Speech Prosody 2024
2-5 July 2024, Leiden, The Netherlands

Speech rate and prosodic phrasing interact in Korean listeners' perception of temporal cues

Jeremy Steffman[1], *Sahyang Kim*[2], *Taehong Cho*[3], *Sun-Ah Jun*[4]

[1]The University of Edinburgh
[2]Hongik University
[3]Hanyang University
[4]University of California, Los Angeles

jeremy.steffman@ed.ac.uk, sahyang@hongik.ac.kr, tcho@hanyang.ac.kr, jun@humnet.ucla.edu

Abstract

This study explores the interaction between contextual speech rate and prosodic phrasing in listeners' perception of temporal cues in Korean. We investigate perception of the aspirated/fortis stop contrast, testing categorization of a Voice Onset Time (VOT) continuum. Aspirated stops have longer VOT and shorter vowel duration relative to fortis stops. There is a strong tendency for vowel durations in both stops to be lengthened at the beginning of a prosodic phrase, and this prosodic strengthening pattern has been shown to influence perception of the stop contrast with temporal context controlled. Building on this previous finding, we manipulate preceding speech rate in a carrier phrase (slower/faster) and cross this with a phrasing manipulation: 1) no prosodic juncture before the target, 2) a preceding intonational phrase boundary (cued by pre-boundary lengthening), and 3) the same boundary with an additional pause. Results show canonical speech rate effects only in the absence of a preceding boundary. Prosodic strengthening effects, which are additive based on boundary strength, are robust only when speech rate is slow. In sum, findings suggest a mutual influence between speech rate effects and prosodic phrasing effects, providing insight into the interplay of these factors in shaping temporal cue perception.

Index Terms: speech perception, speech rate, prosodic phrasing, Korean

1. Introduction

1.1. Speech rate effects

Listeners' uptake of temporal information in speech has been shown to be highly context-dependent, a flexibility which is presumed to bolster spoken language comprehension across temporally variable contexts. "Speech rate normalization" is one well documented phenomenon in which a temporal cue is perceived relative to a preceding or following speech rate context, which impacts both phoneme categorization and word segmentation; e.g., [1, 2, 3], and [4] for a review. These effects are considered contrastive in the sense that the temporal cue in question is contrasted with context. Take the case of voice onset time (VOT) as a cue to stop voicing in American English. An ambiguous voice onset time preceded by faster speech rate is perceived as relatively slow (i.e., long); hence mapped to a voiceless (and aspirated) stop category /p,t,k/. Conversely, the same VOT is perceived as relatively short when preceded by slower speech rate, and thus categorized as /b,d,g/. Effects of this sort have been documented for a wide variety of temporal cues, and on the basis of both local/proximal and distal (that is, temporally removed) rate [2, 5, 6].

A common understanding of rate effects is that they operate within a sort of "moving window" [1] where temporally more local rates exert a primary influence (though see [5, 7]), and that they can further be explained by a low-level auditory mechanism of contrast. This auditory account is in line with findings showing that rate effects occur on the basis of non-speech material [5] and are immune to task effects such as variation in cognitive load [8]. At the same time, however, an interesting tension in the literature is that rate effects also appear to be subject to higher-level information. Certain rate effects occur only with speech stimuli [9], some are dependent on language experience [10], and perceived rate of speech is further dependent on other cues, such as the presence/absence of fast speech processes or F0 variation [11, 12]. In sum, the literature suggests rate-dependent speech perception may involve an interplay between lower-level and higher-level information.

One type of "higher-level" information which has been suggested to play a role in the perception of temporal information in speech is prosodic structure [13, 14], which is conveyed by (among other cues) temporal changes such as phrase-final lengthening at the end of a prosodic phrase. This offers an interesting opportunity to examine the relation between generalized rate-based contrast effects, and effects that are related to listeners' perception of prosodic patterns in speech.

