Effects of recoverability on perception of illusory vowels

A series of studies by Dupoux and colleagues (Dehaene-Lambertz et al., 2000; Dupoux et al., 1999, 2011) found that Japanese listeners tend to perceive a vowel between consonant clusters even in the complete absence of a vocalic segment. The authors argue that this is due to the CVCV phonotactic restriction of Japanese. I provide experimental evidence that it is not just phonotactics but also sensitivity to coarticulatory cues that drive the perceptual epenthesis in Japanese listeners and that the recoverability of the vowel in contexts where high vowel reduction is expected determines how sensitive listeners are to the available cues.

Japanese high vowel reduction is a phenomenon where unaccented high vowels /i, u/ are devoiced or deleted when flanked by two voiceless obstruents. Varden (2010) states that when only one of the two high vowels can follow a given obstruent, the vowel can delete since the vowel is fully predictable and the context alone is sufficient for recovery. Conversely, when both high vowels can follow a given obstruent, the target vowel is less predictable and thus devoices because coarticulatory cues are necessary for recovery. While Japanese speakers have been shown to be sensitive to some coarticulatory cues (Beckman and Shoji, 1984; Dupoux et al., 2011), the effects of predictability on perception have not been tested systematically.

The experiment for this study was a forced-choice labeling task with thirty native Japanese participants. The stimuli were of the form $VC_1(V_T)C_2V$, where V_T was a target vowel ([i, a, u]) and C_2 a voiceless stop ([p, t, k]). The stimuli consisted of three contexts: non-reducing (C_1 = [b, z, g]), high predictability reducing where only one of the two high vowels is allowed in Japanese (C_1 = [ɸ, s, ç]), and low predictability reducing where both high vowels are allowed (C_1 = [p, ʃ, k]). The stimuli were created by recording tokens with full target vowels, then removing parts of the vowels until no vowel remained. There were four answer choices for each stimulus (e.g., [eʃupo] → <eʃpo>, <eʃapo>, <eʃipo>, <eʃupo>).

Linear mixed-effects models were fit to test the effects of context and target vowel on vowel identification. Performance was near ceiling regardless of context and target vowel when a vowel was present. When no vowel was present, performance for the low predictability context was highest at 67% compared to non-reducing (51%; t = -6.486) and high predictability (42%; t = -8.530) contexts. Also, performances for both [i] (71%) and [u] (59%) were significantly higher than for [a], a non-reducing vowel (25%; t = 13.226 and 14.618, respectively).

The overall high accuracy in vowel identification suggests that the perception of "illusory" vowels by Japanese listeners are not simply due to phonotactic repair but sensitivity to coarticulatory cues. Furthermore, listeners seem more reliant on coarticulatory cues when predictability alone is insufficient for recovery. Recoverability-conditioned gestural coordination has been proposed by Chitoran et al. (2002) and Silverman (1995), and this study provides evidence that perception is conditioned by recoverability as well.