ABSTRACT

USING FNIRS AS AN OBJECTIVE MEASURE OF SUSCEPTIBILITY TO INFORMATIONAL MASKING

by
Min Zhang

Resolving complicated auditory scenes is crucial for daily communication where background sound is often present. However, most hearing aid (HA) and cochlear implant (CI) users have difficulties understanding speech when competing sound sources are present, resulting in reduced job opportunities and increased risk for social isolation. Perceptual interference from background sound is called auditory masking. At least two distinct masking phenomena exist, called energetic and informational masking (EM and IM).

In the first masking phenomenon, EM, target and masker energies coincide at the same time and frequency. Computational and physiological models of cochlear auditory processing can reliably predict listeners' performances in EM situations, demonstrating that EM is primarily caused by peripheral processes. Therefore, even ideal HA or CI device could not restore information that is energetically masked at the cochlea. In contrast, the second masking phenomenon, informational masking (IM), occurs when target-like sound is in the background, even when cochlear models do not predict much interference. Because IM is thought to arise from central interference downstream from the cochlea, HA or CI devices could be designed to overcome IM. However, the mechanisms underlying IM are currently not understood. Tools that objectively predict and compensate for an individual's susceptibility to IM do not currently exist in clinical practice. Thus, there is a timely need to elucidate the mechanisms underlying IM and to develop safe, quiet and inexpensive brain imaging tools that could guide the fitting of HA and CI devices.

In this dissertation, psychometrical testing is combined with functional near-infrared spectroscopy (fNIRS) as a brain imaging tool to objectively measure the listeners' susceptibility to IM. Chapter 1 reviews the literature. Using psychometric testing, Chapter 2 then demonstrates that susceptibility to IM negatively correlates with susceptibility to crowding in vision, a superficially similar and better-understood phenomenon. Domain-general selective attention, motivation, effort, or vigilance would have predicted a positive association between the two phenomena. Thus, Chapter 2 demonstrates that additional central processing must underly IM. In search for neural correlates of IM mechanisms, three fNIRS experiments are then conducted. Chapter 3, establishes fNIRS as a viable objective measure for sensing whether a normally hearing listener actively listens to an auditory scene vs. passively hears sound. Extending this method, Chapter 4 shows that listening with IM interference causes auditory-task evoked hemodynamic responses near auditory regions in the lateral frontal cortex and superior temporal gyrus, bilaterally. However, only the hemodynamic responses near the superior temporal gyrus predict individuals' susceptibility to IM (R^2 = 20-43%). Using machine learning techniques, Chapter 5 confirms robust test-retest reliability of the fNIRS protocols used in Chapters 3 and 4. Chapter 6 summarized and discusses the results of this dissertation.

USING FNIRS AS AN OBJECTIVE MEASURE OF SUSCEPTIBILITY TO INFORMATIONAL MASKING

by
Min Zhang

A Dissertation
Submitted to the Faculty of
New Jersey Institute of Technology and
Rutgers Biomedical and Health Sciences – Newark
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy in Biomedical Engineering

May 2021

APPROVAL PAGE

USING FNIRS AS AN OBJECTIVE MEASURE OF SUSCEPTIBILITY TO INFORMATIONAL MASKING

Min Zhang

Dr. Antje Ihlefeld, Dissertation Advisor Date
Assistant Professor of Biomedical Engineering, NJIT

Dr. Sergei Adamovich, Committee Member Date
Professor of Biomedical Engineering, NJIT

Dr. Tara Alvarez, Committee Member Date
Professor of Biomedical Engineering, NJIT

Dr. Carrie Esopenko, Committee Member Date
Assistant Professor of Rehabilitation and Movement Science, Rutgers University, School of Graduate Studies-RBHS Campus

Dr. Jorge Serrador, Committee Member Date
Associate Professor of Pharmacology, Physiology and Neuroscience, Rutgers University, School of Graduate Studies-RBHS Campus

BIOGRAPHICAL SKETCH

Author:	Min Zhang
Degree:	Doctor of Philosophy
Date:	May 2021

Undergraduate and Graduate Education:

- Doctor of Philosophy in Biomedical Engineering,
  New Jersey Institute of Technology and RBHS, Newark, NJ, 2021
- Master of Science in Biomedical Engineering,
  New Jersey Institute of Technology, Newark, NJ, 2014
- Bachelor of Science in Engineering (Bio-engineering),
  Nanyang Technological University, Singapore, 2010

Major: Biomedical Engineering

Presentations and Publications:

Zhang M and Ihlefeld A, “Sensory Resolution Drives Auditory Responses in Lateral Frontal Cortex,” *International Congress on Acoustics*, vol. 23, no. 1, pp. 5659-5663, 2019.

Zhang M, Ying YLM and Ihlefeld A, “Spatial Release from Masking: Evidence from Near Infrared Spectroscopy,” *Trends in Hearing*, doi: https://doi.org/10.1177/2331216518817464,2018.

Zhang M and Ihlefeld A, “Using Functional Near-infrared Spectroscopy to Assess Auditory Attention,” *The Journal of the Acoustical Society of America*, vol. 143, no. 3, pp. 1847, 2018.

Nuti S, Akiyama T, Zhang M, and Ihlefeld A, “Processing Heart Rate Variability from fNIRS to Determine Listening Effort,” *Biomedical Engineering Society (BMES) Annual Meeting*, 2019.

Zhang M and Ihlefeld A, “The Role of Center Frequency in Informational Masking,” *43rd Northeast Bioengineering Conference (NEBEC)*, 2017.

I would like to dedicate this dissertation to my loving wife, Wen. Thank you for the support and understanding you have given me over the years. Without you, I would not have made it this far.

I also dedicate this dissertation to my parents. Thank you for the endless love and encouragement. It was hard living away from you these years, and I miss you.

我把这篇论文献给我的爱人朱文。感谢你多年来给予我的支持和理解。没有你，我不会走得这么远。

我还将这篇论文献给我的父母。感谢他们无尽的爱与鼓励。这些年来，没有陪在你们身边，对不起，我想你们。

ACKNOWLEDGMENT

The years I spent working in Dr. Antje Ihlefeld's Engineering Speech and Hearing (NESH) lab at NJIT were the highlight of my life. Dr. Ihlefeld, thank you so much for your guidance and mentorship; you are the best teacher I have ever had.

To Dr. Tara Alvarez and Dr. Bharat Biswal, thank you for giving us access to the fNIRS scanner and the help you gave me to get started with the recordings. To Dr. Jorge Serrador, thank you for the opportunity to work in your lab for my lab rotation. I gained valuable insight from this experience. I would also like to thank my committee members Dr. Sergei Adamovich and Dr. Carrie Esopenko, for reviewing this dissertation and providing me with valuable feedback.

I would like to express my appreciation to the department of biomedical engineering for their financial support, I would like to extend my thanks to the office of the provost. Finally I would like to thank the national institutes of health for the grant funding (PC 24-18 to AI).

It was a great experience working with the NESH gang. Thank you for the friendship and help over the years.

To Nima, you are the best lab mate one could ever ask for; thank you! To Thuy Tien, Taiga and Sravana, thank you for helping me with data collection. It was a pleasure working with you.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS
(Continued)

Chapter **Page**

TABLE OF CONTENTS (Continued)

Chapter Page

LIST OF TABLES

LIST OF FIGURES

Figure **Page**

LIST OF FIGURES (Continued)

LIST OF FIGURES (Continued)

Figure **Page**

CHAPTER 1

INTRODUCTION

Hearing is vital to daily communication. In most real-life situations, background sound often interferes with the sound we are interested in hearing. The ability to decode complicated auditory scenes is required to hear out the sound of interest (target) from background interference (masker), an ability most normal-hearing listeners took for granted. Even for normal-hearing listeners, there are cases where listeners have difficulties understanding speech with competing sound sources present (Hind et al., 2011; Kujawa and Liberman, 2009). This problem is more common among hearing aid and cochlear implant (CI) users (Bugannim et al., 2019; Caldwell and Nittrouer, 2013; Ching et al., 2018; Eisenberg et al., 2016; Wilson and Dorman 2008).

Due to the cochlea's snail shape, sound waves of different frequencies could propagate to different depths of the cochlea tube and stimulate hair cells at specific areas, resulting in a tonotopic mapping along the basilar-membrane. Fletcher (1940) suggested this tonotopic mapping can be modeled as a filter bank consisting of overlapping bandpass filters or "auditory filters." Energetic masking (EM) occurs when the target and masker coincide in time and frequency, falling into the same frequency band for the auditory filter at the same time, which leads to the neural activities evoked by the target masked by that evoked by the masker. Patterson (1976) used the notched noise method to derive the shape of auditory filters by presenting the pure tone target with two bands of noise maskers flanking the target in frequency. Computational models based on auditory filters have since been developed and can predict performance in EM situations (Heinz et al., 2001; Jepsen et al., 2008). Responses measured from the auditory brainstem showed correlation with the speech in noise performance, supporting EM originates from peripheral auditory processing (Anderson et al., 2010).

The peripheral models of EM can not explain the other and less understood kind of masking. Those masking situations beyond traditional EM were called "perceptual masking" (Carhart et al. 1969) then "informational masking" (IM, Pollack, 1975; Watson and Kelly, 1981; Lutfi, 1989). Early work on IM had found the link between IM and auditory attention (Leek et al., 1991). They also tried to use the filter bandwidth method from EM to quantify IM's effect on listeners by calculating the equivalent rectangular bandwidths (ERB) for attentional filters. Later studies have found large variability between listeners' performances in IM situations and many factors contributing to IM. It is often thought the uncertainty about the masker (Arbogast et al., 2005; Brungart, 2001; Freyman et al., 1999; Kidd et al., 2005) and the similarity between target and masker (Micheyl et al.,2010, Oh and Lutfi, 2000; Martin et al., 2012; Bregman, 1990; Kidd et al., 2008) contribute to IM. A unified framework that can consider different acoustic features contributing to IM and a model that could reliably predict listener performance are still being developed for IM.

This dissertation aims to develop functional near-infrared spectroscopy (fNIRS) as an objective measure of IM susceptibility. In this chapter, the current findings regarding IM were summarized, and the role of auditory attention in alleviating IM's effect was discussed. The fNIRS recording system was then briefly explained, followed by the goals and structure of this thesis.

1.1 Informational Masking

A classic example of an IM situation is a "cocktail party" environment (Cherry, 1953), where the target speaker is masked by complex maskers consisting of noise and speech. It is assumed IM involves not only peripheral processing but also higher-level processing. Uncertainty regarding the masker and similarity between target and masker were thought to be the main contributors to IM.

Uncertainty means the acoustic properties of the masker vary at random. Those acoustic properties could be the frequency, intensity, duration, or location of the masker. In multi-tone pattern discrimination tasks, Lutfi (Lutfi, 1993) used the term "component's relative entropy (CoRE)" to quantify the relative effects of frequency, intensity, and duration. By applying this measure for uncertainty to previous studies, the results showed listeners' performance (the threshold for change detection) worsened with an increase in CoRE, indicating that uncertainty contributes to IM. Later, using a coordinate response measure (CRM) task, the effect of uncertainty in speech recognition tasks was also studied (Brungart and Simpson, 2004; Kidd et al., 2008). They found similar results; increased randomness in masker increased the effect of IM. The acoustic similarity between target and masker includes: a) temporally correlated (Micheyl et al., 2010); b) spatially close (Arbogast et al., 2005); c) adjacent in frequency (Brungart, 2001; Kidd et al., 2008). Studies have shown that increasingly similar the masker is to the target resulting in an IM's increased effect.

Although the acoustic properties that could give rise to IM were studied thoroughly, there has been criticism regarding IM's concept, arguing that the observed effect of similarity and uncertainty on IM situation performance results from different central processing (Watson, 2005). A unified framework that could account for various phenomena in IM has been proposed, such as the information-divergence hypothesis (Lutfi et al., 2013). This hypothesis used a single variable to measure the statistical divergence between target and masker that captures the effect of both uncertainty and similarity. Experimental results of listener performances from multi-tone pattern discrimination task, word recognition task, sound source identification task, and sound localization task showed correlations between performance and statistical divergence, suggesting that the effect of uncertainty and similarity on IM are from the same underlying central processing.

Building on the premise that different effects associated with IM share the same underlying neural mechanism, there have been studies to examine IM from the neural perspective by drawing comparisons between vision and audition regarding object-based attention (Shinn-Cunningham and Wang, 2008; Shinn-Cunningham, 2008). Functional magnetic resonance imaging (fMRI) studies revealed a bottom-up and top-down attentional processing in the vision. Frontal eye field (FEF) and intraparietal sulcus (IPS) control top-down attention, while ventral frontal cortex (VFC) and temporoparietal junction (TPJ) are responsible for the bottom-up processing (Corbetta and Shulman, 2002). A similar top-down and bottom-up attentional processing was proposed for spatial unmasking for auditory processing (Shinn-Cunningham et al., 2005). In addition to traditional bottom-up processing, higher-level attention-based processing also modulates the process to form auditory objects. Differences in perceptual features between target and masker facilitate the segregation of competing streams and the formation of target/masker objects. (Figure 1.1).

Many studies have found higher-level perceptual features, or cues, that could help listeners direct auditory attention selectively and improve listeners' performance in IM situations. Those cues include semantic cues (Zekveld et al., 2013), spatial cues (Carlile and Corkhill, 2015), priming cues (Freyman et al., 2004), and visual cues (Wightman et al., 2006). Those psychophysical experiments supported the theory that any cue that contributes to selective attention can improve top-down attention's modulation effect, resulting in better segregation of competing streams and auditory objects' formation. (Cusack et al., 2004; Ihlefeld and Shinn-Cunningham, 2008, Shinn-Cunningham, 2017).

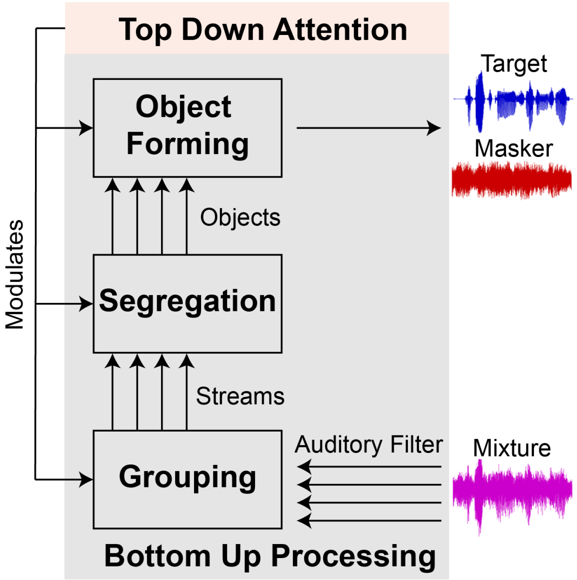

Figure 1.1 The illustration of top-down modulation and bottom-up processing of auditory attention
Source: Shinn-Cunningham, 2008

1.2 Representation of Selective Auditory Attention on the Cortex

Psychophysical studies revealed the role of selective auditory attention in auditory scene analysis in IM situations. Brain imaging studies confirmed the representation of selective attention in the cortex.

Electroencephalogram (EEG) and magnetoencephalography (MEG) studies have found the event-related potential (ERP), alpha oscillation, and frequency-following responses (FFRs) correlate with sound stimuli and selective attention in IM situations. The ERP response is generated from the primary auditory cortex around the superior surface of the temporal lobes (Chen et al., 2011, Flinker et al., 2010), and it peaks after the onset of stimuli. The sensitivity of ERP to the auditory stimuli features supports the bottom-up processing in the auditory attention model (Zhang et al., 2016). Alpha-band power oscillations in the parietal cortex regions had increased when subjects were performing selective auditory attention tasks (Foxe and Snyder, 2011; Wöstmann et al., 2016). Frequency-following responses (FFRs) recorded from subcortical auditory structures showed high temporal and spectral correlation with acoustic properties of sound stimuli. (Aiken and Picton, 2008, Akhoun et al., 2008, Johnson et al., 2005). FFRs encode sound stimuli features in the IM situation that could help listeners direct selective attention (Du et al., 2011; Zhang and Gong, 2019). These findings suggest that alpha oscillations and FFRs are modulated by selective attention, supporting the auditory attention model's top-down modulation.

fMRI offers a much finer spatial resolution of cortex activities and has been used to map out an attention-related network on the cortex. An earlier study had found activations within the posterior parietal, superior temporal, and inferior frontal regions were increased when listeners performed tasks that required selective attention (Pugh et al., 1996). Kong and co-workers (Kong et al., 2014) found that tasks that required auditory spatial attention recruited the superior temporal gyrus (STG) and the supramarginal gyrus (SMG) regions. Later studies using intrinsic functional connectivity (Michalka et al., 2015; Osher, Tobyne, Congden, Michalka, and Somers, 2015) revealed auditory-biased attention networks and visual-biased attention networks consisting of different brain regions. During auditory attention tasks, activities from the transverse gyrus intersecting precentral sulcus (tgPCS) and caudal inferior frontal sulcus (cIFS) correlated with activities recorded STG, suggesting they are part of the same auditory-biased networks.

In addition to brain imaging studies, multi-electrode surface recording on the auditory cortex showed a cortical representation of sound stimuli reflecting attention. A linear classifier was able to identify the word and speaker the listener was attending to (Mesgarani and Chang, 2012), suggesting that the cortical representation of sound includes the listener's selective attention. A lesion study showed patients with unilateral auditory-cortex lesions had worse performance in tone-pattern detection tasks compared to the control group, supporting the hypothesis that there are upper stream neural mechanisms for the effect of IM in addition to peripheral processing (Prilop and Gutschalk, 2019).

These findings from brain imaging, multi-electrode surface recording, and lesion studies support the hypothesis that auditory attention has a cortical representation. However, due to the nature of listed imaging techniques and the ferromagnetic nature of CI divides, the study of underlying neural mechanisms for IM is mainly limited to normal-hearing listeners or a select few qualified for multi-electrode surface recording or lesion studies. Alleviating the effect of IM is important for CI users to perform well in real-life situations. Most CI users demonstrate sizable improvements in speech perception after implantation (Hochberg et al., 1992; Kiefer et al., 1996). However, CI users' speech recognition performance is vulnerable to IM with competing speech or speech-shaped noise as the masker (Müller-Deile et al., 1995; Nelson et al., 2003). The ability to direct selective auditory attention is essential for alleviating the effect of IM. Bilateral CIs (BiCI) intend to restore binaural cues. However, most BiCI users benefit little from spatial cues (Loizou, 2009). Some BiCI users show release from masking when target and interferer are presented in opposite ears instead of the same ear, but many do not benefit at all (Goupell et al., 2016). It is unclear whether BiCI users can direct selective auditory attention. There is a timely need to have a brain imaging technique that is safe for CI users to measure their IM susceptibility.

1.3 Functional Near-infrared Spectroscopy(fNIRS)

Past work using brain imaging techniques confirmed a network of brain regions recruited in IM situations where directing selective attention to the target stream is required. A suitable brain imaging technique is required to study brain functions in IM situations for CI users.

fNIRS utilizes the difference in light absorption spectra for oxy-genated and deoxy-genated hemoglobin (HbO and HbR, respectively), as shown in Figure 1.2B (Villringer and Chance, 1997). If two or more wavelengths of infrared lights are used, it is possible to calculate the HbO and HbR concentrations by applying the modified

Beer-Lambert law (Kocsis et al., 2006). Typical fNIRS systems use infrared lights with wavelengths from 650 to 1000 nm, which can penetrate human tissue and skull to reach the cortex. Infrared light originates from the light source placed on the scalp directing the light into the brain perpendicularly to the scalp surface. The light travels through the tissue, skull, and top level of the cortex following a banana-shaped diffusion path. The paired light detector placed nearby then picks up the emerging light (Figure 1.2A). This banana-shaped diffusion path determines the penetration depth and can be controlled by varying the distance between the paired light source and light detector placed on the scalp.

fNIRS is ideally suited to study brian functions of CI users in IM situation due to it measures infrared-light intensity changes as a result of changes in HbO and HbR concentration while EEG and MEG detect electrical signal from the brain, which is impacted by the electromagnetic artifacts from high-frequency carrier pulses used in CIs. By measuring HbO and HbR concentration changes, fNIRS can produce similar blood-oxygen-level-dependent (BOLD) signals as fMRI to infer the activation status of targeted regions of interest (ROIs). However, fMRI involves a high strength magnetic field and is not safe for CI users due to the ferromagnetic nature of CI. Furthermore, the fMRI scanner produces acoustic noise (McJury and Shellock, 2000) and is very sensitive to motion. fNIRS is quiet and resistant to motion artifacts; thus, it is more suitable for children's studies (Bortfeld et al., 2009).

The fNIRS system is not without its limitations. One limitation of fNIRS is its spatial resolution. fMRI could have millimeter level resolution to study fine structures of the cortex. fNIRS's special resolution is limited by the number of optodes that can be secured on the listener's head. It is in the range of centimeters, which means the regions of interest (ROIs) need to be carefully chosen and functional tasks specifically designed to elicited response to specific ROIs. Caudal inferior frontal sulcus (cIFS)

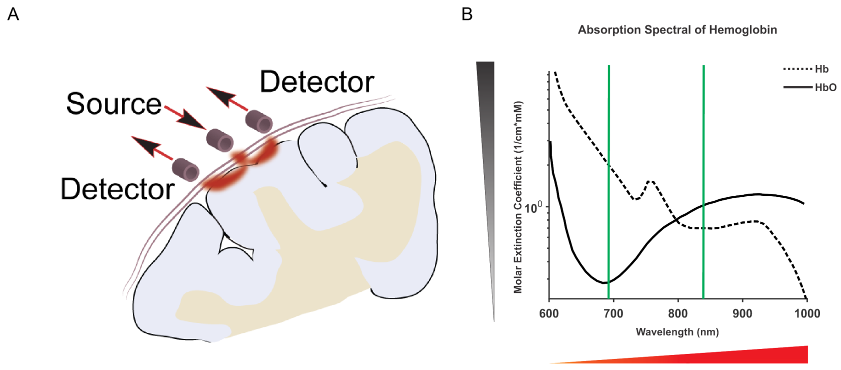
A
Source
Detector
Detector
B
Absorption Spectral of Hemoglobin
Hb
HbO
Molar Extinction Coefficient (1/cm*mM)
10^0
600
700
800
900
1000
Wavelength (nm)

Figure 1.2 (A), near infrared light source and detector pair and the banana-shaped light pathway; (B), near-infrared light has different absorption rates at different wavelengths for HbO and HbR.

and superior temporal gyrus (STG) were chosen based on past fMRI studies. A speech detection task with spatial cues in IM was used as the functional task.

Another limitation of fNIRS is that it is susceptible to psychological fluctuations during recording (Tong et al., 2011; Kirilina et al., 2012). Thus, the fNIRS signal may not accurately represent task-evoked brain activity (Tachtsidis and Scholkmann, 2016). Separation of task-evoked signals from signals of physiological origins can be achieved by incorporating short separation channels which have shorter source-detector distances (less than 1.5 cm) when compared to standard recording channels (3 cm) and are more affected by global physiological fluctuations into the design(Funane et al., 2013). Due to the physical structure of our optodes, we used 1.5 cm in our optode design. The linear mixed effect model (LMEM) was used to regress out not only psychological fluctuations from short separation channels but also take into account listeners-specific attributes.

1.4 Informational Masking and Visual Crowding

Psychophysical experiments and brain imaging studies supported the hypothesis that IM has higher-level neural mechanisms in addition to peripheral processing, and the listener's ability to direct selective attention plays a vital role in separating the target from the masker in IM situations.

Similar to IM in auditory, there is a phenomenon in vision called crowding where nearby maskers clutter or "flank" the visual target causing the target to be unrecognizable. In vision studies, the effect of crowding is quantified by critical spacing, the distance between target and masker required for the target to be recognized. (Whitney and Levi, 2011; Pelli, 2008). There is mapping of cortical representations to unique visual space locations, and there is tonotopic mapping on the auditory cortex to the specific frequency band on the cochlear. Both phenomena involve separating a specific target from similar interference and require object-based attention (Shinn-Cunningham, 2008).

To quantitatively measure listeners' performance under IM situation, we built on the established notched noise method by changing the noise masker to a tonal sequence masker that can be easily confused. By generating the psychometric curve of listeners' performance, we could quantify the listeners' susceptibility to IM. Listeners' susceptibility to IM is compared with their susceptibility to crowding measured in critical spacing. Should there be a relationship between the susceptibilities from two different sensory modalities, it could serve as further support for the idea IM originates from central processing.

1.5 Listening Effort

One potential caveat in the IM task design is the listening effort during auditory attention tasks was not measured and accounted for. Previous studies using behavioral measures and physiological measures have demonstrated the relationship between listening demand listening efforts. The dual-task paradigm is often used to measure listening effort (Picou et al., 2014; Wu et al., 2016). The change in primary task performance can measure listening effort due to increased competition for cognitive resources from more complex secondary tasks. Listening effort can also be quantified via physiologic measurements as well, such as pupil dilation (Piquado et al. 2010; Zekveld and Kramer 2014; Kramer et al. 2016), Electromyography (EMG) activity (Mackersie and Cones 2011), and heart rate variability (Mackersie and Calderon-Moultrie 2016). Experiment results showed listening effort increased with listening task difficulty. In the experiments described in this dissertation, the listening effort was not measured. Ideally, the auditory task should elicit the same listening effort across listeners to make the measured IM susceptibility comparable across listeners.

1.6 Goals of this Dissertation

Restoring speech intelligibility in situations with background sound is an unsolved need for CI users, and objective measures of CI users' susceptibility to IM could be used to facilitate the fitting of CI devices. A non-invasive neuroimaging tool that can be used safely on CI users is required to investigate CI users' ability to direct selective auditory attention in IM tasks.

