially separated configuration. However, converging evidence from recent studies in NH listeners finds physiological correlates of speech intelligibility in the *left* hemisphere and at the level of auditory cortex as opposed to LFCx (Olds et al., 2016; Pollonini et al., 2014; Scott, Rosen, Beaman, Davis, & Wise, 2009). It is possible that here, listeners had to spend more listening effort in the spatially colocated versus separated configurations. However, comparing noise-vocoded versus unprocessed speech in quiet, or in competing background speech, previous work finds that increased effort differentially activates the *left* inferior frontal gyrus (Wiggins, Wijayasiri, & Hartley, 2016a; Wijayasiri, Hartley, & Wiggins, 2017). Moreover, testing NH listeners with a two-back working memory task on auditory stimuli, Noyce and coworkers (2017) confirmed the existence of auditory-biased LFCx regions, suggesting that here, the observed physiological correlates of

spatial release from masking may be caused by differences in utilization of short-term memory across the two spatial configurations. Together, the current findings support a hypothesis already proposed by others (Papesh, Folmer, & Gallun, 2017) that a cortical representation of spatial release from masking exists and suggest that assessment of right LFCx activity is a viable objective physiological measure of spatial release from masking.

Recent work shows that decoding of cortical responses is a feasible measure for determining which talker a listener attends to (e.g., Choi, Rajaram, Varghese, & Shinn-Cunningham, 2013; Mesgarani & Chang, 2012; Mirkovic, Debener, Jaeger, & De Vos, 2015; O'sullivan et al., 2104).

Moreover, previous physiological work on speech perception in situations with EM or IM shows recruitment of frontal–parietal regions when listening to speech with EM (Scott, Rosen, Wickham, & Wise, 2004) and suggests that the left superior temporal gyrus is differentially recruited for IM, whereas recruitment of the right superior temporal gyrus is comparable for both types of masker (Scott et al., 2009). With the current paradigm, LFCx recruitment could be used to predict whether or not a listener attends to spatial attributes of sound, a question to be investigated by future work.

Utility of fNIRS as Objective Measure of Auditory Attention

A growing literature shows that fNIRS recordings are a promising tool for assessing the neurobiological basis of clinical outcomes in cochlear implant users (e.g., Dewey & Hartley, 2015; Lawler, Wiggins, Dewey, & Hartley, 2015; McKay et al., 2106; van de Rijt et al., 2016). Cochlear implants are ferromagnetic devices, and when imaged with magnetic resonance imaging (MRI), electroencephalography, or magnetoencephalography, the implants typically cause large electromagnetic artifacts and are sometimes even unsafe for use inside the imaging device. In contrast to MRI, electroencephalography and magnetoencephalography, fNIRS uses light to measure HbO signals and thus does not produce electromagnetic artifacts when used in conjunction with cochlear implants. Moreover, compared with functional MRI machines, fNIRS scanners are quiet, they do not require the listener to remain motionless and are thus more child friendly (cf. Bortfeld, Wruck, & Boas, 2007), and they are generally more cost effective.

However, previous work using fNIRS for assessing auditory functions found highly variable responses to auditory speech at the group level (Wiggins, Anderson, Kitterick, & Hartley, 2016). To reduce across-listener variability, here, we used the individual's own maximal amplitude during controlled breathing for normalizing the HbO traces during the auditory task, followed by fitting a GLM where we regressed out nuisance signals from a shallow trace that recorded blood oxygenation close to the surface of the skull. Results demonstrate that fNIRS is a feasible approach for characterizing central auditory function in NH listeners.

Objective measures of masked speech identification in IM could, for instance, be used to assess the neurobiological basis for predicting rehabilitative success in newly implanted individuals. A long-term goal of our work is thus to establish an objective measure of auditory attention that could be used to study central nervous function in cochlear implant users. Here, we find that fNIRS is a promising tool for recording objective measures of spatial auditory attention in NH listeners, with potential application in cochlear implant users.

Conclusions

Two experiments demonstrated that when NH listeners are tasked with detecting the presence of target keywords in a situation with IM, bilateral LFCx HbO responses, as assessed through fNIRS, carry information about whether or not a listener is attending to sound. In addition, right LFCx responses were stronger in a spatially separated as compared with a colocated configuration, suggesting that right LFCx activity is associated with spatially directed attention.

