Ongoing VOT merger unmerged in a singing context

Jonny Jungyun Kim, Hyunjung So, Ahjin Ko, Jiyea Heo, and Seoyeong Ahn
Pusan National University

Aspirated stops /p^h, t^h, k^h/ and lenis stops /p, t, k/ in Seoul Korean are undergoing a VOT merger, with younger speakers employing higher pitch in the vowel (instead of longer VOT) to signal aspirated stops [1, 2, 3]. This sound change is prosodically conditioned [3], as segment-dependent F0 cues are primarily used only at the onset of a phrasal unit (e.g., IP-initially, AP-initially). In phrase-medial position, however, VOT continues to hold the feature primacy, as F0 is governed by the intonation-level tone assignment [4]. We explore how VOT realization is mediated in a singing context (as compared to reading), where pitch is fixed independently of the phonemic contrast in all positions. We predict that young speakers would 'unmerge' their phrase-initial VOT overlap, rather than retaining it as a stable variant form, to compensate for musically constrained F0 use.

The 12 sentences in Table 1, each containing an initially-contrasted (near) minimal-pair target word (e.g., p^haraŋ vs. param) were used as critical items, alongside 8 filler items with no initial stop. The initial stops were distributed across 4 prosodic positions in natural Seoul intonation: utterance-initial (P1), word-initial (P2), AP-initial (P3), and IP-initial (P4). The 12 musical tunes in Fig.1 were composed using the 5 notes between C and G in C major, ensuring that their rhythmic structure aligned with the boundary strengths of the 4 positions. Within each tune, the half note before P4 served as preboundary lengthening, a hallmark of an IP boundary. The measure boundary before P3 equated to an AP boundary, and P2 occurred within the measure, inducing a smaller juncture comparable to a word boundary. To imitate LH and HL edge tones with varying pitch range, we counterbalanced the pitch difference between the target syllable and its following syllable across the tunes, resulting in the following syllable rising or falling with 1-4 semitone differences (e.g., compare ① vs. ②, or ③ vs. ④ in Fig. 1). The singing condition was presented in the first 2 blocks (an aspirated block followed by a lenis block). In each trial, 12 participants (age: 21-26, 7 female) listened to one of the tunes in electronic sound with a tempo comparable to normal speech rate, and then sang along with the sentence presented on the screen. Two additional keys (A major, F major) were provided to accommodate their pitch range. The reading block followed, in which participants read the identical sentences with no explicit instruction about prosody.

A total of 6,912 VOT values in the singing context (12 participants x 12 sentences x 12 tunes x 4 positions) and 2,304 VOT values in the reading context (4 repetitions, instead of 12 tunes) were obtained. Excluding 53 tokens due to recoding errors (n=48) or mispronunciation (n=5), 9,163 tokens were analyzed. As shown by the VOT distributions in the left panel of Fig.2, aspirated and lenis stops were realized with the shortest VOT in P2 (Wd-initial). This was also evidenced by the dense cluster of data points near VOT=0 in P2, indicating fully-voiced realizations. Stops in P2 were differentiated by VOT, consistently with the previous finding [3]. Importantly, in positions with a larger prosodic boundary (particularly P1 and P4), aspirated and lenis stops in the reading condition showed largely overlapping VOT distributions, confirming the prosodically conditioned VOT merger. However, the merger was not observed in the singing condition, where even P1 (utterance-initial) and P4 (IP-initial) revealed a substantial difference in VOT distributions between the aspirated and lenis categories. These trends were tested in lmer analysis fit to the difference in VOT (i.e., Δ(asp-lenis)), with a model structure best-supported by the data. As a main effect of Context, the VOT difference was significantly greater in the singing block (β=11.785, p<.001). Context:Position interaction showed that the Context effect was significantly reduced in P2 as compared to P1 (β=-11.970, p<.001), P3 (β=-14.404, p<.001), and P4 (β=-18.006, p<.001).