Due to cochlear implants' ferromagnetic nature, fMRI cannot be used on CI users. Positron Emission Tomography (PET) has a radioactive hazard thus is not suitable for long-term studies and is generally not used in children's studies. CI uses high-frequency pulses to directly stimulate the auditory nerve, which will interfere with EEG or MEG recording. fNIRS uses infrared light to detect hemodynamic responses similar to fMRI but does not have the above shortcomings. Thus, this thesis's primary goal is to develop fNIRS to be an alternative imaging technique. This thesis's secondary gold is to investigate the possible correlation between IM's performance is auditory and crowding performance in vision. If this correlation exists, visual crowding tests could be used to assess or predict CI users' performance prior to implantation.

1.7 Structure of this Dissertation

The first chapter gives an overview of the current findings regarding the role of attention in IM and points out the timely need for a brain imaging technique that is safe and suitable for CI users. The following four chapters are the four experiments conducted.

The first experiment tested the possible correlation between the susceptibility to IM in auditory and the susceptibility to crowding in vision. The results showed a significant negative correlation between the two, which hinted at a common central processing limiting factor for IM and crowding. The following three experiments were about establishing fNIRS as the viable tool to study attention in IM. The first fNIRS experiment built upon the previous fMRI studies and showed bilateral LFCx HbO fNIRS recordings can be used to assess listeners' attentive state when performing the target detection task in IM situations. The second fNIRS experiment refined the data processing pipeline and incorporated LMEM to include listener-specific effects into the model. The results showed that an individual's behavior performance was correlated with the fNIRS response measured from the superior temporal gyrus (STG). The third fNIRS experiment investigated our fNIRS protocol's test-retest reliability by using machine learning and determined the minimum source-detector pair and the shortest recording time required to get robust and reliable results.

CHAPTER 2

INFORMATIONAL MASKING VS. CROWDING —A MID-LEVEL TRADE-OFF BETWEEN AUDITORY AND VISUAL PROCESSING[1]

2.1 Abstract

In noisy or cluttered environments, sensory cortical mechanisms help process auditory or visual target sources into perceived objects. Knowing that individuals vary greatly in their abilities to suppress unwanted sensory information, and knowing that the sizes of auditory and visual cortical regions are correlated, we wondered whether there might be a corresponding relation between an individual's ability to suppress auditory vs. visual interference. In *auditory masking*, background sound makes spoken words unrecognizable. When masking arises due to interference at central auditory processing stages, beyond the cochlea, it is called *informational* masking (IM). A strikingly similar phenomenon in vision, called *visual crowding*, occurs when nearby clutter makes a target object unrecognizable, despite being resolved at the retina. We here compare susceptibilities to auditory IM and visual crowding in the same participants. Surprisingly, across participants, we find a negative correlation (R = -0.67) between IM susceptibility and crowding susceptibility: Participants who have low susceptibility to IM tend to have high susceptibility to crowding, and vice versa. This reveals a mid-level trade-off between auditory and visual processing.

2.2 Main

Hearing-impaired individuals commonly do not understand spoken words when background sound interferes. This often occurs because the same neurons in the auditory nerve respond to both the target of interest and the background sound,

[1] submitted for peer review

swamping target information. However, even when background sound energy flanks, rather than overlaps with, a target's spectrum such that cochlear models predict that the target should not be swamped,(Goldsworthy and Greenberg, 2004) many people, including those with "normal" audiological thresholds, have trouble understanding speech. This second difficulty with background sound is often attributed to an incompletely understood central auditory phenomenon called informational masking (IM) (Watson and Pelli, 1983; Kidd and Colburn, 2017). We here ask whether IM relates to a phenomenologically similar effect in vision: crowding (Pelli et al., 2001). In crowding, clutter in the visual scene prevents object recognition. Using comparable crowding and IM tasks, we are surprised to find that crowding susceptibility correlates negatively with IM susceptibility, revealing a mid-level trade-off between auditory and visual processing.

As perceptual phenomena, IM and crowding are strikingly similar. Both are cortical processes by which target-like maskers can prevent target identification (Scott et al., 2004, 2006, 2009; Gutschalk et al., 2008). Both are refractory to learning in neurotypical individuals (Millin et al., 2014; Scott et al., 2009; Neff et al., 1993; Hussain et al., 2012). In crowding, nearby clutter is perceived as part of the visual target (Figure 2.1A). Analogously, in IM, flanking sound, with little energy inside the target's critical band, hampers individuation of target and flankers (Figure 2.1B). Crowding and IM are stronger when target and masker are perceptually similar. Both phenomena occur even when the masker is much fainter than the target, or when the target is presented to one eye/ear and the masker to the other eye/ear (Pelli et al., 2001; Ihlefeld and Shin-Cunnigham, 2008; Flom et al., 1963; Gallun et al., 2007). In vision, clutter crowds a target when the target and clutter are less than a specific distance apart, called the crowding distance. Crowding distance predicts an individual's crowding susceptibility over a broad range of stimuli (Pelli et al., 2001).

Given these functional similarities, we wondered whether IM and crowding rely on similar mechanisms.

We measured behavioral performance in one crowding and two IM tasks in the same 20 normal-hearing young participants. To measure crowding distance, participants identified a peripheral target letter between two flanking letters (Figure 2.1A, see Supplements). To measure IM, we separately tested target individuation with maskers that were either target-like or noise, which is not target-like, in a speech and a non-speech task, as a function of *target-to-masker broadband energy ratio (TMR)*. IM susceptibility is the difference between threshold TMRs in target-like background sound vs noise. In the "speech task," participants identified target words constructed from narrow spectral bands that were masked by either speech (target-like) or noise. In the “melody task”, participants reported whether a target sequence of eight constant frequency tones was present while spectrally flanked by masking sequences of either target-like tones or noise bursts (Figure 2.1B).

We initially ran this experiment as a pilot study with 20 participants. Prompted by the unexpected findings, we re-ran it as the “main” experiment with a new group of 20 participants and minor modifications (see Supplements). Results of the pilot and main experiments are similar (compare blue vs white symbols in Figure 2.1C-F). IM susceptibility in the melody task predicts IM susceptibility in the speech task (pilot: $R = 0.60$, $p = 0.003$; main: $R = 0.76$, $p < 0.001$, Figure 2.1C), suggesting that IM susceptibility generalizes across speech and non-speech stimuli. Moreover, across participants, crowding susceptibility (i.e. crowding distance) and IM susceptibility are correlated, but, unexpectedly, the correlation is negative. This inverse relation between crowding and IM presents for both the speech task (pilot: $R = -0.63$, $p = 0.002$; main: $R = -0.72$, $p < 0.001$) and the melody task (pilot: $R = -0.68$, $p < 0.001$; main: $R = -0.67$, $p < 0.001$, Figure 2.1D). Equivalent rectangular bandwidth (ERB, Figure 2.1E), estimated from noise masking thresholds, does not

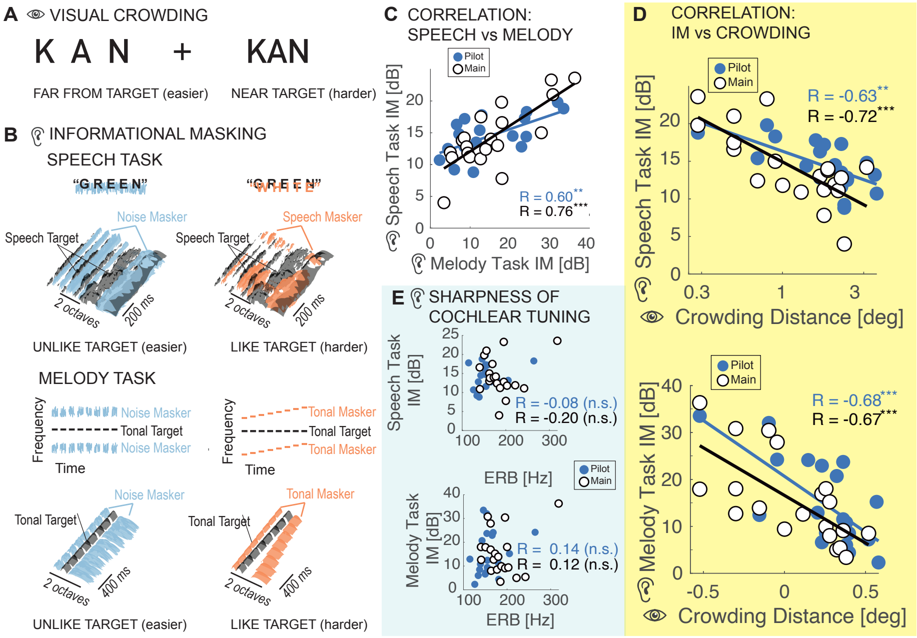

Figure 2.1 In all graphs: Susceptibility increases with increasing IM and crowding distance. Blue and white symbols represent the pilot and main experiments, respectively. (A) To experience crowding, fixate the cross. Your task is to identify the middle letter in each triplet, ignoring the outer letters. When they are closer, the target is harder to identify. The two targets are equidistant from fixation, for equal acuity. (B) Schematics and spectrograms of the sounds we used to study informational masking (IM). The noise masker covered the same spectral range as the speech masker in the speech task or the tonal masker in the melody task, such that masking in the cochlea should be greater or equal in the noise vs. speech or tonal masker conditions; any excess masking in the speech or tonal masker conditions is post-cochlear, i.e., IM. We measure the participant's accuracy in identifying the target sound (denoted in black) as a function of the Target-to-Masker broadband energy ratio (TMR; see Supplements). Target identification is easier when the masker is unlike the target (left spectra). Therefore, threshold TMR for target-like masking is larger than that for noise masking. For each participant, the differences in threshold TMRs measure the individual's IM susceptibility. (C) Participants who are IM-susceptible in the speech task also tend to be IM-susceptible in the melody task. (D) However, IM susceptibility in the speech task is anti-correlated with the participant's crowding susceptibility. Similarly, participants who are less crowding-susceptible tend to be more IM-susceptible in the melody task. (E) Equivalent Rectangular Bandwidths (ERBs) at the target center frequency of 1000 Hz (estimated from noise-masking thresholds in unlike-target conditions of the melody task) confirm that participants had "normal" sharpness of peripheral tuning, with individual ERBs that were smaller than the smallest tested notch width. (Note that a 0.3 octave notch width corresponds to 208 Hz at 1000 Hz. One participant's ERB exceeded 208 Hz, but removing this participant does not affect the conclusions.) (F) ERBs are a poor predictor of IM susceptibility in the speech task. Similarly, ERBs and IM susceptibility in the melody task are uncorrelated.

correlate with IM susceptibility (Speech: pilot: $R = 0.08$, $p = 0.4$; main: $R = 0.20$, $p = 0.6$; Melody: pilot: $R = 0.14$, $p = 0.261$; main: $R = 0.12$, $p = 0.399$, Figure 2.1F), confirming that an individual's sharpness of cochlear tuning does not predict their IM susceptibility (Oxenham et al., 2003).

These results show that IM and crowding, two remarkably similar mid-level sensory processes, are related, but in an inverse fashion. The neural origin of this mid-level trade-off between auditory and visual processing is unclear. Shared mechanisms that induce positive correlations – including generic factors like developmental deprivation , domain-general selective attention, motivation, effort, or vigilance cannot explain this effect. Assuming that crowding is conserved in cortex, the sizes of underlying visual cortical areas should be reciprocally proportional to crowding distance (Pelli, 2008). Crowding distance correlates with size of hV4, but not V1 (Kurzawaski et al., 2021). The correlation between sizes of A1 and hV4 is unknown, but primary visual and auditory cortices tend to covary in size (Song et al., 2011). Visual areas V1, V2, V3, hV4 are in posterior cortex, whereas auditory cortex is more anterior. Patients with posterior cortical atrophy are more susceptible to visual crowding and - surprisingly - less able to perceptually segregate auditory scenes(Hardy et al., 2020), unlike our negative correlation in neurotypical participants. Finally, we wonder if years of prolonged visual or auditory attention might reduce crowding or IM, respectively. People who spend more time looking may listen less, and vice versa. Further work is needed to discover how an individual's ability to recognize a target in clutter develops in each sensory modality.

2.3 Materials and Methods

2.3.1 Participants

A total of 40 participants (ages 19 to 25) took part in the study, 20 per experiment (8 females in the main experiment, and 10 females in the pilot). All participants 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 participants self-reported that they had never learned to play an instrument and never sung in a vocal ensemble. All subjects gave written informed consent to participate in the study. All testing was approved by the Institutional Review Board of the New Jersey Institute of Technology.

2.3.2 Experimental Setup

Throughout testing, participants were seated inside a single-walled sound-attenuating chamber (International Acoustic Company, Inc.) with a quiet background sound level of less than 13 dBA. Acoustic stimuli were generated in Matlab (Release R2016a, The Mathworks, Inc., Natick, MA, USA), D/A converted with a sound card (Emotiva Stealth DC-1; 32 bit resolution, 44.1 kHz sampling frequency, Emotiva Audio Corporation, Franklin, TN, USA) and presented over insert earphones (ER-2, Etymotic Research Company Elk Grove Village, IL, USA). The acoustic setup was calibrated with a 2-cc coupler, 1/2" pressure-field microphone, and a sound level meter (2250-G4, Brüel Kjær, Nærum, Denmark). Visual stimuli were delivered via a 23-inch monitor with 1920 x 1080 resolution. Prior to visual testing, the experimenter positioned the monitor such that it was centered 50 cm away from the center of the participant's nose. Using this setup, three task were administered in an order that was counterbalanced across participants.

2.3.3 Visual Crowding Task

Crowding distance was assessed with a target identification paradigm (Pelli et al., 2016), illustrated in Figure 2.2A. Participants were instructed to fixate their gaze at the center of a cross hair displayed on the monitor in front of them. A target

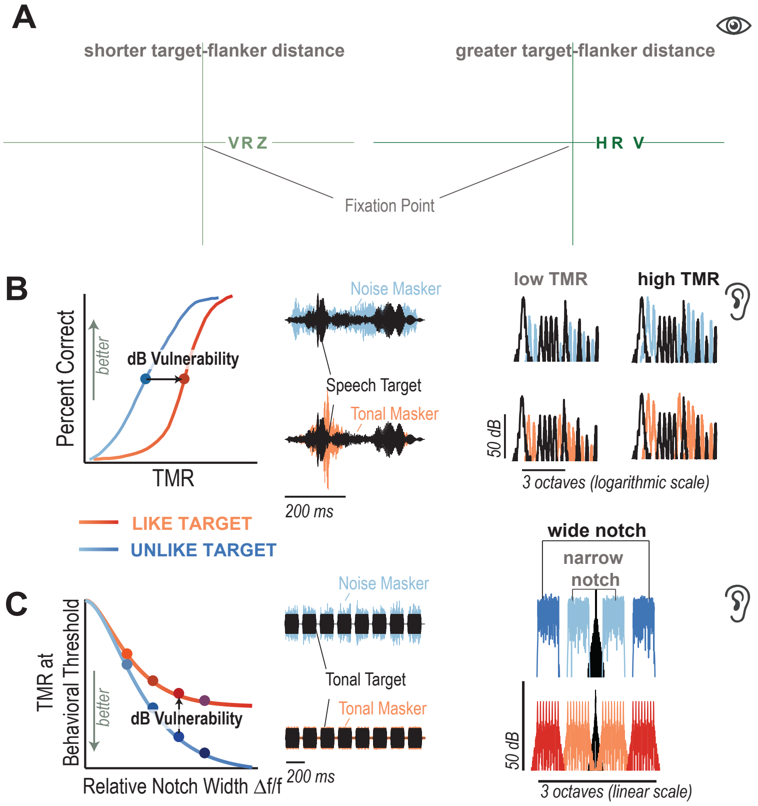

Figure 2.2 (A) In the visual crowding task, participants fixated the cross hair and called out the target letter in the center (here: *R*), while two flankers were cluttering it. Crowding distance was measured by adapatively varying the distance between the target and flankers. In the example here, the target and flanking letters are shown to the right side of the fixation cross hair. In the main experiment, the letters would randomly appear to the left or the right of the cross hair, whereas in the pilot experiment they only appeared to the right. (B) In the speech task, vulnerability to IM was measured by subtracting the TMR where participants correctly identified 50% of target words in the presence of background speech *vs* background noise. (C) Analogously, in the melody task, vulnerability to IM was assessed as the difference in TMRs between target detection with melody maskers *vs* noise maskers, at a notch width of 1 octave.

letter was shown between two flanking letters. The three letters were placed in a row and randomly chosen without replacement from nine possible letters <D, H, K, N, O, R, S, V, Z>. Target and flankers appeared together to either the right or left of the cross hair for 0.5 seconds, then disappeared. Participants were tasked to read out loud the middle letter while continuing to fixate on the cross hair. An experimenter recorded the participant's response on each trial. Using the QUEST method (Watson and Pelli, 1983), the distance between target and flanker letters was adaptively varied in two blocks of 20 trials, aiming for 70% correct and assuming a Weibull psychometric function. Each track started with an initial guess for distance from the center of the middle letter based on neurotypical critical spacing (Song et al., 2014). The crowding distance measured in the first track was a practice run. If estimated crowding distances between first and second track differed by more than 0.2 degrees, a third track of 20 trials was administered. The crowding distance on the final track was recorded as the participant's threshold.

2.3.4 Speech Task

IM vulnerability was measured in a speech task (Figure 2.2B), by subtracting the TMR at 50% correct speech identification threshold with a speech masker from the threshold TMR with a noise masker (Arbogast et al., 2002). Speech identification was assessed using the Coordinate Response Measure matrix task (Bolia et al., 2000). This matrix task uses sentences of the following fixed structure: 'Ready [callsign] go to [color] [number] now.' During testing, a target and a masker sentence were simultaneously presented to the left ear only, constrained to differ from each other in terms of callsign, color and number keywords (Kidd et al., 2008). Target sentences always had the callsign 'Baron.' There were four color keywords <'red','blue','white','green'> and seven possible numbers <'one','two','three','four','five','six','eight'> (excluding the number 'seven' because it has two syllables. Participants were instructed to answer the question: ‘Where did Baron go?’ by pressing the corresponding color and number buttons on a touch screen response interface. A trial was counted as correct if the participant correctly reported the target color and the target number, resulting in a chance performance of 4% ($\frac{1}{4} \cdot \frac{1}{7} = \frac{1}{28} = 0.04$).

To vocode the utterances, raw speech recordings were normalized in root mean square (RMS) value and filtered into 16 sharply tuned adjacent frequency bands using time reversal filtering, resulting in no appreciable phase shift. Each resulting band covered 0.37 mm along the cochlea between 3-dB down-points according to Greenwood’s function (Greenwood, 1990), or approximately 1/10th octave bandwidth, and had a 72 dB/octave frequency roll-off, with center frequencies ranging from 300 to 10,000 Hz. In each narrow speech band, the temporal envelope of that band was then extracted using the Hilbert transform and multiplied by uniformly distributed white noise carriers. To remove the side bands, the resulting amplitude-modulated noises were processed by the same sharply tuned filters that were used in the initial processing stage. Depending on the experimental condition, a subset of these sixteen bands was then added, generating intelligible, spectrally sparse, vocoded speech. Stimuli were generated from utterances by two different male talkers, one for the target and a different talker for the speech masker.

To generate noise maskers that matched the spectrum of the vocoded speech, all processing steps were the same as in the vocoding described above, with one exception. Instead of using the Hilbert envelope to amplitude-modulate the noise carriers, here, the noise carriers were gated on and off with 10 ms cosine-squared ramps that had the same RMS as the Hilbert envelope of the corresponding speech token in that band. A subset of the resulting 16 narrowband pulsed noise sequences was added to generate low-IM noise maskers.

On each trial, nine randomly chosen bands were added to create the target. The masker was comprised of the remaining seven bands and either consisted of vocoded utterances from the same corpus, recorded by a different male talker (target-like) or of noise tokens with similar long-term spectral energy as the vocoded utterances (target-unlike). The center and right panels of Figure 2.2B shows a representative temporal and spectral energy profile for a mixture of target (black) and speech (purple) or noise (brown) maskers.

The masker was presented at a fixed level of 55 dB SPL. The target level varied randomly from trial from 35 to 75 dB SPL, with a 10 dB step size, resulting in five broadband TMRs from -20 dB to 20 dB. For familiarization with the vocoded speech task and to ensure that the vocoded speech stimuli were indeed intelligible, participants were initially tested in 20 trials on nine-band target speech at 35 dB SPL, without masking. Target bands varied randomly from trial to trial. All participants reached at least 90% accuracy during this testing in quiet.

Next, participants were tested in five blocks of 40 trials while target-unlike noise or target-like speech interfered in the background. Thus, each specific combination of the five different TMRs and two masker types was presented 20 times (5 TMRs * 2 masker types *20 trials = 5 blocks *40 trials = 200 trials total). TMR and masker type varied randomly from trial to trial such that all combinations of TMR masker type were presented in random order once before all of them were repeated in a different random order.

To estimate the TMR at the 50% correct threshold for each participant, percent correct scores as a function of TMR were fitted with Weibull-distributed psychometric functions, by using the `psignifit` package (Wichmann and Hill, 2001). IM vulnerability was computed as the difference in TMR at 50% correct between noise *vs* speech masking.

2.3.5 Melody Task

IM vulnerability was also assessed using two-up-one-down adaptive tracking with a non-speech task, by contrasting 70.7% correct thresholds for detecting eight-tone-burst targets across two notched-masker conditions: a noise *vs* a melody masker (Levitt, 1971), illustrated in Figure 2.2C. Target and masker were presented to the left ear only. The target consisted of eight pure tones at a fixed frequency of 1000 Hz. Each tone was 150 ms long (including 10 ms cosine-squared ramps, random phase), with 75 ms gaps between consecutive tones. The target intensity was varied adaptively.

Using a classic paradigm for estimating ERB, in the noise masker condition, two 600-Hz-wide narrow bands of noise were placed symmetrically around the target frequency, creating a symmetrical notch in logarithmic frequency (Patterson, 1976). The notch width was one of the following: <0.3, 0.5, 1, 1.5> octaves. Noise tokens with very steep spectral slopes of over 400 dB/octave were constructed by generating uniformly distributed white noise, transforming it via Fast Fourier Transform and setting the notch frequencies in the spectrum to 0, before transforming the signal via the real portion of the inverse Fast Fourier Transform back into the time domain.

The melody masker condition was designed to closely match the spectral profile of the noise masker. Two eight-tone melodies, each carrying eight possible frequencies that were spaced linearly within 600-Hz-wide bands, flanked the target. One of these melodies was played above, the other below the target frequency, positioned symmetrically around the target frequency along a logarithmic frequency axis. The maskers were chosen from four possible melodies ¡up, down, up-down, down-up¿. Those patterns indicated how the frequency changed for eight pure tones that formed the sequence, for instance 'up' means that each pure tone increased in frequency compared to the previous one in the sequence. The phase of each tone was independently and randomly drawn for each tone, resulting in phases that generally differed across all tones in the target-flanker mixture.

Maskers were played at a fixed spectrum level of 40 dB SPL, equivalent to a broadband level of 68 dB SPL (total level $= 40 + 10 \cdot log_{10}(300) + 10 \cdot log_{10}(2) = 68$). To protect against the possibility of distortion products as a possible task cue in the melody condition, a low-intensity broadband white noise masker was continuously played in the background during both the noise and the melody masker condition at 15 dB SPL.

Under both masker conditions, participants performed a two-alternative forced-choice target detection task, responding with 'yes' or 'no' to indicate whether they heard the target. At the beginning of each adaptive track, the target intensity started at 70 dB SPL. The target intensity was initially decreased by 10 dB for every two consecutive correct answers and increase by 10 dB for every incorrect answer. After every two reversals in the adpative tracks, the step size was halved. Participants completed 12 reversals. Threshold was the average target intensity across the final 12 responses. IM vulnerability was calculated by subtracting thresholds between noise and melody masker at one octave separation.

2.3.6 ERB

To estimate each participant's ERB, noise masked thresholds from the melody task, denoted as W, were minimum-least-square fitted to rounded exponential (roex) functions (command `lsqcurvefit` in Matlab). The roex functions were defined as $W(g) = (1 - r)(1 + pg)e^{-pg} + r$, where p determined the steepness of the roex function's passband, r shaped the stopband, and g denoted the distance between the target frequency f_T and the corner frequencies of the masker notch with $g = \frac{f - f_T}{f_T}$ (Patterson et al., 1982).

2.3.7 Pilot Experiment

The methods used for the pilot experiment were similar to those of the main experiments except for two differences. First, unlike in the main experiments, the tasks in the pilot experiment were not counterbalanced across participants and administered in fixed order instead. Specifically, during pilot testing, participants first completed the crowding task, followed by the IM speech task, followed by the IM melody task. The second difference across the experiments is that in the crowding task during piloting, participants were only tested to the right side of the cross hair. Whereas during the main experiments, target and masker randomly alternated between presented to the left *vs* the right side of the cross hair.

2.3.8 Statistical Analyis

Statistical analyses were performed using linear regression via the command `fitlm` in Matlab 2019b, and adjusted R-squared are reported. Multiple comparisons were adjusted with Bonferroni correction.

2.4 Supplemental Results

2.4.1 Melody Task

As expected, thresholds were much less variable across listeners in the the noise masker condition as compared to the melody condition, in both the main and pilot experiment (compare the spread in the density plots in the top vs bottom panels of Figure 2.3A) . Considering the high across-participant variability in vulnerability to IM, we next used bootstrapping to estimate the confidence intervals the correlation between crowding distance and vulnerability to non-speech IM, as a function of notch width (Figure 2.3B). Adjusted correlation coefficients roughly increased with increasing notch width and were most consistent across the main *vs* pilot experiments at the 1-octave notch width.

