Neuropsychologia 106 (2017) 31–41

Contents lists available at ScienceDirect

Neuropsychologia

journal homepage: www.elsevier.com/locate/neuropsychologia

Investigating the role of temporal lobe activation in speech perception accuracy with normal hearing adults: An event-related fNIRS study

Jessica Defenderfer[a,*], Anastasia Kerr-German[b], Mark Hedrick[a], Aaron T. Buss[b]

[a] University of Tennessee Health Sciences Center, Department of Speech Pathology and Audiology, United States
[b] University of Tennessee, Department of Psychology, United States

ARTICLE INFO

Keywords:
Speech perception
Hearing science
Event-related design
fNIRS

ABSTRACT

Functional near infrared spectroscopy (fNIRS) is a safe, non-invasive, relatively quiet imaging technique that is tolerant of movement artifact making it uniquely ideal for the assessment of hearing mechanisms. Previous research demonstrates the capacity for fNIRS to detect cortical changes to varying speech intelligibility, revealing a positive relationship between cortical activation amplitude and speech perception score. In the present study, we use an event-related design to investigate the hemodynamic response in the temporal lobe across different listening conditions. We presented participants with a speech recognition task using sentences in quiet, sentences in noise, and vocoded sentences. Hemodynamic responses were examined across conditions and then compared when speech perception was accurate compared to when speech perception was inaccurate in the context of noisy speech. Repeated measures, two-way ANOVAs revealed that the speech in noise condition (−2.8 dB signal-to-noise ratio/SNR) demonstrated significantly greater activation than the easier listening conditions on multiple channels bilaterally. Further analyses comparing correct recognition trials to incorrect recognition trials (during the presentation phase of the trial) revealed that activation was significantly greater during correct trials. Lastly, during the repetition phase of the trial, where participants correctly repeated the sentence, the hemodynamic response demonstrated significantly higher deoxyhemoglobin than oxyhemoglobin, indicating a difference between the effects of perception and production on the cortical response. Using fNIRS, the present study adds meaningful evidence to the body of knowledge that describes the brain/behavior relationship related to speech perception.

1. Introduction

Spoken communication is most often challenged by background noise or other acoustic interference. Speech in noise (SIN) perception ability is influenced by a combination of peripheral, subcortical, and cortical processing (Bidelman and Bhagat, 2015; Billings et al., 2009; Kaplan-Neeman et al., 2006; Parbery-Clark et al., 2011; Song et al., 2011; Wild et al., 2012). Not surprisingly, populations with hearing loss exhibit high SIN performance variability, though performance variability also exists in normal hearing populations as well (Anderson et al., 2011; Bidelman and Howell, 2016; Kujawa and Liberman, 2015), indicating that SIN performance cannot be predicted by the audiogram alone. Hearing researchers seek to better understand the brain/behavior relationship between speech understanding and the neural mechanisms that give rise to this process with the goal of advancing hearing device technology, improving rehabilitation strategies, and refining identification of hearing aid candidates. The current study applies an emerging optical imaging tool, functional Near Infrared Spectroscopy (fNIRS) to examine the neural underpinnings of speech recognition.

fNIRS measures brain activation by detecting local changes in oxy- and deoxyhemoglobin (HbO & HbR, respectively) (Jobsis, 1977; Quaresima et al., 2012). Near infrared light is directed from a source optode through the scalp, into the cortical surface and subsequently back-scattered to a nearby detector optode. Concentration changes of oxy- and deoxyhemoglobin are calculated using the modified Beer-Lambert Law to estimate neural activation (Cope et al., 1988). In the interest of hearing science, the utilization of fNIRS to assess neural activation during auditory signal processing demonstrates several advantages over other current imaging techniques. Although recent technological advances have made brain imaging methods more suitable for use with hearing impaired populations, both ethical and physical limitations, including the physical properties of cochlear implants (CI), complicate the research utility of Positron Emission Topography (PET), Magnetic Resonance Imaging (MRI) and electroencephalography (EEG) (Bidelman and Howell, 2016; Carlson et al.,

