Variability attenuates sensitivity to acoustic detail in cross-language speech production

In the processing of non-native consonant clusters, speakers' systematic errors have been attributed to the influence of universal factors or native-language phonotactics (e.g., Scholes 1966, Hallé et al. 1998, Dupoux et al. 1999, Moreton 2002, Berent et al. 2007). However, recent studies of non-native speech production suggest that speakers are also sensitive to fine acoustic details in the stimuli (Wilson and Davidson to appear). In this talk, we examine whether speakers' sensitivity to phonetic detail is modulated by variability in the speech signal, and whether they abstract away from subphonemic detail when sufficient acoustic variability is contained in the input. This was tested by presenting English speakers with ill-formed clusters (e.g. bdafa, tmape, zgade) containing systematically manipulated sub-phonemic acoustic properties. Based on previous pilot results (Wilson and Davidson, to appear), three acoustic cues in the stimuli were manipulated: (1) stop bursts increased in duration from 20ms to 50ms; (2) the amplitude of the stop burst was either higher or lower than average relative to the following segment; and (3) voiced stops and fricatives either contained the simultaneous onset of voicing and frication, or they contained pre-obstruent voicing (POV): a short period of voicing that precedes frication or the stop closure. For the single-talker study, it was predicted that longer bursts and higher burst amplitudes would lead to more epenthesis modifications, whereas low amplitude bursts should give rise to more C1 change and C1 deletion. POV should result in more prothetic modifications. For the multiple talker experiment, it was predicted that similar patterns could be found, but the overall effect should be considerably attenuated.

In Experiment 1, which presented stimuli produced by a single Russian talker, English speakers produced stop-initial clusters with significantly less epenthesis at the shorter duration, and produced more epenthesis overall for voiced clusters. Deletion of the first consonant and consonant change errors were significantly more frequent for lower amplitude stop bursts, and there was significantly more prothesis responses when POV was present. In Experiment 2, the same stimuli as in Experiment 1 were produced by three different talkers, though the same speaker as in Experiment 1 was always presented right before the English speakers' productions were recorded. The results for Experiment 2 showed that the duration manipulation did not significantly affect epenthesis rates, though epenthesis was still more frequent for voiced clusters as compared to voiceless clusters. The amplitude manipulation had no effect on either C1 change or deletion. Unlike duration and amplitude, the presence of POV did have an effect, resulting in significantly more prothesis errors, though the effect was smaller than it was in Experiment 1.

Consistent with previous research (e.g., Davidson 2010), the English speakers in this study have a strong tendency to repair unattested consonant clusters with epenthesis. However, the remaining error patterns found in Experiment 1 demonstrate that, in conditions with low overall variability, speakers' responses are affected by low-level acoustic detail. Experiment 2 provides evidence that many effects of sub-phonemic detail on non-native speech production are notably attenuated when there is a higher level of speaker variability in the stimuli. This finding suggests that listeners are generalizing over the acoustic information provided by the multiple talkers, and that they are able to establish a phonemic representation of the word that is not so heavily dependent on the idiosyncratic properties of one stimulus item or one speaker. Yet, at least some of the effects observed in Experiment 1 are present in the results of Experiment 2, suggesting that increased overall variability decreases sensitivity to low-level acoustic variation but does not cause it to vanish entirely.

References

Berent, I., D. Steriade, T. Lennertz and V. Vaknin (2007). What we know about what we have never heard: Evidence from perceptual illusions. *Cognition* **104**, 591-630.

Davidson, L. (2010). Phonetic bases of similarities in cross-language production: Evidence from English and Catalan. *Journal of Phonetics* **38:2**, 272-288.

Dupoux, E., K. Kakehi, Y. Hirose, C. Pallier and J. Mehler (1999). Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance* **25**, 1568-1578.

Hallé, P., J. Segui, U. Frauenfelder and C. Meunier (1998). Processing of illegal consonant clusters: A case of perceptual assimilation? *Journal of Experimental Psychology: Human Perception and Performance* **24:2**, 592-608.

Moreton, E. (2002). Structural constraints in the perception of English stop-sonorant clusters. *Cognition* **84**, 55-71.

Scholes, R. (1966). *Phonotactic Grammaticality*. The Hague: Mouton.

Wilson, C. and L. Davidson (to appear). Bayesian analysis of non-native cluster production. In *Proceedings of the Northeast Linguistics Society 40*.