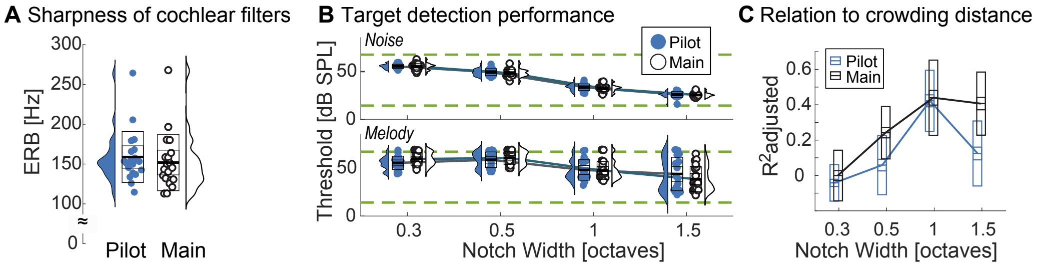

Figure 2.3 (A) Individual variability is much higher in target-like masking than noise across all tested ROIs in experiment 1. For all participants, target detection thresholds generally fall between the broadband level of the notched masker (68 dB SPL) and broadband masker (15 dB SPL), shown by green dashed lines. In the noise masker configuration, target detection thresholds decrease with increasing notch width for all participants (top), but only for some participants in the melody masker configuration (bottom). Note that in both the main and the pilot experiment, the densities in the melody task are bimodal, with one mode close to 68 dB SPL throughout, and the other mode decreasing with increasing notch width. (B) The adjusted correlation coefficient between visual crowding distance and IM vulnerability reveals a coarse tuning curve. Results between the two experiments are congruent at 0.3 and 1 octave notch widths, but appear to diverge at 0.5 and 1.5 octave notch widths. Test-retest variability of R^2, estimated via bootstrapping that sampled 10 out of 20 participants without replacement 100 times, show that, indeed, crowding distance is robustly correlated with IM vulnerability at 1 octave separation, in both experiments. At other notch widths, the relationship is less pronounced.

Visual inspection of the density functions in Figure 2.3A hints that the distribution of melody masking thresholds was bimodal, gradually widening with increasing octave separation. The mean of the lower mode, corresponding to participants who were more resilient to masking, decreased with increasing notch width. The mean of the other mode remained roughly constant as a function of notch width and close to the broadband level of the masker, indicating that the more poorly performing participants chose a strategy to listen for the louder source as opposed to relying on target pitch. In these poorly performing listeners, thresholds did not monotonically improve with increasing notch width. Perhaps as a result, roex functions used to estimate ERB under noise masking did not provide appropriate fits of the data under melody masking.

While we did not originally anticipate this result, in general, approximately a third of normal-hearing listeners have difficulty discerning pitch, and can, for instance, not reliably distinguish between major and minor triads in musical chords, even when given trial-by-trial correct response feedback (Chubb et al., 2013; Mednicoff et al., 2018; Graves and Oxenham, 2019). Note that we here tested the IM melody task at 0.3, 0.5, 1 and 1.5 octave notch width, resulting in center frequencies of the lower and upper flanker bands that were related by factors of 1.231, 1.414, 2.000 and 2.828. Those numbers were originally chosen to cover the range of notch widths that typically result in ERB estimates (Patterson, 1976). However, they had the unintended effect that the constituent flanker frequencies were not perfectly harmonically related, and therefore potentially unfused at 0.3, 0.5 and 1.5 octaves, whereas flanker frequencies were harmonically related at the 1 octave notch. In summary, IM vulnerability in the melody task at these other three notch widths is more weakly correlated or even uncorrelated with crowding distance (as well as IM vulnerability to speech), showing that harmonicity affects IM in this paradigm.

CHAPTER 3

SPATIAL RELEASE FROM INFORMATIONAL MASKING: EVIDENCE FROM FUNCTIONAL NEAR INFRARED SPECTROSCOPY[2]

3.1 Abstract

Informational masking (IM) can greatly reduce speech intelligibility, but the neural mechanisms underlying IM are not under-stood. 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.

[2] Min Zhang, Yu-Lan Mary Ying, and Antje Ihlefeld (2018). "Spatial Release From Informational Masking: Evidence From Functional Near Infrared Spectroscopy" Trends in Hearing Volume 22: 1–12

3.2 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, and Helfer 2001; Jones and Litovsky, 2011; Mattys, Brooks, and Cooke, 2009). EM occurs when sound sources have energy at the same time and frequency (e.g., Brungart, Chang, Simpson, and 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 and Colburn, 2017). IM is thought to be a major factor limiting performance of hearing aid and cochlear implant devices (Marrone, Mason, and Kidd, 2008; Shinn-Cunningham and Best, 2008; Xia, Kalluri,Micheyl, and Hafter, 2017). However, the neural mechan-isms 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 (Goldsworthy and Greenberg, 2004). Consistent with this interpretation, sub-cortical responses are modulated by how well a listener processes speech in EM noise (Anderson and Kraus,2010). However, peripheral models fail to account for speech intelligibility in IM-dominated tasks (e.g., Cooke, Garcia Lecumberri, and 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, and Carlyon, 2004; Ihlefeld and Shinn-Cunningham, 2008a; Jones, Alford, Bridges, Tremblay, and Macken, 1999). Previous work suggests that common onsets and harmonicity determine how short-term segments form (Darwin and Hukin, 1998; Micheyl, Hunter, and 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 and Simpson, 2002; Darwin, Brungart, and Simpson, 2003; Darwin and 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, and 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, and Somers, 2015; Noyce, Cestero, Michalka, Shinn-Cunningham, and 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 co-located, a phenomenon referred to as spatial release from masking (e.g., Carhart, Tillman, and Johnson, 1967; Darwin and Hukin, 1997; Glyde, Buchholz, Dillon, Cameron, and Hickson, 2013; Kidd, Mason, Best, and 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 co-located configuration of the same stimuli. We reasoned that a stronger HbO response in the spatially separated versus co-located configurations would support the view that spatial attention under IM activates LFCx. In contrast, a stronger LFCx response in the co-located configuration would suggest that LFCx does not encode the direction of spatial auditory attention.

Using the setup shown in Figure 3.1A, we recorded hemodynamic responses near cIFS and STG bilaterally, from normal-hearing young individuals. Listeners were instructed to detect when the target voice on the left uttered color keywords while SPEECH vs NOISE maskers interfered from the right side (Figure 3.1B). Behavioral pilot testing confirmed that these spectrally sparse maskers produced high-IM (SPEECH) vs low-IM (NOISE).

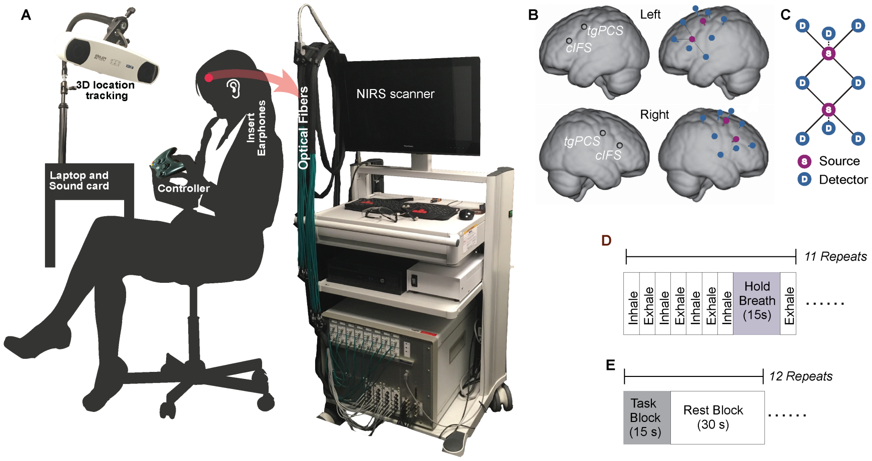

Figure 3.1 (A) Experimental apparatus and setup. (B) ROIs and optode placement for a representative listener. Blue circles show placements of detector optodes and red circles of source optodes. (C) fNIRS optical probes design with deep neurovascular (solid line) and shallow nuisance (dotted line) channels.(D) Block design, controlled breathing task, and (E) Block design, auditory task. S = source; D = detector.

3.3 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.

3.4 Methods

3.4.1 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 3.1(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, and Tournoux, 1988), these four landmark coordinates were then used as reference for locating the four regions of interest (ROIs, locations illustrated in Figure 3.1(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, 1/200 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 ofoptode-detector pairs was interpreted as response chan- nels 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 3.1(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 Figure 2.1(c)), whereas shallow channels, used to estimate physiological noise, had a source-detector distance of 1.5 cm (dotted line in Figure 2.1(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.

3.4.2 Controlled Breathing Task

Variability in skull thickness, skin pigmentation, and other idiosyncratic factors can adversely affect recording quality with fNIRS (Bickler, Feiner, and 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 2.1(d). This nonauditory task consisted of 11 blocks of controlled breathing (Thomason, Foland, and 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).

3.4.3 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 3.3 shows a schematic of the paradigm (a) and spectrograms for two representative stimuli (b).

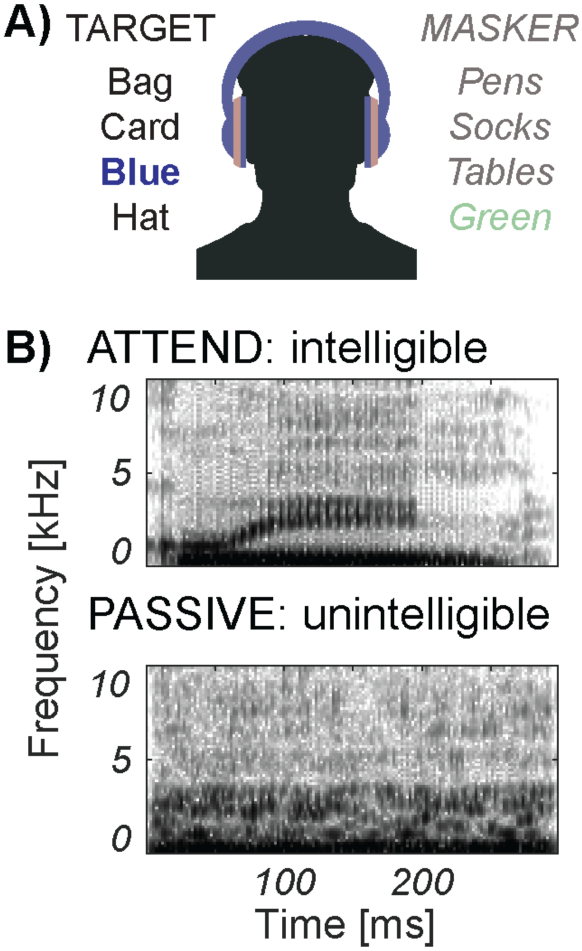

Figure 3.2 (A) Speech paradigm. (B) Spectrograms of the word green. Unprocessed speech in the ATTEND condition (top) and scrambled speech in the PASSIVE condition (bottom).

In Experiment 1, the target stream was always presented with a left-leading interaural time difference (ITD) of 500 ms, while the concurrent masker stream was presented with a right-leading ITD of 500 ms (spatially separated configuration). In Experiment 2, we also tested a spatially colocated configuration, where both the target and the masker had 0 ms 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 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, and 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 and 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 and 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 ms 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.

3.4.4 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 and Dumont, 2012).

We applied the modified Beer–Lambert law (Cope and Delpy, 1988; Kocsis, Herman, and 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.

3.4.5 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:

$$y(t) = x_{task1}(t)\beta_1 + x_{task2}(t)\beta_2 + x_{nuisance}(t)\beta_3 + \varepsilon(t)$$

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, $x_{taski}(t)$ was the regressor of the hemodynamic change attributed to behavioral task i. $x_{nuisance}(t)$ was the HbO concentration in the shallow channel (Brigadoi and Cooper, 2015), and $\varepsilon(t)$ was the residual error of the GLM.

The task regressors xtask 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, and Wager, 2009):

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

where 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 et al, 2017). During pilottesting, 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.

3.4.6 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.

3.5 Results

3.5.1 Controlled Breathing Task

Figure 3.3 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 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).

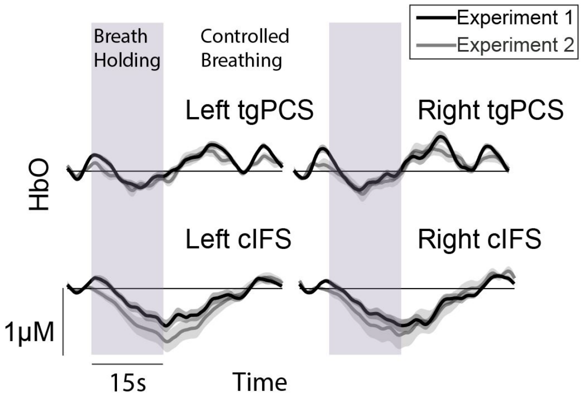

Figure 3.3 HbO concentration change during controlled breathing in Experiments 1 and 2. HbO = oxy-hemoglobin; tgPCS = transverse gyrus intersecting precentral sulcus; cIFS = caudal inferior frontal sulcus.

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.

3.5.2 Experiment 1

Figure 3.4(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 3.4(B) shows HbO activation levels b, 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.

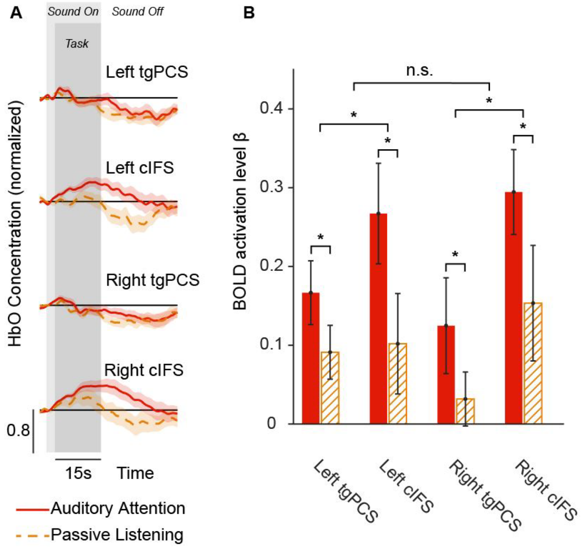

Figure 3.4 Results from Experiment 1. (A) Normalized HbO traces during the direction of auditory attention versus passive listening, at each of the four ROIs in Experiment 1. The ribbons around each trace show one standard error of the mean across listeners. (B) Normalized HbO traces during pitch and 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.

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.

3.5.3 Experiment 2

Figure 3.5(A) and (B) display the HbO traces (red lines denote spatially separated, blue lines co-located configurations) and the across-listener average in HbO activation

β-levels for the spatially separated (red fill) versus co-located configurations (blue fill), at each of the four ROIs; 14 listeners reached criterion performance during behavioral testing and were included in the group analysis.

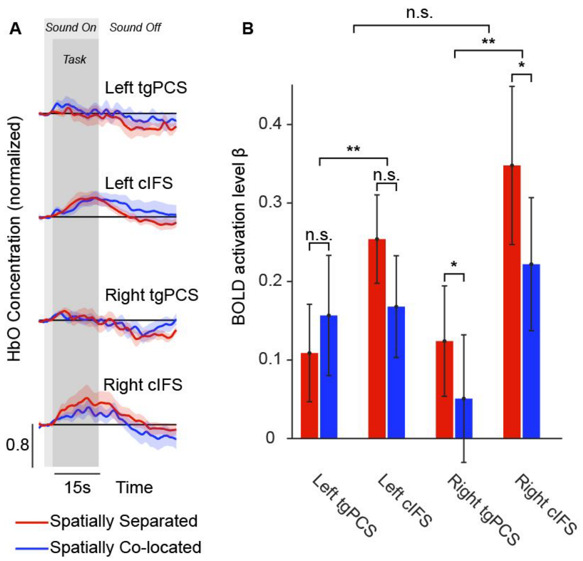

Figure 3.5 Results from experiment 2, formatting similar to Figure 2.4. HbO = oxy-hemoglobin; tgPCS = transverse gyrus intersecting precentral sulcus; cIFS = caudal inferior frontal sulcus.

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 co-located configuration. No difference between spatial configurations was discovered in the HbO concentration changes in the left hemisphere.

3.6 Discussion

3.6.1 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.

3.6.2 Right LFCx Activation Associated With SRM

We wished to disentangle the role of spatial attention on the LFCx HbO response. In Experiment 1, spatial differ- ences 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 and 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 co-located 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 co-located 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, and McAlpine, 2016; von Kriegstein, Griffiths, Thompson, and McAlpine, 2008).

In general, spatial release from masking is thought to arise from three different mechanisms (e.g., Shinn-Cunningham, Ihlefeld, Satyavarta, and 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 and Picton, 2006; Wack et al., 2012;Wong and Stapells, 2004) and spatial attention, assumed to arise at cortical processing levels (Ahveninen et al., 2006; Larson and Lee, 2014; Shomstein and Yantis, 2006; Wu, Weissman, Roberts, and Woldorff, 2007; Zatorre, Mondor, and 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, and 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, and Hartley, 2016a; Wijayasiri, Hartley, and 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, and 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, and Shinn-Cunningham, 2013; Mesgarani and Chang, 2012; Mirkovic, Debener, Jaeger, and 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, and 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.

3.6.3 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 and Hartley, 2015; Lawler, Wiggins, Dewey, and Hartley, 2015; McKay et al., 2016; 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, and 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, and 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.

3.7 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 co-located configuration, suggesting that right LFCx activity is associated with spatially directed attention.

CHAPTER 4

HEMODYNAMIC RESPONSES LINK INDIVIDUAL DIFFERENCES IN INFORMATIONAL MASKING TO THE VICINITY OF SUPERIOR TEMPORAL GYRUS[3]

4.1 Abstract

Suppressing unwanted background sound is crucial for aural communication. Public spaces often contain a particularly disruptive background sound, called informational masking (IM). At present, IM is identified operationally: when a target should be audible, based on suprathreshold target/masker energy ratios, yet cannot be heard because perceptually similar background sound interferes. Here, behavioral experiments combined with functional near infrared spectroscopy identify brain regions that predict individual vulnerability to IM. Results show that task-evoked blood oxygenation changes near the superior temporal gyrus (STG) and behavioral speech detection performance covary for same-ear IM background sound, suggesting that the STG is part of an IM-dependent network. Moreover, listeners who are more vulnerable to IM show an increased metabolic need for oxygen near STG. In contrast, task-evoked responses in a region of lateral frontal cortex, the caudal inferior frontal sulcus (cIFS), do not predict behavioral sensitivity, suggesting that the cIFS belongs to an IM-independent network.

4.2 Introduction

Perceptual interference from background sound, also called auditory masking, has long been known to impair the recognition of aurally presented speech through a

[3] Min Zhang, Nima Alamatsaz, Antje Ihlefeld (2020). "Hemodynamic responses link individual differences in informational masking to the vicinity of superior temporal gyrus" bioRxiv preprint doi:https://doi.org/10.1101/2020.08.21.261222

combination of at least two mechanisms. Energetic masking (EM) occurs when target and masker have energy at the same time and frequency, such that the masker swamps or suppresses the auditory nerve activity evoked by the target (Young and Barta, 1986; Delgutte, 1990). Informational masking (IM) is presently defined operationally. IM occurs when a target is expected to be audible based on EM mechanisms, yet cannot be detected or identified. Listeners experience IM when target and masker are perceptually similar to each other (e.g., hearing two women talk at the same time vs hearing out a female in the background of a male voice; Brungart (2001b)) or when the listener is uncertain about perceptual features of the target or masker (e.g., trying to hear out a target with known vs unexpected temporal patterning, cf. Luti et al. (2013)).

Unlike EM, IM is associated with striking variation in individual vulnerability (Neff and Dethlefs, 1995; Durlach et al., 2003). Moreover, an individual's susceptibility to IM is largely refractory to training (Neff et al., 1993; Oxenham et al., 2003). Identifying brain regions where IM-evoked activation patterns covary with individual differences in behavioral vulnerability to IM may thus hold a key for defining the neural mechanisms underlying IM.

Neuroimaging studies have greatly advanced our understanding of the neural mechanisms of masking. Converging evidence links both EM and IM to recruitment of superior temporal gyrus (STG) and frontal cortex (Davis and Johnsrude, 2003, 2007; Scott et al., 2004, 2006, 2009; Mesgarani and Chang, 2012; Lee et al., 2013; Michalka et al., 2015). For instance, the predominantly activated STG hemisphere can shift depending on the amount of IM in the background sound (Scott et al., 2009). Moreover, for speech that was either spectrally degraded or had impoverished amplitude cues, spanning the range from unintelligible to fully intelligible, activation near STG can account for approximately 40 to 50% of the variance in speech intelligibility (Pollonini et al., 2014; Lawrence et al., 2018).

In addition, lateral frontal cortex engages more strongly with increasing listening effort or increasing recruitment of higher-order semantic processes (Davis and Johnsrude, 2003; Scott et al., 2004; Wild et al., 2012; Wijayasiri et al., 2017). Parts of lateral frontal cortex, including the caudal inferior frontal sulcus (cIFS), are also sensitive to auditory short-term memory load in situations with IM (Michalka et al., 2015; Noyce et al., 2017). Using functional near-infrared spectroscopy (fNIRS), we previously confirmed that the cIFS region engages more strongly when listeners actively attend to speech in IM vs listen passively (Zhou et al., 2018b), making the STG and cIFS promising region of interest (ROIs) for the current study.

Widening an established IM paradigm (Arbogast et al., 2002), we here compare hemodynamic responses to low vs high IM speech. We test two hypotheses. H1: Individual differences in vulnerability to IM are mediated through processing limitations in the vicinity of STG. H2: Individual differences in vulnerability to IM arise near cIFS.

To study the effect of cortical responses on individual differences in behavioral speech comprehension, our goal is to differentiate between brain areas with IM independence (task-evoked responses do not predict vulnerability to IM) vs areas with IM dependence (task-evoked responses predict IM vulnerability). Using fNIRS, we simultaneously quantify behavioral sensitivity and hemodynamic responses in the vicinity of STG and cIFS. In experiment 1, we contrast hemodynamic responses to speech detection in presence of combined low-IM vs high-IM with same-ear masking. To control for EM, in experiment 2, we contrast high-IM with same-ear vs opposite-ear masking. The two experiments serve as their own control, confirming test-retest reliability of the measured cortical traces. Our results support H1 but not H2.

4.3 Results

4.3.1 Experiment 1

Using the setup shown in Figure 4.1A, we recorded hemodynamic responses near cIFS and STG bilaterally, from normal-hearing young individuals. Listeners were instructed to detect when the target voice on the left uttered color keywords while SPEECH vs NOISE maskers interfered from the right side (Figure 4.1B). Behavioral pilot testing confirmed that these spectrally sparse maskers produced high-IM (SPEECH) vs low-IM (NOISE, Supplements 1).

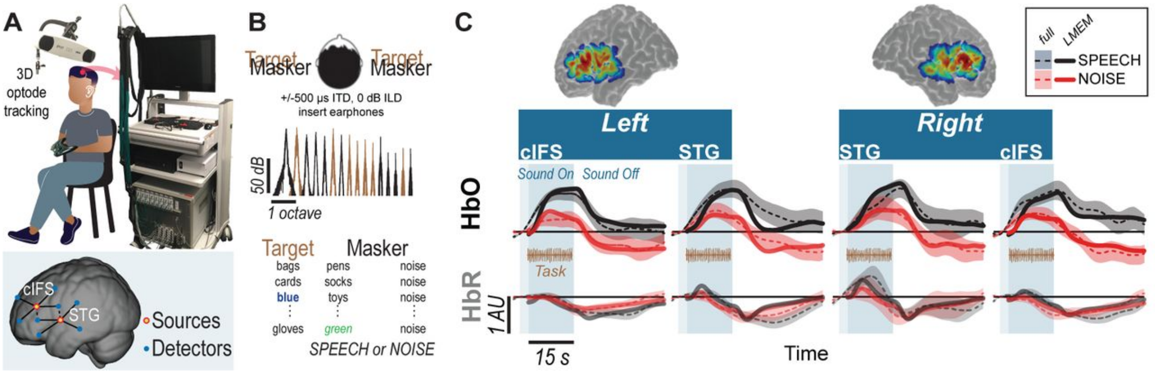

Figure 4.1 High-IM elicits stronger task-evoked responses than low-IM across all tested ROIs in experiment 1. (A) Experimental apparatus and setup and optode placement for a representative listener. Blue circles show placements of detector optodes, red circles of source optodes (deep channels: solid lines; reference channels: dashed lines). (B) Task design for SPEECH vs NOISE. Both target (left-leading ITD of 500 μs) and masker (right-leading ITD of 500 μs) were presented binaurally. Spectral density for target vs masker show mutually flanking, sharply tuned component bands. (C) Sensitivity map. Warmer colors denote increased likelihood that photons will be recorded from these areas. (D) HbO (top) and HbR (bottom) traces. Full hemodynamic responses are denoted by solid lines and error ribbons. Here and elsewhere, ribbons show one standard error of the mean across listeners. The task-evoked hemodynamic responses predicted from the LMEM are shown as dashed lines. Shaded areas mark the task duration.

Accounting for approximately half of the variance in the recorded data ($R^2 = 0.45$), a Linear Mixed Effects Model (LMEM) was then used to predict task-evoked hemodynamic responses, by regressing out reference channels (β_6 and β_7), block number (β_5), and pure tone average (PTA, $\beta_1 1$ and $\beta_1 2$) from the full response

(Supplement 2). Note that the reference channels comprise 44.6% of the total activation levels in the LMEM fits, as calculated via the area under the fitted curve with vs without 6 and 7. Indeed, unlike the full hemodynamic response, the LMEM-estimated task-evoked hemodynamic response aligns well with the task-onset (compare onset of darker shaded area and dashed line throughout 1C).

LMEM fits reveal significant task-evoked responses at all four ROIs (Table 4.1; $\beta_{14} > 0$, p < 0.0001; see Figure 4.1C for HbO (top row) and HbR traces (bottom row). Moreover, all ROIs were sensitive to IM. Activation was stronger in the SPEECH as compared to the NOISE configuration ($\beta_{10} > 0$). The size of the difference between SPEECH (black lines in Figure 4.1C) vs NOISE (red lines) activation varied across ROIs, but these interactions with ROI were small compared to the overall effect size(interaction between masker configuration and cortical structure: $\beta_{13} < 0$; interaction between masker configuration and hemisphere: β_{14} <0; see Supplement 3).

