A CORPUS OF NATIVE AND NON-NATIVE SPEECH FOR SPEECH PRODUCTION RESEARCH

Ruolan Li, Xin Xie, & T. Florian Jaeger,
Department of Brain & Cognitive Sciences, University of Rochester
r.li@rochester.edu

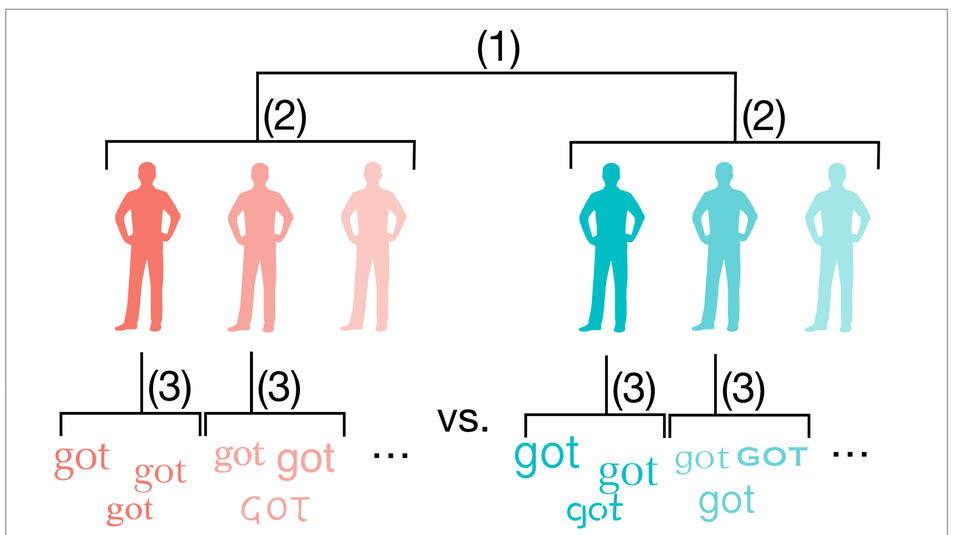
Goal

Aim of database:

- To characterize the structure of variation across multiple socio-phonetic levels, which is difficult for small-sample analysis in previous research [1].

We provide a database of:

1) native + non-native speech
2) annotated words + sentences



Hierarchical structure of phonetic variation

- (1) across accents
- (2) individuals within an accent
- (3) tokens within an individual speaker

