Acoustic cue integration in natural speech: empirical and computational results

Joseph C. Toscano and Bob McMurray
Dept. of Psychology, University of Iowa
Special Session preferred

Multiple acoustic cues in the speech signal often contribute to a single phonetic categorization. For example, in the perception of word-initial voicing in stop consonants, voice onset time (VOT) and vowel length have both been shown to influence voicing judgments (Summerfield, 1981). However, previous research has found disparate results for the effect of vowel length, with some showing a large effect for synthetic stimuli and others showing a diminished or absent effect for more natural-sounding stimuli (Miller & Liberman, 1979; Shinn, Blumstein, and Jongman, 1985).

We examined whether a larger vowel length effect could be observed in natural speech using an eye tracking procedure that is sensitive to listeners' discriminations between small differences in acoustic detail (McMurray, Aslin, & Tanenhaus, 2002). Participants selected the picture of the stimulus they heard from screens containing the target, its phonological competitor, and two filler objects. The stimuli were natural utterances of minimal pairs (e.g. "beach" and "peach") that were spliced to create nine steps of VOT. The length of the following vowel was also modified to create long and short vowel length conditions.

The results showed a small effect of vowel length on voicing, indicated both by the proportion of fixations to competitor objects and by the pattern of identification responses. However, a previous experiment using synthetic speech showed a larger vowel length effect in both measures. To examine this difference, we measured an additional cue to voicing (F1 onset frequency; Summerfield and Haggard, 1977) in the two sets of stimuli. Acoustic measurements showed that F1 onset provided a useful cue in the natural stimuli, but did not in the synthetic stimuli (it was held constant across VOT steps). We hypothesized that the presence of an additional cue that provides information about voicing will decrease the relative influence of vowel length, resulting in a smaller effect of that cue. Thus, the relevant factor that influences the process of cue integration is not the overall naturalness of the stimuli, but rather the number of usefully covarying cues in the speech signal.

We examined this idea using a computational model of cue integration. The model uses a statistical learning mechanism to determine the categories present in the input for particular acoustic cues. It then uses the reliability of those cues to weight and combine them into a single representation of voicing. The results of the simulations indicate that the presence of a reliable third cue (F1 onset) affects the size of the trading relation between VOT and vowel length, consistent with the empirical results. Thus, it appears that by tracking the statistical reliability of individual acoustic cues, the speech system is able to integrate them during the process of speech perception.

References

McMurray, Aslin, & Tanenhaus (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition 86(2),* B33-B42.

Miller, J. L. and Liberman, A. M. (1979). Some effects of later-occurring information on the perception of stop consonant and semivowel. *Perception & Psychophysics, 25(6),* 457-465.

Shinn, P. C., Blumstein, S. E., and Jongman A. (1985). Limitations of context conditioned effects on the perception of [b] and [w]. *Perception & Psychophysics, 38(5),* 397-407.

Summerfield, Q. and Haggard, M. (1977). On the dissociation of spectral and temporal cues to the voicing distinction in initial stop consonants. *Journal of the Acoustical Society of America, 62(2),* 435-448.

Summerfield, Q. (1981). Articulatory rate and perceptual constancy in phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance, 7(5),* 1074-1095.