Subcortical Frequency-Coding Errors Are Linked to Speaker-Variability Intolerance in Normal-Hearing Adults

OHIO UNIVERSITY

Breanna N. Hart, Fuh-Cherng Jeng and Chao-Yang Lee

Communication Sciences and Disorders, Ohio University, Athens, Ohio

INTRODUCTION

- The same word spoken by different talkers can be acoustically very different. Therefore, how listeners process speaker variability is a fundamental issue in speech perception.
- Throughout this study we will be using a common phrase, ***Speaker-Variability Intolerance***, which is defined as a listener's inability to tolerate variation between different speakers. There is ample evidence that processing multi-speaker speech is challenging for normal hearing listeners (Mullennix et al, 1989**).** Effects of speaker variability on speech perception in listeners' with hearing impairment has also been reported (Kirk et al., 1997).
- Behavioral studies (Lee et al., 2012, 2013) have reported that, among the various cues that are used to deal with speaker variability, fundamental frequency (**F0**) of the speakers voice presents the most important cue of all.
- Frequency-following response (FFR) is a scalp-recorded, neurophysiological potential that reflects phase-locked neural activities at the subcortical level that are in sync with the frequency contents of a stimulus (Skoe & Kraus, 2010). Unlike most cortical responses that may be highly variable and affected by sleep, FFR is a reliable response (Song et al., 2011) originating from neural substrates at the subcortical level, primarily in the midbrain area, and thus does not require the listener's attention, alertness, or active participation. Due to these advantages, the FFR has been used to investigate the subcortical neural representation of the various features of speech sounds such as F0 tracking accuracy and timing in normal-hearing adults (Jeng et al., 2016) and individuals with hearing impairment (Anderson et al., 2013).
- Neural correlates to the listeners ability to differentiate sounds vocalized by different speakers remains unknown. The goal of this study is to examine how English-speaking listeners with normal hearing but without prior knowledge of a tonal language process speaker variability behaviorally and neurophysiologically in Mandarin tone perception. The research question is whether the behavioral and neurophysiological measurements are correlated.
- Because neural substrates at the subcortical level provide critical information to the auditory cortex and related areas where executive decisions of the tonal stimuli and speaker identification take place, it was **hypothesized** that subcortical frequency-coding errors would be significantly associated with the listener's ***Speaker-Variability Intolerance***.

METHODS

Participants

- 21 normal hearing American adults from 18 to 40 years old
- Mean Age: 25 years
- 18 Females, 3 Males
- Native American English Speakers

Study Design

- Pretest: hearing screening
- Behavioral training
- Subcortical recording

Behavioral

Stimulus

- /i1, i2, i3, i4/, flat, rising, dip, falling pitch contours; 250 ms; human speech

Procedure

- Testing: Familiarization, Training (must receive 50%+), Blocked-By-Speaker, Mixed-Across-Speakers
- Stimulus through headphones with selection on touch screen monitor

Behavioral Data Analysis

- *Speaker-Variability Intolerance* = Mixed-Across-Speakers – Blocked-By-Speaker
- This formula was used for both percent correct and reaction time.

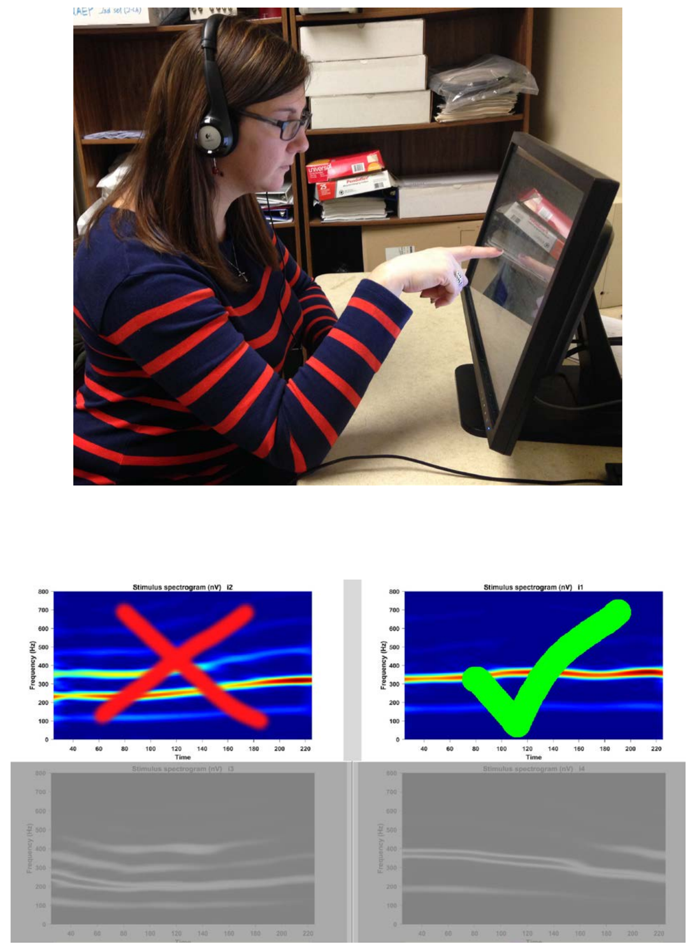

Figure 1: Experimental setup and computer interface utilized in the behavioral portion of the experiments.