Database

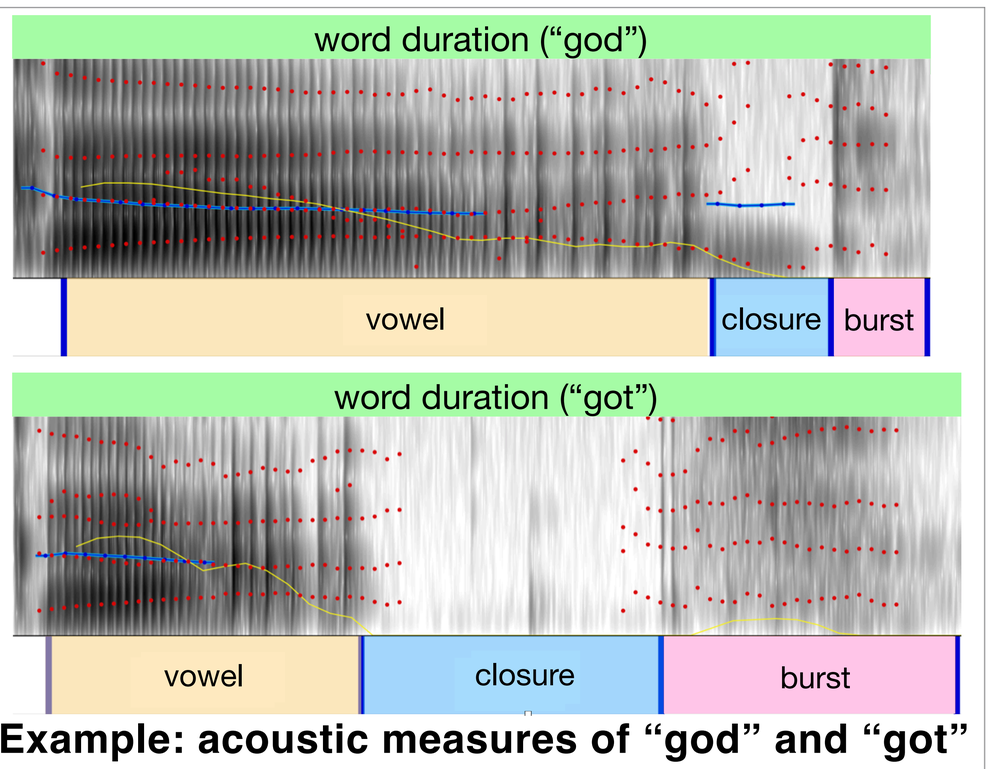
Speaker statistics

	Male	Female	Total
AE (American English)	10	5	15
ME (Mandarin-accented English)	10	5	15

Every speaker has read (1) 180 isolated words
(2) connected sentences*
(each word/sentence is read three times)
*: 81 sentences across 5 phonetically balanced sets

Acoustic annotation

- Stop-final words: manually annotated
  - word, phoneme, and acoustic tiers available
  - word duration
- The rest of the words: automatic alignment [2]



Example: acoustic measures of "god" and "got"

For each stop-final word, temporal duration of vowel, closure and burst are measured. Duration of the whole word is also measured for normalizing purpose.

Sample Analysis

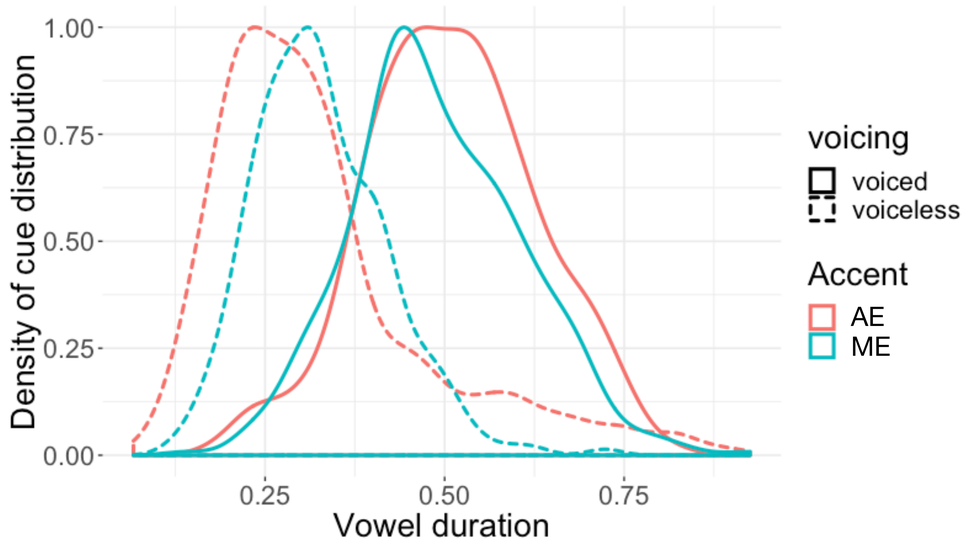
Our interest:

How does ME differ from AE on the final-stop voicing?

- Acoustic measurements: vowel, closure, and burst duration
- Proportional measurement: $= \frac{\text{measurement}}{\text{word duration}}$