References

Ahveninen, J., Jääskeläinen, I.P., Raij, T., Bonmassar, G., Devore, S., Hämäläinen, M., Levänen, S., Lin, F.H., Sams, M., Shinn-Cunningham, B.G. and Witzel, T., 2006. Task-modulated "what" and "where" pathways in human auditory cortex. *Proceedings of the National Academy of Sciences*, *103*(39), pp.14608–14613. doi:10.1073/pnas.0510480103

Anderson, S., & Kraus, N. (2010). Objective neural indices of speech-in-noise perception. *Trends in Amplification*, *14*(2), 73–83. doi:10.1177/1084713810380227

Bickler, P. E., Feiner, J. R., & Rollins, M. D. (2013). Factors affecting the performance of 5 cerebral oximeters during hypoxia in healthy volunteers. *Anesthesia and Analgesia*, *117*, 813–823. doi:10.1213/ANE.0b013e318297d763

Bortfeld, H., Wruck, E., & Boas, D. A. (2007). Assessing infants' cortical response to speech using near-infrared spectroscopy. *Neuroimage*, *34*(1), 407–415. doi:10.1016/j.neuroimage.2006.08.010

Brigadoi, S., & Cooper, R. J. (2015). How short is short? Optimum source–detector distance for short-separation channels in functional near-infrared spectroscopy. *Neurophotonics*, *2*(2), 025005. doi:10.1117/1.nph.2.2.025005

Brungart, D., Chang, P., Simpson, B., & Wang, D. (2006). Isolating the energetic component of speech-on-speech masking with ideal time-frequency segregation. *The Journal of the Acoustical Society of America*, *120*, 4007–4018. doi:10.1121/1.2363929

Brungart, D. S. (2001). Informational and energetic masking effects in the perception of two simultaneous talkers. *The Journal of the Acoustical Society of America*, *109*(3), 1101–1109. doi:10.1121/1.1345696

Brungart, D. S., & Simpson, B. D. (2002). The effects of spatial separation in distance on the informational and energetic masking of a nearby speech signal. *The Journal of the Acoustical Society of America*, *112*(2), 664–676. doi:10.1121/1.1490592

Carhart, R., Tillman, T. W., & Johnson, K. R. (1967). Release of masking for speech through interaural time delay. *The Journal of the Acoustical Society of America*, *42*(1), 124–138. doi:10.1121/1.1910541

Carlyon, R. P. (2004). How the brain separates sounds. *Trends in Cognitive Sciences*, *8*(10), 465–471. doi:10.1016/j.tics.2004.08.008

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

Conway, A. R., Cowan, N., & Bunting, M. F. (2001). The cocktail party phenomenon revisited: The importance of working memory capacity. *Psychonomic Bulletin & Review*, *8*(2), 331–335. doi:10.3758/BF03196169

Cooke, M., Garcia Lecumberri, M. L., & Barker, J. (2008). The foreign language cocktail party problem: Energetic and informational masking effects in non-native speech perception. *The Journal of the Acoustical Society of America*, *123*(1), 414–427. doi:10.1121/1.2804952

Cope, M., & Delpy, D. T. (1988). System for long term measurement of cerebral blood and tissue oxygenation on newborn infants by near infrared transillumination. *Medical and Biological Engineering and Computing*, *26*(3), 289–294. doi:10.1007/BF02447083

Cusack, R., Decks, J., Aikman, G., & Carlyon, R. P. (2004). Effects of location, frequency region, and time course of selective attention on auditory scene analysis. *Journal of Experimental Psychology: Human Perception and Performance*, *30*(4), 643doi:10.1037/0096-1523.30.4.643

Dajani, H. R., & Picton, T. W. (2006). Human auditory steady-state responses to changes in interaural correlation. *Hearing Research*, *219*(1–2), 85–100. doi:10.1016/j.heares.2006.06.003

Darwin, C. J., Brungart, D. S., & Simpson, B. D. (2003). Effects of fundamental frequency and vocal-tract length changes on attention to one of two simultaneous talkers. *The Journal of the Acoustical Society of America*, *114*(5), 2913–2922. doi:10.1121/1.1616924

Darwin, C. J., & Hukin, R. W. (1997). Perceptual segregation of a harmonic from a vowel by interaural time difference and frequency proximity. *The Journal of the Acoustical Society of America*, *102*(4), 2316–2324. doi:10.1121/1.419641

Darwin, C. J., & Hukin, R. W. (1998). Perceptual segregation of a harmonic from a vowel by interaural time difference in conjunction with mistuning and onset asynchrony. *The Journal of the Acoustical Society of America*, *103*(2), 1080–1084. doi:10.1121/1.421221

Darwin, C. J., & Hukin, R. W. (2000). Effectiveness of spatial cues, prosody, and talker characteristics in selective attention. *The Journal of the Acoustical Society of America*, *107*(2), 970–977. doi:10.1121/1.428278

