

The causal relationship between vowel devoicing and duration in Japanese and Korean

Hanyoung Byun¹, Seung Suk Lee²

¹UC Santa Cruz, ²McGill University

Introduction. High vowel devoicing (HVD) is attested in many languages, including Tokyo Japanese (TJ) and Seoul Korean (SK), while a devoicing process that only targets non-high vowels is not [1]. One factor underlying this asymmetry may be that high vowels are intrinsically short, and their voicing is more easily affected by adjacent voiceless consonants [2,3]. It is also known that cross-linguistically vowels before voiceless consonants are shorter than ones before voiced consonants [6]. While these factors are cross-linguistic, HVD is said to be phonological only in certain languages (e.g., TJ [1,4], but not SK [5]). Previously, [4] showed a categorical distinction between high and non-high vowels in TJ—while ÇVs containing high and non-high vowels were both shorter before voiceless consonants than before voiced ones [6], ÇVs containing high vowels shortened much more than ones with non-high vowels. While [4] suggested high vowel shortening in the HVD context as evidence for phonologization of HVD, it is still left unclear if high vowels are shortened because they are devoiced, or if they are devoiced because of extensive shortening. In general, it remains unclear the causal relationship between variables, when a phonological context affects multiple acoustic variables (i.e., vowel devoicing and shortening), and how it changes when a phonetic process (HVD) is *phonologized*. To address these gaps, we conducted a corpus study of HVD in TJ and SK. Specifically, we ask: **(1)** Does likelihood of devoicing correlate with ÇV duration, regardless of vowel height? **(2)** Do high vowels shorten before voiceless consonants, categorically more than non-high vowels? **(3)** Is devoicing causally prior to shortening?

Methods. We examined vowels between a voiceless C₁ and a voiced or voiceless C₂ in two speech corpora [7,8] in TJ and SK. To address **(1)**, we fit logistic mixed-effects models with control predictors such as speech rate. For **(2)**, we fit linear mixed-effects models with ÇV duration as the response variable, as opposed to V duration, following [4]. For TJ, phonemically long vowels and vowels followed by a geminate or pause were excluded. Vowel devoicing in TJ was identified from annotations in [7]. For SK, we used ‘Voice report’ [9] to code vowel devoicing and lenis consonant voicing [10] (“voiced” if >50% of the forced aligned vowel interval was voiced). Most vowel and consonant intervals were either 100% voiced or voiceless. In total, we analyzed 116,876 vowels for TJ and 50,349 for SK. For **(3)**, we fit models of ÇV duration with vowel devoicing as a predictor and ones without it, to see how much the effect on duration for each predictor was mediated by devoicing [10-12].

Results. **Fig. 1** shows the model-predicted likelihood of devoicing per vowel height in each language. The devoicing patterns in TJ are near-categorical **(1A)**. The devoicing likelihood of high vowels is almost at the ceiling and stays above 75% even for fairly long durations. For non-high vowels, the devoicing likelihood is almost at the floor regardless of durations. For SK **(1B)**, both high and non-high vowels devoice less frequently as their CV durations lengthen. **Fig. 2** shows the effect of C₂ voicing on CV duration by vowel height. For TJ **(2A)**, high vowels become much shorter before Ç₂ (subject to HVD) compared to non-high vowels, replicating [4]’s results. For SK **(2B)**, the degree of CV shortening before Ç₂ is not statistically different between the vowel heights—the voicing-induced shortening affects vowels of both heights similarly. Finally, **Table 1** shows that for TJ **(3A)**, devoicing largely mediates the effect of vowel height and C₂ voicing on CV duration, suggesting that devoicing is causally prior to shortening, not the other way around, supporting [4]’s argument. In contrast, for SK **(3B)**, the proportions of mediation by devoicing are much smaller than TJ, suggesting a different causal relationship. In sum, HVD in TJ exhibits the signatures of phonologized HVD, while HVD in SK does not. Our results suggest that when a phonetic process which affects multiple acoustic variables is phonologized, the acoustic variable that is directly affected by the phonological context (e.g., devoicing) becomes causally prior to the other acoustic variable (e.g., duration).

