Rate-dependent speech perception interacts with prosodic phrasing in Korean

Jeremy Steffman, Sahyang Kim, Taehong Cho & Sun-Ah Jun

Introduction: Rate-dependent speech perception is the phenomena by which temporal cues in speech are perceived relative to context. A preceding fast context causes subsequent cues to be perceived as relatively slow/long (e.g. an ambiguous voice onset time (VOT) will be perceived as longer); a preceding slow context makes the same speech sound relatively fast (e.g., a VOT value will be perceived as shorter). These effects are usually understood via auditory *contrast* mechanisms [1,2]. However, prosodic phrasing/grouping, which is often cued by local duration modulations around a prosodic juncture, has also been shown to generate predictive or expectation-based effects on temporal cue processing, based on prosodic strengthening [2,3]. In this study we test how speech rate context *interacts with prosodic phrasing* in Korean, exploring how rate effects operate across prosodic contexts. We test the perception of the contrast between aspirated stops (/kʰ/ with long VOT and a short following vowel) and fortis stops (/k*/ with short VOT and a long following vowel). **Korean prosodic phrasing** induces "domain-initial strengthening" in speech production [4,5]. For aspirated stops, this entails VOT lengthening. For fortis stops, this entails slight VOT shortening, or little change. For *both* stop categories, the subsequent vowel in a CV sequence is lengthened. [3] showed that these patterns have perceptual consequences. When there was a prosodic contextual cue that would lead to an expectation of domain-initial strengthening, the vowel in the CV sequence was perceived as shorter than when there was no such cue (i.e. due to expected phrase-initial vowel lengthening, a vowel "needed to be longer" to cue /k*/). Crucially, this effect of prosodic grouping was observed with temporal context controlled and only F0 cuing prosodic structure. **Method:** 32 speakers of Seoul Korean categorized a VOT continuum as /k*/ (tense) or /kʰ/ (aspirated), in the carrier phrase shown in Figure 1. Six conditions crossed **rate** (normal vs. slow) and prosodic **phrasing** (no boundary vs. boundary vs. boundary+pause). Rate manipulation resulted from the linear expansion of pre-target material (distal context and local context in Fig. 1). The phrasing manipulations involved lengthening only the pre-target syllable, resulting in an intonational phrase (IP) boundary condition, and creating a stronger boundary condition by introducing a pause (IP+pause condition). Responses (252 per participant) were analyzed with Bayesian mixed-effects logistic regression (maximal by-participant random slopes). In Fig. 2 we report estimates and "pd" (prob. of direction) which when > 97.5 is taken as evidence for an effect. We made several **predictions**: 1) There should be a typical rate-based contrast effect in the no boundary condition. 2) Based on [3] there should be a domain-initial strengthening effect across prosodic boundary conditions in terms of /k*/ responses, ordered as no boundary > IP > IP+pause. It is an exploratory question if rate effect persists in the IP conditions, or interact with phrasing effects. **Results:** Prosodic strengthening effects: an interaction between rate and prosodic boundary was observed (pd = 100), reflecting a difference in phrasing conditions, *only* when rate was slow (Fig. 2B, "slow" rate bars ascend left to right). This reflects the predicted phrasing effect. Speech rate effects: A credible rate-based contrast effect is found in the no boundary condition: slow rate makes VOT sound relatively short and *increases* /k*/ responses. No rate effect is found in the IP condition, and an unexpected reversal of the rate effect is present in the IP+pause condition. We speculate that this effect results from a **re-interpretation of rate slow-down cues in the presence of a strong prosodic boundary**, i.e. when a pause is present. No effect of rate in the IP condition is hypothesized to result from a competition of rate and phrasing influences, though this remains to be further tested. **In summary:** We find that 1) prosodic phrasing effects obtain (which defy local contrast effects: longer pre-boundary vowels in slow + IP vs. slow + no boundary favors perception of longer-VOT /kʰ/). However, phrasing effects only occur at slower speech rates. 2) Canonical rate effects obtain when there is no prosodic boundary, but disappear, or are reversed, when a prosodic boundary (i.e, with a pause) is present. These results thus provide novel insight into how rate-dependent perception interacts with phrasing, and show that linguistic structure (prosodic boundaries) plays a mediating role.

