What are you sinking about? Effects of phonetic learning on online lexical processing of accented speech

Yevgeniy Melguy and Keith Johnson
University of California, Berkeley (USA)
ymelguy@berkeley.edu, keithjohnson@berkeley.edu

Speech produced with an unfamiliar accent may pose a challenge for listeners, resulting in delayed processing or lower intelligibility [1]. Such costs may be due to a mismatch between listener expectations about how a given sound category should be phonetically realized, and how it is implemented by non-native speakers. Phonetic mismatches can increase processing time [2], but listeners could avoid them by adjusting their expectations for a given speaker or speech variety. There is evidence that listeners use just such a strategy to perceptually adapt to an unfamiliar accent via *phonetic recalibration* of perceptual category boundaries. For instance, following exposure to an artificial accent involving a realization of /s/ that is phonetically intermediate between [s] and [f] (e.g., the word *moss* realized as *mo*[s/f]), listeners are more likely to categorize ambiguous tokens along a phonetic continuum between /s/ and /f/ as the trained phoneme /s/ [3].

Despite such adaptation being well-attested in the literature (see [4] for a review), the mechanisms involved in such category re-tuning are still underexplored. Namely, it is unclear whether listeners use a targeted mechanism specific to the phonetic patterns they encounter, or whether they use a more general mechanism of "criteria relaxation" that is insensitive to phonetic detail [5]. Recent literature has suggested that recalibration of category boundaries is achieved by a relatively general mechanism by which listeners expand phonetic categories in perceptual space, generalizing beyond the specific phonetic pattern they are exposed to [6]. In this study, the authors found that following exposure to an atypical accent where the dental fricative /θ/ was produced as [θ/s] (e.g., *throat* as [θ/s]*roat*), listeners shifted their category boundary toward /s/ on a /θ/-/s/ continuum. They also generalized learning to a novel contrast involving the same target phoneme /θ/-/ʃ/, classifying more ambiguous tokens as /θ/. However, no shift was observed for /θ/-/f/. The authors explain this finding based on the high degree of phonetic similarity between [s] and [ʃ], as measured by perceptual confusability data. This suggests that phonetic learning involves some sensitivity to phonetic detail, but that it is general enough to allow for transfer to a distinct pronunciation. Given that non-native speakers may be especially variable [7], maintaining this kind of relatively tolerant strategy may be beneficial for achieving speaker-independent accent adaptation.

However, as recent literature has pointed out [8], it is unclear whether adjustments to category boundaries in fact underlie improvements to comprehension and/or processing of accented speech. The current study tests the question of whether the same mechanism found in [6] leads to improvements in lexical processing following accent exposure. Across two experiments, 137 adult listeners recruited on the Prolific web platform completed a cross-modal priming lexical decision task, following exposure to an unfamiliar accent where a target phoneme was manipulated to be ambiguous (/θ/ = [θ/s]). This task involved presentation of an auditory prime followed by a written word, and listeners were asked to decide whether the latter was a real word or not. Critical trials involved the presentation of written /θ/ words (e.g., <therapy>) paired with either (1) ambiguous 'identity' primes either equivalent or highly similar to the exposure accent (e.g., [θ/s]*erapy* or [θ/ʃ]*erapy)* + <therapy>), (2) unambiguous but phonetically similar related primes (e.g., *serapy* or *sherapy* + <therapy>), or (3) unrelated primes (e.g., *banana* + <therapy>). In Exp. 1, results of linear mixed-effects modeling found a significant interaction of group and experimental condition ($\chi 2(2) = 6.08$, $p < 0.05$), indicating that prior accent exposure affected word processing. Both controls and listeners with prior accent exposure saw similarly large 'identity' priming effects with /θ/ = [θ/s] primes (Fig.1), although there was a trend toward faster RTs for the accent exposure group ($b = -0.127$, $SE = 0.11$, $p = 0.25$). This suggests that these words were sufficiently similar to natural /θ/ that they did not pose significant processing problems. However, listeners in the accent exposure group showed a significantly larger difference between related

trials (*serapy* + <therapy>) and identity trials (b = 0.26, SE = 0.11, p < 0.05) compared to controls. In Exp. 2, listeners with prior accent exposure saw significant related priming (e.g., *sherapy* + <therapy>) in the first half of trials, (b = -0.18, SE = 0.09, p < 0.05), whereas controls saw none. However, controls saw significant learning over the course of the task, with an increase in the size of the priming effect for both identity (b = -0.28, SE = 0.08, p < 0.001) and related prime trials (b = -0.23, SE = 0.08, p < 0.01), whereas listeners with prior experience saw no significant learning effect.