In line with the view that phonological representations of a sociophonetic variable would be flexibly fine-tuned across verbal and musical contexts [5], the results suggest that speakers who engage in the innovative cue-shifting trend towards F0-based distinction selectively inhibit their own cue-weighting style to maintain the contrast in accordance with the context. The unmerging process was observed in young speakers who showed prosodically conditioned magnitudes of VOT merger closely aligned with boundary strength (i.e., the merger in the reading block was most pronounced in P1 and P4, followed by P3, and then P2). Relatedly, as a methodological implication, the singing method can be applied in studies examining pitch-associated phonetic variation, as well as prosody studies to prime intended prosodic frames with no direct instructions.

Table 1. Critical sentence items: Target syllables are colored by phonation type (blue=aspirated, red=lenis).

Stop		V	Intonational phrase 1			Phrase2		English translation
			P1 (utt)	P2 (wd)	P3 (AP)	P4 (IP)	Ending	
bilabial	asp.	/a/	pʰaɾaŋ	pʰaɾaŋ	pʰaɾaŋ-i	pʰaɾaŋ	ʃiwonhɛ	Blue-blue, blue is, blue is cool.
	lenis		paɾam	paɾam	paɾam-i	paɾam		Wind-wind, wind is, wind is cool.
	asp.	/ʌ/	pʰʌlʌŋ	pʰʌlʌŋ	pʰʌlʌŋ-i	pʰʌlʌŋ	kʌɾjʌjo	Flap-flap, flapping is, it is flapping.
	lenis		pʌlʌŋ	pʌlʌŋ	pʌlʌŋ-i	pʌlʌŋ		Flip-flip, flipping is, it is flipping.
alveolar	asp.	/a/	tʰaɾak	tʰaɾak	tʰaɾak-in	tʰaɾak	hadʒima	Corrupt-corrupt, corrupt, don't corrupt.
	lenis		taɾak	taɾak	taɾak-in	taɾak	kadʒima	Attic-attic, attic, don't go to the attic.
	asp.	/ʌ/	tʰʌlʌŋ	tʰʌlʌŋ	tʰʌlʌŋ-i	tʰʌlʌŋ	kʌɾjʌjo	Jangle-jangle, jangling is, it's jangling.
	lenis		tʌlʌŋ	tʌlʌŋ	tʌlʌŋ-i	tʌlʌŋ		Jingle-jingle, jingling is, it's jingling.
velar	asp.	/a/	kʰaɾɛ	kʰaɾɛ	kʰaɾɛ-pap	kʰaɾɛ	pap mʌkʌ	Curry-curry, curry rice, eat curry rice.
	lenis		kaɾɛ	kaɾɛ	kaɾɛ-t*ʌk	kaɾɛ	t*ʌk mʌkʌ	kaɾɛ-kaɾɛ, kaɾɛt*ʌk, eat kaɾɛt*ʌk (rice cake)
	asp.	/ʌ/	kʰʌlʌŋ	kʰʌlʌŋ	kʰʌlʌŋ-i	kʰʌlʌŋ	kʌɾjʌjo	Curly-curly, curling is, it's curling.
	lenis		kʌlʌŋ	kʌlʌŋ	kʌlʌŋ-i	kʌlʌŋ		(same, with weakened mimetic meaning)

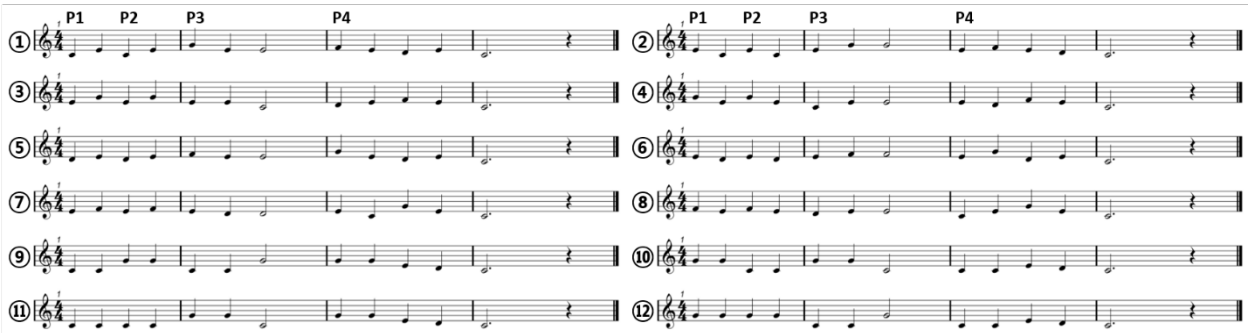

Fig.1. Scores of the 12 musical tune stimuli. Each note corresponds to one syllable in Table 1.

