Hearing Research 333 (2016) 108–117

Contents lists available at ScienceDirect

Hearing Research

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

Research paper

Acoustic richness modulates the neural networks supporting intelligible speech processing

Yune-Sang Lee [a,*], Nam Eun Min [a], Arthur Wingfield [b], Murray Grossman [a], Jonathan E. Peelle [c,**]

[a] Department of Neurology and Penn Frontotemporal Degeneration Center, University of Pennsylvania, Philadelphia, PA, USA
[b] Volen National Center for Complex Systems, Brandeis University, Waltham, MA, USA
[c] Department of Otolaryngology, Washington University in St. Louis, St. Louis, MO, USA

ARTICLE INFO

Article history:
Received 2 October 2015
Received in revised form
7 December 2015
Accepted 10 December 2015
Available online 23 December 2015

Keywords:
Speech
Acoustic
fMRI
Language
Executive function
Listening effort
Vocoding
Hearing

ABSTRACT

The information contained in a sensory signal plays a critical role in determining what neural processes are engaged. Here we used interleaved silent steady-state (ISSS) functional magnetic resonance imaging (fMRI) to explore how human listeners cope with different degrees of acoustic richness during auditory sentence comprehension. Twenty-six healthy young adults underwent scanning while hearing sentences that varied in acoustic richness (high vs. low spectral detail) and syntactic complexity (subject-relative vs. object-relative center-embedded clause structures). We manipulated acoustic richness by presenting the stimuli as unprocessed full-spectrum speech, or noise-vocoded with 24 channels. Importantly, although the vocoded sentences were spectrally impoverished, all sentences were highly intelligible. These manipulations allowed us to test how intelligible speech processing was affected by orthogonal linguistic and acoustic demands. Acoustically rich speech showed stronger activation than acoustically less-detailed speech in a bilateral temporoparietal network with more pronounced activity in the right hemisphere. By contrast, listening to sentences with greater syntactic complexity resulted in increased activation of a left-lateralized network including left posterior lateral temporal cortex, left inferior frontal gyrus, and left dorsolateral prefrontal cortex. Significant interactions between acoustic richness and syntactic complexity occurred in left supramarginal gyrus, right superior temporal gyrus, and right inferior frontal gyrus, indicating that the regions recruited for syntactic challenge differed as a function of acoustic properties of the speech. Our findings suggest that the neural systems involved in speech perception are finely tuned to the type of information available, and that reducing the richness of the acoustic signal dramatically alters the brain's response to spoken language, even when intelligibility is high.

1. Introduction

During everyday communication, the acoustic richness of speech sounds is commonly affected by many factors including background noise, competing talkers, or hearing impairment. Ordinarily, one might expect that when a speech input is lacking in sensory detail, greater processing resources would be needed for successful recognition of that signal (Rönnberg et al., 2013). Less certain, however, is the effect on neural activity when two intelligible speech signals are presented, but with one signal lacking in spectral detail—conceptually similar to what might be heard with a hearing aid or cochlear implant.

The acoustic quality of the speech signal has been of long-standing interest because acoustic details help convey paralinguistic information such as talker sex, age, or emotion (Gobl and Chasaide, 2003), as well as prosodic cues that can aid in spoken communication. We use the term acoustic richness instead of vocal quality to emphasize that changes to acoustic detail of the speech signal can arise from many sources. Although many behavioral studies have assessed speech perception by systematically manipulating voice quality (Chen and Loizou, 2011, 2010; Churchill et al., 2014; Loizou, 2006; Maryn et al., 2009), relatively little

* Corresponding author.
** Corresponding author.
E-mail addresses: yslee@mail.med.upenn.edu (Y.-S. Lee), peellej@ent.wustl.edu (J.E. Peelle).

http://dx.doi.org/10.1016/j.heares.2015.12.008

neuroimaging research has investigated the neural consequence of acoustic richness in intelligible speech. Here we examine how acoustic clarity affects the neural processing of intelligible speech. We focus on sentence comprehension, where the acoustic richness of the speech might interact with computational demands at the linguistic level.

Neuroanatomically, speech comprehension is supported in large part by a core network centered in bilateral temporal cortex (Hickok and Poeppel, 2007; Rauschecker and Scott, 2009), frequently complemented by left inferior frontal gyrus during sentence processing (Adank, 2012; Peelle, 2012). These regions are more active when listening to intelligible sentences than when hearing a variety of less intelligible control conditions (Davis and Johnsrude, 2003; Evans et al., 2014; Obleser et al., 2007; Rodd et al., 2012; Scott et al., 2000). There is increasing evidence that as speech is degraded to the point that its intelligibility is compromised, listeners rely on regions outside of this core speech network, particularly in frontal cortex. Regions of increased activity during degraded speech processing include the cingulo-opercular network (Eckert et al., 2009; Erb et al., 2013; Vaden et al., 2013; Wild et al., 2012) and premotor cortex (Davis and Johnsrude, 2003; Hervais-Adelman et al., 2012). The fact that these regions are more active for degraded speech than for acoustically rich speech suggests that listeners are recruiting additional cognitive resources to compensate for the loss of acoustic detail.