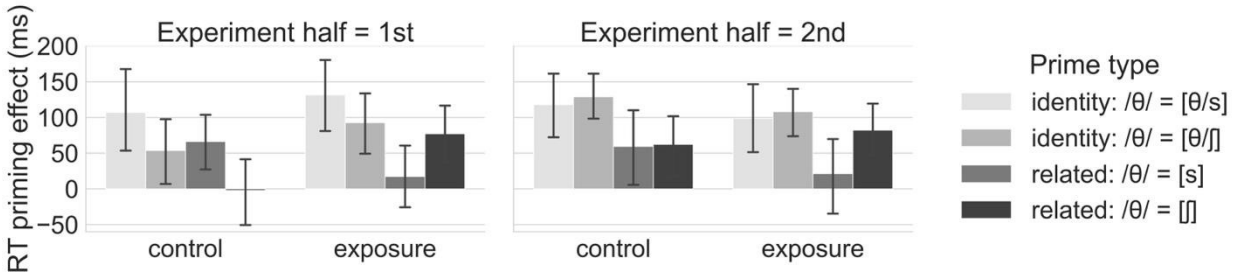

Fig. 1. Cross-modal priming results from Experiments 1 and 2 for listeners with and without prior exposure to a phonetically ambiguous /θ/ = [θ/s] pronunciation. Priming effect calculated by subtracting RTs from trials with 'identity' or related primes from trials with unrelated primes. Error bars indicate boot-strapped 95% confidence intervals.

Together this set of findings shows that phonetic detail plays a complex role in perceptual learning for speech. Although trained listeners showed a trend for stronger 'identity' priming with the ambiguous /θ/ primes vs. controls, accent experience did not yield a significant processing advantage. However, trained listeners did show changes to lexical processing elsewhere, as illustrated in weaker /s/-word priming (*serapy* + <therapy>) but stronger /ʃ/-word priming (*sherapy* + <therapy>) compared to untrained controls, suggesting that training resulted in listeners becoming more tolerant of atypical productions of the target phoneme in some cases (Exp.2) but less tolerant in others (Exp.1). This suggests that the learning mechanism is sensitive to phonetic detail and similarity to previously encountered speech, but that listeners can abstract over differences, facilitating lexical processing in certain novel contexts.

References

[1] Clarke, C. M., & Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *JASA*, *116*(6), 3647–3658.

[2] Whalen, D. H. (1984). Subcategorical phonetic mismatches slow phonetic judgments. *Percept. & Psychophys. 35*(1), 49–64.

[3] Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, *47*(2), 204–238.

[4] Samuel, A. G., & Kraljic, T. (2009). Perceptual learning for speech. *Attent., Perception, & Psychophysics*, *71*(6), 1207–1218.

[5] Schmale, R., Cristia, A., & Seidl, A. (2012). Toddlers recognize words in an unfamiliar accent after brief exposure. Developmental Science, 15(6), 732-738.

[6] Melguy, Y. V., & Johnson, K. (2022). Perceptual adaptation to a novel accent: Phonetic category expansion or category shift? *JASA*, *152*(4), 2090–2104.

[7] Wade, T., Jongman, A., & Sereno, J. (2007). Effects of Acoustic Variability in the Perceptual Learning of Non-Native-Accented Speech Sounds. *Phonetica*, *64*(2–3), 122–144.

[8] Zheng, Y., & Samuel, A. G. (2020). The relationship between phonemic category boundary changes and perceptual adjustments to natural accents. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(7), 1270–1292.