1.2. Prosodic context effects

In addition to considering contrast-based rate effects, the present study also examines the impact of prosodic context in mediating listeners' uptake of segmental cues. The general idea builds on a body of speech production research which documents that segmental articulations and acoustics are fine-tuned systematically by prosodic organization, i.e., prosodic boundaries and prominence (e.g.,[15, 16]). Mirroring this intricate relationship between more global/utterance-level prosodic features and granular segmental detail, perception research has shown that listeners do indeed incorporate phrasal prosodic context in their perception of segmental contrasts, for example, the presence/absence of a phrasal prosodic boundary, or the phrasal prominence of a segment (e.g., [17, 18, 19, 20]). Most relevant to the present study, [21] documented that the presence of a phrasal boundary in Korean, signaled by a non-temporal cue, modulated perception of subsequent temporal cues to stop laryngeal contrasts (the same that we test here). We briefly review necessary background and that study here.

Korean prosodic structure [22] is described as consisting of

 10.21437/SpeechProsody.2024-220

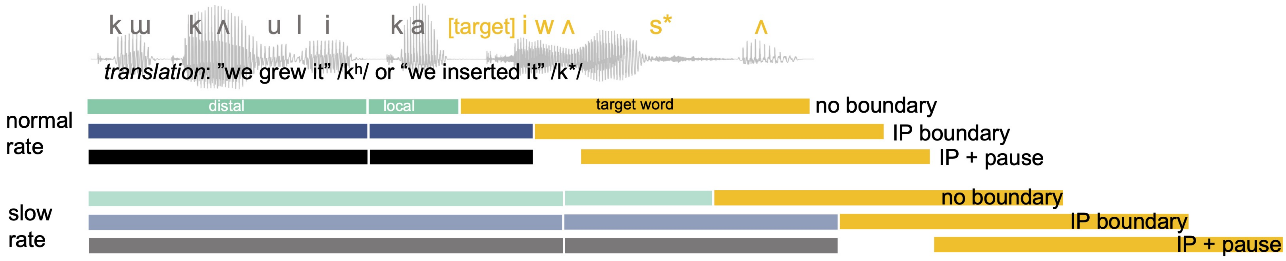

Figure 1: *Schema of stimuli used in the study. The relative durations of the pre-target material are shown schematically for the rate conditions (normal rate at top, slow rate at bottom), and prosodic phrasing conditions (labeled at right).*

smaller "accentual phrases" (APs) which are demarcated primarily by F0 (higher F0 phrase-finally), and also serve as the domain for phonological processes and domain-initial strengthening [22, 23, 24]. Larger "intonatonal phrases" (IPs) can be made up of multiple APs and are marked by tonal variation but also robust phrase-final lengthening. Relevant here, temporal cues to stop laryngeal contrasts are modified by phrasing. Korean has three contrastive laryngeal categories which we will refer to as lenis, aspirated, and fortis. F0 has developed as a cue distinguishing the lenis category (low F0) from the other categories (dependent on speaker age, prosodic factors, etc. [25]). However, fortis and aspirated stops are not distinguished by following F0; two cues for the fortis/aspirated contrast are voice onset time (VOT) and following vowel duration. VOT is longer for aspirated stops than fortis stops; vowel duration is shorter following aspirated stops as compared to fortis stops [26, 27, 24, 23]. Post-stop vowel duration (in a CV sequence) is subject to domain-initial lengthening for both fortis and aspirated stops, generally becoming longer in phrase-initial position (in an additive fashion, when comparing AP to IP-initial stops) [23].[1]

Given that vowel duration serves as a cue for the stop contrast [28] and is subject to prosodically-conditioned modulation, [21] focused on the perceptual effects of domain-initial strengthening for the contrast in AP-initial (stronger) vs. AP-medial (weaker) position. In [21], temporal context was identical across conditions, with the presence/absence of an AP boundary cued by only F0 on the syllable preceding the target segment. They found that a preceding AP boundary generated shifts in categorization of the aspirated-fortis stop contrast which mirrored domain-initial strengthening patterns. Listeners effectively required longer vowel duration to perceive an AP-initial stop as fortis, perceptually compensating for domain-initial vowel lengthening. This result shows that prosodic structure can serve to generate expectations for the realization of cues to segmental contrasts, in line with previous research to this effect [17, 29, 20].

