**Targetless /u/ in Tokyo Japanese**

Shigeto Kawahara (Keio U.), Jason A. Shaw (Western Sydney U. & Keio U.), James Whang (NYU)

kawahara@icl.keio.ac.jp & J.Shaw@westernsydney.edu.au

**Introduction:** The high, non-front vowel in Tokyo Japanese (TJ), variably transcribed as [u], [ʊ], [ɯ], shares with English schwa an ensemble of “default vowel” properties: shortest and most central vowel in the TJ system (Campbell, 1999; Nogita et al., 2013), deleted (or at least devoiced) in some environments (Kawakami, 1977; c.f., Vance (1987:52-55)), inserted in others, e.g., English loanword “spring” is [ˈsuːprɪŋɡu], and perceived between consonants with negligible acoustic evidence (Dupoux et al., 1999). Given this cluster of properties, we explored the possibility that TJ /u/ lacks a lingual target, as has been explored for predictable schwas in English (e.g., Lammert et al., 2014).

**Methods:** We used an NDI Wave Electromagnetic Articulography System to transduce lingual articulatory movements during the speech of two female native TJ speakers. Sensors were attached to the tongue tip (TT), body (TB), dorsum (TD), lips, jaw, nasian and mastoids. Acoustic data were recorded simultaneously with a shotgun microphone. Lingual motion was examined after computationally correcting for head movement. Stimuli consisted of 10 words containing /u/ in C1-C2 environments, where C1 = {/s/, /k/, /ʃ/, /ɸ/, /ts/}. For each C1, we included one word with a voiceless C2 to create the environment for high-vowel devoicing and another word with a voiced C2 counterpart forming a minimal, or near minimal, pair, e.g., /ɸusoku/ ‘shortage’ ~ /ɸuzoku/ ‘enclosed’, /ʃutaisei/ ‘individuality’ ~ /ʃudaika/ ‘theme song’. Words were read 9-10 times each in the carrier phrase: ookee ______ to itte ‘Ok, say ______’.

**Predictions:** We hypothesized that there would be differences in TD trajectories in voiced and devoiced vowel contexts, owing to the deletion of devoiced vowels in (at least) some contexts, possibly just those contexts in which /u/ is predictable from the preceding consonant (Whang, 2014).

**Results:** In accordance with the literature on high vowel devoicing, we observed complete devoicing of /u/ between voiceless consonants. However, we found small effects of vowel voicing on TD trajectories. The identity of C1, on the other hand, had a substantial influence on TD trajectory as well as tongue shape as Figures 1 and 2 illustrate. For clarity in presentation, the figures show just 5 words (9-10 reps each) from one of our speakers, although both speakers showed the same pattern. Figure 1 shows the change over time (x-axis) of the TD (upper panel) and TT (lower panel) sensors in the vertical dimension (y-axis) 200 ms before and after /u/, as determined by the point of minimum velocity in the TD sensor between flanking consonants. Each line represents a different token. The zero point on the y-axis is the bite between the front incisors. The colors of the lines represent different C1-C2 environments (note that we have collapsed over C2 voicing distinctions). The key observation is that the vertical position of the TD for /u/ varies in accordance with the articulatory demands of the flanking gestures. This is particularly clear for /ɸu{s,z}/ and /ʃu{t,d}/ environments where the TD begins high for the /e/ of the carrier phrase. Instead of rising from /e/ to /u/, the TD lowers from /e/ and follows an almost linear trajectory towards /{o,a}/ targets found later in the word. The prospect of an /u/ target in /katsudo/ (blue line) is somewhat more promising, as TD rises from its position for /a/ to /u/ before lowering again for /o/, but this rise is possibly just a consequence of /ts/ production, as coronal consonants receive substantial support from the TD in articulation (Iskarous et al., 2010). Moreover, the rise observed from /a/ to /u/ still does not bring the TD to the height of /u/ in other environments (where it lowers from /e/). Converging evidence for the impact of surrounding consonants on /u/ articulation can be found in Figure 2, which shows the spatial location of the three tongue sensors at the 200ms timestamp from Figure 1.
To estimate tongue shape, a polynomial was fit to the mean position of each lingual sensor according to preceding consonantal environment. The estimated tongue shapes reflect the demands of the flanking consonants. In /ʃ\_\{s,z\}/, where the TD is unconstrained by the preceding consonant, we see the most canonical tongue shape for /u/ (c.f., Nogita et al. 2013), but, as observed in Figure 1, TD height is lower in this environment than /e/ in the carrier phrase. In the /ʃ\_\{t,d\}/ context, the tongue shape approximates a palatal constriction with the TB higher than TD. In /ts\_\_d/, the tongue shape for /u/ is flat, suiting the articulatory requirements of the coronal affricate. Thus, in three different environments, we observe three distinct tongue shapes, each of which constitutes an intermediate position between flanking gestures.

**Discussion:** The articulation of TJ /u/ in this data appears largely determined by the articulation of surrounding gestures, an observation which holds across voicing environments. This result has implications for the interpretation of “perceptual illusions”, i.e., Japanese listeners’ propensity to perceive /u/ between non-native consonant clusters (Dupoux et al., 1999). Japanese listeners may perceive /u/ not so much because of the phonotactic illegality of C₁C₂ sequences but because the articulatory dynamics of /C₁C₂/ sequences in other languages may closely resemble C₁uC₂ sequences in TJ, a similar explanation to that offered for “illusory” schwa perceived by English speakers (Davidson and Shaw, 2012).

**Selected References:**