

Factors affecting perceptual learning: Synthesized speech and orthographic support

Charlotte Vaughn – University of Oregon

Molly Babel – University of British Columbia

Michael McAuliffe – McGill University

Perceptual learning of novel pronunciations is a seemingly robust and efficient process for adapting to unfamiliar speech patterns. In this study we examine two factors which may affect listeners' ability to successfully adapt to a novel pronunciation: the use of synthesized speech and the presence of orthographic support. To do this, we investigate the perceptual learning of /s/-words where a medially occurring /s/ is substituted with /ʃ/, rendering, for example, *castle* from /kæsl/ to /kæʃl/. Exposure to the novel pronunciations is presented in the guise of a lexical decision task. Perceptual learning is then assessed in a categorization task where listeners are presented with minimal pair continua (e.g., *sock-shock*).

Previous perceptual learning studies have used a variety of ways to introduce ambiguity of variability to listeners, and have found that listeners can and do perceptually learn from both naturally produced (e.g., Bradlow & Bent, 2008; Witteman, Weber, & McQueen, 2013) and synthetically produced (e.g., Kraljic & Samuel, 2005; Maye, Aslin, & Tanenhaus, 2008; Norris, McQueen, & Cutler, 2003; Weatherholtz, 2015) variation in spoken language. But, to our knowledge, no study has directly compared adaptation processes in synthesized and naturally-produced speech. This is especially important given recent suggestions that generalization of perceptual learning is robust with natural as opposed to synthesized speech (cf. Francis et al., 2016; Weatherholtz, 2015). To better understand the potential effects of synthetic speech on perceptual learning, in this study we compared perceptual learning in groups that either receive natural /s/-to-/ʃ/ words, or /s/-to-/ʃ/ words that have been resynthesized using STRAIGHT (Kawahara et al., 2008).

Further, previous work using this paradigm has not given much attention to the role of explicit training in perceptual learning (though cf. Mitterer & McQueen, 2009 for non-native speech; and Liss et al., 2002 for dysarthric speech). In this study, we give extra information to some participants about the lexical status of the ambiguous words they hear by presenting the orthographic transcription of the word on a screen along with each auditory token for the first half of the exposure trials. Given that the critical tokens in this study are not ambiguous between /s/ and /ʃ/, but instead are a wholesale remapping of /s/ to /ʃ/, listeners may need more information in order to categorize critical items as words (e.g. /kæʃl/ = castle) and not non-words (/kæʃl/ = cashle). This is especially important since previous work has established that perceptual learning effects rely on items being perceived as words by listeners (e.g. Clarke-Davidson et al., 2008).

References

Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, 106(2), 707-729.

- Clarke-Davidson, C. M., Luce, P. A., & Sawusch, J. R. (2008). Does perceptual learning in speech reflect changes in phonetic category representation or decision bias?. *Perception & psychophysics*, *70*(4), 604-618.
- Francis, A. L., MacPherson, M. K., Chandrasekaran, B., & Alvar, A. M. (2016). Autonomic nervous system responses during perception of masked speech may reflect constructs other than subjective listening effort. *Frontiers in psychology*, *7*.
- Liss, J. M., Spitzer, S. M., Caviness, J. N., & Adler, C. (2002). The effects of familiarization on intelligibility and lexical segmentation in hypokinetic and ataxic dysarthria. *The Journal of the Acoustical Society of America*, *112*(6), 3022-3030.
- Kawahara, H., Morise, M., Takahashi, T., Nisimura, R., Irino, T., & Banno, H. (2008, March). TANDEM-STRAIGHT: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, F0, and aperiodicity estimation. In *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on* (pp. 3933-3936). IEEE.
- Kraljic, T., & Samuel, A. G. (2005). Perceptual learning for speech: Is there a return to normal? *Cognitive Psychology*, *51*, 141-178.
- Maye, J., Aslin, R. N., & Tanenhaus, M. K. (2008). The weckud wetch of the wast: Lexical adaptation to a novel accent. *Cognitive Science*, *32*(3), 543-562.
- Mitterer, H., & McQueen, J. M. (2009). Foreign subtitles help but native- language subtitles harm foreign speech perception. *PLoS One*, *4*, e7785. doi:10.1371/journal.pone.0007785
- Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, *47*(2), 204-238.
- Weatherholtz, K. (2015) Perceptual learning of systematic cross category vowel variation, unpublished Ph.D. dissertation, The Ohio State University.
- Witteman, M. J., Weber, A., & McQueen, J. M. (2013). Foreign accent strength and listener familiarity with an accent co-determine speed of perceptual adaptation. *Attention, Perception, & Psychophysics*, *75*, 537– 556. doi:10.3758/s13414-012-0404-y