Subcortical Measurement (Frequency Following Response, FFR)

Stimulus

- /i2/, rising pitch contour; 250 ms; human speech

Procedure

- 3 gold-plated electrodes (high forehead, low forehead, right mastoid)
- Participant resting or fast asleep prior to recording
- Stimulus Presentation: 75 dB SPL in the right ear
- 8002 accepted sweeps were collected

FFR Data Analysis

- All data was analyzed through MATLAB
- Frequency spectrograms were utilized to estimate the f0 contours of the recordings
- Frequency Error, defined as how well the brain is able to follow the intonation of sound, was computed from the brain waves of each participant to represent the pitch-encoding accuracy during the stimulus presentation.

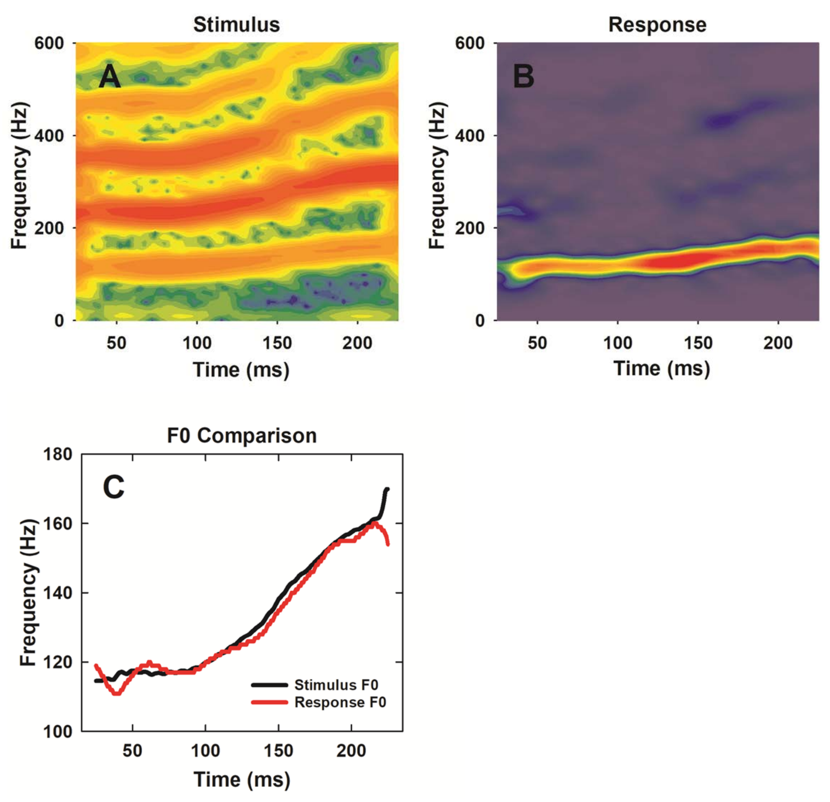

Figure 2: Estimates of subcortical frequency-coding errors (*Frequency Error*). A. Amplitude spectrogram of the Tone 2 stimulus with a rising F0 contour. B. A typical FFR spectrogram obtained from a normal-hearing participant (subject S003). C. F0 contours of the stimulus (black curve) and an FFR recording (red curve). Frequency error is computed by finding the mean of the absolute values of the F0 differences between the stimulus and a recording.