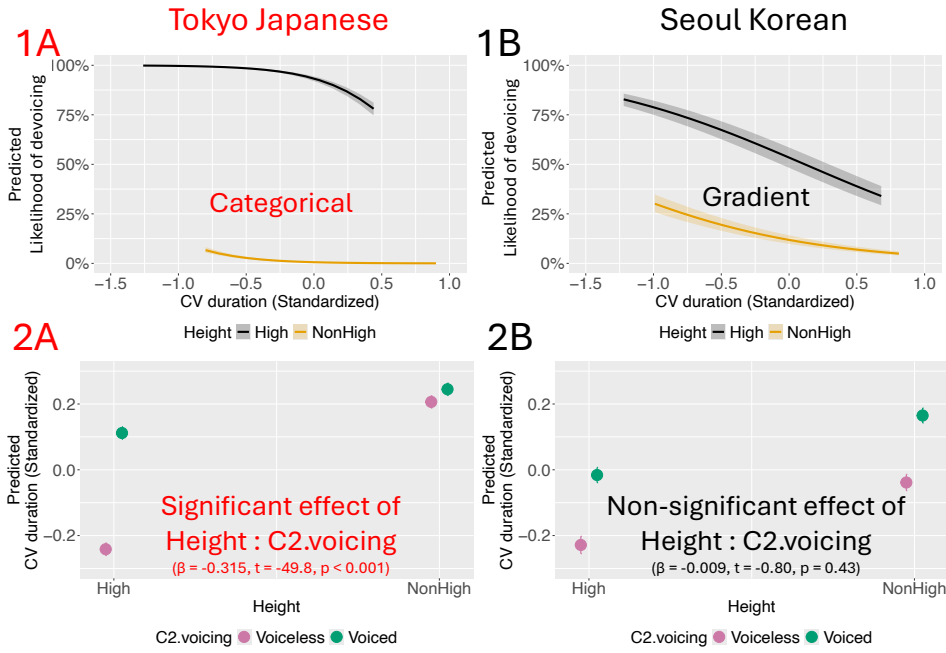


Fig 1. Model-predicted rate of vowel devoicing by CV duration, showing model predictions for vowels in the $[C_1_C_2]$ context only. CV duration is centered on mean CV duration and divided by 2 standard deviations (following [4]) in each language.

Fig 2. Model-predicted standardized CV duration by height and C_2 voicing.

Table 1. Mediation analyses results. Total effects are coefficients of factors in the model predicting CV duration *without* vowel devoicing as a predictor, direct effects are coefficients of factors in the model predicting CV duration *with* vowel devoicing as a predictor [10-12]. The last column shows how much the effect of duration for each factor is mediated by devoicing.

	Factor	Total effect (T)	Direct effect (D)	Indirect effect (I = T - D)	Proportion of mediation (I/T)
TJ (3A)	Height	-0.29 ($p < 0.01$)	-0.10 ($p < 0.01$)	-0.19	0.67
	C2.voicing	-0.20 ($p < 0.01$)	-0.06 ($p < 0.01$)	-0.14	0.71
	Height:C2.voicing	-0.32 ($p < 0.01$)	-0.06 ($p < 0.01$)	-0.26	0.82
SK (3B)	Height	-0.19 ($p < 0.01$)	-0.13 ($p < 0.01$)	-0.05	0.28
	C2.voicing	-0.21 ($p < 0.01$)	-0.16 ($p < 0.01$)	-0.05	0.22
	Height:C2.voicing	-0.01 ($p = 0.43$)			

References. [1] Gordon, M. (1998). The phonetics and phonology of non-modal vowels: a cross-linguistic perspective. *Berkley Linguist. Soc.* 24. [2] Lehiste, I. (1970) *Suprasegmentals*. MIT Press. [3] Solé, M. J., and Ohala, J. J. (2010). What is and what is not under the control of the speaker. Intrinsic vowel duration. In *Papers in Laboratory Phonology 10*. [4] Tanner, J., et al. (2019). Durational evidence that Tokyo Japanese vowel devoicing is not gradient reduction. *Front. Psychol.* 10:821. [5] Jun, S., et al. (1998). Electromyographic evidence for a gestural-overlap analysis of vowel devoicing in Korean. *UCLA Working Papers in Phonetics* 96. [6] Chen, M. (1970). Vowel length variation as a function of the voicing of the consonant environment. *Phonetica* 22. [7] Maekawa, K., et al. (2000). Spontaneous speech corpus of Japanese. Proceedings of the Second International Conference of Language Resources and Evaluation Vol. 2. [8] Yun, W. et al. (2015). The Korean corpus of spontaneous speech. *Phonetics and Speech Sciences* 7. [9]. Boersma, P. and Weenink, D. (2023). Praat: Doing phonetics by computer. [10] Lee, S. (2025). *Simultaneous Encoding and Decoding of Laryngeal Contrasts and Prosodic Structure in Seoul Korean*. UMass Ph.D. thesis. [11] Cohen Priva, U. and Gleason, E. (2010). The causal structure of lenition: A case for the causal precedence of durational shortening. *Language* 96. [12] Katz, J. and Pitzanti, G. (2019). The phonetics and phonology of lenition: A Campidanese Sardinian case study. *LabPhon* 10(1).