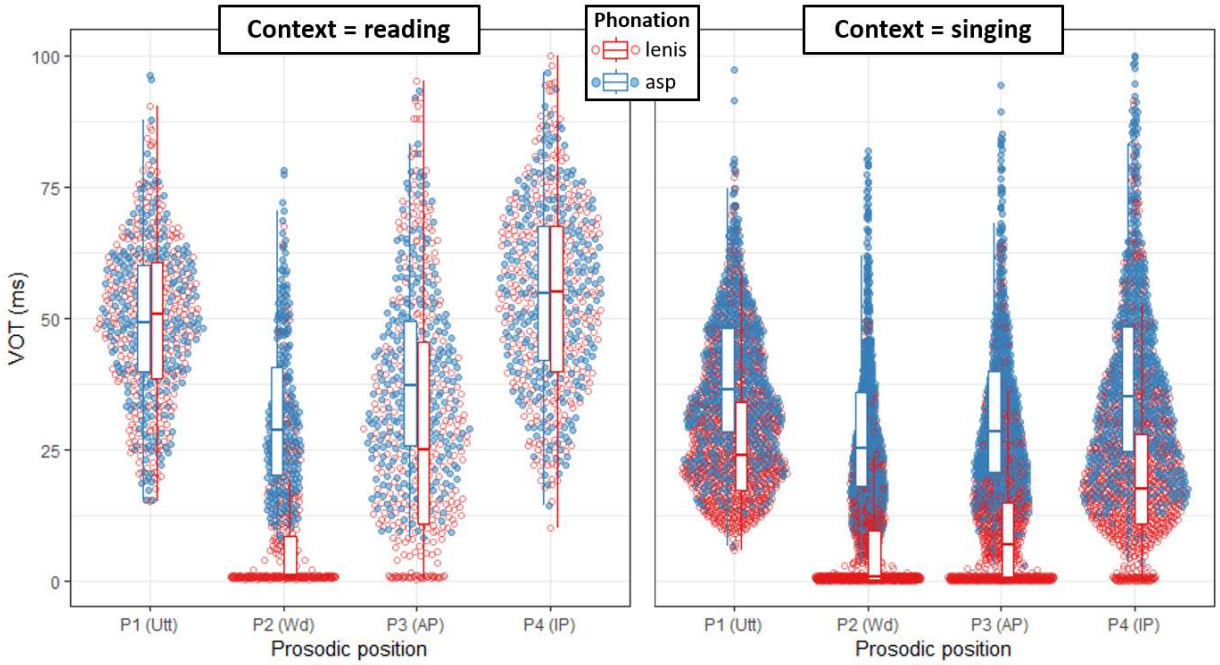

Fig.2. Distribution of VOT (all data points) conditioned by Phonation, Position, and Context

References

[1] Silva, D. J. (2006). Acoustic Evidence for the Emergence of Tonal Contrast in Contemporary Korean. *Phonology*, 23(2), 287-308.

[2] 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.

[3] Choi, J., Kim, S. & Cho, T. (2020). An apparent-time study of an ongoing sound change in Seoul Korean: A prosodic account. *PLoS ONE*, 15(10), e0240682.

[4] Jun, S-A. (2000). K-ToBI (Korean ToBI) Labelling Conventions. *Speech Sciences*, 7(1), 143-170.

[5] Gibson, A. M. (2019). *Sociophonetics of Popular Music: Insights from Corpus Analysis and Speech Perception Experiments* (doctoral dissertation), University of Canterbury, New Zealand.