Dewey, R. S., & Hartley, D. E. (2015). Cortical cross-modal plasticity following deafness measured using functional near-infrared spectroscopy. *Hearing Research*, *325*, 55–63. doi:10.1016/j.heares.2015.03.007

Ellis, D. P. (2010). *Time-domain scrambling of audio signals in Matlab*. Retrieved from http://www.ee.columbia.edu/~dpwe/resources/matlab/scramble/

Ferrari, M., & Quaresima, V. (2012). A brief review on the history of human functional near-infrared spectroscopy (fNIRS) development and fields of application. *NeuroImage, 63*(2), 921–935. doi:10.1016/j.neuroimage.2012.03.049

Freyman, R. L., Balakrishnan, U., & Helfer, K. S. (2001). Spatial release from informational masking in speech recognition. *The Journal of the Acoustical Society of America*, *109*, 2112–2122. doi:10.1121/1.1354984

Glyde, H., Buchholz, J. M., Dillon, H., Cameron, S., & Hickson, L. (2013). The importance of interaural time differences and level differences in spatial release from masking. *The Journal of the Acoustical Society of America*, *134*(2), EL147–EL152. doi:10.1121/1.4812441

Goldsworthy, R. L., & Greenberg, J. E. (2004). Analysis of speech-based speech transmission index methods with implications for nonlinear operations. *The Journal of the Acoustical Society of America*, *116*(6), 3679–3689. doi:10.1121/1.1804628

Hejna, D., & Musicus, B. R. (1991). *The SOLAFS time-scale modification algorithm* (BBN Technical Report). Retrieved from: http://www.ee.columbia.edu/~dpwe/papers/HejMus91-solafs.pdf

Huppert, T. J., Diamond, S. G., Franceschini, M. A., & Boas, D. A. (2009). HomER: A review of time-series analysis methods for near-infrared spectroscopy of the brain. *Applied Optics*, *48*(10), D280–D298. doi:10.1364/ao.48.00d280

Ihlefeld, A., & Shinn-Cunningham, B. (2008a). Spatial release from energetic and informational masking in a selective speech identification task. *The Journal of the Acoustical Society of America*, *123*(6), 4369–4379. doi:10.1121/1.2904826

Ihlefeld, A., & Shinn-Cunningham, B. (2008b). Disentangling the effects of spatial cues on selection and formation of auditory objects. *The Journal of the Acoustical Society of America*, *124*(4), 2224–2235. doi:10.1121/1.2973185

Jones, D., Alford, D., Bridges, A., Tremblay, S., & Macken, B. (1999). Organizational factors in selective attention: The interplay of acoustic distinctiveness and auditory streaming

in the irrelevant sound effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(2), 464. doi:10.1037/0278-7393.25.2.464

Jones, G. L., & Litovsky, R. Y. (2011). A cocktail party model of spatial release from masking by both noise and speech interferers. *The Journal of the Acoustical Society of America*, 130(3), 1463–1474. doi:10.1121/1.3613928

Kidd, G., Best, V., & Mason, C. R. (2008). Listening to every other word: Examining the strength of linkage variables in forming streams of speech. *The Journal of the Acoustical Society of America*, 124(6), 3793–3802. doi:10.1121/1.2998980

Kidd, G. Jr., & Colburn, H. S. (2017). Informational masking in speech recognition. In J. C. Middlebrooks, J. Z. Simon, A. N. Popper, & R. R. Fay (Eds.), *The auditory system at the cocktail party* (pp. 75–109). New York, NY: Springer Naturedoi:10.1007/978-3-319-51662-2_4

Kidd, G. Jr., Mason, C. R., Best, V., & Marrone, N. (2010). Stimulus factors influencing spatial release from speech-on-speech masking. *The Journal of the Acoustical Society of America*, 128(4), 1965–1978. doi:10.1121/1.3478781

Knauth, M., Heldmann, M., Münte, T. F., & Royl, G. (2017). Valsalva-induced elevation of intracranial pressure selectively decouples deoxygenated hemoglobin concentration from neuronal activation and functional brain imaging capability. *NeuroImage*, 162, 151–161. doi:10.1016/j.neuroimage.2017.08.062

Kocsis, L., Herman, P., & Eke, A. (2006). The modified Beer–Lambert law revisited. *Physics in Medicine and Biology*, 51(5), N91–N98. doi:10.1088/0031-9155/51/5/n02

Larson, E., & Lee, A. K. (2014). Switching auditory attention using spatial and non-spatial features recruits different cortical networks. *NeuroImage*, 84, 681–687. doi:10.1016/j.neuroimage.2013.09.061