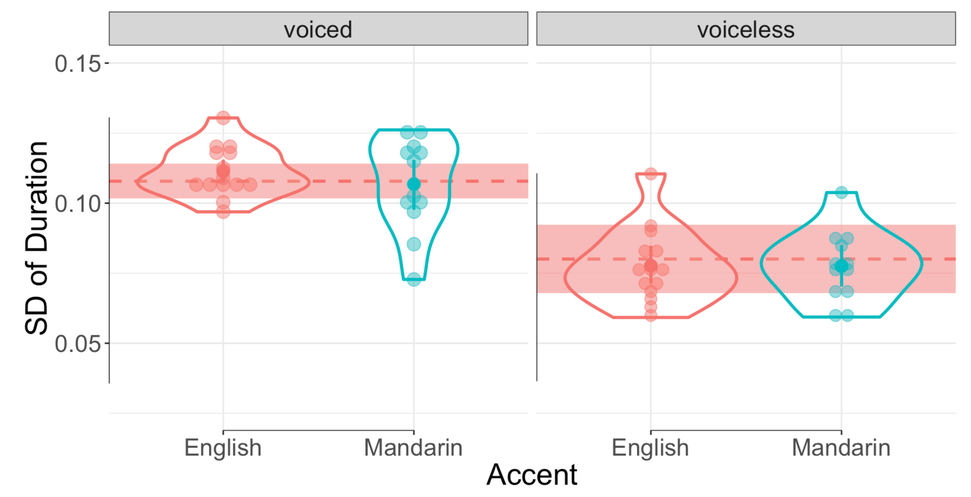
(1) Across accents: difference in distribution

Plotted are proportional measures of vowel duration. Note the difference between ME and AE distribution, as well as the greater overlap of two distributions in ME, which makes listeners harder to distinguish voicing in ME if they rely mostly on vowel duration.



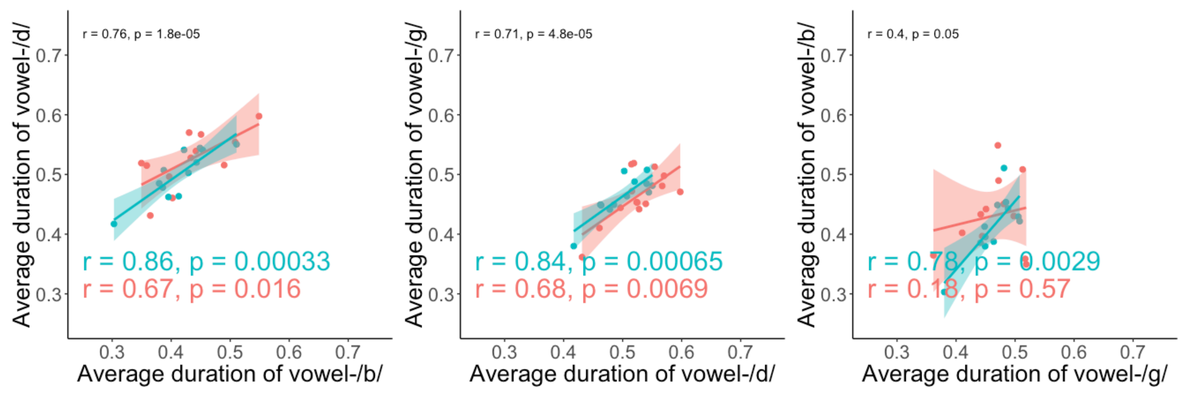
(2) Individuals within an accent: variability

Each dot represents standard deviation of one speaker's vowel duration (proportional measure). Results on all three acoustic measures suggest that ME, as a foreign-accented speech, is not necessarily more variable than native speech (AE). Dashed lines: mean of AE; error bar: SD of AE.



(3) Tokens within an individual speaker

Vowel duration of different phonemes (proportional measures, voiced stops shown only). Stops with different places of articulation are paired up. Overall, correlation between cues is strong in most conditions, suggesting that variation is structured within speakers. Note that ME has correlation patterns similar to AE.



Work in Progress:

1. Compare human behavioral data with computational models;
2. Examine the structure of variation of other phoneme contrasts;
3. Expand the annotation to include continuous sentences.

Access to Corpus:

Please contact the authors by email.

Acknowledgement:

This work was supported by an NIHCD R01 (HD075797) to TFJ.

The conference presentation was supported by the Research Presentation Award (University of Rochester) to RL.

Reference:

[1] Flege et al. (1995; 1998), etc. (production studies); Bradlow (1995); Bradlow & Bent (2008), etc. (perception studies).

[2] Yuan & Liberman. 2008. Speaker identification on the SCOTUS corpus. Proceedings of Acoustics '08.

Link to poster: