Effects of Speech Rate on Korean Stop Perception

Yoonjung Kang[1,2] & Na-Young Ryu[3]

[1]University of Toronto Scarborough, [2]University of Toronto, [3]Pennsylvania State University

Introduction

- **Speech rate and Perception**: Contextual speech rate can affect the perception of durational contrasts and identical target stimuli can be perceived as "long" when embedded in fast speech and as "short" in slow speech (Miller, Grosjean, & Lomanto, 1984) but studies also suggest that speech rate effects are not consistently found across different types of contrasts or different speech rate manipulations (Heffner, Newman, & Idsardi, 2017).
- **Korean stops**: The three-way laryngeal contrasts of Korean stops between aspirated, lenis, and fortis stops are signaled by a combination of durational (VOT) and spectral (F0) cues (Cho, Jun, & Ladefoged, 2002). Korean stops present an interesting case study given the reported merger of VOT (a durational cue) between lenis and aspirated stops in younger Seoul speakers' speech (Kang, 2014). Korean stops also present an analytical challenge given its three-way distinction (Schertz, Kang, & Han, 2019).
- **Goals:** Current study examines if and how the speech rate affects the perception of lenis vs. aspirated stop contrasts in younger Korean listeners for whom the VOT, a durational cue, plays a limited role.

Methods

- **Stimuli speaker production**
  - Stimuli speaker: A young female native speaker of Seoul Korean (26 yrs old)
  - 27 monosyllabic CV words (where C ∈ {p, t, k, p', t', k', p^h, t^h, k^h} V ∈ {a, o, u}) embedded in a carrier sentence 내가 하려는 말은___ ('What I want to say is ___').
  - The VOT and F0 at the onset of following vowel were measured to define the speaker's range of production values (Figure 1)

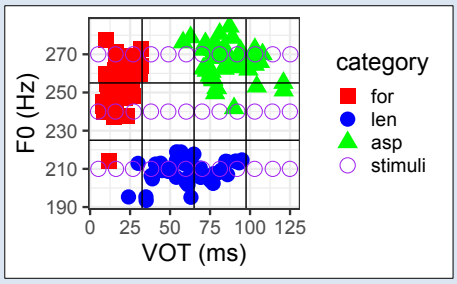

Figure 1: Distribution of produced tokens & manipulated stimuli

- **Stimuli manipulation**
  - A baseline token was created by splicing together a representative token of aspirated /p^ha/ (Carrier sentence + Closure + Aspiration) + fortis /p'a/ (Vowel). (Figure 2)
  - **Manipulation** (Figure 1)
    - **VOT:** 12 equal steps (5ms ~ 125ms)
    - **F0** at onset : 3 steps (210, 240, & 270 Hz)
    - **Speech rates**: 2 steps of carrier sentence duration (fast = 930 ms, slow = 1700 ms)

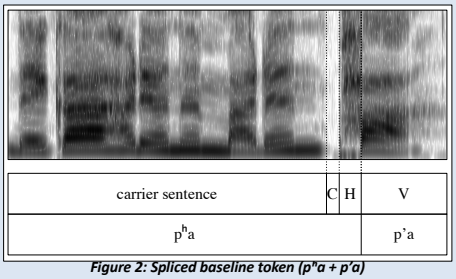

Figure 2: Spliced baseline token (p^ha + p'a)

- **Participants**
  - 30 young Seoul Korean listeners recruited online (mean age = 27 yrs, range = 20 ~ 38)
- **Task**
  - Online experiment built using jsPsych (De Leeuw, 2015)
  - Experiment link: http://phonlab1.utsc.utoronto.ca/~phonlab/TWC_complete/M2/M2.html
  - Identification by a mouse click (바<pa>, 빠<p'a>, 파<p^ha>)
  - 144 trials = 12 VOT steps * 3 F0 steps * 2 speech rates * 2 repetitions
- **Statistical analyses**
  - Mixed-effects logistic regression in R
  - **Response variables:** The responses were converted into three binary variables (aspirated vs. others; fortis vs. others; lenis vs. others) for three separate analyses.
  - **Predictors** as categorical variables due to anticipated non-linear effects:
    - **VOT** (short: 5~27ms, mid: 38~60ms, long: 70~92ms, super-long: 103~125ms)
    - **F0** (low: 210Hz, mid: 240Hz, and high: 270Hz)
    - **Rate** (slow vs. fast)
    - Interactions
  - **Random effect:** random intercept for participants and by-rate slope where converging
    `glmer(Resp ~ VOT*F0*Rate+(Rate|participant), "binomial")`
  - **Follow-up models:**
    - Separate analyses for 12 VOT/F0 conditions
    - **Rate** as a predictor and random intercept and slope for participant
      `glmer(Resp ~ Rate+(Rate|participant), "binomial")`