4.3.2 Experiment 2

The sharply tuned, mutually flanking bands of target and masker in experiment 1 were presented to both ears, and were designed to produce high-vs low IM, with little EM. However, IM can also occur when target and masker are presented to opposite ears. It is unclear whether the neural mechanisms underlying IM are similar when target and masker are presented to the same vs opposite ears. Thus, we next wished to examine whether the pattern of STG and cIFS recruitment would generalize to a dichotic IM configuration.

Testing a new group of 14 listeners, experiment 2 contrasted SPEECH with SPEECH-oppo, a stimulus configuration that was identical to SPEECH, except that target and masker were now presented to opposite ears (Figure 4.2). Mirroring results from experiment 1, an LMEM fitting all HbO and HbR traces from experiment 2

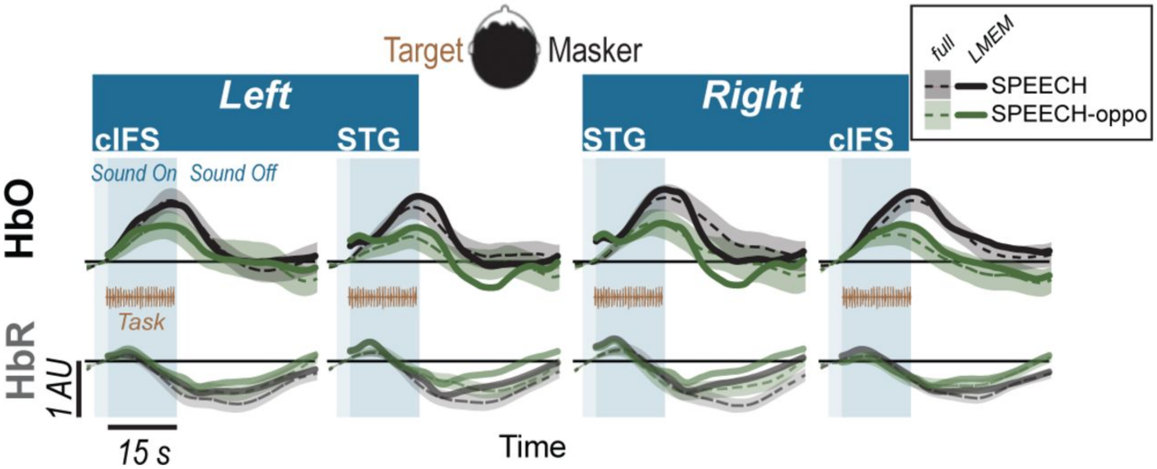

Figure 4.2 Hemodynamic responses for SPEECH (Black) vs SPEECH-oppo (green) show robust task-evoked recruitment of all ROIs in experiment 2, even when target and masker are presented to opposite ears. Solid lines and error ribbons denote raw recordings; dashed lines show LMEM fits.

accounted for approximately half of the variance in the recorded data ($R^2 = 0.52$), with 60.2% of the full hemodynamic activation attributed to reference channels. Moreover, LMEM fits confirmed that task-evoked responses in all four ROIs occurred in both masker configurations, even when target and masker were presented to opposite ears (Table 4.2; $\beta_{14} > 0$, p $<$ 0.0001). All ROIs engaged more strongly in the SPEECH as compared to the SPEECH-oppo configuration ($\beta_{10} > 0$), with effect size depending somewhat on ROI (see Supplement 3).

4.3.3 Vulnerability to masking and hemodynamic responses

To test the core hypotheses, we next examined STG and cIFS for IM-dependence. We reasoned that in an IM-dependent ROI, the hemodynamic activation strength should predict behavioral sensitivity.

For each ROI, planned adjusted coefficients of determination, R2, between behavioral speech detection sensitivity and the peak of the HbO response were calculated. In experiment 1, individual behavioral thresholds were significantly anti-correlated with peak HbO only in the SPEECH configuration in the vicinity of left or right STG, where hemodynamic responses explained 23% (left STG) and 31% (right STG) of the behavioral variance (black square symbols in Figure 4.3A). In contrast, behavioral NOISE thresholds were uncorrelated with hemodynamic responses (Figure 4.3B). Note that this pattern was observed despite the fact that the behavioral speech detection performance, measured during the fNIRS recordings, was comparable between NOISE and SPEECH [paired t-test: t(13) = 1.14, p = 0.27]. Furthermore, activity levels near cIFS (Figure 4.3C) were not correlated with behavioral thresholds in SPEECH or NOISE.

Testing a different group of listeners, experiment 2 confirmed the finding from experiment 1 that HbO peaks near left or right STG were significantly anti-correlated with behavioral sensitivity for the SPEECH configuration. Moreover, activity levels in cIFS were again uncorrelated with behavioral thresholds. Identical SPEECH configurations were assessed in experiments 1 and 2. Therefore, the converging results across two groups of listeners confirm high test-retest reliability of the current fNIRS approach. Specifically, in experiment 2, STG HbO peak activation explained 43% and 34% of the behavioral variance in left and right STG respectively (blue square symbols in Figure 4.3A). In contrast, hemodynamic responses for SPEECH-oppo did not predict behavioral sensitivity (Figure 4.3C).

A caveat, unlike in experiment 1, in experiment 2, task difficulty differed across masking conditions. Specifically, behavioral speech detection thresholds were better for SPEECH-oppo than SPEECH [paired t-test: t(13) = 3.13, p = 0.008; compare green symbols in Figure 4.3C falling to the right of the red, blue and black symbols in Figure 4.3A,B]. However, even for the more poorly performing listeners in experiment 2, no obvious trend links behavioral sensitivity to peak HbO levels in left or right STG.

Of note, behavioral responses were not predicted from HbR activity levels, across any of the tested conditions, in either of the two experiments. As expected, task-evoked HbO and HbR responses were robustly anti-correlated (in Figure 4.1C,

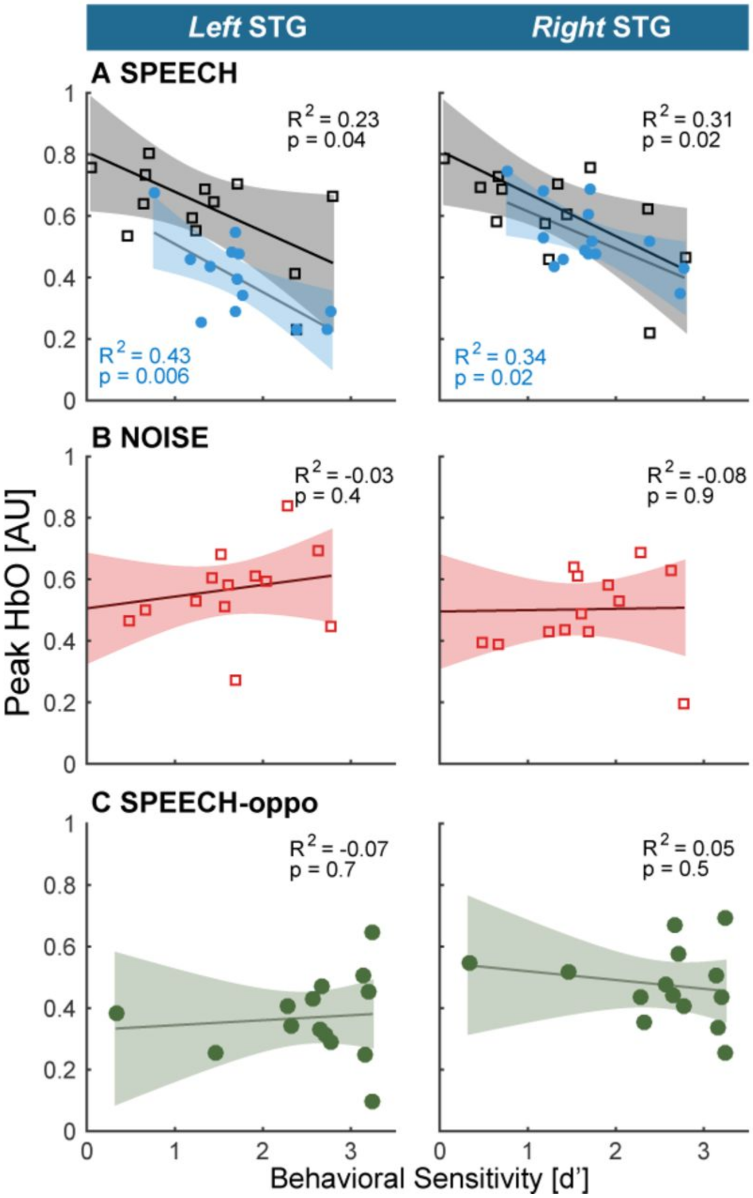

Figure 4.3 Hemodynamic responses link individual differences in vulnerability towards IM to the vicinity of STG (A) STG activity and behavioral vulnerability to the high-IM SPEECH condition are robustly anti-correlated, across both hemispheres in experiments 1 and 2 (black vs blue symbols, respectively). (B) There was no appreciable association between HbO peaks and the low-IM NOISE condition. (C) When target and masker were presented to opposite ears in the SPEECH-oppo configuration, HbO peaks did not predict psychophysical thresholds.

and 3.2, compare dark dashed lines in the top row to the lighter dashed lines of the same color in the bottom row) This anti-correlation would predict that HbR responses mirror the correlation patterns between HbO peaks and behavioral sensitivity. However, in general, HbR response magnitudes were very small, approximately 20% of HbO magnitudes, hinting that here, the HbR responses may have been contaminated by the noise floor of the recording system.

4.4 Discussion

The goal of the current work was to identify brain regions where individual differences in IM vulnerability emerge. To that end, we sought to differentiate between IM-independent parts of the brain whose activation levels are equivalently driven by low- or high-IM, vs IM-dependent regions whose activation levels correlate with individual IM-vulnerability.

4.4.1 Hemodynamic correlates of IM

The current data confirm that cortical regions at or near STG and cIFS engage during masked speech comprehension tasks (Scott et al., 2004, 2006, 2009; Rowland et al., 2018; Kerlin et al., 2010; Mesgarani and Chang, 2012; Ding and Simon, 2012; Michalka et al., 2015; Noyce et al., 2017). Robust task-evoked hemodynamic responses in STG and cIFS occurred, in both brain hemispheres, when the listener was engaged in a speech detection task, in either high- or low-IM. Task-evoked bilateral responses in STG and cIFS were even observed when target and high-IM masker were presented to opposite ears (SPEECH-oppo in experiment 2).

SPEECH masking recruited a stronger task-evoked response than NOISE masking in both left and right STG, consistent with prior work (Scott et al., 2004). Activation levels during SPEECH masking consistently predicted 20-43% of individual differences in vulnerability in left or right STG, in both experiments. Moreover, STG recruitment did not predict vulnerability to masking for the low-IM masker (NOISE condition in experiment 1). Together, these results show that recruitment in the vicinity of STG was IM-dependent. In contrast, while cIFS also showed task-evoked responses that were stronger in SPEECH than in NOISE, cIFS activation strength did not significantly correlate with individual vulnerability in any tested masking configuration, suggesting that the vicinity of cIFS was IM-independent.

IM is thought to be a central auditory mechanism. However, IM generally interferes much more strongly when target and masker are presented to the same ear(s), as compared to being presented to opposite ears (Brungart and Simpson, 2002, 2007; Kidd Jr et al., 2003; Gallun et al., 2005; Wightman and Kistler, 2005). Indeed, prior behavioral evidence suggests that interference from a nontarget ear can be attributed to a combination of a failure to attend to the target ear as well as increased listening effort (Gallun et al., 2007). However, it is unclear whether these mechanisms are similar for same-ear vs opposite ear IM.

Here, SPEECH-oppo evoked bilateral responses in STG and cIFS. If identical STG-based networks were activated for same-ear-IM (SPEECH) and opposite-ear-IM (SPEECH-oppo), STG activity should have been a negative predictor of behavioral SPEECH-oppo sensitivity, but this was not observed. A caveat, speech identification thresholds in SPEECH-oppo were close to ceiling for a few of the listeners. However, even for poorly performing listeners, no trend emerged linking the peak HbO response and behavioral sensitivity (Figure 4.3C). Moreover, the interpretation that contralateral IM recruits different brain networks than ipsilateral IM is also supported by prior evidence from research in children, where the ability to suppress a masker ipsilateral to the target matures more slowly than the ability to suppress a masker on the contralateral side (Wightman et al., 2010).

For same-ear IM, listeners reached comparable speech detection thresholds in low-IM and high-IM, but had marked individual difference during IM speech identification during behavioral pilot testing. This observation is consistent with the idea that more IM-vulnerable listeners exerted more listening effort (Pichora-Fuller et al., 2016). A cortical marker for listening effort was previously located in lateral inferior frontal gyrus, a brain area which shows attention-dependent increase in frontal brain activation during listening to degraded speech (Wild et al., 2012; Wijayasiri et al., 2017). The current study did not target the lateral inferior frontal gyrus, nor did we record alternative measures of listening effort, such as pupilometry (Zekveld and Kramer, 2014; Parthasarathy et al., 2020), precluding any direct test of this possibility.

Together, the results show that even with comparable behavioral sensitivities and similar long-term acoustic energy, high-IM in the same ear increased HbO peaks near STG and cIFS, as compared to low-IM. This effect was observed separately for same-ear as well as opposite-ear IM. Moreover, the observed anti-correlation between HbO peak levels and individual task performance in same-ear high-IM is consistent with the interpretation that left and right STG are part of a same-ear-IM-dependent network. In contrast, the vicinity of cIFS engaged in an IM-independent manner.

4.4.2 Emergence of IM

Listeners with higher cognitive abilities comprehend masked speech better (Mattys et al., 2012; Rönnberg et al., 2008), but prior work shows no evidence that cognitive ability contributes differently to IM vs EM. For instance, cognitive scores poorly predict how well an individual can utilize an auditory scene analysis cue to suppress IM (Füllgrabe et al., 2015). Consistent with this, here, task-evoked responses near cIFS were IM-independent, unlike in the vicinity of STG.

Inded, prior work hints that IM emerges at the level of auditory cortex, a part of the STG (Gutschalk et al., 2008). We here tested maskers that were spectrally interleaved with the target, designed to produce either high IM (SPEECH) or low IM

(NOISE). EM, when present, was limited to spectral regions outside the frequency bands that comprised most of the target energy. Consistent with this, for speech detection, behavioral thresholds were comparable between SPEECH and NOISE. However, our behavioral pilot results also confirmed that speech identiication was much more difficult in the presence of SPEECH than NOISE (Freyman et al., 1999; Arbogast et al., 2002; Brungart et al., 2006; Wightman et al., 2010).

This behavioral pattern parallels a behavioral phenomenon in vision, called Crowding. In Crowding, the presence of visual flankers decreases target identification performance, even when target and maskers are processed with peripheral (as opposed to foveal) vision (Strasburger et al., 1991). Moreover, analogous to the current behavioral results, flanking maskers that cause Crowding in target identification tasks do not typically impair target detection (Pelli et al., 2001). Furthermore, using a behavioral paradigm that is comparable to the current speech identification task, prior work shows that IM can occur even when the masker is softer than the target (Brungart, 2001a; Ihlefeld and Shinn-Cunningham, 2008). Analogously, Crowding can occur even when the flankers are smaller in size than the target (Pelli et al., 2001). Of importance to the current work, Crowding is currently thought to emerge at cortical processing levels (Millin et al., 2014; Zhou et al., 2018a). Together, the apparent congruence in stimulus design and behavioral outcomes raises the possibility that analogous canonical principles of sensory processing may underlie IM and Crowding, further supporting the prior notion that IM arises at the level of cortex.

4.4.3 Cortical Mechanisms of IM

The current results show that for similar behavioral sensitivities and similar long-term acoustic energy, individual differences in vulnerability to high-IM in the same ear correlated with increased need for supply of oxygen in the vicinity of STG and cIFS, as compared to low-IM. This is consistent with the idea that the metabolic needs of an individual's STG contribute to one's ability to filter out unwanted IM. Indeed, recent cortical recordings in humans demonstrate that neural tuning properties of the STG flexibly shift in gain, temporal sensitivy and spectrotemporal tuning, depending on the stimulus (Keshishian et al., 2020). This raises the the possibility that an individual's metabolic need for adapting the neural code in STG plays a role in shaping vulnerability to informational masking. Converging evidence also shows that the temporal fidelity by which cortical responses encode sound is a strong predictor of masked speech intelligibility. Even listeners with audiologically normally hearing can vary dramatically in their ability to resolve and utilize temporal fine structure cues (Ruggles et al., 2011; Bharadwaj et al., 2019). In addition, an individual's sensitivity to monaural or binaural temporal fine structure predicts masked speech intelligibility, especially in temporally fluctuating background sound (Lorenzi et al., 2006; Papesh et al., 2017). Intriguingly, this mechanism is thought to be of subcortical origin (Parthasarathy et al., 2020), hinting that temporal coding fidelity does not differentially affect listening in EM vs IM backgrounds. However, future work is needed to explore how metabolic need and the fidelity of cortical temporal coding interact.

4.4.4 Spatial Specificity

The spacing of fNIRS optodes determines both the depth of the brain where recorded traces originate, as well as their spatial resolution along the surface of the skull. Here, optode sources and detectors were spaced 3 cm apart and arranged cross-wise around the center of each ROI (Figure 4.1A). To estimate the hemodynamic activity in each ROI, we averaged across the four channels of each ROI. This averaging greatly improved test-retest reliability of each ROI's activation trace during pilot testing, both here and in our prior work (Zhang et al., 2018). A caveat of this approach is that it reduces the spatial resolution of the recordings. Thus, it is unclear whether increased hemodynamic activity near STG is due to increased STG recruitment, or due to a more broadly activated brain network in the vicinity of STG. For instance, there is precedence for activation of additional brain regions as a compensatory strategy for coping with age-related cognitive decline (Presacco et al., 2016; Jamadar, 2020). Listeners who are more vulnerable may use either a broadened brain network or increase STG recruitment, two possibilities that the current data cannot differentiate. However, either interpretations is consistent with the idea that a central processing limitation exists that includes STG and shapes vulnerability to IM.

4.4.5 Diagnostic Utility

The current results bear clinical relevance. A technique we here used to design our stimuli, vocoding, is currently the core principle of speech processing with current cochlear implants. A pressing issue for the majority of cochlear implant users is that they cannot hear well in situations with masking, an impairment in part attributed to cortical dysfunction (Anderson et al., 2017; Zhou et al., 2018b). Sending target and masker sound to opposite ears can improve target speech identification in some, but not all, bilateral cochlear implant users of comparable etiology, suggesting that central auditory processing contributes to clinical performance outcomes (Goupell et al., 2016). However, a challenge for imaging central auditory function in cochlear implant users is that cochlear implants are ferromagnetic devices. Thus, cochlear implants often either unsafe for use in magnetic resonance imaging (MRI) scanners and/or cause sizeable artifacts when imaged with MRI or EEG (Hofmann and Wouters, 2010). Moreover, when imaged under anesthesia, cochlear implant stimulation can fail to elicit cortical responses (Nourski et al., 2013). In contrast, fNIRS, a quiet and light-based technology, is safe to use with cochlear implants. The current paradigm demonstrates that fNIRS-recorded cortical responses to masked speech with impoverished, cochlear-implant-like qualities, can explain approximately a third of the variance in individual vulnerability to IM - an approach that, it is hoped, may prove useful in future clinical practice.

4.5 Methods and Materials

4.5.1 Participants

Our sample size (14 participants for each of the two fNIRS experiments and 11 participants for a behavioral pilot control) was selected a priori using effect size estimates from prior work on IM (Zhang et al., 2018; Arbogast et al., 2002). In total, we recruited 40 paid listeners, who were right-handed native speakers of English, and between 19 and 25 years old (17 females). Assessment of pure-tone audiometric detection thresholds (PTAs) at all octave frequencies from 250 Hz to 8 kHz of 20 dB HL or better verified that all listeners had normal hearing. Specifically, the across-ear differences in pure tone thresholds was 10 dB or less, at all of the audiometric frequencies. All listeners gave written informed consent prior to participating in the study. All testing was administered according to the guidelines of the Institutional Review Board of the New Jersey Institute of Technology.

4.5.2 Speech Stimuli

There were 16 possible English words, each utterance recorded without co-articulation by each of two male talkers (Kidd, et al. 2008). The words consisted of the colors <red, white, blue, and green> and the objects <hats, bags, cards, chairs, desks, gloves, pens, shoes, socks, spoons, tables, and toys>. The colors were designated as keywords. Target word sequences were generated by picking a total of 25 random words from the overall set of 16, including between three and five target words, and concatenating them in random order with replacement (a set of more than 10^{26} possible permutations for the target sequence, $\binom{27}{3} \cdot 12^{22} \cdot 4^3 + \binom{28}{4} \cdot 12^{21} \cdot 4^4 + \binom{29}{5} \cdot$

$12^{20} \cdot 4^5 > 1.6 \cdot 10^{16}$). Similarly, masker sequences were made by picking 25 random words from the overall set of 16, constrained such that target and masker words always differed from each other, for any given word position in the target and masker sequence. One talker was used for the target, the other for the masker. Prior to concatenation, each utterance was initially time-scaled to a duration of 300 ms (Hejna and Musicus, 1991). In addition, 300 ms silences were included between consecutive words, such that the total duration of each target sequence equaled 15 s.

4.5.3 Vocoding

Next, the target word sequences were vocoded through an analysis-, followed by a synthesis-filtering stage. For the analysis stage, each word sequence was filtered into 16 adjacent spectral bands, with center frequencies from 300 to 10 kHz. These spectral bands were spaced linearly along the cochlea according to Greenwood's scale, with a distance of more than one equivalent rectangular cochlear bandwidth between neighboring filters (Greenwood, 1990; Chen et al., 2011). Analysis filters had a simulated spectral width of 0.37 mm along the cochlea (Greenwood, 1990) or approximately 1/10th octave bandwidth, had a 72 dB/octave frequency roll-off and were implemented via time reversal filtering, resulting in zero-phase distortion. In each narrow speech band, the temporal envelope of that band was then extracted using Hilbert transform. Broadband uniformly distributed white noise carriers were multiplied by these envelopes. For the synthesis stage, these amplitude-modulated noises were then processed by the same filters that were used in the analysis stage. Depending on the experimental condition, a subset of these sixteen bands was then added, generating an intelligible, spectrally sparse, vocoded target sequence.

4.5.4 Target/Masker Configurations

A target sequence was always presented simultaneously with a masker sequence. Analogous to an established behavioral paradigm for assessing IM, we used two different masker configurations, consisting of different-band-speech or different-band-noise (Arbogast et al., 2002). In the SPEECH condition, the masker sequence was designed similarly to the target except that it was constrained such that 1) the target and masker words were never equal at the same time and 2) the masker was constructed by adding the remaining seven spectral bands not used to build the target sequence. In the NOISE condition, the masker sequence consisted of 300-ms long narrowband noise bursts that were centered at the seven spectral bands not used to build the target sequence. All processing steps were identical to the SPEECH condition, expect that, instead of being multiplied with the Hilbert envelopes of the masker words, the noise carriers were multiplied by 300-ms long constant-amplitude envelopes that were ramped on and off with the target words (10 ms cosine squared ramps). Figure 4.1B shows a representative spectral energy profile for a mixture of target (brown) and SPEECH (black) sequences. Note that the spectrum of a mixture of target and NOISE samples comprised of similar frequency bands would look visually indistinguishable from target in SPEECH and is thus not shown here (c.f. Arbogast et al. (2002)).

In experiment 1, target and either a different-band speech or a different-band-noise masker were presented binaurally (Figure 4.1B). The target had a left-leading interaural time difference (ITD) of 500 µs. The masker sequence had a right-leading 500 µs ITD, resulting in two possible target/masker configurations, called SPEECH (different-band-speech with 500 µs ITD) vs NOISE (different-band-noise with 500 µs ITD). The target and masker were each presented at 59 dBA, as calibrated with a 1-kHz tone that was presented at the same root mean square as the target and masker and recorded with KEMAR microphones (Knowles Electronics model

KEMAR 45BB). As a result, the broadband Target-to-masker energy ratio (TMR) equaled 0 dB. However, at each of the center frequencies of the nine vocoded spectral bands that made up the target, the TMR equaled 93 dB or more.

In experiment 2, the masker always consisted of a different-band-speech sequence. Target and masker sequences were presented in two possible configurations. The first configuration was identical to the SPEECH condition of experiment 1, with the target presented binaurally with a 500 μs ITD and a SPEECH masker at 500 μs ITD. In the second "SPEECH-oppo" configuration, a target and different-band-speech masker were presented to opposite ears, with the target presented monaurally to the left, and a different-band-speech masker monaurally to the right ear (Figure 4.2).

4.5.5 Behavioral Task

The auditory task consisted of twelve 45-second long blocks. To familiarize the listener with the target voice, at the beginning of each block, we presented a 3-second long cue sentence with the target talker's voice and instructed the listeners to direct their attention to this talker. The cue sentence was "Bob found five small cards," and was processed identically to the target speech for that block (same spectral bands, same binaural configuration). Each block then consisted of a 15-second long acoustic mixture of one randomly generated target and one randomly generated masker sequence, followed by a rest period of 30 seconds of silence. Moreover, at the end of each auditory task block, we added a random silent interval (mean: 3.8 s, variance: 0.23 s, uniform distribution). In experiment 1, we randomly interleaved six SPEECH blocks with six NOISE blocks, whereas in experiment 2, we randomly interleaved six SPEECH blocks with six SPEECH-oppo blocks. The spectral bands of the vocoded target and masker were fixed within each block and randomly interleaved across blocks.

