Consonants versus vowels: Phonetic changes under acoustic feedback transformation

When the unit of phonological or phonetic analysis is assumed to be segmental, as opposed to syllabic or morphemic, the consonant/vowel distinction forms a critical divide in linguistic sound systems. Prosodic behavior differs, with consonants appearing at syllable margins, for example, and depending often on vocalic syllable nuclei for perceptual transmission and phonotactic well-formedness, while vowels may occur without associated consonants. Phonological sonority, the number of attested phones, and the trends of historical and sociolinguistic variation all differ between the consonant and vowel segment classes.

Finding empirical evidence for the dissociation between consonants and vowels is not always easy, however, and the line between the two classes is not always clear. This paper describes a novel methodology for phonetic research which promises to open new avenues of research in the area, investigating the role of real-time acoustic feedback in speech production. The importance of acoustic feedback (as opposed to somatosensory feedback) in speech production has been hypothesized to differ for consonants and vowels (Guenther et al., 1998), making feedback manipulation methods well-suited to investigations of the psychological reality of these distinct segmental classes.

In this experiment, a portable, real-time vocoder (PRTV) was used to transform acoustic speech feedback. The device captures acoustic signals and alters them in a simulation of the processing done by cochlear implants (Casserly et al., 2011). As a result, spectral resolution is degraded substantially (see Fig. 1) and very high- and low-frequency signals are eliminated (cf. Shannon et al., 1995). We hypothesized that the substantial loss in phonetic detail, particularly for fine-grained frequency contrasts, would have a substantial impact on speech production (cf. Matthies et al, 1996; Lane et al., 2007).

Speech samples were recorded from seven subjects, both under normal speaking/listening conditions and while wearing the PRTV. In each condition, speakers produced 12 “phoneme specific sentences” (Huggins & Nickerson, 1985) and 114 English words in isolation, containing 10 tokens each of [i, æ, ɑ, u], 16 of [s, ʃ] and 8 of [p, t, k]. Phonetic analysis of the consonants and vowels produced in isolated words revealed significant changes as a result of real-time feedback transformation, but of very distinct types. Fricatives and stops experienced global, general changes, both in duration (MANOVA, $p < 0.01$) and in the frequency of frication in sibilants ($p < 0.001$). Vowel quality, conversely, was affected asymmetrically: high vowels [i, u] were unchanged as a result of feedback transformation, but the low vowels [æ, ɑ] were consistently raised (significant Condition*Vowel interaction in F1, $p < 0.01$).

Moreover, when subjects were exposed to the feedback transformation, they appeared to focus more heavily on the articulation of consonants as opposed to vowels; the ratio of consonant to vowel articulation duration (for stops, fricatives, and affricates) increased significantly across speaking conditions ($p < 0.01$). The tendency for subjects to spend more time articulating consonants than vowels under conditions of challenging acoustic feedback control seems to support claims that somatosensory and acoustic feedback operate in a trading relationship across these two classes of segments. When faced with uncertain acoustic feedback, speakers rely more on the unperturbed consonant feedback. We hope to explore this and other possible accounts in future work investigating the nature of these segments and their sensory goals. Regardless of the theoretical account invoked in their interpretation, however, these results point to differential behavioral control of consonants and vowels during production. Accounts which predict phonological units at the segmental level in cognition and speech motor control, therefore, are supported.

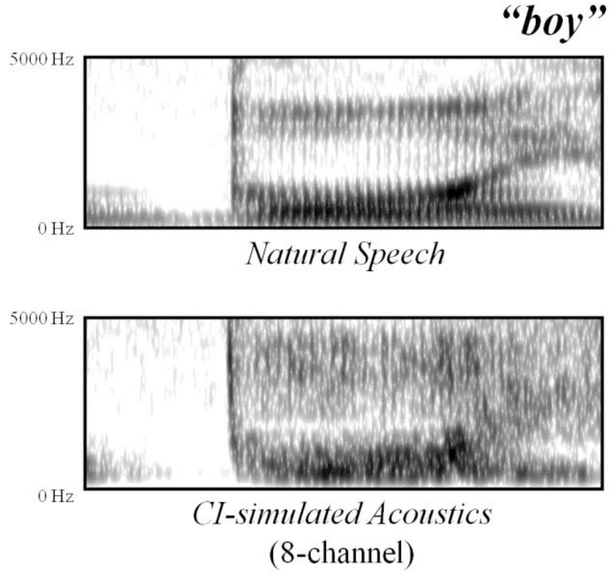

Figure 1 - Natural and "vocoded" (transformed) speech acoustics

References

Casserly, E., Pisoni, D. B., Smalt, C., & Talavage, T. (2011). A portable, real-time vocoder: Technology and preliminary perceptual learning findings. *Journal of the Acoustical Society of America, 129*, 2527.

Guenther, F. H., Hampson, M., & Johnson, D. (1998). A theoretical investigation of reference frames for the planning of speech movements. *Psychological Review, 105*, 611-633.

Huggins, A. W. F. & Nickerson, R. S. (1985). Speech quality evaluation using "phoneme-specific" sentences. *Journal of the Acoustical Society of America, 77*(5), 1896-1906.

Matthies, M. L., Svirsky, M., Perkell, J., Lane, H. (1996). Acoustic and articulatory measures of sibilant production with and without auditory feedback from a cochlear implant. *Journal of Speech and Hearing Research, 39*, 936-946.

Lane, H., Matthies, M. L., Guenther, F. H., Denny, M., Perkell, J. S., Stockmann, E., Tiede, M., Vick, J., & Zandipour, M. (2007). Effects of short- and long-term changes in auditory feedback on vowel and sibilant contrasts. *Journal of Speech, Language, and Hearing Research, 50*, 913-927.

Shannon, R.V., Zeng, F., Kamath, V., Wygonski ,J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science, 270*, 303-304.