RESULTS

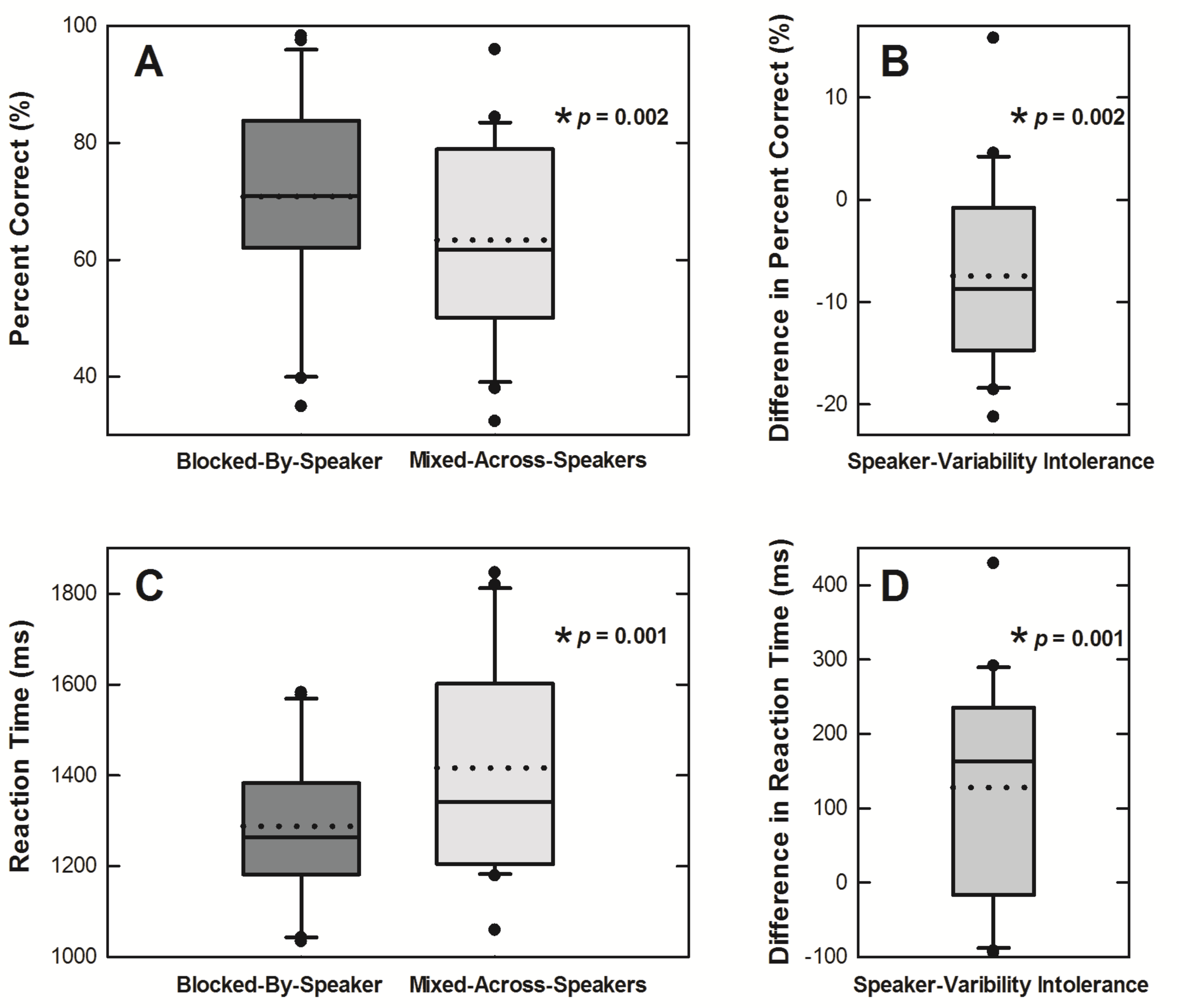

Figure 3: *Speaker-Variability Intolerance* in terms of percent correct (A) and reaction time (B). For percent correct, the listeners' performance scores were significantly smaller in the mixed-across-speakers condition than those obtained in the Blocked-By-Speaker condition (mean difference = -7.427%, p = 0.002). Reaction times obtained in the Mixed-Across-Speakers condition were significantly longer than those obtained in the Blocked-By-Speaker condition (mean difference = 128 ms, p = 0.001). *Speaker-Variability Intolerance* = Mixed-Across-Speakers – Blocked-By-Speaker scores.