2. The present study

Thus far we have outlined two influences (speech rate and prosodic phrasing) on listeners' perception of temporal information in speech. The basic goal of the present study was to test the extent to which rate effects persisted across prosodic contexts and the extent to which prosodic effects persisted across rate contexts. Further, as described in [13, 14], it can be difficult to disentangle effects related to speech rate (and linked to auditory contrast mechanisms) from effects related to prosodic parsing of utterance, as measured with listeners' behavioral categorization responses. In the case of English domain-initial lengthening of VOT, compensatory perceptual shifts for listeners are in the same direction as perceptual contrast effects due to speech rate, (see [14] for details). We thus also test a potential case in Korean in which the two effects may potentially be dissociated. Specific predictions are detailed following the description of the task and stimuli.

2.1. Methods and materials

We carried out a forced-choice categorization task, in which Seoul Korean listeners categorized a phonetic continuum ranging between a velar aspirated stop /k^h/ and velar fortis stop /k*/. The target sound was placed in contexts which manipulated preceding temporal information, and which was intended to convey both speech rate and prosodic phrasing. The carrier phrase placed the target segment as the initial segment in the final word in the phrase, for which there was a minimal pair for the relevant stop contrast (shown in Figure 1).

Stimuli were created by the manipulation of naturally produced speech, recorded by a male speaker of Seoul Korean, using PSOLA synthesis in Praat for the contextual manipulation [30, 31], and a VOT resynthesis script [32] which used the "cutback" method to create lengthened aspiration. The continuum ranged in nine steps between 35 ms (step 1, a clear /k*/) and 135 ms (step 9, a clear /k^h/). To reduce the number of continuum repetitions step 2 and 8 were then dropped, such that there were a total of seven steps including clear endpoints, and five relatively ambiguous mid-continuum steps. The creation of the contexual conditions altered only the pre-target material, and took what will be referred to as the "normal rate + no boundary" condition as the starting point (shown as a waveform in Figure 1). First, the immediately pre-target syllable was lengthened, with the amount of lengthening determined by the speaker's own natural productions of a prosodic boundary. This condition, which differed from the no boundary condition in only this syllable, is referred to as the "normal rate + IP" condition. To create the "normal rate + IP+pause" condition, silence was inserted preceding the target for a total of 320 ms, as based on the speaker's productions of a prosodic boundary with a pause (notably, the stop closure even in the absence of a pause was 100 ms in the normal condition). The "slow" rate conditions, which were also based on the speakers' natural production of a slower speech rate, resulted from the linear expansion of each of the normal rate conditions by approximately 168%, for all of the pre-target material. Thus, both local context (with the pre-

[1] VOT in aspirated stops tends to be lengthened in phrase-initial position, while it remains relatively stable for fortis stops, effectively enhancing the contrast in terms of VOT [24].

target syllable also lengthened in the IP conditions) and distal context were longer in the slow rate conditions.

2.2. Predictions

Predictions: Given the stimulus design, we considered two competing predictions regarding the phrasing manipulation. The first possibility is linked to the fact that a prosodic boundary leads to local temporal modifications in the preceding context, resulting in preboundary lengthening: a localized modification in speech rate near the boundary. Under this assumption, one can expect an additive influence of the IP boundary effect (with or without a pause) in conjunction with the global speech rate manipulation. This would lead to increased /k*/ responses in both the IP and IP+pause conditions compared to the no boundary condition. These results could be taken as reflecting a general rate-related contrast effect, stemming from both the global speech rate and the boundary-induced local rate modifications that affect VOT perception. However importantly, this effect may not be solely an auditory-perceptual contrast adjustment. Instead, it could still be attributed to the computation of prosodic structure: a boundary percept could possibly create an expectation of VOT lengthening in the target consonants, especially for the aspirated stops, resulting in increased /k*/ responses in the IP and IP+pause conditions [13]. This thorny ambiguity, as it pertains to VOT perception is discussed in detail in [14].