Lawler, C. A., Wiggins, I. M., Dewey, R. S., & Hartley, D. E. (2015). The use of functional near-infrared spectroscopy for measuring cortical reorganisation in cochlear implant users: A possible predictor of variable speech outcomes? *Cochlear Implants International*, 16, S30–S32. doi:10.1179/1467010014Z.000000000230

Lindquist, M. A., Loh, J. M., Atlas, L. Y., & Wager, T. D. (2009). Modeling the hemodynamic response function in fMRI: Efficiency, bias and mis-modeling. *NeuroImage*, 45(1), S187–S198. doi:10.1016/j.neuroimage.2008.10.065

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

Mattys, S. L., Brooks, J., & Cooke, M. (2009). Recognizing speech under a processing load: Dissociating energetic from informational factors. *Cognitive Psychology*, 59(3), 203–243. doi:10.1016/j.cogpsych.2009.04.001

McKay, C. M., Shah, A., Seghouane, A. K., Zhou, X., Cross, W., & Litovsky, R. (2016). Connectivity in language areas of the brain in cochlear implant users as revealed by fNIRS. In: P. van Dijk, D. Başkent, E. Gaudrain, E. de Kleine, A. Wagner, C. Lanting (Eds.), *Physiology, psychoacoustics and cognition in normal and impaired hearing* (pp. 327–335). Cham, Switzerland: Springer.

Mesgarani, N., & Chang, E. F. (2012). Selective cortical representation of attended speaker in multi-talker speech perception. *Nature*, 485(7397), 233. doi:10.1038/nature11020

Michalka, S., Kong, L., Rosen, M., Shinn-Cunningham, B., & Somers, D. (2015). Short-term memory for space and time flexibly recruit complementary sensory-biased frontal lobe attention networks. *Neuron*, 87(4), 882–892. doi:10.1016/j.neuron.2015.07.028

Micheyl, C., Hunter, C., & Oxenham, A. J. (2010). Auditory stream segregation and the perception of across-frequency synchrony. *Journal of Experimental Psychology: Human Perception and Performance*, 36(4), 1029doi:10.1037/a0017601

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

Molavi, B., & Dumont, G.A. (2012). Wavelet-based motion artifact removal for functional near-infrared spectroscopy. *Physiological Measurement*, 33(2), 259.

Noyce, A. L., Cestero, N., Michalka, S. W., Shinn-Cunningham, B. G., & Somers, D. C. (2017). Sensory-biased and multiple-demand processing in human lateral frontal cortex. *The Journal of Neuroscience*, 37(36), 8755–8766. doi:10.1523/jneurosci.0660-17.2017

Olds, C., Pollonini, L., Abaya, H., Larky, J., Loy, M., Bortfeld, H., & Oghalai, J. S. (2016). Cortical activation patterns correlate with speech understanding after cochlear implantation. *Ear and Hearing*, 37(3), e160–e172. doi:10.1097/aud.0000000000000258

O'sullivan, J. A., Power, A. J., Mesgarani, N., Rajaram, S., Foxe, J. J., Shinn-Cunningham, B. G.,...Lalor, E. C. (2014). Attentional selection in a cocktail party environment can be decoded from single-trial EEG. *Cerebral Cortex*, 25(7), 1697–1706. doi:10.1093/cercor/bht355

Papesh, M. A., Folmer, R. L., & Gallun, F. J. (2017). Cortical measures of binaural processing predict spatial release from masking performance. *Frontiers in Human Neuroscience*, 11, 124. doi:10.3389/fnhum.2017.00124

Patterson, R. D., & Holdsworth, J. (1996). A functional model of neural activity patterns and auditory images. *Advances in Speech, Hearing and Language Processing*, 3(Part B), 547–563.

Pei, Y., Wang, Z., & Barbour, R. L. (2007). *NAVI-SciPort solution: A problem solving environment (PSE) for NIRS data analysis*. Poster at Human Brain Mapping, Chicago, IL.

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. *Hearing Research*, 309, 84–93. doi:10.1016/j.heares.2013.11.007

Scott, S. K., Rosen, S., Beaman, C. P., Davis, J. P., & Wise, R. J. (2009). The neural processing of masked speech: Evidence for different mechanisms in the left and right temporal lobes. *The Journal of the Acoustical Society of America*, 125(3), 1737–1743. doi:10.1121/1.3050255

Scott, S. K., Rosen, S., Wickham, L., & Wise, R. J. (2004). A positron emission tomography study of the neural basis of informational and energetic masking effects in speech perception. *The Journal of the Acoustical Society of America*, 115(2), 813–821. doi:10.1121/1.1639336

Shinn-Cunningham, B. G. (2008). Object-based auditory and visual attention. *Trends in Cognitive Sciences*, *12*(5), 182–186.