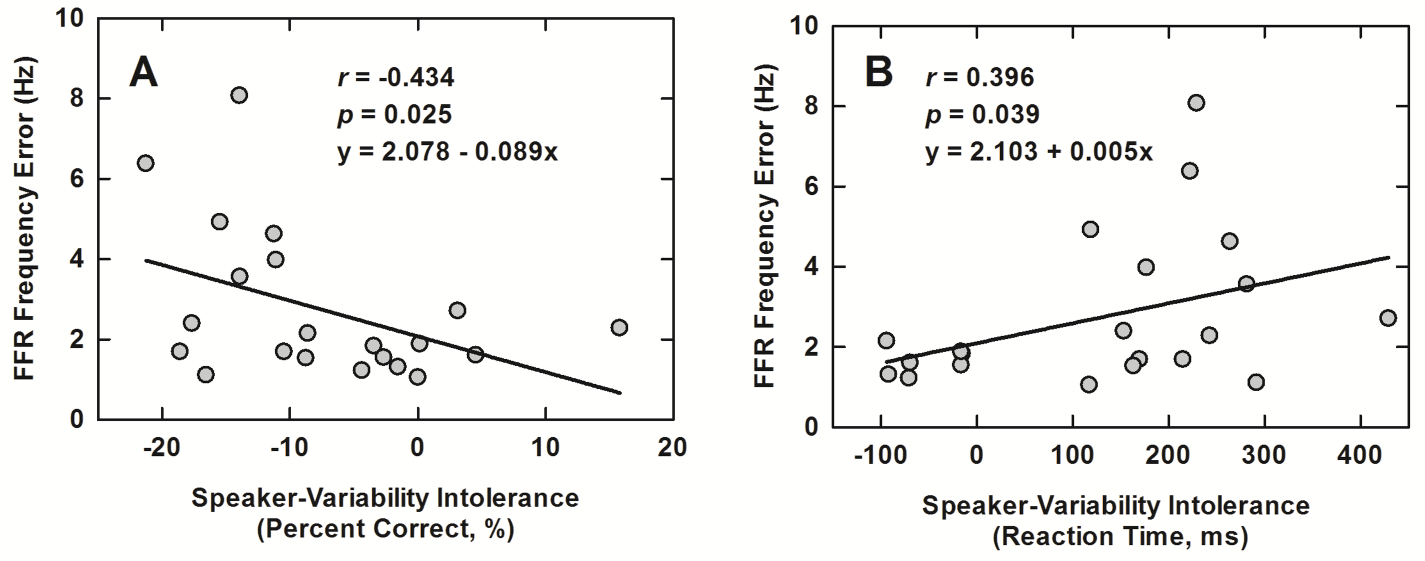

Figure 4: Frequency Error is **negatively correlated** with *Speaker-Variability Intolerance* in terms of **percent correct** and **positively correlated** with *Speaker-Variability Intolerance* in terms of **reaction time**.

Corresponding author: Breanna Hart bh130015@ohio.edu

CONCLUSION

- Behavioral: The Blocked-By-Speaker condition produced significantly better percent correct and shorter reaction time than the Mixed-Across-Speakers condition. Findings of *Speaker-Variability Intolerance* was observed through the behavioral methods used in this study.
- FFR: Larger *Frequency Errors* (i.e., more frequency-coding errors at the subcortical level) were associated with poorer *Speaker-Variability Intolerance* (i.e., more negative numbers) in terms of percent correct.
- FFR: The more *Frequency Errors* exhibited at the subcortical level, the longer reaction times there were in the Mixed-Across-Speakers session, and thus a poorer *Speaker-Variability Intolerance* in terms of reaction time (i.e., more positive numbers).

ACKNOWLEDGEMENTS

The authors would like to thank the participants who participated in this study and the six speakers who contributed their voice. The authors would also like to thank Tiffany McDonald, Elizabeth Teets, and Hallie Ganch for their assistance in participant recruitment and data collection.

REFERENCES

J. W. Mullennix, D. B. Pisoni, and C. S. Martin, "Some effects of talker variability on spoken word recognition," J. Acoust. Soc. Am. **85**, 365–378 (1989).

K. I. Kirk, D. B. Pisoni, and R. C. Miyamoto, "Effects of stimulus variability on speech perception in listeners with hearing impairment," J. Speech Lang. Hear. Res. JSLHR **40**, 1395–1405 (1997).

C.-Y. Lee, Y. Zhang, X. Li, L. Tao, and Z. S. Bond, "Effects of speaker variability and noise on Mandarin fricative identification by native and non-native listeners," J. Acoust. Soc. Am. **132**, 1130–1140 (2012).

C.-Y. Lee, L. Tao, and Z. S. Bond, "Speaker variability and context in the identification of fragmented Mandarin tones by native and non-native listeners," J. Phon. **37**, 1–15 (2009).

E. Skoe and N. Kraus, "Auditory brain stem response to complex sounds: A tutorial," Ear Hear. **31**, 302–324 (2010).

J. H. Song, T. Nicol, and N. Kraus, "Test–retest reliability of the speech-evoked auditory brainstem response," Clin. Neurophysiol. **122**, 346–355 (2011).

F.-C. Jeng, C.-D. Lin, J. T. Sabol, G. R. Hollister, M.-S. Chou, C.-H. Chen, J. E. Kenny, and Y.-A. Tsou, "Pitch perception and frequency-following responses elicited by lexical-tone chimeras," Int. J. Audiol. **55**, 53–63 (2016).

S. Anderson, A. Parbery-Clark, T. White-Schwoch, S. Drehobl, and N. Kraus, "Effects of hearing loss on the subcortical representation of speech cues," J. Acoust. Soc. Am. **133**, 3030–3038 (2013).