Conversely, the opposite effect might occur when we consider domain-initial strengthening on the vowel in a CV sequence. As demonstrated in [21], a prosodic boundary can lead to perceptual shifts that reflect post-stop vowel lengthening, causing vowel duration (as a cue for the stop contrast) to be perceived as relatively shorter in the IP (and IP+pause) conditions, thus decreasing /k*/ responses. Unlike the effect mentioned in the preceding paragraph, this outcome would be more clearly related to prosodic structure, and would comport with [21] in suggesting vowel duration is the locus of the perceptual domain-initial strengthening effect for the contrast.

In addition to these specific predictions, the present study is also suited to examine several exploratory questions. We did not have concrete predictions about the extent to which our predicted prosodic effects would persist across both rate contexts, which can be assessed by examining the presence/magnitude of phrasing-based differences across rates. We considered that an intervening pause in the IP+pause conditions might also diminish rate effects due to the temporal separation of the context and target by the pause.

2.3. Participants, Procedure and Analysis

32 native speakers of the Seoul dialect of Korean were recruited. Each took part in the study seated in a sound-attenuated booth fitted with Sony MDR-1R headphones. Response choices were presented visually using Hangul orthography and participants were instructed that their task was simply to listen to spoken sentences and identify which word they heard. Each of the 42 unique stimuli was presented 6 times in randomized order for a total of 252 trials. Responses were submitted to a Bayesian mixed-effects logistic regression, implemented in brms [33]. The model was fit to predict categorization response (aspirated mapped to 0, fortis mapped to 1) as a function of speech rate (contrast coded with normal mapped to -0.5, and slow mapped to 0.5), prosodic phrasing, coded with the IP condition as the reference level, and continuum step (treated as continuous and scaled). Priors were set as weakly informative normal priors.[2] We used emmeans [34] to examine marginal effects and test interactions between prosodic phrasing and rate conditions. When reporting effects we provide the estimated posterior median and 95% Credible intervals (CrI). When CrI exclude zero we take this as credible evidence for an effect. We also report the percentage of the posterior which shows a given sign, referred to as pd for "probability of direction" and computed using [35]. This provides graded evidence for the strength of an effect, ranging between 50 (no effect) and 100 (a clearly non-zero estimate). pd $>$ 97.5 corresponds to 95% CrI excluding zero.

3. Results

Results are shown in Figure 2. We found no marginal effect of rate (across all phrasing conditions), where pd = 78. At the reference level in the model, the IP phrasing condition, there was also no credible effect of rate (pd = 70). Importantly however, the interaction between rate and the no-boundary phrasing condition was credible (pd = 99), prompting us to examine the marginal effects for rate at each phrasing condition. This revealed that, in contrast to the IP condition, the no-boundary condition showed a credible effect of speech rate ($\hat{\beta}$ = -0.90, CrI = [-1.57, -0.21]), whereby slow speech rate increased /k*/ responses. This is in line with a typical/expected rate-related contrast effect whereby slower preceding speech should make VOT perceptually shorter (and hence, more /k*/-like). Contrary to expectations, a credible rate effect, with the reverse directionality was found in the IP+pause condition ($\hat{\beta}$ = 0.43, CrI = [0.01, 0.86]). We will return to this effect momentarily.

There was a marginal effect of phrasing (across both rate conditions) whereby no boundary showed fewer /k*/ responses than IP (pd = 100), and IP showed fewer /k*/ responses than IP+pause (pd = 96; credible intervals narrowly including zero). Importantly, though, the credible interactions between speech rate and phrasing in the model suggest that the phrasing effect should be examined at each speech rate. This result is most apparent in Figure 2, Panel B. The estimates from the model show no credible difference across phrasing conditions when speech rate is normal (pds = 57 and 73 for comparing no boundary to IP and IP to IP+pause respectively). In the slow condition, however, there was credible evidence for an ordinal decrease in /k*/ responses, first from the no boundary to IP condition ($\hat{\beta}$ = -1.05, CrI = [-1.60, -0.56]), and second, from the IP to IP+pause condition ($\hat{\beta}$ = -0.46, CrI = [-0.88,-0.04]). Notably, the directionality of differences aligns in the normal rate condition, but the main effect is clearly driven by the slow rate condition. This phrasing effect parallels what was found in [21] for phrasing at the level of the AP: A domain-initial stop was categorized less often as fortis, in line with perceptual compensation for domain-initial lengthening of the post-stop vowel. Two aspects of the results are notable: first, the presence of an intervening pause strengthens the effect. Under the assumption that the pause increased the perceived strength of the prosodic boundary, this suggests that stronger boundaries induce larger perceptual shifts. Secondly, the main effect of phrasing (and in particular the effect at slow rates) is notably in defiance of local durational contrast. In the IP condition the preceding vowel