Results

- **Overall results – full models**
  - **Fortis**: no main effect of speech rate or interaction
  - **Lenis**: a significant main effect of Rate and no interaction of Rate and VOT/F0
    - Fast~***less*** lenis responses: dotted lines are lower than solid lines in Figure 3.
  - **Aspirated**: a significant main effect of Rate and no interaction of Rate and VOT/F0
    - Fast~***more*** aspirated responses: dotted lines are lower than solid lines in Figure 3.
  - Lenis and Aspirated are trading off of each other.
  - Overall, the effect is minimal and localized. The lack of significant interaction of Rate and VOT/f0 is likely due to lack of power.

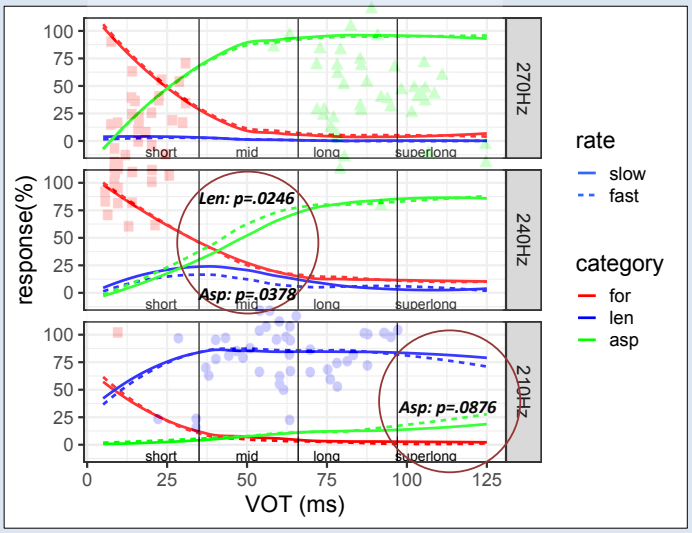

Figure 3: Identification response by the acoustic condition and the speech rate, aggregated over all participants, with stimuli speaker's production distribution from Figure 1 overlaid for reference. The brown circles indicate VOT/F0 contexts where rate effects are significant for Lenis and Aspirated stops based on the follow-up tests.

- **Follow-up local models**
  - Separate analyses of each VOT/F0 conditions, to probe the local effects of rate closely.
  - These local analyses show that the rate effects are found only in very specific parts of the acoustic space (indicated by brown circles in Figure 3), where none of the three stops categories are attested in natural production.

Conclusions

- **Speech rate and Korean stop perception**
  - Our results show that speech rates have little discernable effect in the perception of Korean stops and any effect we found is much smaller in size than may be expected if perception mirrors the extent of overlap in VOT, especially between lenis and aspirated stops (as shown in Figure 1) and the speech variation in production (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 finding 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 in Korean to probe how the changing status of durational cues interacts with speech rate effects, by varying the age of the stimuli speaker and as well as the listeners.

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.

De Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior research methods, 47*(1), 1-12.

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.

Kang, Y. (2014). Voice Onset Time merger and development of tonal contrast in Seoul Korean stops: A corpus study. *Journal of Phonetics, 45*, 76-90.

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), 1-32.

Acknowledgements: The research is funded by the University of Toronto Scarborough Research Competitiveness Fund and SSHRC Grant # 435-2020-0209. Thanks to Hyoung Seok Kwon for help with experiment development.