* Corresponding author.
E-mail address: jdefende@uthsc.edu (J. Defenderfer).

http://dx.doi.org/10.1016/j.neuropsychologia.2017.09.004
Received 20 February 2017; Received in revised form 29 August 2017; Accepted 4 September 2017
Available online 06 September 2017
0028-3932/

2015; Gallagher et al., 2012; Hinton, 2002; Kim et al., 2015; Posner, 1997; Saliba et al., 2016; Seghier et al., 2005; Tang and Li, 2012). fNIRS is noninvasive, quiet, safe for repeated use, unrestrictive and tolerant of movement artifact (Dieler et al., 2012; Hoshi, 2003; Kiguchi et al., 2007). In terms of signal resolution, fNIRS falls between MRI and EEG: temporally superior to fMRI and spatially superior to EEG (Huppert et al., 2009). For its adaptability and convenience, the use of fNIRS in hearing research is a burgeoning field of study.

Importantly, fMRI and fNIRS measurements have similar underlying biophysical mechanisms, and previous research has reported correlations between the blood oxygenation level dependent (BOLD) response of fMRI and the deoxyhemoglobin measurement of fNIRS (Buxton et al., 1998; Huppert et al., 2006; Wijeakumar et al., 2016). Fewer studies report finding an equally strong or stronger correlation with the oxy- and total-hemoglobin measurements (Cui et al., 2011; Okamoto et al., 2004). This relationship allows us to make similar inferences across imaging modalities. Foundational fMRI research demonstrated activation in the superior temporal gyrus (STG) to speech sounds (as opposed to no sounds) (Binder et al., 1994; Jancke et al., 1998; Mazoyer et al., 1993) while later studies revealed hemispheric and/or area-specific responses in the temporal lobe to various auditory stimuli (i.e. words, sentences, pseudo-words, reversed speech, tones, noise, etc.) (Binder et al., 2000; McGettigan et al., 2012; Specht et al., 2009).

Results of fNIRS studies are analogous to these findings, reporting robust increases of HbO in the temporal lobe in response to various types of auditory & speech stimuli with both adults and children (Chen et al., 2015; Gallagher et al., 2012; Minagawa-Kawai et al., 2002; Sato et al., 1999; Sevy et al., 2010; Wiggins et al., 2016). Recently, four conditions (natural speech, vocoded speech, scrambled speech, environmental sounds) were designed to assess whether fNIRS could detect possible variations in temporal lobe activation during a passive-listening task. Using a block design, fNIRS measurements revealed significant differences in the cortical response to varying degrees of speech intelligibility. Interestingly, vocoded speech elicited the strongest ΔHbO relative to the other conditions in the left hemisphere, whereas ΔHbO was greatest to normal speech in the right hemisphere. The data also revealed that ΔHbR, not ΔHbO, was most sensitive to listening condition, and the hemodynamic response to normal speech exhibited significantly higher ΔHbR compared to the other conditions. Therefore, the authors concluded that normal speech evoked the highest activation (Pollonini et al., 2014). A follow-up study revealed that natural speech-induced activation was positively associated with speech recognition score in CI recipients. In addition, CI users with good speech perception demonstrated similar cortical patterns of activation to that of normal hearing control subjects where activation increased with increasing intelligibility of the speech stimulus (Olds et al., 2016).