[2] We encoded a prior expectation that increasing VOT should result in a decrease in the log-odds of a fortis response by setting the mean for the prior for this effect to be -3 (log odds) with a standard deviation of 1.5. This value was determined via prior predictive simulations. Priors for other fixed effects and the intercept were set with a mean of 0.)

was longer as compared to the no boundary condition, which under a standard contrast prediction would result in shorter perceived VOT, and hence more /k*/ responses; the opposite of the observed effect. This provides some evidence that interpretation of durational patterns as signaling prosodic structure can be disentangled from otherwise-expected contrast effects.

Finally, let us revisit the reversal of the rate effect found in the IP+pause condition. This finding contradicts the typical rate-related contrast effect, where slow speech rates are expected to result in increased /k*/ responses. We offer a speculative explanation that warrants validation in future studies. The presence of a pause is likely to prompt perception of a clear prosodic boundary in both rate conditions. With the salient pause-induced boundary percept, the preceding temporal variation caused by speech rate, even though it was "global", might be re-interpreted as reinforcing the upcoming boundary strength in the slow speech rate condition. While the enhanced boundary strength could, in principle, modulate perception of both VOT and the following vowel, the above-mentioned phrasing effects suggest its impact on the latter. The temporal gap created by the pause between the preceding context and the target consonant may also reduce the immediate influence on the most-proximal VOT cue, while inducing boundary-related perceptual modification of the distal vowel. This is consistent with the previous prosodic strengthening effects in the perception of the stop contrast, attributed to vowel duration [21], and the phrasing effects described above. This speculation could be further investigated by considering other boundary cues, including tonal cues, to assess the extent to which they result in a similar hypothesized reinterpretation of temporal slowing in the presence/absence of a pause.

4. Conclusions

This study set out to examine the inter-relationship between speech rate and prosodic phrasing in listeners' perception of temporal cues. Notably, we replicated canonical rate-related contrast effects in the no boundary condition. However, in defiance of expected contrast effects, we documented "prosodic" effects across the three phrasing conditions, especially at slow speech rates. These effects showed additive patterning as a function of boundary strength and were in line with previously shown effects at the level of the Korean AP [21]. The fact that canonical contrast effects did not obtain here suggests that they may be overridden when a prosodic parse of temporal patterns is available to listeners [13]. Finally, we found an unexpected reversal of the rate effect in the IP+pause condition, which will be a fruitful area for future studies to explore. In sum, our results point to a mutual influence of prosodic phrasing and speech rate in listeners' perception of temporal information, and the importance of considering both in segmental perception and lexical processing [19, 36].

