Effects of speech rate on Korean stop perception

Yoonjung Kang & Na-Young Ryu
University of Toronto Scarborough & The Pennsylvania State University

Introduction: Studies have shown that contextual speech rate can affect the perception of durational contrasts and that identical target stimuli can be perceived as "long" when embedded in fast speech and as "short" in slow speech (Miller, et al. 1984). However, studies also suggest that speech rate effects are not consistently found across different types of contrasts or different speech rate manipulations (Heffner, et al. 2017). The current study aims to examine if and how the speech rate affects the perception of three-way laryngeal contrasts of Korean, which are signaled by a combination of durational (VOT) and spectral (F0) cues (Cho, et al. 2002). Korean stops present an interesting case study given the reported merger of VOT between lenis and aspirated stops in younger Seoul speakers' speech. Korean stops also present an analytical challenge given its three-way distinction (Schertz, et al. 2019).

Methods: A young female native speaker of Seoul Korean (26 yrs) produced 27 monosyllabic CV words ($C \in \{p, t, k, p', t', k', p^h, t^h, k^h\}$ $V \in \{a, o, u\}$) embedded in a carrier sentence ('What I want to say is ___') in a randomized order 5 times. The VOT and F0 at the onset of the following vowel were measured to define the speaker's range of production values. The carrier sentence, the closure, and the aspiration are spliced from a token of /p^ha/, and the post-stop vowel /a/ is taken from a token of /p'a/. This baseline token was manipulated to vary orthogonally in VOT (12 equal steps: 5ms ~ 125ms), F0 (3 steps: 210 Hz, 240 Hz, & 270 Hz), and two speech rates (carrier sentence duration: fast = 930 ms vs. slow = 1700 ms) (See Figure 1) and stimuli were repeated twice. 30 Korean speakers (mean age =27, range=20~38) participated in an online experiment where they identified the target syllable (<pa>, <p'a>, <p^ha>).

Results: Figure 2 summarizes that response patterns by the VOT, f0 and speech rate conditions and the responses are minimally affected only in a few VOT and f0 conditions. Logistic mixed-effects models were used for analysis. The three-way responses were converted into three binary variables (aspirated vs. others; fortis vs. others; lenis vs. others) for three separate analyses. In the initial models, the predictors included VOT, F0, speech rate, and their interactions. As the effects of VOT and F0 are expected to be non-linear, the variables were converted into categorical variables (shown by the grid in Figure 1). The results show significant main effects of both VOT and F0 for all three models and a significant effect of speech rate for aspirated and lenis stop models—aspirated responses were more likely in fast than slow speech and lenis response were more likely in slow than fast speech—while fortis responses did not differ across speech rate conditions. Although the interaction of speech rate and acoustic predictors were not significant, this seems to be due to lack of power rather than an indication that the speech rate affects the entire acoustic space for aspirated and lenis stop perception. So, follow-up tests of rate effect were conducted separately for each of the 12 acoustic areas. These local analyses show that the rate effects are found only in specific parts of the acoustic space (brown circles in Figures 1 and 2), where the stops are sparsely attested in natural production.

Discussion: The results show that speech rates have little effect on the perception of Korean stops, and the effect found is much smaller in size than expected given the extent of VOT overlap in production and the speech rate variation (Oh, 2009). It is tempting to interpret the results as an indication of the reduced role of VOT in younger Seoul Korean's speech given the sound change. However, this is in contrast to findings from Dutch /ɑ/-/a:/ contrast, where the contrast is signalled by both duration and spectral cues, and speech rate effects on duration perception were found regardless of spectral manipulation of the stimuli (Bosker, 2017). Future studies will examine the generational difference in speech rate effects to probe how the changing status of durational cues interacts with speech rate effects.

Figure 1: Distribution of produced tokens and manipulated stimuli

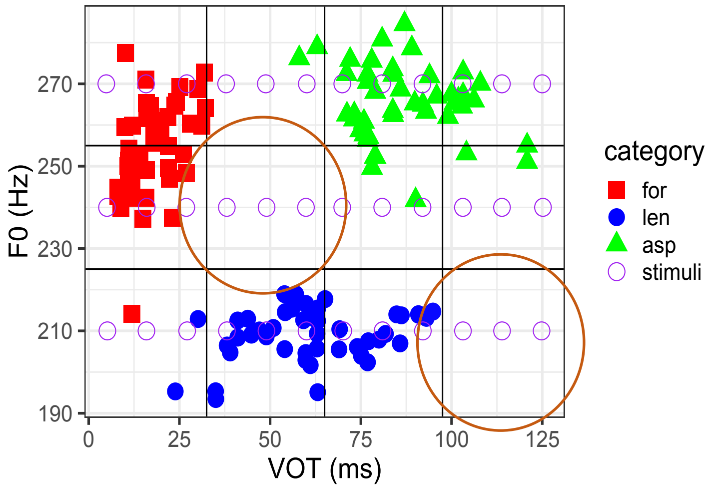
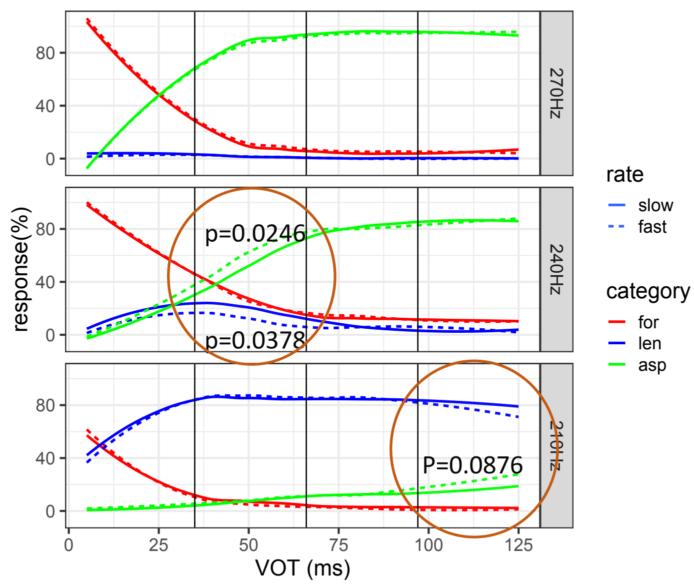


Figure 2: Identification response distribution by the acoustic condition and the speech rate, aggregated over all 30 participants



References

Bosker, H. R. (2017). Accounting for rate-dependent category boundary shifts in speech perception. *Attention, Perception, & Psychophysics, 79*(1), 333-343.

Cho, T., Jun, S.-A., & Ladefoged, P. (2002). Acoustic and aerodynamic correlates of Korean stops and fricatives. *Journal of Phonetics, 30*, 193-228.

Heffner, C. C., Newman, R. S., & Idsardi, W. J. (2017). Support for context effects on segmentation and segments depends on the context. *Attention, Perception, Psychophysics, 79*(3), 964-988.

Miller, J. L., Grosjean, F., & Lomanto, C. (1984). Articulation rate and its variability in spontaneous speech: A reanalysis and some implications. *Phonetica, 41*(4), 215-225.

Oh, E.-J. (2009). Voice onset time of Korean stops as a function of speaking rate. *Phonetics and Speech Sciences, 1*(3), 39-48.

Schertz, J., Kang, Y., & Han, S. (2019). Sources of variability in phonetic perception: The joint influence of listener and talker characteristics on perception of the Korean stop contrast. *Laboratory Phonology: Journal of the Association for Laboratory Phonology, 10*(1).