Acoustic features associated with similarity judgments of voices

Kristina Furbeck, Emily Thurston, Jessica Tin, & Tyler Perrachione

Department of Speech, Language, & Hearing Sciences, Boston University

Summary

- Native speakers of English and Mandarin heard 820 pairs of speech recordings from English and Mandarin talkers and rated the perceived dissimilarity of the pairs of voices.
- The *language-familiarity effect*[6] (greater dissimilarity for different-voice pairs, greater similarity for same-voice pairs in one's native language)[2] was not reliably observed.
- Dissimilarity ratings of pairs of voices were highly consistent across listener language background and between forward and time-reversed speech.
- Across talker language, listener language, and time-reversal, dissimilarity ratings of voice pairs were most closely related to trial-by-trial differences in mean F0.

Methods

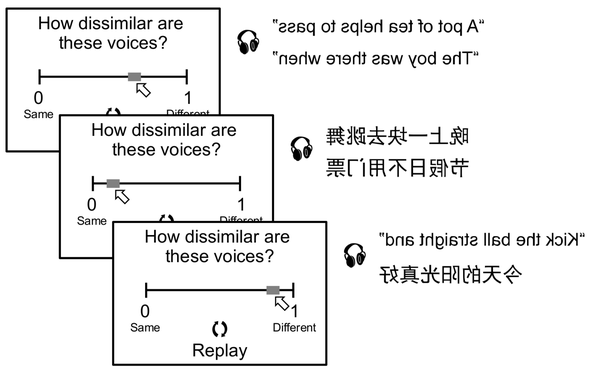
Stimuli: Recordings of female native speakers of Mandarin (n=20) and English (n=20) reading 10 phonetically balanced sentences in their native language [3,4].

Acoustic measurements: For each recording: fundamental frequency mean and variation ($s/\bar{x}$), jitter, harmonics-to-noise ratio (HNR), speech rate, and formant dispersion [1].

Participants: Native speakers of English (n=40) and Mandarin (n=40). Half of each group was assigned to rate voices from either time-reversed [2] or time-forward speech.

Procedure: Participants heard pairs of recordings and indicated subjective dissimilarity from 0 (certain these are the same talker) to 1 (certain these are different talkers) [2].

Perceptual dissimilarity rating paradigm



Perceptual Dissimilarity Judgments

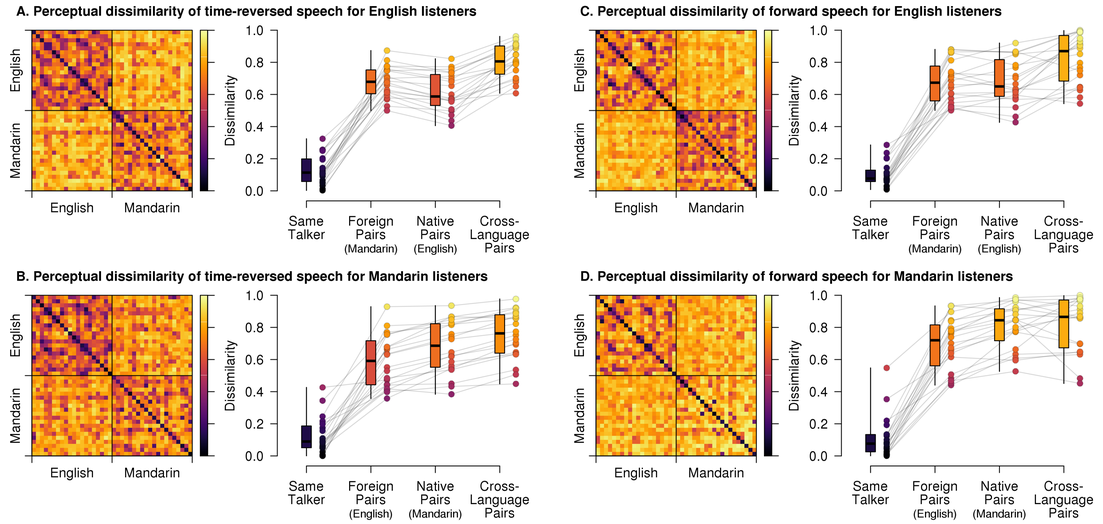


Fig. 1: Mean perceptual dissimilarity judgments for all pairs of voices by listener group, talker language, and time-reversal (matrices); mean ratings by listener and type of voice pairs (boxplots).

Language-Familiarity Effect

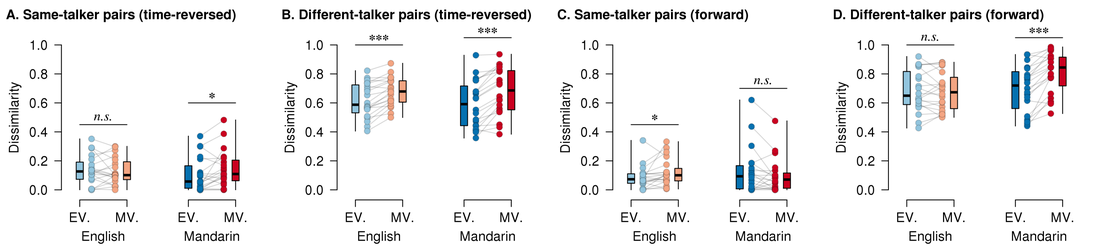


Fig. 2: Comparison of perceptual dissimilarity judgments by listener and talker language frequently revealed listener judgments that were inconsistent with the language-familiarity effect [2,6].