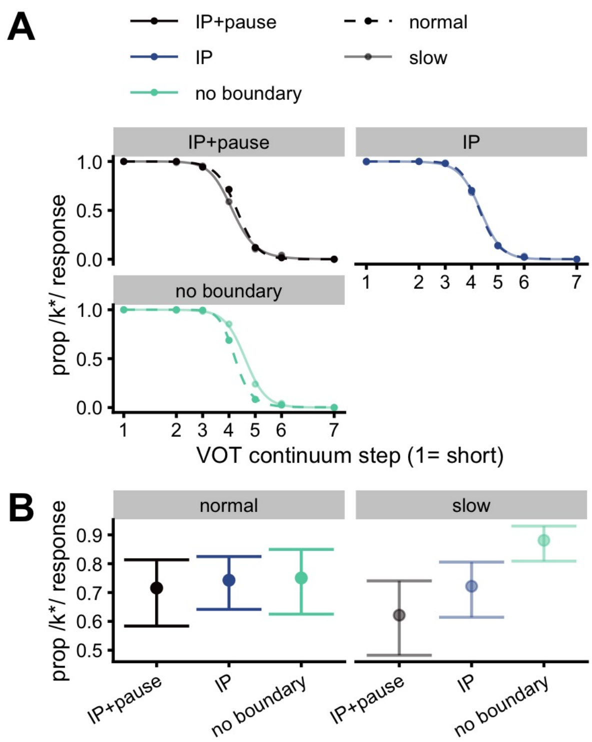

Figure 2: *Panel A: categorization responses across the continuum paneled by prosodic boundary condition. Line type indicates speech rate. Points show empirical responses, lines are psychometric curves. Panel B: overall /k*/ responses for the six rate and phrasing conditions, as estimated by the model for the middle of the VOT continuum (scaled VOT = 0). Error bars are 95%CrI.*

5. References

[1] R. S. Newman and J. R. Sawusch, "Perceptual normalization for speaking rate: Effects of temporal distance," *Perception & Psychophysics*, vol. 58, no. 4, pp. 540–560, 1996.

[2] E. Reinisch, A. Jesse, and J. M. McQueen, "Speaking rate from proximal and distal contexts is used during word segmentation." *Journal of Experimental Psychology: Human Perception and Performance*, vol. 37, no. 3, p. 978, 2011.

[3] L. C. Dilley and M. A. Pitt, "Altering context speech rate can cause words to appear or disappear," *Psychological Science*, vol. 21, no. 11, pp. 1664–1670, 2010.

[4] C. Stilp, "Acoustic context effects in speech perception," *Wiley Interdisciplinary Reviews: Cognitive Science*, vol. 11, no. 1, p. e1517, 2020.

[5] H. R. Bosker, "Accounting for rate-dependent category boundary shifts in speech perception," *Attention, Perception, & Psychophysics*, vol. 79, pp. 333–343, 2017.

[6] C. C. Heffner, R. S. Newman, and W. J. Idsardi, "Support for context effects on segmentation and segments depends on the context," *Attention, Perception, & Psychophysics*, vol. 79, pp. 964–988, 2017.

[7] J. Steffman, "Rhythmic and speech rate effects in the perception of durational cues," *Attention, Perception, & Psychophysics*, vol. 83, no. 8, pp. 3162–3182, 2021.

[8] H. R. Bosker, E. Reinisch, and M. J. Sjerps, "Cognitive load makes speech sound fast, but does not modulate acoustic context effects," *Journal of Memory and Language*, vol. 94, pp. 166–176, 2017.

[9] M. A. Pitt, C. Szostak, and L. C. Dilley, "Rate dependent speech processing can be speech specific: Evidence from the perceptual disappearance of words under changes in context speech rate," *Attention, Perception, & Psychophysics*, vol. 78, pp. 334–345, 2016.

[10] H. R. Bosker and E. Reinisch, "Foreign languages sound fast: Evidence from implicit rate normalization," *Frontiers in Psychology*, vol. 8, p. 1063, 2017.

[11] E. Reinisch, "Natural fast speech is perceived as faster than linearly time-compressed speech," *Attention, Perception, & Psychophysics*, vol. 78, pp. 1203–1217, 2016.

[12] J. Steffman, "Intonational structure mediates speech rate normalization in the perception of segmental categories," *Journal of Phonetics*, vol. 74, pp. 114–129, 2019.

[13] S. Kim and T. Cho, "Prosodic boundary information modulates phonetic categorization," *The Journal of the Acoustical Society of America*, vol. 134, no. 1, pp. EL19–EL25, 2013.

[14] H. Mitterer, T. Cho, and S. Kim, "How does prosody influence speech categorization?" *Journal of Phonetics*, vol. 54, pp. 68–79, 2016.

[15] P. Keating, "Phonetic encoding of prosodic structure," in *Speech Production: Models, Phonetic Processes, and Techniques*. Psychology Press, 2006, pp. 167–186.

[16] T. Cho, "Prosodic Boundary Strengthening in the Phonetics–Prosody Interface," *Language and Linguistics Compass*, vol. 10, no. 3, pp. 120–141, 2016.

