Natural variations in speech intelligibility: An fMRI study

Han-Gyol Yi, Rajka Smiljanic, Bharath Chandrasekaran

*SoundBrain Lab (http://soundbrainlab.wordpress.com), Department of Communication Sciences and Disorders
+UTSoundLab (http://utsoundlab.wordpress.com), Department of Linguistics
The University of Texas at Austin, Austin, TX, United States

INTRODUCTION

Everyday we encounter speech of varying intelligibility due to talker dependent factors such as non-native accents and talker independent factors such as noise or incorporation of visual cues. Despite the inherent variability in speech, everyday conversations are largely unprompted. How does the brain resolve natural variations in speech intelligibility?

Studies on intelligibility have focused on using stimuli that are controlled in acoustical properties, yet reduced in intelligibility. This body of research has mostly indicated the left anterior superior temporal sulcus as the main area for intelligibility processing. This is in contrast to clinical studies that indicate the posterior region in the left superior temporal sulcus as the main lesion site of receptive aphasia.

However, the spectrally rotated speech signal, albeit providing good acoustic control, does not resemble the types of signal that are to be resolved in real life. For instance, there is no guarantee that the cortical areas behind rotated speech processing are equivalent to those involved in non-native speech perception, the latter of which is becoming more and more prevalent owing to the increasing diversity in this society. In this fMRI study, participants listened to:

- *Native vs. non-native speakers’ sentences*
- *With or without visual cues of face movements*

Our aims are to (a) demonstrate the neural correlates for processing of natural variations in intelligibility, (b) assess the role of visual cues, and (c) ultimately resolve the lack of consensus on the nature of the speech intelligibility region in the brain.

METHODS

Participants: Twenty-two participants (13 male, 9 female) were recruited from the University of Texas at Austin. All participants were bilingual in English and another non-native language. Their native languages were Spanish, Korean, French, Mandarin, Arabic, Hebrew, and Mandarin. Spanish was spoken aloud of all sentence stimuli. The native language of each participant was assessed using the International Language and Communication Assessment Battery.

Stimuli:

- 56 English sentences
- Sounded by a female native English speaker.
- Low pass filtered at 4kHz
- 28 presented intact; 28 presented after spectral rotation at 4kHz

**fMRI Task 1: Speech vs. Rotated Speech**

- Four blocks per each stimulus type; seven sentences per block
- **Task**
  - After each block, button press: (1) intelligible or (2) unintelligible

**fMRI Task 2: Native vs. Non-Native Speech with Visual Cues**

- 80 English sentences
- Spoken by native (N=4; 2 f) or non-native (N=4; 2 f; native Korean speakers) English speakers
- 40 presented with visual cues; 40 presented without visual cues (fixation cross)

**Task**

- Event-related design with pseudorandomized sequence
- **Design**
  - Multi-talker babble noise (SNR: -12 dB)
  - Type- perceived sentences; scored on proportion of keywords reported correct
  - Behavioral Analogue Task
  - Participants completed a similar task outside the scanner

RESULTS

**Behavioral Indices of Speech Perception**

A repeated measures ANOVA was run with two within-subjects factors (visual cues; nativeness of talkers) on the intelligibility ratings and proportion of words correct in the sentence perception tasks conducted inside and outside the scanner.

**Neural Correlates of Speech Perception Tasks**

**SUMMARY**

**DISCUSSION**

Validity of the Speech vs. Rotated Speech Comparison

This contrast yielded the entire language network spanning both dorsal (articulatory) and ventral (comprehension) streams. This is not indicative of intelligibility processing proper but language processing, as rotated speech constitutes non-linguistic stimuli, and is not comparable to actual speech in terms of semantics, syntax, or phonology.

Neural Correlates for Native vs. Non-Native Speech Perception

Non-native speech, judged less intelligible, required both dorsal and ventral streams of the language network, since more processing is necessary for challenging tasks. This activity was comparable to the traditional Speech-Rotated contrast. In contrast, native speech processing which should be more automatic and effortless involved limited recruitment of posterior portions of the right temporooccipital junction.

Role of Visual Cues

Visual cues aided speech processing, especially for native speech than for non-native speech. The reduced extent of neural involvement for the non-native/native contrast with visual cues indicate that the level of neural efficiency became more equivalent across two speech types as more external cues were available.

Future Directions

1. Assess sources of individual variability in degrees of visual cues incorporation and native-bias in speech intelligibility (McGurk effect susceptibility; Implicit Association Task)
2. Functional connectivity analysis to complement the univariate subtraction approach in assessing the nature of dual streams activation in speech perception.

ACKNOWLEDGEMENTS

The authors would like to thank Kristen Steenkeste, Joanna E. P. Phelps, and Rachel (for her significant contributions in data collection and processing) and the staff of the Hearing Research Center at the University of Texas at Austin for technical support and advice.