A voicing asymmetry in nonnative cluster epenthesis: perception vs. production

Colin Wilson, Johns Hopkins University Lisa Davidson, New York University

Introduction. A long-standing goal of research on sound systems is to understand the phonetic bases of phonological patterns (e.g., Ohala, 1983; Flemming, 1995[2002]; Blevins, 2004; Hayes et al., 2004). Many processes and static restrictions, such as consonant place assimilation (e.g., Steriade, 2001) and consonant deletion (e.g., Côté, 2004) as well as the distribution of voicing and other features (e.g., Steriade, 1997; White, 2014), have been shown to mirror perceptual similarity relations. Vowel epenthesis into a consonant cluster is also perceptually grounded: epenthesis is more frequent when there is a strong 'perceptual break' between the two consonants (e.g., Fleischhacker 2005; Zuraw, 2007).

On the basis of evidence from a previous study of nonnative cluster production, and new perceptual identification results collected for the same clusters, we argue that speech production makes an independent contribution to detailed epenthesis patterns. Spoken and identification responses are sensitive to many of the same properties, but only speech production shows a strikingly higher rate of epenthesis into clusters that begin with a voiced stop (e.g., /bn/ vs. /pn/).

Production study. As previously reported in Wilson et al. (2014), English speakers (N=24) listened to and produced CCVCV nonwords beginning with a range of nonnative consonant clusters, including stop-nasal and stop-stop clusters of interest here (e.g., /km/, /gm/, /kp/, /gb/). Several coders analyzed waveforms and spectrograms to identify instances of epenthesis (77% of all errors) and other modifications. The most extreme finding was that clusters beginning with *voiced* stops showed higher rates of epenthesis (46% of all responses) than those with initial *voiceless* stops (only 21%). Additionally, epenthesis was more frequent for *stop-nasal* clusters than for *stop-stop* clusters (37% vs. 30%), an anti-sonority-sequencing pattern that has been replicated in similar experiments (e.g., Davidson, 2010). Epenthesis rate also tracked the *release duration* of the initial stop release (20ms: 44% epenthesis responses vs. 50ms: 52% epenthesis).

Identification studies. The effects found previously could reflect asymmetries in perceptual epenthesis across clusters (e.g., Berent et al., 2007) or arise in the speech production process. Two new perception experiments, conducted on Mechanical Turk, sought to distinguish these possible phonetic origins. In Experiment 1, English-speaking participants (N=90) completed a forced-choice identification task on exactly the same stimuli used in the production study. On each trial four options were presented in pseudo-English orthography, representing the target nonnative cluster, epenthesis, prothesis, and deletion or one-feature change of the initial stop (e.g., *bdazo*, *bedazo*, *ebdazo*, *dazo*/*gdazo*). The arrangement of response options was randomized across participants. The entire stimulus set was divided into lists of 28 items; each list was used for two participants and contained a balance of cluster types and release durations. The lists also included filler items beginning with CəC and əCC sequences consonants matched to the critical clusters. The fillers elicited very high accuracy in the production study; in the present experiment they provide a measure of participants' ability to perform the web-based identification task.

The results of Experiment 1 and those of the production experiment were combined into a single data set and submitted to a mixed-effects logistic regression with epenthesis as the binary dependent variable. Note that the prothesis and deletion/feature change response options were chosen infrequently (< 5% of total responses each), justifying a focus on the epenthesis repair. Rate of epenthesis was significantly influenced by cluster type (SN > SS; $\beta = 0.76$, SE = 0.19, $p < .001$), cluster voice (vcd > vcl; $\beta = 1.07$, SE = 0.20, $p < .001$), and release duration (20 ms > 50 ms; $\beta = 0.70$, SE = 0.10, $p < .001$). There was one significant interaction, which indicated that the voice effect differed in the two studies (β = 2.03, SE = 0.24, $p < .001$).

Subsequent analysis of the results from Experiment 1 alone indicated significant effects on epenthesis of cluster type (SN 44% > SS 34%; β = 0.76, SE = 0.21, $p < .001$) and release duration (longer 44% > shorter 33%; β = 0.89, SE = 0.16, $p < .001$)—but *no effect of cluster voice* (vcd 38.7% ≈ 39.2% vcl; $p = .89$), contrary to the strong asymmetry in the production experiment. The disparity with production cannot be due to difficulties in performing the transcription task: the SN vs. SS and release duration effects were found in both experiments, and furthermore the identification of filler items was quite accurate (> 93% correct).

A second identification experiment was performed to ensure replicability of these findings and control for possible strategic effects. Experiment 2 contained the same critical items, but there were more fillers per list (20/36 items) and the set of fillers included items that matched the deletion and change (as well as epenthesis and prothesis) response options. The results confirmed the interaction of response type (production vs. identification) and voice (β = 1.87, SE = 0.20, $p < .001$) and the absence of a significant voice effect in identification ($p = .72$).

Discussion. This pattern of findings supports a modular account of nonnative cluster processing in which perception, phonology, and production make separable contributions. The release duration effect plausibly results from a combination of perceptual similarity—longer transitions between consonant closures are more acoustically consistent with a reduced vowel—and phonological bias against nonnative clusters. Perhaps the greater epenthesis rate on SN clusters has a similar source, with nasal formants (cf. oral stop closure) being misparsed as vocalic material. However, the effect of voice appears to emerge downstream, in production: all stop-initial nonnative clusters are subject to error in gestural timing (Davidson, 20006); an interval of inaccurate vocal tract opening that is accompanied by voicing is more likely to produce vowel-like formant structure. Preliminary simulations with an articulatorily-based synthesizer are consistent with the idea that equivalent levels of gestural mistiming can lead to clearer formant structure after voiced (vs. voiceless) stops. Our results converge with previous experiments that have not found voicing asymmetries in perceptual tasks (e.g., Davidson & Shaw, 2012). They also provide a novel type of evidence for restrictive theories (e.g., Steriade, 1997; Lombardi, 2001) according to which the [voice] feature cannot participate in conditioning vowel epenthesis within the phonological component. The perceptual grounding and phonological triggering of vowel epenthesis appear to be restricted to supralaryngeal properties.

References. **Berent, I.**, Steriade, D., Lennertz, T., & Vaknin, V. (2007). What we know about what we have never heard: Evidence from perceptual illusions. *Cognition*, 104(3), 591-630. **Blevins, J.** (2004). *Evolutionary Phonology: The Emergence of Sound Patterns*. Cambridge: Cambridge University Press. **Côté, M.-H.** (2004). Syntagmatic distinctness in consonant deletion. *Phonology*, 21(1), 1-41. **Flemming, E.** (2002). *Auditory Representations in Phonology*. New York: Routledge. **Hayes, B.**, Kirchner, R., & Steriade, D. (Eds.). (2004). *Phonetically-Based Phonology*. Cambridge: Cambridge University Press. **Lombardi, L.** (2001). Why Place and Voice are different: constraint-specific alternations in Optimality Theory. In Lombardi, L. (ed.), *Segmental Phonology in Optimality Theory: constraints and representations*. Cambridge: Cambridge University Press, 13-45. **Steriade, D.** (2001). Directional asymmetries in place assimilation: a perceptual account. In Hume, E., and Johnson, K. (eds). *The Role of Speech Perception in Phonology*, 219-250. San Diego: Academic Press. **Zuraw, K.** (2007). The role of phonetic knowledge in phonological patterning: corpus and survey evidence from Tagalog infixation. *Language*, 83(2), 277-316. **White, J.** (2014). Evidence for a learning bias against saltatory phonological alternations. *Cognition*, 130(1), 96-115.