[17] H. Mitterer, S. Kim, and T. Cho, "The glottal stop between segmental and suprasegmental processing: The case of Maltese," *Journal of Memory and Language*, vol. 108, p. 104034, 2019.

[18] S. Kim, J. Kim, and T. Cho, "Prosodic-structural modulation of stop voicing contrast along the VOT continuum in trochaic and iambic words in American English," *Journal of Phonetics*, vol. 71, pp. 65–80, 2018.

[19] T. Cho, J. M. McQueen, and E. A. Cox, "Prosodically driven phonetic detail in speech processing: The case of domain-initial strengthening in English," *Journal of Phonetics*, vol. 35, no. 2, pp. 210–243, 2007.

[20] J. Steffman, *Prosodic prominence in vowel perception and spoken language processing*. University of California, Los Angeles, 2020.

[21] J. Steffman, S. Kim, T. Cho, and S.-A. Jun, "Prosodic phrasing mediates listeners' perception of temporal cues: Evidence from the korean accentual phrase," *Journal of Phonetics*, vol. 94, p. 101156, 2022.

[22] S.-A. Jun, *The phonetics and phonology of Korean prosody: Intonational phonology and prosodic structure*. Taylor & Francis, 1996.

[23] T. Cho and P. Keating, "Articulatory and acoustic studies on domain-initial strengthening in Korean," *Journal of Phonetics*, vol. 29, no. 2, pp. 155–190, 2001.

[24] T. Cho and S.-A. Jun, "Domain-initial strengthening as enhancement of laryngeal features: Aerodynamic evidence from Korean," *UCLA Working Papers in Phonetics*, pp. 57–70, 2000.

[25] J. Choi, S. Kim, and T. Cho, "An apparent-time study of an ongoing sound change in Seoul Korean: A prosodic account," *Plos one*, vol. 15, no. 10, p. e0240682, 2020.

[26] H. Chung, K. Kim, and M. Huckvale, "Consonantal and prosodic influences on korean vowel duration," in *Proceedings of Eurospeech*, vol. 99. Citeseer, 1999.

[27] H. Choi, "Interface between speech production and perception: Analysis of stop and vowel duration," *Korean Journal of Linguistics*, vol. 36, no. 3, pp. 815–842, 2011.

[28] ——, "Vowel duration as a perceptual cue for preceding stop laryngeal contrast in korean." in *ICPhS*, vol. 17, 2011, pp. 17–21.

[29] S. Kim, H. Mitterer, and T. Cho, "A time course of prosodic modulation in phonological inferencing: The case of Korean post-obstruent tensing," *PloS one*, vol. 13, no. 8, 2018.

[30] P. Boersma and D. Weenink, "Praat: doing phonetics by computer (version 6.1.09)," 2020. [Online]. Available: http://www.praat.org

[31] E. Moulines and F. Charpentier, "Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones," *Speech Communication*, vol. 9, no. 5-6, pp. 453–467, 1990.

[32] M. B. Winn, "Manipulation of voice onset time in speech stimuli: A tutorial and flexible praat script," *The Journal of the Acoustical Society of America*, vol. 147, no. 2, pp. 852–866, 2020.

[33] P.-C. Bürkner, "brms: An R package for Bayesian multilevel models using Stan," *Journal of Statistical Software*, vol. 80, no. 1, pp. 1–28, 2017.

[34] R. V. Lenth, *emmeans: Estimated Marginal Means, aka Least-Squares Means*, 2021, r package version 1.7.1-1. [Online]. Available: https://CRAN.R-project.org/package=emmeans

[35] D. Makowski, M. S. Ben-Shachar, and D. Lüdecke, "bayestestr: Describing effects and their uncertainty, existence and significance within the bayesian framework." *Journal of Open Source Software*, vol. 4, no. 40, p. 1541, 2019. [Online]. Available: https://joss.theoj.org/papers/10.21105/joss.01541

[36] J. McQueen and L. C. Dilley, "Prosody and spoken-word recognition," in *The Oxford Handbook of Language Prosody*, C. Gussenhoven and A. Chen, Eds. Oxford: Oxford University Press, 2020, ch. 36, pp. 509–522.