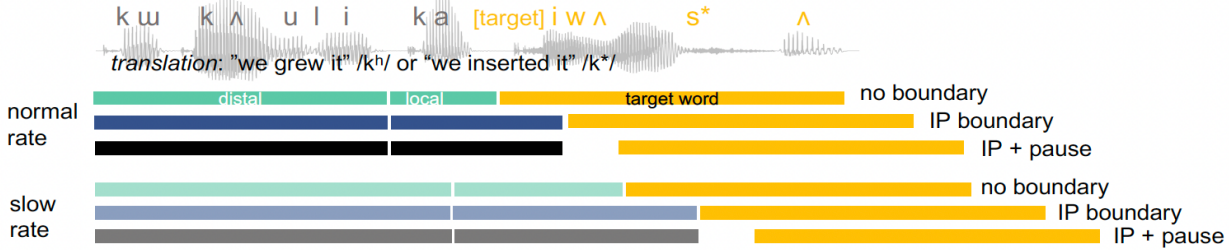

Figure 1: Schematic speech rate conditions and prosodic boundary conditions in the stimuli.

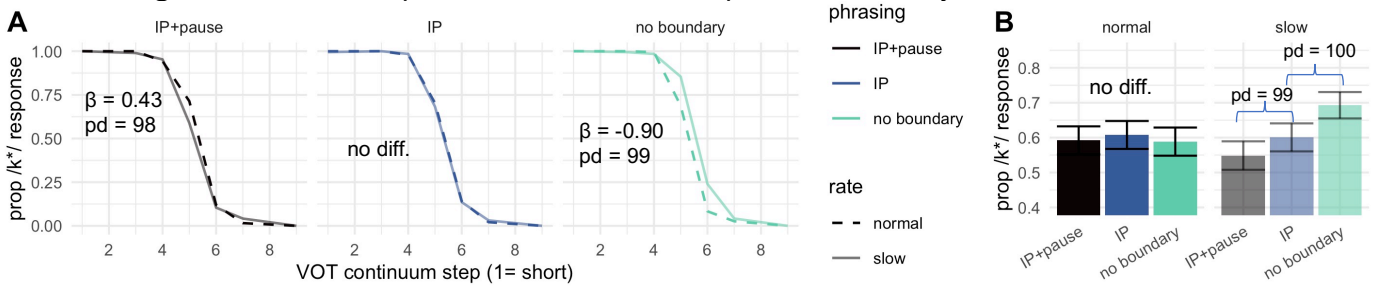

Figure 2: Categorization for the continuum (A) and pooled for steps 3-8 to show context effects (B)

References [1] Newman, R. S., & Sawusch, J. R. (1996). Perceptual normalization for speaking rate: Effects of temporal distance. *Perception & Psychophysics*, *58*(4), 540-560. [2] Mitterer, H., Cho, T., & Kim, S. (2016). How does prosody influence speech categorization?. *JPhon*, *54*, 68-79. [3] Steffman, J., Kim, S., Cho, T., & Jun, S.- A. (2022). Prosodic phrasing mediates listeners' perception of temporal cues: Evidence from the Korean Accentual Phrase. *JPhon*, *94*, 101156. [4] Cho, T., & Jun, S. A. (2000). Domain-initial strengthening as enhancement of laryngeal features: Aerodynamic evidence from Korean. *UCLA working papers*, 57-70. [5] Cho, T., & Keating, P. A. (2001). Articulatory and acoustic studies on domain-initial strengthening in Korean. *JPhon*, *29*(2), 155-190.]