Mounting evidence supports the theory that intelligible speech reliably produces activation in the STG (Abrams et al., 2013; Evans et al., 2014; Narain et al., 2003; Osnes et al., 2011; Scott et al., 2000). At the phonetic level, Osnes et al. documented stronger STG engagement to phonetic stimuli than to non-phonetic sounds (musical instruments) (Osnes et al., 2011). When presented with sentence stimuli, bilateral anterior and posterior regions of the Superior Temporal Sulcus (STS) were shown to be particularly sensitive to differences between unintelligible and intelligible sentences. Specifically, unintelligible sentences demonstrated weaker activation (Okada et al., 2010). Moreover, when speech intelligibility is diminished (yet, remaining intelligible), activation is still observed in the STG with engagement of additional surrounding regions, presumably to aid the primary auditory cortex in achieving accurate signal recognition (Davis and Johnsrude, 2003; Eckert et al., 2016; Salvi et al., 2002). Initial interpretations suggested that this increased activation was the requisite response to a challenging listening condition (Parbery-Clark et al., 2011; Scott et al., 2004; Wong et al., 2008). However, more recent theories suggest that additional neural resources are recruited in difficult listening conditions up to a threshold where discernibility is possible and the listener is motivated to achieve successful recognition (Eckert et al., 2016; Pichora-Fuller, 2016; Vaden et al., 2013).

In the present report, we examine a critical gap in our understanding of the neural basis of speech perception. Previously, studies using fNIRS employed a block design method, contributing critical results that describe the overall quality of the response to auditory stimuli in the cortex. The present study addresses how cortical response varies at the individual trial level. In our design, not only can we examine differences between listening conditions, but we can also examine differences between trials in which speech is accurately perceived versus trials in which speech perception is inaccurate. Using an event-related design, we non-invasively measured temporal lobe activity in young adult volunteers with normal hearing. This method time-locks the cortical response to the onset of short duration trials (3 s). Specifically, we examined cortical activity and speech perception during three listening conditions: speech in quiet (SIQ), speech in noise (SIN) & eight-channel vocoded speech (VOC). In order to obtain results that reflected our daily experience with speech communication (Hervais-Adelman et al., 2012), the perception task was performed in sound field using a loudspeaker. The speech in quiet condition allowed us to observe a baseline measure of the temporal lobe response to speech stimuli alone. The speech in noise condition presents a challenging listening condition in which we expected variability in perception accuracy from trial to trial across participants. In this condition, we can further examine neural responses on trials during which speech perception is accurate compared to trials during which speech perception is inaccurate. Vocoded sentences lack fine spectral detail while still conveying the slow temporal features of speech. Therefore, these sentences are highly intelligible to individuals with normal hearing (Shannon et al., 1995). Vocoding speech diminishes speech intelligibility, but unlike the SIN condition, does not significantly affect behavioral measures of perception. This will provide information about the cortical response when speech is modified in a way that mimics the input that a CI user receives. Based on the findings of previous research in this area, we anticipate that the temporal lobe response will vary according to speech condition (SIQ, SIN, VOC) with stronger activation observed in the more difficult listening condition. Based on speech-induced suppression results from previous studies (Devor et al., 2007; Jenson et al., 2015; Lin et al., 2011), we also expect the temporal lobe response to vary by trial phase with weaker activation during production relative to listening. It is unclear how the hemodynamic patterns will differ between correct and incorrect listening trials of the SIN condition. Some studies show that unintelligible stimuli evoke weaker activation, while other studies demonstrate increased activation to diminished speech intelligibility. An event-related design is critical to help resolve these gaps in our knowledge of speech perception.

2. Methods & analyses

2.1. Participants

The Institutional Review Board of the University of Tennessee Knoxville approved the experimental protocol and plan of research. All participants completed a consent form, handedness questionnaire, and demographic information sheet prior to the experiment. Thirty-one normal-hearing listeners, right handed, between the ages of 18 and 30 years (mean age 24.97 years, 20 females) participated in this study. Normal hearing subjects were native English-speakers and passed a hearing screening with auditory thresholds better than or equal to 25 dB HL at 500, 1000, 2000 and 4000 Hz. The NIR wavelengths of interest are susceptible to absorption characteristics of hair color and density (Strangman et al., 2002); however, pilot data revealed inconsequential effects of hair/skin color variations. Therefore, subjects were not selected with regard to hair or skin color.