Listeners were instructed to press a button each time the target talker to their left side uttered any of the four color keywords, 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.

4.5.6 Behavioral Detection Threshold

Throughout each block we counted N_B, the number of intervals that the listener pushed the button of the response interface. If a button push occurred within 200 to 600 ms after the onset of a target keyword, the response was scored as a hit. Absence of any button push response in the same time period was scored as a miss. The observed percent correct was calculated by dividing the number of hits by the total number of target keywords during that block.

The baseline guessing rate was estimated via a bootstrapping analysis that calculated the chance percent correct that a simulated listener would have obtained by randomly pushing a button N times throughout that block. Specifically, to estimate the chance percent of keywords guessed correctly via random button push, for each particular listener and block, we randomly shuffled N_B button push intervals across the duration of that particular block's target sequence and counted the number of keywords guessed correctly, then repeated the process by randomly shuffling again for a total of 100 repetitions. To correct for bias, the observed vs chance percent correct scores were then converted to d'-scores, by calculating the difference in z-scores of observed percent correct vs chance percent correct (Klein, 2001).

4.5.7 Behavioral Pilot Control

Behavioral pilot testing established the presence of IM in our stimuli, while also verifying that the high-vs low-IM conditions tested via fNIRS resulted in comparable speech intelligibility. Inside a double-walled sound-attenuating booth (Industrial Acoustic Company), we tested 11 normal-hearing listeners using the same auditory testing equipment and the same speech detection task that we used during the fNIRS recordings, except that listeners had their eyes open during this pilot testing.

In addition, using vocoded stimuli that were recorded by the same talkers as the stimuli used for the speech detection task, we assessed speech identification thresholds by using the coordinate response measure task (Brungart, 2001b; Kidd et al., 2008). Briefly, this task presents listeners with the following sentence structure: "Ready [call sign] go to [color] [number] now." There were eight possible call signs < Arrow, Baron, Charlie, Eagle, Hopper, Laker, Ringo, Tiger>, the same four colors as in the detection task <red, blue, white, green>, and seven numbers (numbers one through eight, except "seven" because, unlike the other numbers, it consists of two syllables). The target sentence was spoken by the same talker for every trial and always had "Baron" as call sign; the masker was either SPEECH or NOISE from a different talker, and using a different call sign than "Baron." Listeners were instructed to answer the question "Where did Baron go?" by identifying the color in the target sentence. The masker was held fixed at 65 dB SPL, whereas the target level varied randomly from trial to trial from 45 to 85 dB SPL, resulting in five possible TMRs from 20, 10, 0, 10, and 20 dB. The target levels were randomized such that all five TMRs were tested in random order before all of them were repeated in different random order. Listeners competed 20 trials per TMR, both in SPEECH and in NOISE. In addition, to verify that all listeners could understand the vocoded speech in quiet at the softest target level, prior to testing masked thresholds, listeners completed 20 trials in quiet at 45 dB SPL.

In quiet, all listeners scored at or near ceiling in the identification task (Figure 4.4A), consistent with previous results that nine-band speech stimuli remain highly intelligible despite vocoding (Shannon et al., 1995). Speech identification thresholds were much worse in SPEECH than NOISE thresholds (Figure 4.4B), confirming that the current stimulus processing produces IM (Arbogast et al., 2002). Using Bayesian inference, each listener's SPEECH and NOISE percent correct speech identification curves were fitted with sigmoidally shaped psychometric functions, as a function of TMR (Matlab toolbox: psignifit; (Wichmann and Hill, 2001)). Identification thresholds were defined as the TMR at 50% correct of these fitted functions. Paired t-tests comparing speech identification thresholds between SPEECH and NOISE found that performance was significantly worse in SPEECH [paired t-test, t(10) = 25.4, p<0.001]. The effect size, calculated as the Cohen's d ratio of the difference in SPEECH and NOISE thresholds divided by the pooled standard deviation across listeners, equaled 4.6. Similarly, speech keyword detectability was better in NOISE than SPEECH, by an average 0.4 d'-units [Figure 4.1C; paired t-test, t(10)=-2.6, p = 0.027]. Cohen's d equaled 1.0.

We wished to eliminate the possibility of artifacts from eye movements and visual attention in our hemodynamic traces. Moreover, we wished to have comparable task difficulty across the tested conditions with fNIRS. Therefore, we next selected the keyword detection task for neuroimaging, because listeners could perform it with minimal body movement and closed eyes. Moreover, task performance was more comparable across maskers for speech detection vs the identification task.

4.5.8 Neuroimaging Procedure

For both experiments, each listener completed one session of behavioral testing while we simultaneously recorded bilateral hemodynamic traces in the vicinity of STG and cIFS, using fNIRS. Throughout testing listeners held their eyes closed. Traces were acquired in 23-minute sessions, consisting of 11 blocks of a controlled breathing task (9 minutes), followed by a brief break (ca. 2 minutes) and twelve blocks of auditory assessment (12 minutes). The controlled breathing task was identical to our prior

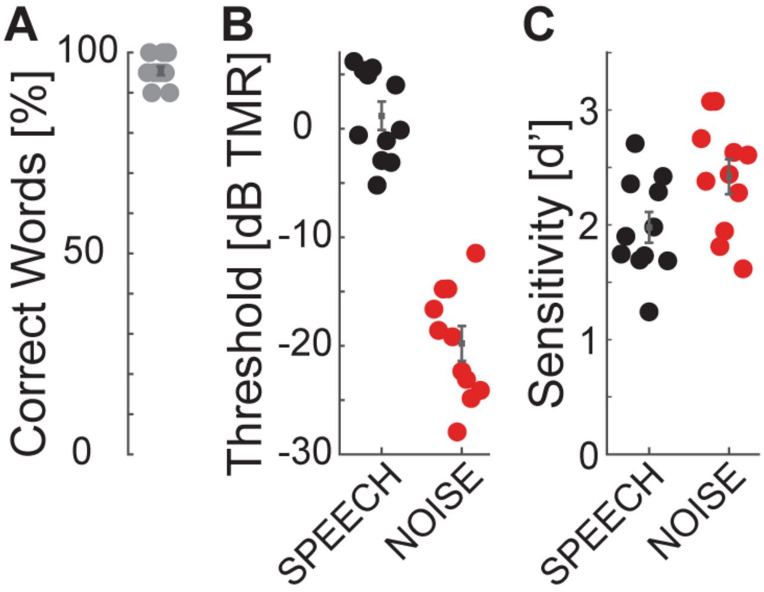

Figure 4.4 Speech identification and detection performance during pilot testing for SPEECH vs NOISE confirm that the SPEECH masker causes IM. The target had a left-leading ITD of 500 μss; the masker a right-leading ITD of 500 μs. (A) Speech identification task. Percent correct keywords identified without masker. (B) Speech identification task. Percent correct keywords identified with SPEECH (black) or NOISE (red) masking. (C) Speech detection task. Sensitivity to keywords with SPEECH (black) or NOISE (red) masking.

methods (see details in Zhang et al. (2018)). Briefly, the task consisted of eleven 45-second-long blocks. In each block, listeners were instructed to breathe in for 5 seconds, breathe out again for 5 seconds. This breathe-in-breathe-out pattern repeated for 6 times (30 seconds in total) before the listeners were instructed to hold breath for 15 seconds. The hemodynamic traces collected during this task establish a baseline dynamic range, from baseline to saturation, over which the optical recordings could vary for each particular listener, recording day and ROI. The auditory assessment was the behavioral detection task described above (see Behavioral Pilot Control).

4.5.9 Recording Setup for fNIRS

The listener wore insert earphones (Etymotic Research ER-2) and a custom-made fNIRS head-cap and held a wireless response interface in the lap (Microsoft Xbox 360 Wireless Controller; Figure 4.1A). Acoustic stimuli were generated on a laptop (Lenovo ThinkPad T440P) with Matlab (Release R2016a, The Mathworks, Inc., Natick, MA, USA), D/A converted 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 (BruelKjaer 2250-G4). The testing suite had intermittent background sound level with peak levels of 44 dBA (moderately quiet university hallway with noise from staff walking by). Together with the ER-2 insert earphones, which provide approximately 30 dB attenuation, the effective background noise level reaching the listener's tympanic membrane was 14 dB A, i.e., moderately quiet.

A camera-based 3D-location tracking and pointer tool system (Brainsight 2.0 software and hardware by Rogue Research Inc., Canada) was used to place the optodes above the left and right cIFS and STG, referenced to standardized brain coordinates (Talairach Atlas; Lancaster et al. (2000)). A custom-built head cap, fitted to the listener's head via adjustable straps, embedded the optodes and held them in place.

Hemodynamic traces were recorded with a 4-source and 16-detector continuous-wave fNIRS system (690 nm and 830 nm optical wavelengths, 50 Hz sampling frequency; CW6, TechEn Inc). The spatial layout of the optical source-detector pairs was custom-designed to cover each of the four ROIs using cross-wise deep quadruple channels with source-detector distances of 3 cm (solid lines in the bottom insert in Figure 4.1A) and one short separation channel with a source-detector distance of 1.5 cm (dashed lines in bottom insert of Figure 4.1A). For each of the resulting 16 deep and 4 shallow source-detector pairs, we then used simulated photon paths to estimate a sensitivity map across the surface of brain by mapping the light paths through a standardized head (Figure 4.1C, AtlasViewer; (Aasted et al., 2015)).

4.5.10 Signal Processing of the fNIRS Traces

Raw fNIRS traces were processed to estimate hemodynamic activation strength (Supplement 2 Figure 4.5A). We first used HOMER2 to process the raw recordings during both the breath holding and auditory tasks, at each of the 16 deep and four shallow source-detector channels (Huppert et al., 2009). Specifically, the raw recordings were band-pass filtered between 0.01 and 0.1 Hz, using time-reversal filtering with a fifth order zero-phase Butterworth filter for high pass filtering and time-reversal filtering with a third order zero-phase Butterworth filter for low pass filtering (commands iltilt and butter in Matlab 2016). Next, we removed slow temporal drifts in the band-pass filtered traces by de-trending each trace with a 20th-degree polynomial (Pei et al., 2007). To suppress artefacts due to sudden head movement, these de-trended traces were then transformed with Daubechies-2 base wavelet functions. Wavelet coefficients outside the one interquartile range were removed, before the remaining coefficients were inversely transformed (Molavi and Dumont, 2012). We then applied a modified Beer-Lambert law to these processed traces, resulting in the estimated oxygenated hemoglobin (HbO) and deoxygenated hemoglobin (HbR) concentrations for each channel (Cope and Delpy, 1988; Kocsis et al., 2006). To obtain hemoglobin changes relative to the maximum dynamic recording range for each individual listener and recording site, we then applied a normalization step. Specifically, for each listener and each of the 20 source-detector channels, we divided the HbO and HbR concentration from the task conditions by the peak of the HbO concentration change during the controlled breathing task, resulting in normalized HbO and HbR traces for each channel. Finally, we averaged the crosswise quadruple deep channels at each ROI, resulting in a total of four task-evoked raw hemoglobin traces per ROI and listener (deep and shallow, HbO and HbR). We previously found that this dynamic range normalization step helps reduce across-listener variability in our listener population with a diverse range of skin pigmentations, hair consistencies and skull thicknesses (Zhang et al., 2018).

4.5.11 Hemodynamic Activation

To estimate auditory-task-evoked neural activity predicted by fixed effects of high-vs low-IM, for each of the two experiments, we next fitted a linear mixed effect model (LMEM) to the pre-processed deep HbO and HbR traces (see Supplement 2 for details on the equations). The LMEM model assumes that three main sources of variance shape the HbO and HbR traces: 1) a task-evoked response with IM independence (significant task-evoked activation that does not covary with IM vulnerability), 2) a task-evoked response with IM dependence (significant task-evoked activation that covaries with IM vulnerability), and 3) nuisance signals, deemed to be unlikely of neural origin. In addition, the LMEM includes the following factors that are known to drive neural response changes in STG and cIFS: audibility as modelled through left and right across-frequency average PTAs, and plasticity as modelled through change in output attributed to block number. To allow direct comparison of the masker evoked responses across different ROIs, alli were referenced relative to the SPEECH recordings in left cIFS.

To estimate whether a neural response captures behavioral phenotypes for vulnerability to IM, for each listener, masker configuration and ROI, we calculated the predicted total HbO and HbR responses from the LMEM weights, ignoring nuisance signals, PTA and plasticity. Using the peak height of the reconstructed HbO or HbR traces as a measure of that ROI's neural recruitment for that masker, we then evaluated whether that ROI's hemodynamic recruitment correlated with the listener's behavioral d' sensitivity to IM.

4.6 Supplement 1

4.6.1 Differences Between Behavioral Pilot vs fNIRS Testing

During behavioral pilot testing, a significant but small effect of masker emerged in the speech detection task. However, during fNIRS testing, any differences between SPEECH vs NOISE in the same behavioral task were too small to reach statistical significance. Specifically, averaged across listeners, speech detection performance in SPEECH equaled 1.97 (S.E. 0.11) during pilot testing as compared to 1.26 (S.E. 0.22) in experiment 1 and 1.71 (S.E. 0.16) in experiment 2. Across-listener average speech detection performance in NOISE equaled 2.41 (S.E. 0.13) during pilot testing vs 1.56 (S.E. 0.21) in experiment 1.

The acoustic delivery of stimuli was identical for fNIRS testing and the behavioral pilot, except that testing happened in different rooms. The fNIRS testing suite had environmental background sound, but it was modest. Indeed, the energy reaching the ears from environmental sound in the fNIRS suite was 50 dB softer than either the masker or target source, presumably only subtly worsening EM or not at all, as compared to the behavioral pilot suite. This hints that the overall reduced performance during fNIRS testing is due to listeners being either more distracted and/or having to put more effort into performing the behavioral task when wearing fNIRS head caps.

4.7 Supplement 2

4.7.1 LMEM

For each experiment, listener and source-detector pair, full hemodynamic traces were preprocessed (Supplement 2 Figure 4.5A) before task-evoked responses were estimated via LMEM (Supplment 2 Figure 4.5B).

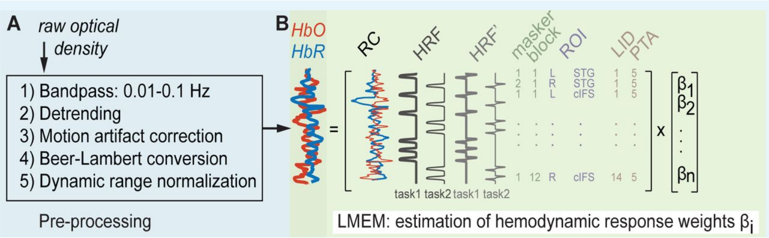

Figure 4.5 For each experiment, all recorded traces were fit with one LMEM. (A) Signal pre-processing steps. (B) Illustration of default effects in the LMEM.

The hemodynamic response function (HRF) is described by:

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

A single LMEM per experiment then fitted these normalized full HbO and HbR traces as follows:

$HG_d =$ (Intercept $\cdot\beta_0 +$ HRF$_{HbO} \cdot \beta_1 +$ HRF'$_{HbO} \cdot \beta_2 +$ #default effects

HRF$_{HbR}$ $\cdot\beta_3 +$ HRF'$_{HbR} \cdot \beta_4 +$ Block number $\cdot\beta_5 +$

Reference channel$_{HbO} \cdot \beta_6 +$ Reference channel$_{HbR}$ $\cdot\beta_7 +$

Hemisphere $\cdot\beta_8 +$ Cortical structure$\cdot\beta_9 +$ Masker configuration $\cdot\beta_{10} +$

R Audio threshold $\cdot\beta_{11} +$ L Audio threshold $\cdot\beta_{12}) +$

(Cortical structure : Masker configuration $\cdot\beta_{13} +$ # interactions

Hemisphere : Masker configuration $\cdot\beta_{14}+$

Hemisphere : Cortical structure $\cdot\beta_{15}+$

HbO HRF : Masker configuration $\cdot\beta_{16}+$

HbO HRF : Cortical structure $\cdot\beta_{17}+$

HbO HRF : Hemisphere $\cdot\beta_{18}+$

HbO HRF' : Masker configuration $\cdot\beta_{19}+$

HbO HRF' : Cortical structure $\cdot\beta_{20}+$

HbO HRF': Hemisphere $\cdot\beta_{21}+$

HbR HRF : Masker configuration $\cdot\beta_{22}+$

HbR HRF : Cortical structure $\cdot\beta_{23}+$

HbR HRF : Hemisphere $\cdot\beta_{24}+$

HbR HRF' : Masker configuration $\cdot\beta_{25}+$

HbR HRF' : Cortical structure $\cdot\beta_{26}+$

HbR HRF' : Hemisphere $\cdot\beta_{27}+$

listener-dependent (Task condition + Cortical Structure + Hemisphere) # random effects

where HG_d is a two-dimensional vector of the normalized full hemoglobin concentrations, HbO and HbR, recorded from the deep source-detector channels. Thei weights associated with each term are a linear measure of how much the term affected the recorded hemoglobin concentration change, relative to the reference condition of SPEECH in left cIFS.

To adjust the onset of the fitted functions to each individual, the LMEM included HRF', the first derivative of HRF (Uga et al., 2014).

Moreover, the LMEM default effects Hemisphere, Cortical structure, and Masker configuration each were two-level categorical variables representing two hemispheres (left vs right, 8), two cortical structures (cIFS vs STG, 9) and two task conditions per experiment (SPEECH vs NOISE in experiment 1; SPEECH vs SPEECH-OPPO in experiment 2, 10). Together, these default effects estimated task-evoked responses in the HbO and HbR traces. In addition, the LMEM included factors that are known to drive neural response changes in STG and cIFS: plasticity, modelled through block number (5), as well as peripheral hearing, modelled through each individual listener's across-frequency average left and right PTA (β_{12} and β_{13}). Finally, cardiovascular nuisance signals unlikely to be of neural origin were regressed out via the shallow source-detector Reference channels (RC;β_{6-7}) in the default model.

As a result, this LMEM implicitly considered that HbR hemodynamic responses are generally much smaller in amplitude and build up more slowly, as compared to HbO (Watanabe et al., 1996; Sato et al., 2004). Specifically, HRFs for HbR and HbO were of the same overall canonical functional form. However, to capture potentially different amplitudes and temporal onsets of HbO and HbR, the LMEM fitted HRF and HRF' amplitudes and their interactions with Masker configuration, Hemisphere and Cortical structure separately for HbO vs HbR (β_{1-4};(Niioka et al., 2018)).

Finally, to regress out idiosynchratic listener-dependent effects on HbO and HbR traces (Sato et al., 2005; Minati et al., 2011), the LMEM included random effects for each listener of Masker configuration, Cortical structure and Hemisphere. Note that we initially explored a range of statistical models. We deemed this LMEM model best in terms of explanatory power and parsimony, because it yielded low overall Akaike's Information Criterion and Bayesian Information Criterion scores (Anderson and Burnham, 2002).

4.8 Supplement 3

4.8.1 Temporal Buildup

Using PET, prior work discovered stronger bilateral STG activation for speech masked by speech relatively to a speech masked by speech baseline (Scott et al., 2009), a finding confirmed by the current results via fNIRS. That prior work assessed speech identification while participants listened passively (Scott et al., 2009). In contrast, here, hemodynamic responses were recorded while listeners were actively engaged in a speech detection task. Of note, the prior study also showed that the left STG was more strongly recruited than right STG under IM (Scott et al., 2009). To examine hemispheric differences, we compared the LMEM predicted hemodynamic response across left and right hemisphere for STG, and, separately for STG. However, for the stimuli tested here during active listening, no robust hemispheric differences in STG activation were revealed.

Furthermore, frontal cortex peak responses during an EM task were found to lag behind STG responses, by approximately 1.5 seconds, when normally-hearing listeners were assessed with fNIRS while listening to vocoded speech in noise (Wijayasiri et al., 2017). To analyze the temporal buildup of the task-evoked responses, we subtracted the responses attributed by the LMEM to the STG from those attributed to the cIFS.

In both experiments, masker-evoked differences in overall recruitment of STG vs cIFS varied over time. In experiment 1, STG was slightly more strongly recruited during the first 2 seconds of the task interval, before the recruitment between STG and cIFS became more balanced, in both the left and right hemispheres (Supplement 3 Figure 4.6A). Similarly, in experiment 2, within each hemisphere, STG was relatively more engaged than cIFS during the first 2-5 seconds of the task, followed by stronger HbO and HbR recruitment in the cIFS region (Supplement 3 Figure 4.6B). Thus, the current temporal buildup results are consistent with prior findings that STG activates before frontal regions, for both EM and IM (Wijayasiri et al., 2017).

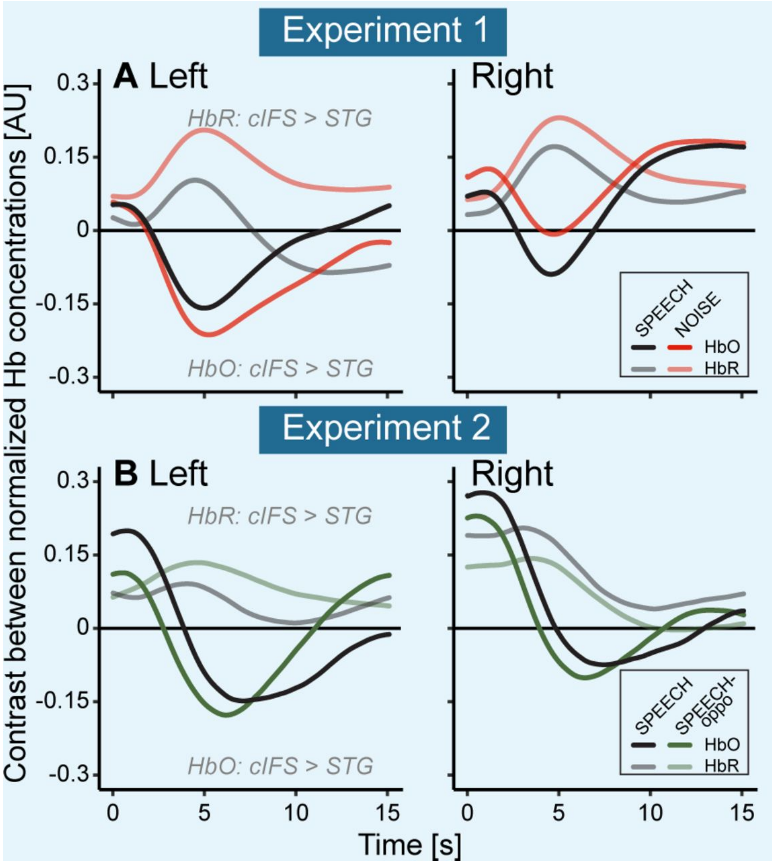

Figure 4.6 For the first 2-5 seconds of the task, STG was more active than cIFS, whereas cIFS responded more strongly afterwards, with both regions being balanced in their relative activations near the end of the task interval (15 s). This temporal buildup was observed in both hemispheres, for all tested masker configurations, and for both HbO and HbR. Note that HbO (darker lines) and HbR (lighter lines) are anti-correlated, here, as expected. (A) Relative to each ROIs own peak activation levels, STG is slightly more strongly activated than cIFS, for both SPEECH (black) and NOISE in experiment 1(red), in both the Left and the Right hemisphere. (B) The temporal buildup of dominant STG vs cIFS activity in experiment 2 are qualitatively comparable to the results in experiment 1 (compare black lines in top vs bottom plots. Moreover, the pattern where early STG activity emerges prior to stronger cIFS recruitment also holds for SPEECH-oppo.

Table 4.1 Results of LMEM, Experiment 1

	Term		Estimate	S.E.	t	p	
β_0	Intercept		-0.35	0.092	-3.8	<0.0001	***
β_1	HRF$_{HbO}$		0.55	0.004	138.3	<0.0001	***
β_2	HRF'$_{HbO}$		0.17	0.004	39.4	<0.0001	***
β_3	HRF$_{HbR}$		0.02	0.004	5.8	<0.0001	***
β_4	HRF'$_{HbR}$		0.11	0.043	26.8	<0.0001	***
β_5	Block Number		0.01	0.000	76.6	<0.0001	***
β_6	Reference Channel$_{HbO}$		0.42	0.000	1342.0	<0.0001	***
β_7	Reference Channel$_{HbR}$		0.44	0.001	580.8	<0.0001	***
β_8	Hemisphere		0.04	0.028	1.5	0.14	
β_9	Cortical Structure		0.08	0.026	3.0	0.003	**
β_{10}	Masker		0.14	0.061	2.2	0.025	*
β_{11}	R ear PTA		0.02	0.008	1.8	0.08	.
β_{12}	L ear PTA		-0.01	0.005	-0.9	0.38	
β_{13}	Masker Configuration	: Cortical Structure	-0.03	0.003	-12.8	<0.0001	***
β_{14}	Masker Configuration	: Hemisphere	-0.05	0.003	-21.1	<0.0001	***
β_{15}	Cortical Structure	: Hemisphere	-0.01	0.003	-5.4	<0.0001	***
β_{16}	HRF$_{HbO}$	: Masker Configuration	-0.19	0.004	-46.5	<0.0001	***
β_{17}	HRF$_{HbO}$	: Cortical Structure	0.17	0.004	41.6	<0.0001	***
β_{18}	HRF$_{HbO}$	: Hemisphere	-0.43	0.004	-10.8	<0.0001	***
β_{19}	HRF'$_{HbO}$	: Masker Configuration	0.02	0.004	5.6	<0.0001	***
β_{20}	HRF'$_{HbO}$	: Cortical Structure	-0.22	0.004	-51.6	<0.0001	***
β_{21}	HRF'$_{HbO}$	: Hemisphere	-0.04	0.004	-9.5	<0.0001	***
β_{22}	HRF$_{HbR}$	: Masker Configuration	-0.12	0.004	-30.2	<0.0001	***
β_{23}	HRF$_{HbR}$	: Cortical Structure	-0.01	0.004	-1.0	0.3	
β_{24}	HRF$_{HbR}$	: Hemisphere	0.05	0.004	11.9	<0.0001	***
β_{25}	HRF'$_{HbR}$	: Masker Configuration	-0.10	0.004	-22.4	<0.0001	***
β_{26}	HRF'$_{HbR}$	: Cortical Structure	0.16	0.004	36.6	<0.0001	***
β_{27}	HRF'$_{HbR}$	: Hemisphere	0.04	0.004	9.3	<0.0001	***

All estimates are referenced to a default condition in left cIFS for Attentive.

