Phonetic Convergence after Auditory Exposure towards Native and Nonnative Speakers: Perceptual and Acoustic Assessments on Sentences

Midam Kim, midamkim@gmail.com, Linguistics, Northwestern University (oral presentation)

This study investigates phonetic convergence on sentences by native English speakers after auditory exposure to speech by a native speaker of English or a Korean learner of English. Previous studies on phonetic convergence assessed phonetic convergence patterns by either conducting XAB perception tests (e.g. Kim et al, 2011) or measuring specific acoustic cues at the word level, such as VOT (e.g. Nielsen, 2011), vowel formants (e.g. Babel, 2009), and speaking rate of words (e.g. Pardo et al., 2010). In the current study we explore how language distance between a model and an imitator influence the imitators’ phonetic convergence pattern on sentences, which is a larger and more ecologically valid linguistic unit than words, by using both perceptual and acoustic assessments.

We recorded two native speakers of English and two Korean nonnative speakers of English (Models), reading 64 English sentences. A separate group of participants (Imitators) 1) read the 64 sentences, 2) heard the sentences read by one of the model speakers, and 3) read the sentences again. Two types of assessments on phonetic convergence were conducted: XAB perception tests and acoustic measurements. For the XAB perception tests, another separate group of participants (Listeners) heard a model speech sample (X), a pre-exposure speech sample of an imitator (A or B), and a post-exposure speech sample of the same imitator (B or A), and selected either A and B as a better match to the model (X). For the acoustic assessment, we measured “global” acoustic parameters, including duration and F0 range of the sentences. The dependent measure for duration and F0 range was the difference in the Euclidean distance from the model value to the pre-exposure value and from the model value to the post-exposure value. We also applied a dynamic time warping (DTW) algorithm to the original sentence recordings and to hums that were synthesized based on F0 contours extracted from the original sentence recordings. The DTW algorithm aligns the spectrograms of each time frame in each of two signals and returns a value that reflects the “cost” of this alignment process. The DTW dependent measure was thus the difference in cost of alignment of the model and pre-exposure sentences and the model and post-exposure sentences.

Results from the XAB perception tests showed that imitators converged towards both nonnative models and towards one native model. Results from the acoustic assessments showed that all acoustic parameters were significant predictors of the XAB perception test results. However, there were significant model speaker by acoustic parameter interactions. Therefore, we conclude that 1) perceived interlocutor language distance does not necessarily lead to phonetic convergence (closer interlocutor language distance can lead to phonetic maintenance), 2) perceived phonetic convergence can be traced to global acoustic features at the sentence level, and 3) the precise acoustic variations that constitute perceived phonetic convergence vary across different model speakers and their imitators.
References