In these and related studies, however, acoustic richness and intelligibility are frequently correlated, such that the degraded speech has also been less intelligible. The relationship between intelligibility and acoustic richness makes it impossible to disentangle changes in neural processing due to reduced intelligibility from changes due to reduced acoustic richness. To address this issue, in the current study we used 24 channel noise vocoded speech that reduced the spectral detail of speech while allowing for good intelligibility. We refer to these stimuli as acoustically less-detailed speech because of the reduction of spectral resolution, compared to the acoustically rich original signal. Furthermore, to determine how resource demands for cognitive and auditory processes interact, we independently manipulated linguistic challenge by varying syntactic complexity. Because we have clear expectations for brain networks responding to syntactic challenge, including a syntactic manipulation also allowed us to validate the efficacy of our fMRI paradigm and data analysis approach.

One possibility is that, even when speech is intelligible, decreasing the acoustic richness of the speech signal would lead listeners to recruit a set of compensatory frontal networks. In this case, we would expect increased frontal activity for acoustically less-detailed speech, which may be shared or different from that required to process syntactically complex material. An alternative possibility is that removing acoustic detail from otherwise intelligible speech would reduce the quality of the paralinguistic information (e.g., sex and age of the speaker) available to listeners, and thus limit the neural processing for non-linguistic information. In this case, we would expect to observe reduced neural processing for acoustically less-detailed speech.

2. Material and methods

2.1. Subjects

Twenty-six adults (age: 20–34 years, mean = 24.9 years; 12 females) were recruited from the University of Pennsylvania community. All reported themselves being right-handed native speakers of American English and in good health, with no history of neurological disorders or hearing difficulty. Based on pure tone audiometry, all participants' hearing acuity fell within a clinically normal range, with pure tone averages (1, 2, and 4 kHz) of 25 dB HL or less. Fig. 1A shows individual audiometric profiles up to 8 kHz. All participants provided written consent as approved by the Human Subjects Institutional Review Board of the University of Pennsylvania and were paid for their participation.

2.2. Stimuli

Our experimental stimuli consisted of 96 6-word English sentences, half of which contained a subject-relative center-embedded clause and half an object-relative center-embedded clause (Peelle et al., 2010b, 2004). The syntactic manipulation was accomplished by rearranging word order, and thus lexical characteristics were identical across subject-relative and object-relative sentences (e.g., subject-relative: "Kings that help queens are nice"; object-relative: "Kings that queens help are nice"). Each sentence contained a male and female character, but only one character

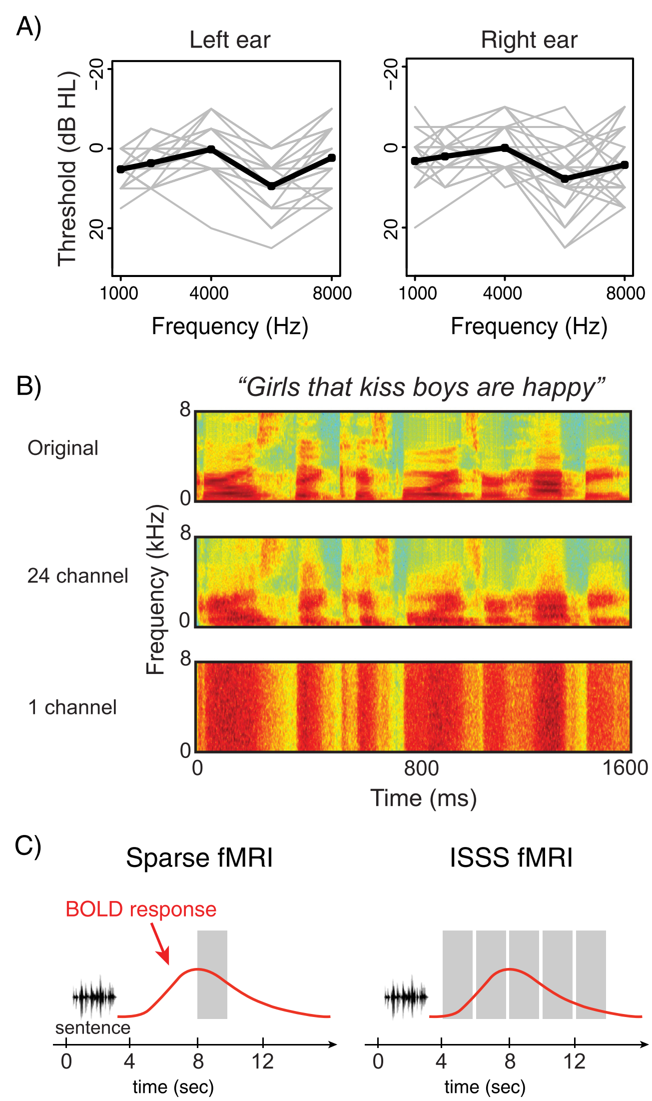

Fig. 1. A. Pure-tone hearing acuity for participants' left and right ears. Individual listeners' profiles are shown in gray lines, with the group mean in black. **B.** Spectrograms of a representative sentence in the three acoustic conditions tested: unprocessed speech (acoustically rich), vocoded with 24 channels (acoustically less-detailed but fully intelligible), or vocoded with 1 channel (unintelligible). **C.** Schematic comparison between a traditional sparse fMRI protocol and the ISSS protocol used in the current study.