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

Shinn-Cunningham, B. G., Ihlefeld, A., Satyavarta, S., & Larson, E. (2005). Bottom-up and topdown influences on spatial unmasking. *Acta Acustica*, *91*, 967–979.

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

Talairach, J., Rayport, M., & Tournoux, P. (1988). *Co-planar stereotaxic atlas of the human brain: 3-dimensional proportional system: An approach to cerebral imaging*. Stuttgart, Germany: Thieme.

Thomason, M. E., Foland, L. C., & Glover, G. H. (2006). Calibration of BOLD fMRI using breath holding reduces group variance during a cognitive task. *Human Brain Mapping*, *28*(1), 59–68. doi:10.1002/hbm.20241

Undurraga, J. A., Haywood, N. R., Marquardt, T., & McAlpine, D. (2016). Neural representation of interaural time differences in humans—An objective measure that matches behavioural performance. *Journal of the Association for Research in Otolaryngology*, *17*(6), 591–607 doi:10.1007/s10162-016-0584-6

van de Rijt, L. P., van Opstal, A. J., Mylanus, E. A., Straatman, L. V., Hu, H. Y., Snik, A. F., & van Wanrooij, M. M. (2016). Temporal cortex activation to audiovisual speech in normal hearing and cochlear implant users measured with functional near-infrared spectroscopy. *Frontiers in Human Neuroscience*, *10*, 48. doi:10.3389/fnhum.2016.00048

von Kriegstein, K., Griffiths, T. D., Thompson, S. K., & McAlpine, D. (2008). Responses to interaural time delay in human cortex. *Journal of Neurophysiology*, *100*(5), 2712–2718. doi:10.1152/jn.90210.2008

Wack, D. S., Cox, J. L., Schirda, C. V., Magnano, C. R., Sussman, J. E., Henderson, D., & Burkard, R. F. (2012). Functional anatomy of the masking level difference, an fMRI study. *PLoS One*, *7*(7), e41263. doi:10.1371/journal.pone.0041263

Wiggins, I. M., Anderson, C. A., Kitterick, P. T., & Hartley, D. E. (2016). Speech-evoked activation in adult temporal cortex measured using functional near-infrared spectroscopy (fNIRS): Are the measurements reliable? *Hearing Research*, *339*, 142–154. doi:10.1016/j.heares.2016.07.007

Wiggins, I. M., Wijayasiri, P., & Hartley, D. (2016). Shining a light on the neural signature of effortful listening. *The Journal of the Acoustical Society of America*, *139*(4), 2074–2074. doi:10.1121/1.4950150

Wijayasiri, P., Hartley, D. E., & Wiggins, I. M. (2017). Brain activity underlying the recovery of meaning from degraded speech: A functional near-infrared spectroscopy (fNIRS) study. *Hearing Research*, *351*, 55–67. doi:10.1016/j.heares.2017.05.010

Wong, W. Y., & Stapells, D. R. (2004). brain stem and cortical mechanisms underlying the binaural masking level difference in humans: An auditory steady-state response study. *Ear and Hearing*, *25*(1), 57–67. doi:10.1097/01.aud.0000111257.11898.64

Wu, C., Weissman, D., Roberts, K., & Woldorff, M. (2007). The neural circuitry underlying the executive control of auditory spatial attention. *Brain Research*, *1134*, 187–198. doi:10.1016/j.brainres.2006.11.088

Xia, J., Kalluri, S., Micheyl, C., & Hafter, E. (2017). Continued search for better prediction of aided speech understanding in multi-talker environments. *The Journal of the Acoustical Society of America*, *142*(4), 2386–2399. doi:10.1121/1.5008498

Yoshitani, K., Kawaguchi, M., Miura, N., Okuno, T., Kanoda, T., Ohnishi, Y., & Kuro, M. (2007). Effects of hemoglobin concentration, skull thickness, and the area of the cerebrospinal fluid layer on near-infrared spectroscopy measurements. *Anesthesiology*, *106*(3), 458–462. doi:10.1097/00000542-200703000-00009

Zatorre, R. J., Mondor, T. A., & Evans, A. C. (1999). Auditory attention to space and frequency activates similar cerebral systems. *NeuroImage*, *10*(5), 544–554. doi:10.1006/nimg.1999.0491.