

Effects of talker accent and listener language background on lexical retuning and phonetic categorization

Emily J. Clare^{1,2} and Jessamyn Schertz^{1,2}

¹*Department of Language Studies, University of Toronto Mississauga;*

²*Department of Linguistics, University of Toronto*

Background: Listeners' perception is influenced by top-down (signal-external) information: for example, listeners use lexical knowledge to adjust their phonetic category representations in response to idiosyncratic accents (e.g., [1], [2]). This work explores how this sort of "lexical retuning" is influenced by the accent of the talker and the language background of the listener. We test categorization and retuning of the English /d/-/t/ contrast by two groups of listeners (L1 English and L1 Mandarin listeners) hearing two talkers: an L1 Mandarin talker with a perceptible non-native accent, and an L1 English talker. We focus on two questions: 1) *Does the extent of perceptual retuning differ based on the accent of the talker and/or listeners' language background?*; and 2) *How do these factors affect baseline categorization?* We expect there may be more retuning when listening to a non-native accent, based on previous work examining use of different kinds of top-down information: for example, semantic context exerts more influence on listeners' phonetic categorization when the talker is a non-native speaker [3]. Furthermore, since Mandarin voiceless stops have higher VOT than English stops in production [4], we expect L1 Mandarin listeners to have an overall higher perceptual category boundary (i.e. lower /t/ response rate across a /d-t/ continuum) than L1 English listeners. Finally, if familiarity with an accent influences categorization, we would expect that L1 Mandarin listeners, who have familiarity with both accents, would show differentiated perception for the two talkers, with a lower /t/ response rate for the L1 Mandarin than the L1 English talker.

Participants: We report data from 246 North American listeners who passed attention checks. Half of the listeners (*L1 Mandarin*) learned Mandarin in the home, and the other half (*L1 English*) learned English in the home and did not report any knowledge of Mandarin.

Tasks: The study was completed online. In an exposure phase, listeners heard real English words, including 36 critical words containing /d/ or /t/, and were asked to count the number of words they heard [5]. For half of the participants (*d-ambig group*), all of the /d/ sounds were replaced by an ambiguous [d~t] sound, and for the other half (*t-ambig group*), the /t/ sounds were replaced by the ambiguous sound. In a test phase, listeners heard nonwords from a 5-step VOT series ('damu'-'tamu') and were asked whether they heard 'damu' or 'tamu' (100 trials). Stimuli were created from natural recordings of two talkers: an L1 English talker and an L1 Mandarin talker. Each listener heard stimuli from a single talker, the same for both tasks.

Results: In a logistic mixed-effects regression model (Table 1), a significant three-way interaction indicated that the effect of lexical retuning was not consistent across listener/talker groups (Fig. 1). Follow-up tests indicated that the effect of lexical retuning was significant for English listeners hearing the Mandarin talker, (in the expected direction, with more /t/ response for the *t-ambig* than the *d-ambig* group), but not for the other groups. This is consistent with previous work finding greater use of top-down semantic information when hearing a non-native talker, perhaps due to less certainty about the expected pronunciation of specific sounds [3]. The fact that Mandarin listeners did not show this retuning may be due to their greater familiarity with Mandarin-accented English, resulting in stronger prior beliefs about expected pronunciation and therefore less inclination to adjust their representations.

We also found different patterns of baseline categorization based on both listener and talker group (Fig. 2). Mandarin listeners showed fewer /t/ responses than English listeners for both talkers, providing perceptual evidence for a systematic phonetic difference between Mandarin vs. English stops, with longer VOT for Mandarin stops [4]. Furthermore, the Mandarin listeners showed talker-specific categorization, with significantly fewer /t/ responses for the Mandarin than for the English talker, whereas English listeners showed no talker-related differences. This provides an example of the use of talker-based expectations during speech perception: Mandarin listeners presumably have more exposure to Mandarin-accented English, and they use this knowledge of accent-specific characteristics to inform their categorization.

