Background

- Plain versus clear speech modifications involve articulatory and acoustic modifications\(^4,5\), which can enhance auditory and visual speech intelligibility in native (L1) and non-native (L2) languages\(^6,7\). However, excessively exaggerated speech resulting in overlap of phonemic categories, or attention to incorrect auditory or visual cues, may inhibit intelligibility\(^8,9\).
- English tense-lax vowel contrasts and plain-to-clear vowel modifications share similar articulatory and acoustic features\(^9,10\). These overlapping cues may affect intelligibility of tense contrasts in clear speech, especially for L2 perceivers whose L1 does not have vowel tension contrasts (e.g., Mandarin Chinese\(^11\)).

Present Study

- Examines auditory-visual perception of clearly versus plainly produced English tense and lax vowels by native English and Mandarin perceivers.
- Our articulatory study showed greater articulatory movements in clear than plain productions for both tense and lax vowels\(^4\). Our acoustic study further revealed greater temporal modifications for tense vowels and greater spectral modifications for lax vowels in clear speech\(^1\).
- Based on these findings, we ask:
  1. How does vowel tension interact with clear speech in auditory and visual perception?
  2. How do clear speech modifications influence L1 versus L2 vowel tension perception?

Methods

Participants:
- 23 native English perceivers
- 30 native Mandarin perceivers, late intermediate-level learners of English

Stimuli:
- Vowels [i, ɪ, a, æ, u, o] in [k, d, z, l, j] minimal pairs: `laid`, `laid`, `cool`, `cud`, `coed`, `could`.
- Produced by 6 native English speakers (3F) in plain and clear speech\(^4\).
- Elicitation of stimuli involves a simulated computer speech recognition program\(^4,5\).
- Presented in audio (A), audio-visual (AV), and visual (V) modes\(^6,7\).
- Audio was embedded in cafeteria noise (\(-15\) dB SNR).

Task:
- Six-alternative forced choice (6AFC)

Figure 1: Elicitation of plain and clear speech

Figure 2: Presentation of stimuli

Results: Accuracy

- Mixed-effects logistic regression run on Accuracy with Subject as a random effect in English and Mandarin perceivers groups

Relative Cue Contributions

- F1 was the most important\(^*\) parameter in Audio models for both English and Mandarin perceivers.
- No visual parameters were predictive of English Audio-Visual responses, whereas all four visual parameters contributed to the Mandarin model.
- Lip rounding was more predictive in the English Visual model, while Jaw Displacement appeared to be a more prominent cue for Mandarin perceivers.

Figure 3: Overall accuracy of English perceivers by Stimulus Mode (A, AV, V). Speech Style (clear, plain), and Vowel Tensity (tense, lax).

Figure 4: Overall accuracy of Mandarin perceivers by Stimulus Mode (A, AV, V). Speech Style (clear, plain), and Vowel Tensity (tense, lax).

Results: Errors

- Separate mixed-effects negative binomial regression models run on Error Counts within Mode and Language group
- Tensity errors predominate; consistent with the accuracy results of the clear < plain pattern in visual lax stimuli

Figure 5: Distribution of errors among English perceivers by Stimulus Mode, Speech Style, Vowel Tensity, and Feature Difference characterizing the error (tensity, backness, height, height+backness).

Figure 6: Distribution of errors among Mandarin perceivers by Stimulus Mode, Speech Style, Vowel Tensity, and Feature Difference characterizing the error (tensity, backness, height, height+backness).

Acknowledgements

We thank Dr. Rob McMurray for his input on the C-Cure analysis. We also thank graduate students Tao Tang and members of the Language and Brain Lab.

References

1. Ferguson & Kewley-Port 2002.
2. Tang et al. 2015.
10. Lang Linguist Compass, 3 45-58.
12. JAISL, 26 3: 230-239.
14. Charles Redmon, Allard Jongman, and Joan Sereno
Department of Linguistics, Simon Fraser University
The 5th Joint Meeting of the Acoustical Society of America and the Acoustical Society of Japan: Honolulu, 2016

Figure 7: Predicting English perception from acoustic\(^2\) and articulatory\(^3\) cues.

Figure 8: Predicting Mandarin perception from acoustic and articulatory cues.

Conclusions

- Clear speech benefits perception when its cues are compatible with those that characterize phonemes (e.g., tense vowels), but inhibits perception when its hyper-articulated features are in conflict with phoneme-unique characteristics (e.g., lax vowels).
- Non-native benefit from clear speech for familiar (tense) vowels in their L1, but perception of non-familiar L2 (lax) vowels can be more impaired by incompatible clear-speech and phonemic features.

*Parameter separation was determined from relative AIC reduction in the full model.