Significance codes: ‘***’ $p < 0.001$, ‘**’ $p < 0.01$, ‘*’ $p < 0.05$, ‘.’ $p < 0.1$, ‘ ’ p 0.1.

Note: “Int” = “intercept”; “S.E.” = standard error of the mean

Table 4.2 Results of LMEM, Experiment 2

	Term		Estimate	S.E.	t	p	
β_0	Intercept		0.02	0.065	0.3	0.75	
β_1	HRF$_{HbO}$		0.29	0.003	89.4	<0.0001	***
β_2	HRF'$_{HbO}$		0.07	0.003	19.5	<0.0001	***
β_3	HRF$_{HbR}$		-0.04	0.003	-10.9	<0.0001	***
β_4	HRF'$_{HbR}$		0.07	0.004	20.8	<0.0001	***
β_5	Block Number		0.01	0.000	39.1	<0.0001	***
β_6	Reference Channel$_{HbO}$		0.67	0.001	1490.4	<0.0001	***
β_7	Reference Channel$_{HbR}$		0.73	0.001	802.0	<0.0001	***
β_8	Hemisphere		-0.02	0.025	-0.7	0.46	
β_9	Cortical Structure		0.04	0.034	1.2	0.23	
β_{10}	Masker		0.00	0.025	0.04	0.97	
β_{11}	R ear PTA		-0.01	0.011	-0.97	0.33	
β_{12}	L ear PTA		0.00	0.009	0.3	0.79	
β_{13}	Masker Configuration	: Cortical Structure	0.06	0.002	26.3	<0.0001	***
β_{14}	Masker Configuration	: Hemisphere	-0.03	0.00	-14.5	<0.0001	***
β_{15}	Cortical Structure	: Hemisphere	0.08	0.002	40.3	<0.0001	***
β_{16}	HRF$_{HbO}$	: Masker Configuration	-0.1	0.003	-31.8	<0.0001	***
β_{17}	HRF$_{HbO}$	: Cortical Structure	0.04	0.003	11.1	<0.0001	***
β_{18}	HRF$_{HbO}$	: Hemisphere	0.03	0.003	8.5	<0.0001	***
β_{19}	HRF'$_{HbO}$	: Masker Configuration	-0.01	0.003	-1.8	0.072	.
β_{20}	HRF'$_{HbO}$	: Cortical Structure	-0.19	0.003	-53.9	<0.0001	***
β_{21}	HRF'$_{HbO}$	: Hemisphere	-0.06	0.003	-16.63	<0.0001	***
β_{22}	HRF$_{HbR}$	: Masker Configuration	0.003	0.003	1.1	0.29	
β_{23}	HRF$_{HbR}$	: Cortical Structure	-0.05	0.003	-14.4	<0.0001	***
β_{24}	HRF$_{HbR}$	: Hemisphere	-0.04	0.003	-11.9	<0.0001	***
β_{25}	HRF'$_{HbR}$	: Masker Configuration	0.01	0.003	3.0	0.0031	**
β_{26}	HRF'$_{HbR}$	: Cortical Structure	0.06	0.003	17.8	<0.0001	***
β_{27}	HRF'$_{HbR}$	: Hemisphere	-0.01	0.003	-3.5	0.0006	***

All estimates are referenced to a default condition in left cIFS for Attentive.

Significance codes: ‘***’ $p < 0.001$, ‘**’ $p < 0.01$, ‘*’ $p < 0.05$, ‘.’ $p < 0.1$, ‘ ’ p 0.1.

Note: “Int” = “intercept”; “S.E.” = standard error of the mean

CHAPTER 5

ASSESSING TEST-RETEST RELIABILITY USING MACHINE LEARNING ON FUNCTIONAL NEAR-INFRARED SPECTROSCOPY DATA

5.1 Abstract

Previous work demonstrates functional near-infrared spectroscopy (fNIRS) recordings over the lateral frontal cortex are a viable marker for dissociating whether a person is attentively or passively listening at the group level. Here, we assessed our fNIRS protocol's test-retest reliability to facilitate this work's transition into clinical practice. Specifically, we established how many fNIRS channels and task blocks are needed for a reliable estimation of the listener's attentive state.

5.2 Introduction

For people with severe-to-profound hearing loss for whom hearing aids are not effective, cochlear implants (CIs) can restore hearing by electrically stimulating the auditory nerve fibers. Most CI users demonstrate sizable improvements in speech perception after implantation (Hochberg et al., 1992; Kiefer et al., 1996). However, CI users' speech recognition performance is vulnerable to background noise, especially when the interfering sound is a competing speech or speech-shaped noise (Müller-Deile et al., 1995; Nelson et al., 2003). In these situations where background interfering sound is present, normal-hearing listeners can hear out target speech by utilizing spatial auditory cues, a phenomenon called spatial release from masking (SRM). Three main factors are thought to contribute to SRM: 1) binaural masking level differences (BMLDs) thought to be caused by binaural decorrelation processing in the brainstem; 2) head shadow effects where the signal to noise ratio (SNR) is increased in one ear due to attenuation of the noise from the listener's head when target speech and masker are spatially separated, and 3) spatial attention, the ability to focus auditory perception to a location in space (Arbogast et al., 2005; Ihlefeld et al., 2008; Jones et al., 2011). Bilateral CIs (BiCI) intend to restore binaural cues. However, most BiCI users benefit little from spatial cues (Loizou, 2009). Some BiCI users show release from masking when target and interferer are presented in opposite ears instead of the same ear, but many do not benefit at all (Goupell et al., 2017). It is unclear whether BiCI users can direct selective auditory attention.

In our previous study (Zhang et al., 2018), we have shown that functional near-infrared spectroscopy (fNIRS) can be used as an objective measure of auditory attention in a rapid-serial target detection task of target speech that was spatially separated from background masking speech by interaural time difference (ITD). This experiment was adapted from Michalka and coworkers' work (Michalka et al., 2015), specifically the sustained visual and auditory spatial attention experiment. However, instead of delivering competing auditory streams separately to the left and right ear, we delivered our competing auditory stream with an ITD. Our fNIRS recordings confirmed that the vicinity of the caudal inferior frontal sulcus (cIFS) showed increased activity when normally hearing participants actively engaged in the task compared to a passive listening control condition where they simply heard the sounds played back without task engagement. One potential caveat of our prior approach is that we tested different stimuli in the attentive vs. passive listening conditions, potentially contributing to the observed difference in hemodynamic activation across those two conditions. Specifically, the speech tokens in the passive conditions were processed to have similar magnitude spectra and similar temporal-fine structure characteristics as the original speech tokens used in the attentive listening condition (Ellis, 2010). However, unlike in the attentive condition, the passive-listening tokens were scrambled temporally such that they were unintelligible. To rule out a potential confound due to token intelligibility, we here conducted a follow-up experiment. This follow-up experiment was similar to our previous work, except that identical stimuli were tested in the attentive and passive listening conditions.

fNIRS can non-invasively record hemodynamic changes in brain tissue, which predict the underlying neural responses due to neurovascular coupling (Huneau et al., 2015; Malonek et al., 1997). Using the Beer-Lambert law, oxygenated (HbO) and deoxygenated (HbR) hemoglobin concentration changes at regions of interest in the brain (ROIs) can be estimated and used to estimate neuronal recruitment (Cauli et al., 2010; Nemoto et al., 1997). Current fNIRS technology can image at depths of up to 3 cm from the skull's surface, making it suitable for assessing cortical responses. Unlike functional magnetic resonance imaging (fMRI), fNIRS uses near-infrared light to detect hemodynamic responses. Thus it is safe to use on CI users, and it is not affected by high-frequency carrier pulses used in CI. Many auditory studies have incorporated fNIRS as the neuroimaging tool to study listening effort (Rovetti et al., 2019; Rowland et al., 2018; Wijayasiri et al., 2017) and perceived loudness (Weder et al.,2018; 2020).

Previous fMRI studies have identified the superior temporal cortex as part of the auditory attention network (Riecke et al., 2016; Tobyne et al., 2017). fNIRS recordings from normal-hearing listeners have shown activation from superior temporal regions increased as speech intelligibility improved (Defenderfer et al., 2017; Lawrence et al., 2018). fNIRS data from CI users performing speech understanding tasks showed a negative correlation between activation in the superior temporal cortex and auditory speech understanding performance (Zhou et al., 2018). These findings suggest that the superior temporal cortex is a promising region of interest (ROI) to study auditory attention tasks. We previously confirmed that the cIFS region engages more strongly when listeners actively attend to speech. Here we widened ROIs to include the superior temporal region as well.

Prior works often find high inter-individual variability in fNIRS responses, both for auditory and non-auditory recordings (Huppert et al., 2006; Maggioni et al., 2015; Sato et al., 2005; Wiggins et al., 2016). The reason for this could be the different scalp-cortex distances among subjects that resulted in different gray matter volumes reached by photons (Haeussinger et al., 2011; Heinzel et al., 2013). Some studies found the effect of skin blood flow on fNIRS data (Klaessens et al., 2005; Takahashi et al., 2011), and transient blood pressure could also contribute to interindividual variability (Minati et al., 2011). The current work's secondary goal is to confirm our fNIRS protocol's test-retest reliability and assess how many fNIRS optode channels and how many task blocks are necessary to obtain a robust estimate of a listener's attentive state. First, we estimate the task-related oxygenated hemoglobin (HbO) response from the fNIRS recording by applying a breath-holding normalization across the recorded traces at all ROIs and fitting the normalized data with LMEM (Zhang et al., 2020). Second, we applied a support vector machine (SVM) on the extracted HbO response during the auditory task period as features to classify an individual listener's attentive or passive listening state and plot the classification accuracy as a function of the number of channels and recording blocks used to build the training datasets. The classification performance was then used as the measure to assess test-retest reliability.

5.3 Methods

5.3.1 Participants

In the previous experiment, we recorded 14 listeners (ages 19 to 25, 4 females). In the current experiment, we recorded from a different group of 14 listeners (ages 19 to 25, 6 females). Only native speakers of English were recruited for both studies. All listeners were tested for audiometric pure-tone detection thresholds from 250 Hz to 8 kHz and were confirmed to have tone detection thresholds of 20 dB HL or better, and the thresholds did not differ by more than 10dB across ears. According to the Institutional Review Board of the New Jersey Institute of Technology guidelines, both studies were administered, and listeners gave written informed consent to participate in the study.

5.3.2 Recording Setup

The current experiment uses an identical recording setup in our previous work, except we used a different optodes design targeting cIFS and STG as the ROI instead of the cIFS and transverse gyrus intersecting precentral sulcus (tgCPS) in the previous work. Briefly, we used a continuous-wave diffuse optical NIRS system (CW6; TechEn Inc., Milford, MA) with a sampling frequency of 50 Hz. A laptop (Lenovo ThinkPad T440P) placed approximately 0.8m in front of the listener was used to display the experiment instructions. Listeners wore inserted earphones (Etymotic Research ER-2) to receive acoustic stimuli. Microsoft Xbox 360 Wireless Controller was used as the wireless response interface. The experiment was conducted in a room with a moderately quiet background sound level of less than 44 dBA. We used Brainsight 2.0, a camera-based three-dimensional positioning system (Rogue Research Inc., Canada), to locate ROIs and place the optodes accordingly. Optodes were embedded in a custom-built head cap, and the head cap was secured on listeners' heads via adjustable straps.

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

Figure 5.1 top shows the ROIs and optodes design of the current experiments. In the figure, solid lines (source-detector distance of 3 cm) are the deep recording channels. The dotted line (source-detector distance of 1.5 cm) are short separation channels used to pick up physiological noises.

5.3.3 Controlled Breathing Task

In the current experiment, listeners performed the controlled breathing task as detailed in the previous paper before the auditory task. They were instructed to follow a sequence of visual instructions on the laptop screen during each task block. The instructions were (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. Each sequence consisted of 6 interleaved inhales and exhales. Each lasted 5 seconds for a total of 30 seconds, followed by a held breath of 15 seconds. Each listener completed 11 blocks of controlled breathing tasks.

The maximum oxygenated blood (HbO) concertation change during the breath-holding task was used to normalize fNIRS recordings for auditory tasks to account for individual variability in skull thickness, skin pigmentation, and other idiosyncratic factors that can adversely affect recording quality with fNIRS data.

5.3.4 Auditory Tasks

After the controlled breathing task, listeners completed 24 blocks of auditory tasks split evenly between attentive and passive conditions, with a counter-balanced task order (attentive and passive) across listeners. In each block, listeners were presented with two competing auditory stimuli of 15 seconds duration each. In both attentive listening and passive listening conditions, the same closed-set speech tokens, uttered in isolation by two different male talkers, were presented at 59 dBA for each source. The two competing streams had different Interaural Time Differences (ITDs): -500 μs (left) for the target stream and 500 μs (right) for the masker stream. The speech tokens consisted of 16 possible words, including the colors <red, white, blue, and

green> and the objects <hats, bags, card, chairs, desks, gloves, pens, shoes, socks, spoons, tables, and toys>. Each original word was time-scaled to make each word last 300 ms. 25 of those processed words were randomly chosen with replacement to form the target and masker stream, and 300 ms of silence was added between each word so that each stream had a total duration of 15 seconds for the task period. A 3 seconds long cue sentence was spoken by the target speaker, "Bob found five small cards" was played at the beginning of each block to familiarize listeners with the target voice.

Although the sound stimuli are the same for attentive and passive conditions, the intrusions given to the listeners for those two conditions were different: for attentive listening condition, listeners were instructed to "Press the response button each time the target talker utters one of the four color words." (Keywords: <"red," "green," "blue," "white">); while for the passive listening condition, listeners were instead instructed to "Press the response button each time the Xbox controller vibrates."

5.3.5 Preprocessing of the fNIRS Data

Similar preprocessing steps as our previous paper were used to process the data. HOMER2 (Huppert et al. 2009) was used to analyze the fNIRS channels' raw recordings, and we only used HbO concertation to train and test the classifiers. First, the recordings were band-pass filtered between 0.01 and 0.1 Hz, using a third-order zero-phase Butterworth filter for the low pass filter and a fifth-order zero-phase Butterworth filter for the high pass filter. Next, we removed slow temporal drifts by de-trending each trace with a 20th-degree polynomial (Pei et al., 2007). The traces were then wavelet transformed using Daubechies 2 (db2) base functions to remove motion artifacts such as sudden head movement during the recording. We removed wavelet coefficients outside of one interquartile range (IQR) (Molavi et al., 2012). Finally, the modified Beer-Lambert law (Cope and Delpy, 1988; Kocsis et al.,

2006) was applied to convert these processed traces to the estimated oxygenated and deoxygenated hemoglobin (HbO, HbR) concentrations for each channel at each ROI. To reduce across-listener variability, we then normalized the HbO concentrations from each ROI by dividing them by the peak of the HbO concentration in the same ROI during controlled breathing tasks.

5.3.6 Linear Mixed Effect Model (LMEM)

We used an LMEM to fit our recordings and assess whether task condition (attentive vs. passive listening) affects hemodynamic response. Briefly, the LMEM model took into account the primary sources of variance that contributed to the observed HbO and HbR concentration changes: hemodynamic response function (HRF) evoked by task condition and nuisance signals of non-neural origin that was recorded from shallow channels. In addition. This model also included the random effects from individual listeners.

The details of LMEM fitted to normalized HbO, and HbR concentration changes are shown as follows:

HG_d = (Intercept $\cdot\beta_0$+ HRF$_{HbO}\cdot\beta_1$+ HRF'$_{HbO}\cdot\beta_2$+ #default effects

HRF$_{HbR}$ $\cdot\beta_3$+ HRF'$_{HbR}\cdot\beta_4$+ Block number $\cdot\beta_5$+

Reference channel$_{HbO}\cdot\beta_6$+ Reference channel$_{HbR}$ $\cdot\beta_7$+

Hemisphere $\cdot\beta_8$+ Cortical structure$\cdot\beta_9$+ Masker configuration $\cdot\beta_{10}$+

R Audio threshold $\cdot\beta_{11}$+ L Audio threshold $\cdot\beta_{12}$)+

(Cortical structure : Masker configuration $\cdot\beta_{13}$+ # interactions

Hemisphere : Masker configuration $\cdot\beta_{14}$+

Hemisphere : Cortical structure $\cdot\beta_{15}$+

HbO HRF : Masker configuration $\cdot\beta_{16}$+

HbO HRF : Cortical structure $\cdot\beta_{17}$+

HbO HRF : Hemisphere $\cdot\beta_{18}$+

HbO HRF' : Masker configuration $\cdot\beta_{19}+$

HbO HRF' : Cortical structure $\cdot\beta_{20}+$

HbO HRF': Hemisphere $\cdot\beta_{21}+$

HbR HRF : Masker configuration $\cdot\beta_{22}+$

HbR HRF : Cortical structure $\cdot\beta_{23}+$

HbR HRF : Hemisphere $\cdot\beta_{24}+$

HbR HRF' : Masker configuration $\cdot\beta_{25}+$

HbR HRF' : Cortical structure $\cdot\beta_{26}+$

HbR HRF' : Hemisphere $\cdot\beta_{27}+$

listener-dependent (Task condition + Cortical Structure + Hemisphere) # random effects

HG_d is a vector of the normalized HbO and HbR concentration changes recorded from the deep source-detector channels. After the model fit, β_i weights indicated how much each term linearly affects the observed concentration changes, using attentive condition in left cIFS as the comparison baseline. To account for different listeners' different HRF onset times, we also added HRF', the first derivative of HRF, to the LMEM (Uga et al., 2014).

5.3.7 Training and Testing of Classifier

We built the training data-sets using parametrically varied combinations of different channels from left/right cIFS/STG and a different number of task blocks (added successively from 1 to 12 blocks for each condition). For example, one training dataset could be formed from the recording of 1 channel from left cIFS, and only the first two blocks of the recording were used. There were 360 datasets in total to cover all combinations of channels and blocks. 15 (1 optode channels used + 2 optode channels used + 3 optode channels used + 4 optode channels used) x 12 (task blocks) x 2 (left/right) = 360.

For the testing data, we want to classify attentive vs. passive at an individual level, so for each listener, the block average of all 12 blocks from each task condition (attentive vs. passive) and the average of 4 channels from left and right cIFS was used to generate the testing data-set. Thus, at each ROI, one listener will contribute one data point to be classified.

During the 15 seconds task period of each block, we noticed the HbO concentration differences between attentive vs. passive conditions were typically largest in the last second before the end of each task period. To further improve accuracy, we built an ensemble classifier that consisted of 50 linear support vector machines (SVM), each trained using the features constructed using data points of HbO concentration from 14 to 15 seconds (fNIRS recording had 50Hz sampling frequency) and the HbO concentration average from the reference channel. The classification result of this ensemble classifier was determined by the majority vote of all linear SVMs.

For each of the 360 training datasets we prepared, one ensemble classifier was trained. Each ensemble classifier was then tested on testing data. First, we trained and tested the classifier using the current experimental data. 10 from the 14 listeners from the current experiment were randomly chosen as the training and the remaining 4 as testing, and then we repeated the same procedure 100 times to validate the ensemble classifier.

Next, we trained the classifier on the current experiment data (14 listeners). We tested it on the previous experiment data, randomly choosing 10 from the 14 listeners of the previous experiment to be tested. The process was repeated 100 times to validate the ensemble classifier.

5.4 Results

5.4.1 Attentive vs. Passive

As shown in Figure 5.1, the repeated experiment using intelligible speech as the stimuli during the passive condition still produced the stronger activation during the attentive condition. The LMEM fitted to the data could account for 36% (R^2 = 0.36) of the recorded data variance. From LMEM fitting results, the coefficient for the term: interaction between HRF_{HbO} and task condition is negative (β_{16} = -0.77, p < 0.0001) using the attentive listening condition as the comparison reference. This negative value coefficient meant the activation during the passive condition was smaller compared to the attentive condition. The detailed LMEM fitting results are shown in Table 5.1.

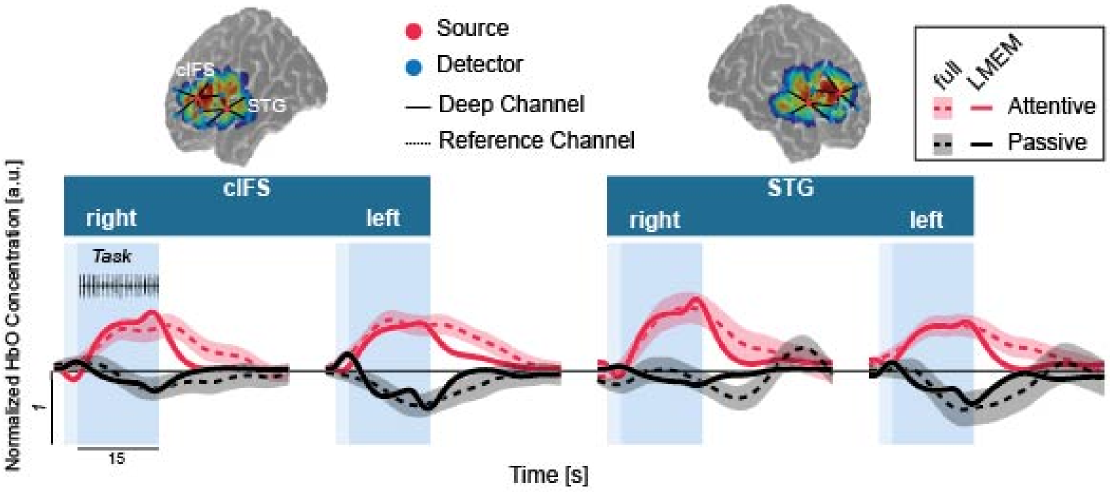

Figure 5.1 Results from the current experiment where the identical intelligible speech was played for active and passive conditions. Top: Sensitivity map. Warmer colors denote an increased likelihood that photons will be recorded from these areas; Bottom: Normalized HbO traces during the attentive vs. passive listening at each of the four ROIs. The ribbons around each trace show one standard error of the mean across listeners.

5.4.2 Classification Accuracy

Classification accuracy was shown as a function of blocks used to build training datasets for different channel combinations in Figure 5.2. The classification accuracy improved as more channels from each ROI were averaged and as more blocks were used to build the training datasets. To determine the number of task blocks required for the accuracy increasing to plateau, we used R package strucchange (Zeileis et al., 2002, 2003) to detect the structural change in the linear model fit to the accuracy vs. block number data. The linear model fits were plotted with classification data as dotted horizontal lines, while dotted vertical lines marked the breakpoints at which the accuracy started to plateau. For the current experiment (Figure 5.2A), six task blocks were required to reliably capture the attentive state information from fNIRS recordings by conservative estimate (median of all breakpoints). With more channels averaged from each ROI, the accuracy vs. block number converged faster. Based on the blocks required to reach the plateau, two channels from each ROI averaged could achieve the same classification results as when all channels from each ROI were used. Classification accuracy on the previous experiment data collected from a different group of listeners reached plateau faster with a median of two task blocks required (Figure 5.2B). A three-channel average from each ROI is required to have a comparable result with a four-channel average from each ROI where accuracy plateaued around two task blocks used to build the training data sets.

5.5 Discussion

The main challenge to fNIRS is its high sensitivity to physiological fluctuations, present during both resting and functional task periods (Tong et al., 2011; Kirilina et al., 2012). Thus, the fNIRS signal may not accurately represent task-evoked brain activity (Tachtsidis and Scholkmann, 2016). Separation of task-evoked signals from signals of physiological origins is required to estimate brain activity in response to

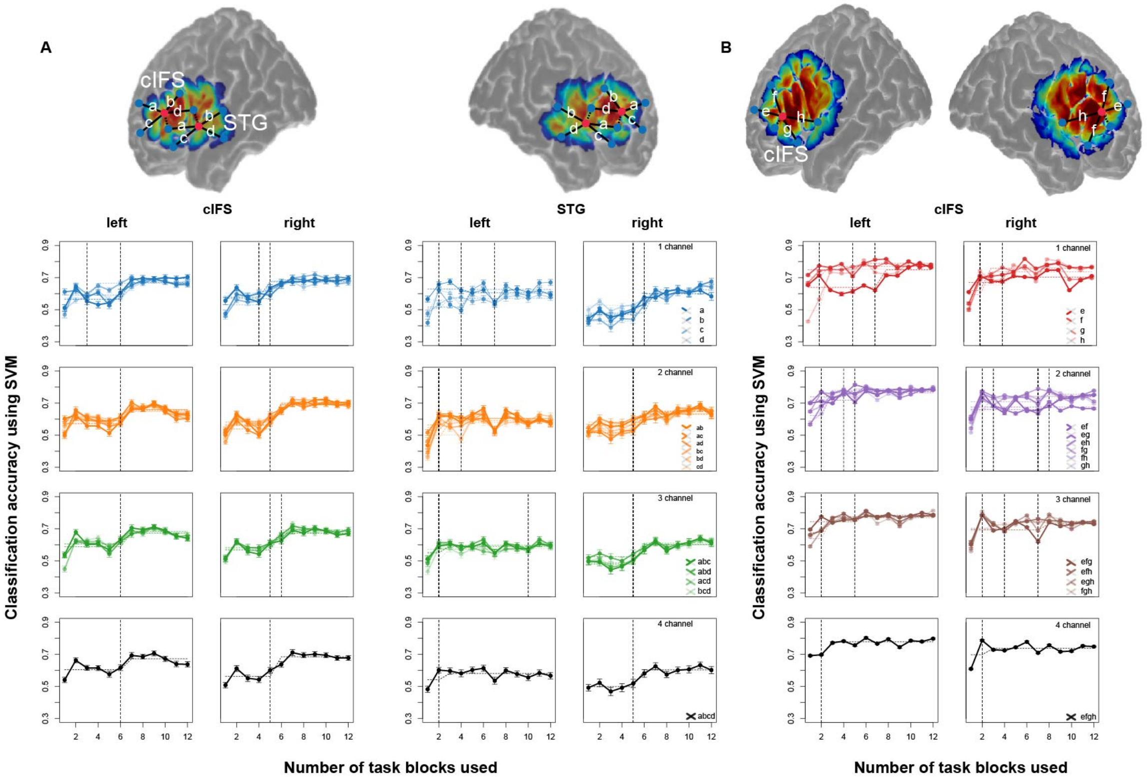

Figure 5.2 Classification accuracy was tested on the current experiment (A) and on the previous experiment (B) with parametrically varied task blocks and recording channels used to build the classifiers' training datasets. Solid lines and circles are the classification accuracy. The dotted line is the linear regression model fit with the vertical dotted line marking the change in linear models.

the task accurately. One way to separate functional and systemic components is by assuming oxygenated hemoglobin (HbO) and deoxygenated hemoglobin (HbR) responses are anti-correlated during the task period (Yamada, 2012). Alternatively, short separation channels can be incorporated into the design. Those short separation channels have shorter source-detector distances (less than 1.5 cm) as compared to standard recording channels (3 cm) and are more affected by global physiological fluctuations (Funane et al., 2013). Contrast-to-noise ratios (CNRs) can be improved by fitting short separation channels into a general linear model (GLM) as regressors and subtracting them from the recording (Sato et al., 2016). A direct comparison between the anti-correlation method and the short-channel subtraction method has shown that short-channel subtraction is better at improving CNRs (Zhou et al., 2020). More studies on how to design short separation channels have found that short separation channels should be placed as close as 1.5 cm from the standard recording channels (Gagnon et al., 2012), and CNRs can be further improved with shorter source-detector distance for short separation channels (Brigadoi et al., 2015; Goodwin et al., 2014). Due to our optodes' physical design, we could only use 1.5 cm for short separation channels. Instead of GLM, we fitted a linear mixed effect model (LMEM) to the fNIRS recording, which takes into account listener-specific random effects as well. After the LMEM, we calculate the percentage by which our short separation channel contributed to the overall variance observed in fNIRS data by calculating the area under the curve of short separation channels. We confirmed that short separation channels contributed to 48% percent of the total variance of hemodynamic activation by calculating the area under the fitted curve with vs. without β_6 and β_7.

Though the fNIRS signal has a weaker signal-to-noise ratio (SNR) than fMRI, the BOLD signal collected using fNIRS is consistent with fMRI studies (Olds et al., 2016; Pollonini et al., 2014). fNIRS still has limitations compared to fMRI: 1) limited spatial resolution; 2) limited penetration depth (Cui et al., 2011). In our fNIRS experiments, the spatial resolution is limited by the source-detector pairs available on the fNIRS recording hardware. To ensure our fNIRS recording could cover the cortex's relevant ROI regarding auditory attention, we used previous fMRI studies and recent fNIRS studies as the guides to choose cIFS and STG as our ROIs. Due to the fNIRS scanner's limited penetration depth, one primary concern is whether the detected hemoglobin concentration change results from neural activities or physiological fluctuations. In the current study, we designed the control condition (passive listening) to have identical sound stimuli and the same finger movement during the task; the only difference between attentive vs. passive condition is the task engagement. Results showed a significant difference between activation patterns in fNIRS recordings from attentive vs. passive state, confirming the previous experiment's findings that the observed difference is of neural activity origin. It could be generalized to the situation where an intelligible speech was presented during the passive listening condition. We used this task as a benchmark for estimating the test-retest reliability of our fNIRS procedure. To get a robust estimate of the listener's attentive state, we found a conservative estimate of six recording task blocks required. Classification accuracy converges faster with more channels averaged from respective ROI, and a two-channel average from each ROI is enough to capture the attentive vs. passive state information. The classification accuracy plateaued when the recording consisted of six task blocks.

5.6 Conclusions

This study confirmed our previous finding that attentive listening evoked a different brain activation pattern compared to passive listening even when the identical speech stimuli were played for both conditions, suggesting fNIRS can be used as an objective measure for auditory attention. In addition, we assessed our fNIRS protocol's test-retest reliability to facilitate this work's transition into clinical practice. We established that two fNIRS channels averaged from each ROI and six task blocks of fNIRS recording are needed for a reliable estimation of the listener's attentive state.

Table 5.1 Results of LMEM, Experiment 3

	Term		Estimate	S.E.	t	p	
β_0	Intercept		0.0692	0.10100.68	0.494		
β_1	HRF$_{HbO}$		0.5178	0.0042	123.28	<0.0001	***
β_2	HRF'$_{HbO}$		0.17	0.004	39.4	<0.0001	***
β_3	HRF$_{HbR}$		0.02	0.004	5.8	<0.0001	***
β_4	HRF'$_{HbR}$		0.11	0.043	26.8	<0.0001	***
β_5	Block Number		0.01	0.000	76.6	<0.0001	***
β_6	Reference Channel$_{HbO}$		0.42	0.000	1342.0	<0.0001	***
β_7	Reference Channel$_{HbR}$		0.44	0.001	580.8	<0.0001	***
β_8	Hemisphere		0.04	0.028	1.5	0.14	
β_9	Cortical Structure		0.08	0.026	3.0	0.003	**
β_{10}	Masker		0.14	0.061	2.2	0.025	*
β_{11}	R ear PTA		0.02	0.008	1.8	0.08	.
β_{12}	L ear PTA		-0.01	0.005	-0.9	0.38	
β_{13}	Masker Configuration	: Cortical Structure	-0.03	0.003	-12.8	<0.0001	***
β_{14}	Masker Configuration	: Hemisphere	-0.05	0.003	-21.1	<0.0001	***
β_{15}	Cortical Structure	: Hemisphere	-0.01	0.003	-5.4	<0.0001	***
β_{16}	HRF$_{HbO}$	: Masker Configuration	-0.19	0.004	-46.5	<0.0001	***
β_{17}	HRF$_{HbO}$	: Cortical Structure	0.17	0.004	41.6	<0.0001	***
β_{18}	HRF$_{HbO}$	: Hemisphere	-0.43	0.004	-10.8	<0.0001	***
β_{19}	HRF'$_{HbO}$	: Masker Configuration	0.02	0.004	5.6	<0.0001	***
β_{20}	HRF'$_{HbO}$	: Cortical Structure	-0.22	0.004	-51.6	<0.0001	***
β_{21}	HRF'$_{HbO}$	: Hemisphere	-0.04	0.004	-9.5	<0.0001	***
β_{22}	HRF$_{HbR}$	: Masker Configuration	-0.12	0.004	-30.2	<0.0001	***
β_{23}	HRF$_{HbR}$	: Cortical Structure	-0.01	0.004	-1.0	0.3	
β_{24}	HRF$_{HbR}$	: Hemisphere	0.05	0.004	11.9	<0.0001	***
β_{25}	HRF'$_{HbR}$	: Masker Configuration	-0.10	0.004	-22.4	<0.0001	***
β_{26}	HRF'$_{HbR}$	: Cortical Structure	0.16	0.004	36.6	<0.0001	***
β_{27}	HRF'$_{HbR}$	: Hemisphere	0.04	0.004	9.3	<0.0001	***

All estimates are referenced to a default condition in left cIFS for Attentive.

Significance codes: ‘***’ $p < 0.001$, ‘**’ $p < 0.01$, ‘*’ $p < 0.05$, ‘.’ $p < 0.1$, ‘ ’ p 0.1.

Note: “Int” = “intercept”; “S.E.” = standard error of the mean

CHAPTER 6

SUMMARY

This thesis established fNIRS as a viable tool to objectively measure the susceptibility to IM by correlating listeners' STG activation level during target detection tasks with listeners' task performances. At the individual level, an increased STG activation is corresponding to worse task performance.

First, in the IM susceptibility vs. visual crowding susceptibility experiment, the results showed an inverse correlation between the two, suggesting a trade-off between mid-level auditory and visual processing.

Toward the goal of defining mid-level mechanisms that shape IM susceptibility, fNIRS recording from bilateral LFCx were then established as a viable bio-maker for selective auditory attention. There was a significant difference in brain activity between attentive condition vs. passive condition on the group level. Additionally, an overall more robust activation in the right hemisphere was observed when the target-masker is spatially separated configuration than that of the target-masker co-located configuration.

Third, the fNIRS data analysis protocol was improved by using LMEM to account for individual-specific effects. Results showed a correlation between tasked-evoked responses near the superior temporal gyrus (STG) and behavioral performance in a target detection task with an IM background. In contrast, task-evoked responses in the caudal inferior frontal sulcus (cIFS) did not correlate with behavioral performance, suggesting that the cIFS belongs to an IM-independent network.

Finally, the test-retest reliability of the fNIRS protocol was examined using machine learning. Results showed fNIRS recordings from attentive vs. passive state generalized well between data collected from two different listeners. Furthermore, for a robust estimate of the listener's attentive state, Six task blocks of recording are required. Two channels averaged from each ROI is enough to capture the BOLD signal reliably. Classification accuracy converges faster with more channels averaged from respective ROI.

BIBLIOGRAPHY

Aasted, C. M., Yücel, M. A., Cooper, R. J., Dubb, J., Tsuzuki, D., Becerra, L., Petkov, M. P., Borsook, D., Dan, I., & Boas, D. A. (2015). Anatomical guidance for functional near infrared spectroscopy atlasviewer tutorial. *Neurophotonics*. https://doi.org/10.1117/1.nph.2.2.020801

Ahveninen, J., Jääskeläinen, I. P., Raij, T., Bonmassar, G., Devore, S., Hämäläinen, M., Levänen, S., Lin, F.-H., Sams, M., Shinn-Cunningham, B. G., Witzel, T., & Belliveau, J. W. (2006). Task modulated what and where pathways in human auditory cortex. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.0510480103

Alain, K., C. ,. Du, Y. ,. Bernstein, L. J. ,. Barten, T. ,. &. Banai. (2018). Listening under difficult conditions: An activation likelihood estimation meta-analysis. *Human Brain Mapping*, *39*(7), 2695–2709. https://doi.org/10.1002/hbm.24031

Anderson, C. A., Wiggins, I. M., Kitterick, P. T., & Hartley, D. E. H. (2017). Adaptive benefit of cross modal plasticity following cochlear implantation in deaf adults. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.1704785114

Anderson, S., & Kraus, N. (2010). Objective Neural Indices of Speech-in-Noise Perception. *Trends in Amplification*, *14*(2), 73–83. https://doi.org/10.1177/1084713810380227

Arbogast, T. L., Mason, C. R., & Kidd, G. (2002). The effect of spatial separation on informational and energetic masking of speech. *Journal of the Acoustical Society of America*. https://doi.org/10.1121/1.1510141

Balakrishnan, U., & Freyman, R. L. (2008). Speech detection in spatial and nonspatial speech maskers. *The Journal of the Acoustical Society of America*, *123*(5), 2680–2691.

Bendor, D. (2015). The role of inhibition in a computational model of an auditory cortical neuron during the encoding of temporal information. *PLOS Comput Biol*, *11*(4), e1004197.

Benson, N. C., Yoon, J. M., Forenzo, D., Kay, K. N., Engel, S. A., & Winawer, J. (2021). Variability of the Surface Area of the V1, V2, and V3 Maps in a Large Sample of Human Observers. *BioRxiv*, 2020–12.

Bharadwaj, H. M., Khan, S., Hämäläinen, M., & Kenet, T. (2016). Electrophysiological correlates of auditory object binding with application to autism spectrum disorders. *The Journal of the Acoustical Society of America*, *140*(4), 3045–3045.

Bharadwaj, H., Mai, A., Simpson, J. M., Choi, I., Heinz, M. G., & Shinn-Cunningham, B. G. (2019). Non invasive assays of cochlear synaptopathy candidates and considerations. *Neuroscience*. https://doi.org/10.1016/j.neuroscience.2019.02.031

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where Is the Semantic System? A Critical Review and Meta-Analysis of 120 Functional Neuroimaging Studies. *Cerebral Cortex*, *19*(12), 2767–2796. https://doi.org/10.1093/cercor/bhp055

Bolia, R. S., Nelson, W. T., Ericson, M. A., & Simpson, B. D. (2000). A speech corpus for multitalker communications research. *The Journal of the Acoustical Society of America*, *107*(2), 1065–1066.

Bortfeld, H., Fava, E., & Boas, D. A. (2009). Identifying Cortical Lateralization of Speech Processing in Infants Using Near-Infrared Spectroscopy. *Developmental Neuropsychology*, *34*(1), 52–65. https://doi.org/10.1080/87565640802564481

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

Bouma, H. (1970). Interaction effects in parafoveal letter recognition. *Nature*, *226*(5241), 177–178.

Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound*. The MIT Press. https://doi.org/10.7551/mitpress/1486.001.0001

Brungart, D. S. (2001). Evaluation of speech intelligibility with the coordinate response measure. *Journal of the Acoustical Society of America*. https://doi.org/10.1121/1.1357812

Brungart, D. S. (2006). *Informational and Energetic Masking Effects in Multitalker Speech Perception*. Defense Technical Information Center. https://doi.org/10.21236/ADA456677

Brungart, D. S., Chang, P. S., Simpson, B. D., & 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*(6), 4007–4018.

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.

Brungart, D. S., Simpson, B. D., Darwin, C. J., Arbogast, T. L., & Kidd Jr, G. (2005). Across-ear interference from parametrically degraded synthetic speech signals in a dichotic cocktail-party listening task. *The Journal of the Acoustical Society of America*, *117*(1), 292–304.

Bugannim, Y., Roth, D. A.-E., Zechoval, D., & Kishon-Rabin, L. (2019). Training of Speech Perception in Noise in Pre-Lingual Hearing Impaired Adults With Cochlear Implants Compared With Normal Hearing Adults. *Otology & Neurotology*, *40*(3), e316–e325. https://doi.org/10.1097/MAO.0000000000002128

Caldwell, A., & Nittrouer, S. (2013). Speech Perception in Noise by Children With Cochlear Implants. *Journal of Speech, Language, and Hearing Research*, *56*(1), 13–30. https://doi.org/10.1044/1092-4388(2012/11-0338)

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

Carlile, S., & Corkhill, C. (2015). Selective spatial attention modulates bottom-up informational masking of speech. *Scientific Reports*, *5*(1), 8662. https://doi.org/10.1038/srep08662

Carney, L. H. (2018). Supra-threshold hearing and fluctuation profiles: Implications for sensorineural and hidden hearing loss. *Journal of the Association for Research in Otolaryngology*, *19*(4), 331–352.

Cauli, B. (2010). Revisiting the role of neurons in neurovascular coupling. *Frontiers in Neuroenergetics*, *2*. https://doi.org/10.3389/fnene.2010.00009

Chen, C.-M. A., Mathalon, D. H., Roach, B. J., Cavus, I., Spencer, D. D., & Ford, J. M. (2011). The Corollary Discharge in Humans Is Related to Synchronous Neural Oscillations. *Journal of Cognitive Neuroscience*, *23*(10), 2892–2904. https://doi.org/10.1162/jocn.2010.21589

Chen, W.-L., Wagner, J., Heugel, N., Sugar, J., Lee, Y.-W., & Conant, L. (2020). Functional Near-Infrared Spectroscopy and Its Clinical Application in the Field of Neuroscience: Advances and Future Directions. *Frontiers in Neuroscience*, *14*. https://doi.org/10.3389/fnins.2020.00724

Cherry, E. C. (1953). Some Experiments on the Recognition of Speech, with One and with Two Ears. *The Journal of the Acoustical Society of America*, *25*(5), 975–979. https://doi.org/10.1121/1.1907229

Ching, T. Y., Zhang, V. W., Flynn, C., Burns, L., Button, L., Hou, S., McGhie, K., & Van Buynder, P. (2018). Factors influencing speech perception in noise for 5-year-old children using hearing aids or cochlear implants. *International Journal of Audiology*, *57*(sup2), S70–S80. https://doi.org/10.1080/14992027.2017.1346307

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*. https://doi.org/10.3389/fnhum.2013.00115

Choi, I., Wang, L., Bharadwaj, H., & Shinn-Cunningham, B. (2014). Individual differences in attentional modulation of cortical responses correlate with selective attention performance. *Hearing Research*, *314*, 10–19.

Chubb, C., Dickson, C. A., Dean, T., Fagan, C., Mann, D. S., Wright, C. E., Guan, M., Silva, A. E., Gregersen, P. K., & Kowalsky, E. (2013). Bimodal distribution of performance in discriminating major/minor modes. *The Journal of the Acoustical Society of America*, *134*(4), 3067–3078.

Clayton, K. K., Swaminathan, J., Yazdanbakhsh, A., Zuk, J., Patel, A. D., & Kidd Jr, G. (2016). Executive function, visual attention and the cocktail party problem in musicians and non-musicians. *PloS One*, *11*(7), e0157638.

Cooke, M. (2006). A glimpsing model of speech perception in noise. *The Journal of the Acoustical Society of America*, *119*(3), 1562–1573.

Cooke, M., Lecumberri, M. L. G., & Barker, J. (2008). The foreign language cocktail party problem energetic and informational masking effects in non native speech perception. *Journal of the Acoustical Society of America*. https://doi.org/10.1121/1.2804952

Cui, X., Bray, S., Bryant, D. M., Glover, G. H., & Reiss, A. L. (2011). A quantitative comparison of NIRS and fMRI across multiple cognitive tasks. *NeuroImage*, *54*(4), 2808–2821. https://doi.org/10.1016/j.neuroimage.2010.10.069

Dajani, H. R., & Picton, T. W. (2006). Human auditory steady state responses to changes in interaural correlation. *Hearing Research*. https://doi.org/10.1016/j.heares.2006.06.003

Darwin, C., & Hukin, R. (1999). Auditory objects of attention: The role of interaural time differences. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(3), 617.

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. *Journal of the Acoustical Society of America*. https://doi.org/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. *Journal of the Acoustical Society of America*. https://doi.org/10.1121/1.419641

Defenderfer, J., Kerr-German, A., Hedrick, M., & Buss, A. T. (2017). Investigating the role of temporal lobe activation in speech perception accuracy with normal hearing adults: An event-related fNIRS study. *Neuropsychologia*, *106*, 31–41. https://doi.org/10.1016/j.neuropsychologia.2017.09.004

Delpy, D. T., Cope, M., Zee, P. van, Arridge, S., Wray, S., & Wyatt, J. (1988). Estimation of optical pathlength through tissue from direct time of flight measurement. *Physics in Medicine and Biology*, *33*(12), 1433–1442. https://doi.org/10.1088/0031-9155/33/12/008

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

Du, Y., Kong, L., Wang, Q., Wu, X., & Li, L. (2011). Auditory frequency-following response: A neurophysiological measure for studying the “cocktail-party problem.” Neuroscience & Biobehavioral Reviews. *Neuroscience & Biobehavioral Reviews*, *35*(10), 2046–2057. https://doi.org/10.1016/j.neubiorev.2011.05.008

Dubno, J. R., Dirks, D. D., & Morgan, D. E. (1984). Effects of age and mild hearing loss on speech recognition in noise. *The Journal of the Acoustical Society of America*, *76*(1), 87–96.

Eisenberg, L. S., Fisher, L. M., Johnson, K. C., Ganguly, D. H., Grace, T., & Niparko, J. K. (2016). Sentence Recognition in Quiet and Noise by Pediatric Cochlear Implant Users: Relationships to Spoken Language. *Otology & Neurotology*, *37*(2), e75–e81. https://doi.org/10.1097/MAO.0000000000000910

Ferrari, M., & Quaresima, V. (2012). A brief review on the history of human functional near infrared spectroscopy fnirs development and fields of application. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2012.03.049

Flinker, A., Chang, E. F., Kirsch, H. E., Barbaro, N. M., Crone, N. E., & Knight, R. T. (2010). Single-Trial Speech Suppression of Auditory Cortex Activity in Humans. *Journal of Neuroscience*, *30*(49), 16643–16650. https://doi.org/10.1523/JNEUROSCI.1809-10.2010

Flom, M. C., Heath, G. G., & Takahashi, E. (1963). Contour interaction and visual resolution: Contralateral effects. *Science*, *142*(3594), 979–980.

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

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*(5), 2112–2122.

Freyman, R. L., Balakrishnan, U., & Helfer, K. S. (2004). Effect of number of masking talkers and auditory priming on informational masking in speech recognition. *The Journal of the Acoustical Society of America*, *115*(5), 2246–2256. https://doi.org/10.1121/1.1689343

Freyman, R. L., Helfer, K. S., McCall, D. D., & Clifton, R. K. (1999). The role of perceived spatial separation in the unmasking of speech. *The Journal of the Acoustical Society of America*, *106*(6), 3578–3588.

Funane, T., Atsumori, H., Katura, T., Obata, A. N., Sato, H., & Tanikawa, Y. (n.d.). Quantitative evaluation of deep and shallow tissue layers' contribution to fNIRS signal using multi-distance optodes and independent component analysis. *NeuroImage*, *85*, 150–165. https://doi.org/10.1016/j.neuroimage.2013.02.026

Furman, A. C., Kujawa, S. G., & Liberman, M. C. (2013). Noise-induced cochlear neuropathy is selective for fibers with low spontaneous rates. *Journal of Neurophysiology*, *110*(3), 577–586. https://doi.org/10.1152/jn.00164.2013

Gagnon, L., Cooper, R. J., Yücel, M. A., Perdue, K. L., Greve, D. N., & Boas, D. A. (2012). Short separation channel location impacts the performance of short channel regression in NIRS. *NeuroImage*, *59*(3), 2518–2528. https://doi.org/10.1016/j.neuroimage.2011.08.095

Gallun, F. J., Mason, C. R., & Kidd, G. (2005). Binaural release from informational masking in a speech identification task. *Journal of the Acoustical Society of America*. https://doi.org/10.1121/1.1984876

Gates, G. A. (1995). Cochlear Implants in Adults and Children. *JAMA: The Journal of the American Medical Association*, *274*(24), 1955. https://doi.org/10.1001/jama.1995.03530240065043

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. *Journal of the Acoustical Society of America*. https://doi.org/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.

Goodwin, J. R., Gaudet, C. R., & Berger, A. J. (2014). Short-channel functional near-infrared spectroscopy regressions improve when source-detector separation is reduced. *Neurophotonics*, *1*(1), 015002. https://doi.org/10.1117/1.nph.1.1.015002

Graves, J. E., & Oxenham, A. J. (2019). Pitch discrimination with mixtures of three concurrent harmonic complexes. *The Journal of the Acoustical Society of America*, *145*(4), 2072–2083.

Greenwood, D. D. (1990a). A cochlear frequency position function for several species 29 years later. *Journal of the Acoustical Society of America*. https://doi.org/10.1121/1.399052

Greenwood, D. D. (1990b). A cochlear frequency-position function for several species—29 years later. *The Journal of the Acoustical Society of America*, *87*(6), 2592–2605.

Haeussinger, F. B., Heinzel, S., Hahn, T., Schecklmann, M., Ehlis, A.-C., & Fallgatter, A. J. (2011). Simulation of Near-Infrared Light Absorption Considering Individual Head and Prefrontal Cortex Anatomy: Implications for Optical Neuroimaging. *PLoS ONE*, *6*, 10. https://doi.org/10.1371/journal.pone.0026377

Hardy, C. J., Yong, K. X., Goll, J. C., Crutch, S. J., & Warren, J. D. (2020). Impairments of auditory scene analysis in posterior cortical atrophy. *Brain*, *143*(9), 2689–2695.

Heinz, M. G., Zhang, X., Bruce, I. C., & Carney, L. H. (2001). Auditory nerve model for predicting performance limits of normal and impaired listeners. *Acoustics Research Letters Online*, *2*(3), 91–96. https://doi.org/10.1121/1.1387155

Heinzel, S., Haeussinger, F. B., Hahn, T., Ehlis, A.-C., Plichta, M. M., & Fallgatter, A. J. (2013). Variability of (functional) hemodynamics as measured with simultaneous fNIRS and fMRI during intertemporal choice. *NeuroImage*, *71*, 125–134. https://doi.org/10.1016/j.neuroimage.2012.12.074

Hejna, D., & Musicus, B. R. (1991). The SOLAFS time-scale modification algorithm. *Bolt, Beranek and Newman (BBN) Technical Report*.

Helfer, K. S., & Freyman, R. L. (2008). Aging and speech-on-speech masking. *Ear and Hearing*, *29*(1), 87.

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, *8*(5), 393–402.

Himmelberg, M. M., Kurzawaski, J. W., Benson, N. C., Pelli, D. G., Carrasco, M., & Winawer, J. (2021). *Cross-dataset reproducibility of population receptive field (pRF) estimates and retinotopic map structure*.

Hind, S. E., Haines-Bazrafshan, R., Benton, C. L., Brassington, W., Towle, B., & Moore, D. R. (2011). Prevalence of clinical referrals having hearing thresholds within normal limits. *International Journal of Audiology*, *50*(10), 708–716. https://doi.org/10.3109/14992027.2011.582049

Hochberg, I., Boothroyd, A., Weiss, M., & Hellman, S. (1992). Effects of Noise and Noise Suppression on Speech Perception by Cochlear Implant Users. *Ear and Hearing*, *13*(4), 263–271. https://doi.org/10.1097/00003446-199208000-00008

Hoddinott, J., Alderman, H., Behrman, J. R., Haddad, L., & Horton, S. (2013). The economic rationale for investing in stunting reduction. *Maternal & Child Nutrition*, *9*, 69–82.