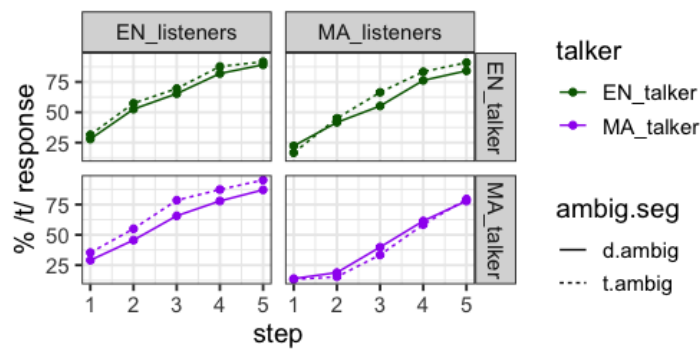


Fig. 1. % /t/ (vs. /d/) response by talker and listener group, across 5 steps varying in VOT.

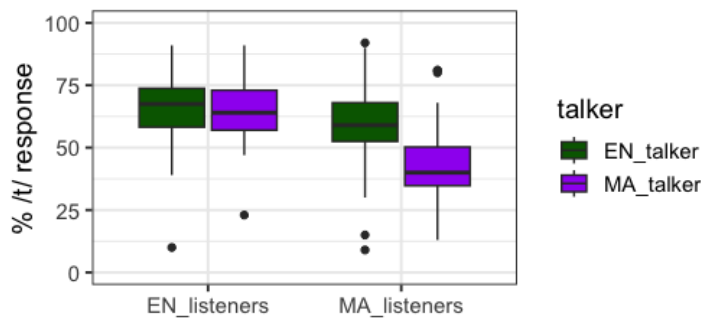


Fig. 2. Boxplots showing distribution of by-participant mean /t/ response.

Table 1. Results from a logistic mixed-effects model: $glmer(t.response \sim VOT.step + (Ambig. segment * Talker accent * Listener group + (step|participant), family=binomial)$. All categorical factors were simple-coded, and reference levels are in italics.

| | β | SE | t | p | |
|--|---------|------|-------|--------|-----|
| Intercept | 0.49 | 0.06 | 8.10 | < .001 | *** |
| VOT step | 1.43 | 0.04 | 34.62 | < .001 | *** |
| Ambiguous segment (<i>ambig.t</i> vs. <i>ambig.d</i>) | 0.22 | 0.12 | 1.78 | 0.075 | . |
| Talker accent (<i>L1 Mandarin</i> vs. <i>L1 English</i>) | -0.61 | 0.12 | -5.09 | < .001 | *** |
| Listener (<i>L1 Mandarin</i> vs. <i>L1 English</i>) | -1.05 | 0.12 | -8.73 | < .001 | *** |
| Ambig.seg * Talker | -0.14 | 0.24 | -0.60 | 0.547 | |
| Ambig.seg * Listener | -0.44 | 0.24 | -1.85 | 0.064 | . |
| Talker * Listener | -1.20 | 0.24 | -5.03 | < .001 | *** |
| Ambig.seg * Talker * Listener | -0.98 | 0.48 | -2.06 | 0.039 | * |

References

- [1] Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, 47(2), 204–238.
- [2] Reinisch, E., & Holt, L. L. (2014). Lexically guided phonetic retuning of foreign-accented speech and its generalization. *Journal of Experimental Psychology. Human Perception and Performance*, 40(2), 539–555.
- [3] Schertz, J., & Hawthorne, K. (2018). The effect of sentential context on phonetic categorization is modulated by talker accent and exposure. *The Journal of the Acoustical Society of America*, 143(3), EL231-EL236.
- [4] Chao, K.-Y., & Chen, L.-M. (2008). A Cross-Linguistic Study of Voice Onset Time in Stop Consonant Productions. *International Journal of Computational Linguistics & Chinese Language Processing*, Volume 13, Number 2, June 2008, 215–232.
- [5] McQueen, J. M., Norris, D., & Cutler, A. (2006). The dynamic nature of speech perception. *Language and Speech*, 49, 101–112.