Acoustic Representational Similarity

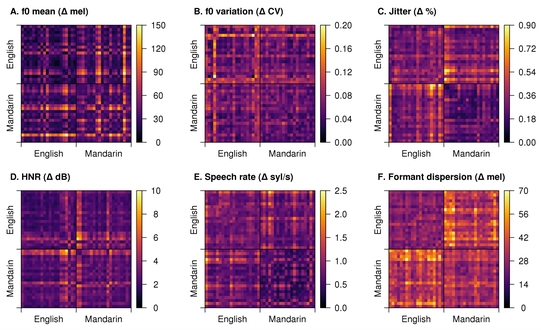


Fig. 3: Mean acoustic differences between pairs of talkers. Comparison of trial-by-trial acoustic differences[5] between talkers and perceptual dissimilarity judgments by listeners revealed perceived dissimilarity to be most closely related to the difference in mean F0 for both forward and time-reversed English and Mandarin speech. Differences in jitter were also related to dissimilarity judgments of Mandarin voices by both listener groups. Formant dispersion played a larger role in judgments of forward speech.

Consistent Ratings

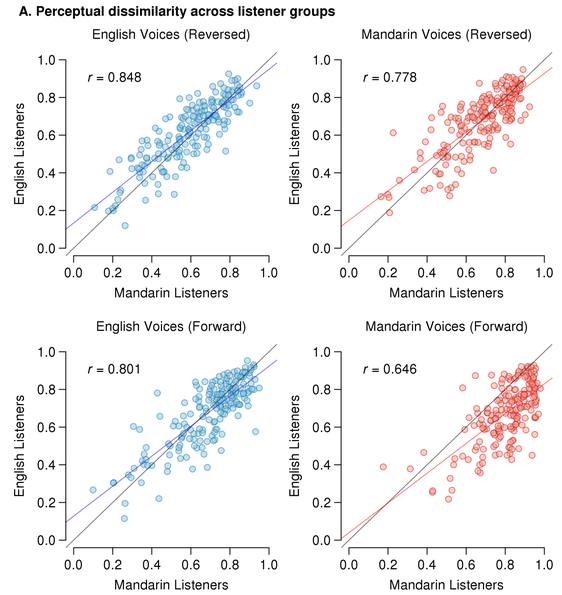
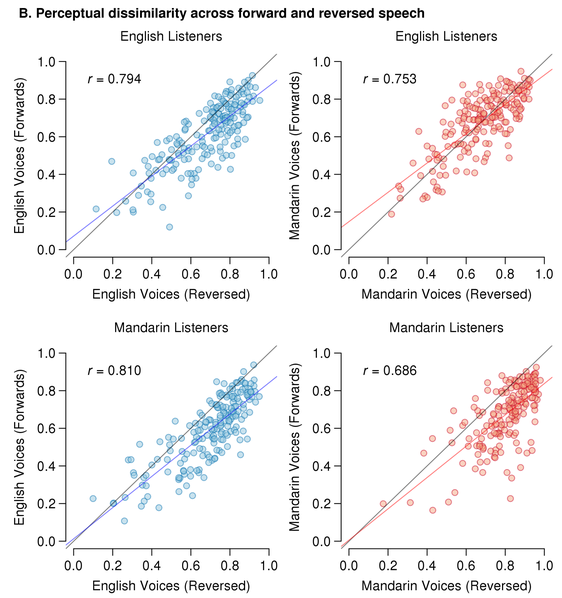


Fig. 4: Relationship between mean perceived dissimilarity of voice pairs across listener group and time-reversal. (A) Voice pairs rated as more dissimilar by English listeners also tended to be rated as more dissimilar by Mandarin listeners. **(B)** Voice pairs rated as more dissimilar from time-reversed speech also tended to be rated as more dissimilar from forward speech.



Acknowledgments

We thank D. McLaughlin, G. Cler, C. Stepp, Y. Carter, S. Dougherty, J. Golditch, A. Chang, C. Cheng, and S.-J. Lim. This research was supported by NIH NIDCD R03DC014045 and a NARSAD Young Investigator Award from the Brain and Behavior Research Foundation to T.P.

References

[1] Fitch (1997). *J. Acoust. Soc. Am.*
[2] Fleming *et al.* (2014). *Proc. Nat. Acad. Sci.*
[3] Fu *et al.* (2011). *J. Acoust. Soc. Am.*
[4] IEEE (1969). *IEEE Trans. Audio Electroacoust.*
[5] Kriegeskorte *et al.* (2008). *Front. Sys. Neurosci*
[6] Perrachione & Wong (2007). *Neuropsychologia.*

BU Communication Neuroscience Research Laboratory

tkp@bu.edu
http://sites.bu.edu/cnrlab/