Huneau, C., Benali, H., & Chabriat, H. (2015). Investigating Human Neurovascular Coupling Using Functional Neuroimaging: A Critical Review of Dynamic Models. *Frontiers in Neuroscience*, *9*. https://doi.org/10.3389/fnins.2015.00467

Huppert, T. J., Hoge, R. D., Diamond, S. G., Franceschini, M. A., & Boas, D. A. (2006). A temporal comparison of BOLD, ASL, and NIRS hemodynamic responses to motor stimuli in adult humans. *NeuroImage*, *29*(2), 368–382. https://doi.org/10.1016/j.neuroimage.2005.08.065

Hussain, Z., Webb, B. S., Astle, A. T., & McGraw, P. V. (2012). Perceptual learning reduces crowding in amblyopia and in the normal periphery. *Journal of Neuroscience*, *32*(2), 474–480.

Ihlefeld, A., & Shinn-Cunningham, B. (2008). 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.

Jepsen, M. L., Ewert, S. D., & Dau, T. (2008). A computational model of human auditory signal processing and perception. *The Journal of the Acoustical Society of America*, *124*(1), 422–438. https://doi.org/10.1121/1.2924135

Kerlin, J. R., Shahin, A. J., & Miller, L. M. (2010a). Attentional gain control of ongoing cortical speech representations in a “cocktail party.” *Journal of Neuroscience*, *30*(2), 620–628.

Kerlin, J. R., Shahin, A. J., & Miller, L. M. (2010b). Attentional gain control of ongoing cortical speech representations in a cocktail party. *The Journal of Neuroscience*. https://doi.org/10.1523/jneurosci.3631-09.2010

Khalighinejad, B., Herrero, J. L., Mehta, A. D., & Mesgarani, N. (2019). Adaptation of the human auditory cortex to changing background noise. *Nature Communications*, *10*(1), 1–11.

Kidd, G., & Colburn, H. S. (2017). Informational masking in speech recognition. *Null*. https://doi.org/10.1007/978-3-319-51662-2_4

Kidd, G., Mason, C. R., Best, V., & Marrone, N. L. (2010). Stimulus factors influencing spatial release from speech on speech masking. *Journal of the Acoustical Society of America*. https://doi.org/10.1121/1.3478781

Kidd, G., Mason, C. R., Richards, V. M., Gallun, F. J., & Durlach, N. I. (2008). Informational Masking. In W. A. Yost, A. N. Popper, & R. R. Fay (Eds.), *Auditory Perception of Sound Sources* (Vol. 29, pp. 143–189). Springer US. https://doi.org/10.1007/978-0-387-71305-2_6

Kidd, G. R., Watson, C. S., & Gygi, B. (2007). Individual differences in auditory abilities. *The Journal of the Acoustical Society of America*, *122*(1), 418–435.

Kidd Jr, 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.

Kidd Jr, G., Streeter, T. M., Ihlefeld, A., Maddox, R. K., & Mason, C. R. (2009). The intelligibility of pointillistic speech. *The Journal of the Acoustical Society of America*, *126*(6), EL196–EL201.

Kiefer, J., Gall, V., Desloovere, C., Knecht, R., Mikowski, A., & von Ilberg, C. (1996). A follow-up study of long-term results after cochlear implantation in children and adolescents. *European Archives of Oto-Rhino-Laryngology*, *253*(3), 158–166. https://doi.org/10.1007/bf00615114

Kirilina, E., Jelzow, A., Heine, A., Niessing, M., Wabnitz, H., Brühl, R., Ittermann, B., Jacobs, A. M., & Tachtsidis, I. (2012). The physiological origin of task-evoked systemic artefacts in functional near infrared spectroscopy. *NeuroImage*, *61*(1), 70–81. https://doi.org/10.1016/j.neuroimage.2012.02.074

Klaessens, J. H., Hopman, J. C., Liem, K. D., van Os, S. H., & Thijssen, J. M. (2005). Effects of skin on bias and reproducibility of near-infrared spectroscopy measurement of cerebral oxygenation changes in porcine brain. *Journal of Biomedical Optics*, *10*(4), 044003. https://doi.org/10.1117/1.1989315

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*. https://doi.org/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. https://doi.org/10.1088/0031-9155/51/5/N02

Kong, L., Michalka, S. W., Rosen, M. L., Sheremata, S. L., Swisher, J. D., Shinn-Cunningham, B. G., & Somers, D. C. (2014). Auditory Spatial Attention Representations in the Human Cerebral Cortex. *Cerebral Cortex*, *24*(3), 773–784. https://doi.org/10.1093/cercor/bhs359

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

Kujawa, S. G., & Liberman, M. C. (2009). Adding Insult to Injury: Cochlear Nerve Degeneration after "Temporary" Noise-Induced Hearing Loss. *Journal of Neuroscience*, *29*(45), 14077–14085. https://doi.org/10.1523/JNEUROSCI.2845-09.2009

Kurzawaski, J. W., and Pelli, D. G., & Winawer, J. (2021). *Conservation across individuals of cortical crowding distance in human V4*.

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

Lawrence, R. J., Wiggins, I. M., Anderson, C. A., Davies-Thompson, J., & Hartley, D. E. (2018). Cortical correlates of speech intelligibility measured using functional near-infrared spectroscopy (fNIRS). *Hearing Research*, *370*, 53–64.

Leek, M. R., Brown, M. E., & Dorman, M. F. (1991). Informational masking and auditory attention. *Perception & Psychophysics*, *50*(3), 205–214. https://doi.org/10.3758/BF03206743

Lesenfants, D., Vanthornhout, J., Verschueren, E., Decruy, L., & Francart, T. (2019). Predicting individual speech intelligibility from the cortical tracking of acoustic- and phonetic-level speech representations. *Hearing Research*, *380*, 1–9.

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *The Journal of the Acoustical Society of America*, *49*(2B), 467–477.

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.

Loizou, P. C., Hu, Y., Litovsky, R., Yu, G., Peters, R., Lake, J., & Roland, P. (2009). Speech recognition by bilateral cochlear implant users in a cocktail-party setting. *The Journal of the Acoustical Society of America*, *125*(1), 372–383. https://doi.org/10.1121/1.3036175

Lutfi, R. A. (1988). Informational processing of complex sound. *The Journal of the Acoustical Society of America*, *84*(S1), S142–S143. https://doi.org/10.1121/1.2025830

Lutfi, R. A., Gilbertson, L., Heo, I., Chang, A.-C., & Stamas, J. (2013). The information-divergence hypothesis of informational masking. *The Journal of the Acoustical Society of America*, *134*(3), 2160–2170.

Mackersie, C., & Calderon-Moultrie, N. (2016). Autonomic Nervous System Reactivity During Speech Repetition Tasks: Heart Rate Variability and Skin Conductance. *Ear and Hearing*. https://doi.org/10.1097/AUD.0000000000000305

Mackersie, C. L., & Cones, H. (2011). Subjective and Psychophysiological Indexes of Listening Effort in a Competing-Talker Task. *Journal of the American Academy of Audiology*, *22*(02), 113–122. https://doi.org/10.3766/jaaa.22.2.6

Maggioni, E., Molteni, E., Zucca, C., Reni, G., Cerutti, S., & Triulzi, F. M. (2015). Investigation of negative BOLD responses in human brain through NIRS technique. A visual stimulation study. *NeuroImage*, *108*, 410–422. https://doi.org/10.1016/j.neuroimage.2014.12.074

Malonek, D., Dirnagl, U., Lindauer, U., Yamada, K., Kanno, I., & Grinvald, A. (1997). Vascular imprints of neuronal activity: Relationships between the dynamics of cortical blood flow, oxygenation, and volume changes following sensory stimulation. *Proceedings of the National Academy of Sciences*, *94*(26), 14826–14831. https://doi.org/10.1073/pnas.94.26.14826

Marrone, N. L., Mason, C. R., & Kidd, G. (2008). Evaluating the benefit of hearing aids in solving the cocktail party problem. *Trends in Amplification*. https://doi.org/10.1177/1084713808325880

Mattys, S. L., Brooks, J. L., & Cooke, M. (2009). Recognizing speech under a processing load dissociating energetic from informational factors. *Cognitive Psychology*. https://doi.org/10.1016/j.cogpsych.2009.04.001

McJury, M., & Shellock, F. G. (2000). Auditory noise associated with MR procedures: A review. *Journal of Magnetic Resonance Imaging: JMRI*, *12*(1), 37–45. https://doi.org/10.1002/1522-2586(200007)12:1<37::aid-jmri5>3.0.co;2-i

Mednicoff, S., Mejia, S., Rashid, J. A., & Chubb, C. (2018). Many listeners cannot discriminate major vs minor tone-scrambles regardless of presentation rate. *The Journal of the Acoustical Society of America*, *144*(4), 2242–2255.

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

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*. https://doi.org/10.1037/a0017601

Micheyl, C., Kreft, H., Shamma, S., & Oxenham, A. J. (2013). Temporal coherence versus harmonicity in auditory stream formation. *The Journal of the Acoustical Society of America*, *133*(3), EL188–EL194. https://doi.org/10.1121/1.4789866

Minati, L., Kress, I. U., Visani, E., Medford, N., & Critchley, H. D. (2011). Intra-and extra-cranial effects of transient blood pressure changes on brain near-infrared spectroscopy (NIRS) measurements. *Journal of Neuroscience Methods*, *197*(2), 283–288.

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

Müller-Deile, J., Schmidt, B. J., & Rudert, H. (1995). Effects of noise on speech discrimination in cochlear implant patients. *Laryngo-Rhino-Otologie*, *73*, 06. https://doi.org/10.1055/s-2007-997136

Neff, D. L. (1995). Signal properties that reduce masking by simultaneous, random-frequency maskers. *The Journal of the Acoustical Society of America*, *98*(4), 1909–1920.

Neff, D. L., & Callaghan, B. P. (1988). Effective properties of multicomponent simultaneous maskers under conditions of uncertainty. *The Journal of the Acoustical Society of America*, *83*(5), 1833–1838.

Neff, D. L., Dethlefs, T. M., & Jesteadt, W. (1993). Informational masking for multicomponent maskers with spectral gaps. *Journal of the Acoustical Society of America*. https://doi.org/10.1121/1.407217

Nelson, P. B., & Jin, S.-H. (2002). Understanding speech in single-talker interference: Normal-hearing listeners and cochlear implant users. *The Journal of the Acoustical Society of America*, *111*(5), 2429. https://doi.org/10.1121/1.4778318

Nemoto, M., Nomura, Y., Tamura, M., Sato, C., Houkin, K., & Abe, H. (1997). Optical Imaging and Measuring of Local Hemoglobin Concentration and Oxygenation Changes During Somatosensory Stimulation in Rat Cerebral Cortex. *Advances in Experimental Medicine and Biology*, *521–531*. https://doi.org/10.1007/978-1-4615-5399-1_74

Niioka, K., Uga, M., Nagata, T., Tokuda, T., Dan, I., & Ochi, K. (2018). Cerebral hemodynamic response during concealment of information about a mock crime: Application of a general linear model with an adaptive hemodynamic response function. *Japanese Psychological Research*, *60*(4), 311–326.

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. *Journal of Neuroscience*, *37*(36), 8755–8766.

Obleser, J., Wise, R. J., Dresner, M. A., & Scott, S. K. (2007). Functional integration across brain regions improves speech perception under adverse listening conditions. *Journal of Neuroscience*, *27*(9), 2283–2289.

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

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.

Patterson, R. D. (1976). Auditory filter shapes derived with noise stimuli. *The Journal of the Acoustical Society of America*, *59*(3), 640–654.

Patterson, R. D., Nimmo-Smith, I., Weber, D. L., & Milroy, R. (1982). The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram, and speech threshold. *The Journal of the Acoustical Society of America*, *72*(6), 1788–1803.

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

Pelli, D. G. (2008). Crowding: A cortical constraint on object recognition. *Current Opinion in Neurobiology*, *18*(4), 445–451.

Pelli, D. G., Palomares, M., & Majaj, N. J. (2004). Crowding is unlike ordinary masking distinguishing feature integration from detection. *Journal of Vision*. https://doi.org/10.1167/4.12.12

Pelli, D. G., Waugh, S. J., Martelli, M., Crutch, S. J., Primativo, S., Yong, K. X., Rhodes, M., Yee, K., Wu, X., Famira, H. F., & others. (2016). A clinical test for visual crowding. *F1000Research*.

Perez, R., M. ,. &. Macias. (n.d.). Criteria of Candidacy for Unilateral Cochlear Implantation in Postlingually Deafened Adults III: Prospective Evaluation of an Actuarial Approach to Defining a Criterion. *Ear and Hearing*, *25*(4), 361–374. https://doi.org/10.1097/01.aud.0000134551.13162.88

Pinti, P., Scholkmann, F., Hamilton, A., Burgess, P., & Tachtsidis, I. (2019). Current Status and Issues Regarding Pre-processing of fNIRS Neuroimaging Data: An Investigation of Diverse Signal Filtering Methods Within a General Linear Model Framework. *Frontiers in Human Neuroscience*, *12*. https://doi.org/10.3389/fnhum.2018.00505

Piquado, T., Isaacowitz, D., & Wingfield, A. (2010). Pupillometry as a measure of cognitive effort in younger and older adults. *Psychophysiology*, *47*(3), 560–569. https://doi.org/10.1111/j.1469-8986.2009.00947.x

Pollack, I. (1975). Auditory informational masking. *The Journal of the Acoustical Society of America*, *57*(S1), S5–S5.

Pugh, K. R., Shaywitz, B. A., Shaywitz, S. E., Fulbright, R. K., Byrd, D., Skudlarski, P., Shankweiler, D. P., Katz, L., Constable, R. T., Fletcher, J., Lacadie, C., Marchione, K., & Gore, J. C. (1996). Auditory Selective Attention: An fMRI Investigation. *NeuroImage*, *4*(3), 159–173. https://doi.org/10.1006/nimg.1996.0067

Reeves, A., Amano, K., & Foster, D. H. (2004). Color constancy stimulus or task. *Journal of Vision*. https://doi.org/10.1167/4.11.12

Riecke, L., Peters, J. C., Valente, G., Kemper, V. G., Formisano, E., & Sorger, B. (2016). *Frequency-Selective Attention in Auditory Scenes Recruits Frequency Representations Throughout Human Superior Temporal Cortex*. Cerebral Cortex. https://doi.org/10.1093/cercor/bhw160

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

Rosen, Sarah, Chakravarthi, R., & Pelli, D. G. (2014). The Bouma law of crowding, revised: Critical spacing is equal across parts, not objects. *Journal of Vision*, *14*(6), 10–10.

Rosen, Stuart, Cohen, M., & Vanniasegaram, I. (2010). Auditory and cognitive abilities of children suspected of auditory processing disorder (APD). *International Journal of Pediatric Otorhinolaryngology*, *74*, 594–600. https://doi.org/10.1016/j.ijporl.2010.02.021

Rovetti, J., Goy, H., Pichora-Fuller, M. K., & Russo, F. A. (2019). Functional Near-Infrared Spectroscopy as a Measure of Listening Effort in Older Adults Who Use Hearing Aids. *Trends in Hearing*, *23*(23312), 1651988672. https://doi.org/10.1177/2331216519886722

Rowland, S. C., Hartley, D. E., & Wiggins, I. M. (2018). Listening in naturalistic scenes: What can functional near-infrared spectroscopy and intersubject correlation analysis tell us about the underlying brain activity? *Trends in Hearing*, *22*, 2331216518804116.

Salo, E., Salmela, V., Salmi, J., Numminen, J., & Alho, K. (2017). Brain activity associated with selective attention, divided attention and distraction. *Brain Research*, *1664*, 25–36.

Sato, H., Kiguchi, M., Kawaguchi, F., & Maki, A. (2004). Practicality of wavelength selection to improve signal-to-noise ratio in near-infrared spectroscopy. *Neuroimage*, *21*(4), 1554–1562.

Sato, T., Nambu, I., Takeda, K., Aihara, T., Yamashita, O., & Isogaya, Y. (2016). Reduction of global interference of scalp-hemodynamics in functional near-infrared spectroscopy using short distance probes. *NeuroImage*, *141*, 120–132. https://doi.org/10.1016/j.neuroimage.2016.06.054

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.

Scott, S. K., Rosen, S., Lang, H., & Wise, R. J. S. (2006). Neural correlates of intelligibility in speech investigated with noise vocoded speech a positron emission tomography study. *Journal of the Acoustical Society of America*. https://doi.org/10.1121/1.2216725

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.

Shinn-Cunningham, B. (2017). Cortical and Sensory Causes of Individual Differences in Selective Attention Ability Among Listeners With Normal Hearing Thresholds. *Journal of Speech, Language, and Hearing Research*, *60*(10), 2976–2988. https://doi.org/10.1044/2017_JSLHR-H-17-0080

Shinn-Cunningham, B. G., Ihlefeld, A., Satyavarta, S., & Larson, E. (2005). Bottom-up and Top-down Influences on Spatial Unmasking. *Acta Acustica United with Acustica*, *91*(6), 967–979.

Shinn-Cunningham, Barbara G. (2008). Object-based auditory and visual attention. *Trends in Cognitive Sciences*, *12*(5), 182–186. https://doi.org/10.1016/j.tics.2008.02.003

Shinn-Cunningham, Barbara G., & Best, V. (2008). Selective attention in normal and impaired hearing. *Trends in Amplification*. https://doi.org/10.1177/1084713808325306

Shomstein, S., & Yantis, S. (2006). Parietal cortex mediates voluntary control of spatial and nonspatial auditory attention. *The Journal of Neuroscience*. https://doi.org/10.1523/jneurosci.4408-05.2006

Singh, G., Pichora-Fuller, M. K., & Schneider, B. A. (2008). The effect of age on auditory spatial attention in conditions of real and simulated spatial separation. *The Journal of the Acoustical Society of America*, *124*(2), 1294–1305.

Somers, B., Long, C. J., & Francart, T. (2020). EEG-based diagnostics of the auditory system using cochlear implant electrodes as sensors. *BioRxiv*.

Song, C., Schwarzkopf, D. S., Kanai, R., & Rees, G. (2011). Reciprocal anatomical relationship between primary sensory and prefrontal cortices in the human brain. *Journal of Neuroscience*, *31*(26), 9472–9480.

Song, S., Levi, D. M., & Pelli, D. G. (2014). A double dissociation of the acuity and crowding limits to letter identification, and the promise of improved visual screening. *Journal of Vision*, *14*(5), 3–3.

Stefanovic, B., Warnking, J. M., & Pike, G. B. (2004). Hemodynamic and metabolic responses to neuronal inhibition. *Neuroimage*, *22*(2), 771–778.

Tachtsidis, I., Leung, T. S., Chopra, A., Koh, P. H., Reid, C. B., & Elwell, C. E. (2009). False Positives In Functional Nearinfrared Topography. *Advances in Experimental Medicine and Biology*, 307–314. https://doi.org/10.1007/978-0-387-85998-9_46

Takahashi, T., Takikawa, Y., Kawagoe, R., Shibuya, S., Iwano, T., & Kitazawa, S. (2011). Influence of skin blood flow on near-infrared spectroscopy signals measured on the forehead during a verbal fluency task. *NeuroImage*, *57*(3), 991–1002. https://doi.org/10.1016/j.neuroimage.2011.05.012

Taylor, S., & Brown, D. (1972). Lateral visual masking: Supraretinal effects when viewing linear arrays with unlimited viewing time. *Perception & Psychophysics*, *12*(1), 97–99.

Thomason, M. E., Foland, L. C., & Glover, G. H. (2007). Calibration of bold fmri using breath holding reduces group variance during a cognitive task. *Human Brain Mapping*. https://doi.org/10.1002/hbm.20241

Tobyne, S. M., Osher, D. E., Michalka, S. W., & Somers, D. C. (2017). Sensory-biased attention networks in human lateral frontal cortex revealed by intrinsic functional connectivity. *NeuroImage*, *162*, 362–372. https://doi.org/10.1016/j.neuroimage.2017.08.020

Tong, Y., Hocke, L. M., & Frederick, B. deB. (2011). Isolating the sources of widespread physiological fluctuations in functional near-infrared spectroscopy signals. *Journal of Biomedical Optics*, *16*(10), 106005. https://doi.org/10.1117/1.3638128

Tripathy, S. P., & Levi, D. M. (1994). Long-range dichoptic interactions in the human visual cortex in the region corresponding to the blind spot. *Vision Research*, *34*(9), 1127–1138.

Undurraga, J., Haywood, N. R., Marquardt, T., & McAlpine, D. (2016). Neural representation of interaural time differences in humans an objective measure that matches behavioural performance. *Jaro-Journal of The Association for Research in Otolaryngology*. https://doi.org/10.1007/s10162-016-0584-6

Vazquez, A. L., Fukuda, M., & Kim, S.-G. (2018). Inhibitory neuron activity contributions to hemodynamic responses and metabolic load examined using an inhibitory optogenetic mouse model. *Cerebral Cortex*, *28*(11), 4105–4119.

Villringer, A., & Chance, B. (1997). Non-invasive optical spectroscopy and imaging of human brain function. *Trends in Neurosciences*, *20*(10), 435–442. https://doi.org/10.1016/S0166-2236(97)01132-6

Viswanathan, V., Bharadwaj, H. M., & Shinn-Cunningham, B. G. (2019). Electroencephalographic signatures of the neural representation of speech during selective attention. *Eneuro*, *6*(5).

Wack, D. S., Cox, J. L., Schirda, C., Magnano, C., Sussman, J. E., Henderson, D., & Burkard, R. (2012). Functional anatomy of the masking level difference an fmri study. *PLOS ONE*. https://doi.org/10.1371/journal.pone.0041263

Watson, A. B., & Pelli, D. G. (1983). QUEST: A Bayesian adaptive psychometric method. *Perception & Psychophysics*, *33*(2), 113–120.

Watson, C. S. (2005). Some Comments on Informational Masking. *Acta Acustica United with Acustica*, *91*(3), 502–512.

Watson, Charles S., & Kelly, W. J. (2017). The Role of Stimulus Uncertainty in the Discrimination of Auditory Patterns. In D. J. Getty & J. H. Howard (Eds.), *Auditory and Visual Pattern Recognition* (1st ed., pp. 37–59). Routledge. https://doi.org/10.4324/9781315532615-3

Weder, S., Shoushtarian, M., Olivares, V., Zhou, X., Innes-Brown, H., & McKay, C. (2020). Cortical fNIRS Responses Can Be Better Explained by Loudness Percept than Sound Intensity. *Ear & Hearing*, *41*(5), 1187–1195. https://doi.org/10.1097/aud.0000000000000836

Weder, S., Zhou, X., Shoushtarian, M., Innes-Brown, H., & McKay, C. (2018). Cortical Processing Related to Intensity of a Modulated Noise Stimulus—A Functional Near-Infrared Study. *Journal of the Association for Research in Otolaryngology*, *19*(3), 273–286. https://doi.org/10.1007/s10162-018-0661-0

Whitney, D., & Levi, D. M. (2011). Visual crowding: A fundamental limit on conscious perception and object recognition. *Trends in Cognitive Sciences*, *15*(4), 160–168. https://doi.org/10.1016/j.tics.2011.02.005

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

Wiggins, Ian M., Wijayasiri, P., & Hartley, D. E. H. (2016). Shining a light on the neural signature of effortful listening. *Journal of the Acoustical Society of America*. https://doi.org/10.1121/1.4950150

Wilson, B. S., & Dorman, M. F. (2008). Cochlear implants: A remarkable past and a brilliant future. *Hearing Research*, *242*(1–2), 3–21. https://doi.org/10.1016/j.heares.2008.06.005

Wong, W. Y. S., & 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*. https://doi.org/10.1097/01.aud.0000111257.11898.64

Wöstmann, M., Herrmann, B., Maess, B., & Obleser, J. (2016). Spatiotemporal dynamics of auditory attention synchronize with speech. *Proceedings of the National Academy of Sciences*, *113*(14), 3873–3878. https://doi.org/10.1073/pnas.1523357113

Xia, J., Kalluri, S., Micheyl, C., & Hafter, E. R. (2017). Continued search for better prediction of aided speech understanding in multi talker environments. *Journal of the Acoustical Society of America*. https://doi.org/10.1121/1.5008498

Zan, P., Presacco, A., Anderson, S., & Simon, J. Z. (2019). Exaggerated Cortical Representation of Speech in Older Listeners: Mutual Information Analysis. *BioRxiv*.

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

Zekveld, A. A., & Kramer, S. E. (2014). Cognitive processing load across a wide range of listening conditions: Insights from pupillometry. *Psychophysiology*, *51*(3), 277–284.

Zhang, M., Alamatsaz, N., & Ihlefeld, A. (2020). Hemodynamic responses link individual differences in informational masking to the vicinity of superior temporal gyrus. *BioRxiv*. https://doi.org/10.1101/2020.08.21.261222

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

Zhang, X., & Gong, Q. (2019). Frequency-Following Responses to Complex Tones at Different Frequencies Reflect Different Source Configurations. *Frontiers in Neuroscience*, *13*, 130. https://doi.org/10.3389/fnins.2019.00130

Zhou, X., Sobczak, G., McKay, C. M., & Litovsky, R. Y. (2020). Comparing fNIRS signal qualities between approaches with and without short channels. *PLOS ONE, 15*, 12. https://doi.org/10.1371/journal.pone.0244186

Zhou, Xin, Seghouane, A.-K., Shah, A., Innes-Brown, H., Cross, W., Litovsky, R., & McKay, C. M. (2018). Cortical speech processing in postlingually deaf adult cochlear implant users, as revealed by functional near-infrared spectroscopy. *Trends in Hearing*, *22*, 2331216518786850.

Zouridakis, G., Simos, P. G., & Papanicolaou, A. C. (1998). Multiple bilaterally asymmetric cortical sources account for the auditory N1m component. *Brain Topography*, *10*(3), 183–189. https://doi.org/10.1023/a:1022246825461

(2021). personal communication.