Articulation of the tongue dorsum in Seoul Korean bilabial obstruents

Daejin Kim and Caroline L. Smith

Department of Linguistics, University of New Mexico (USA)

Using ultrasound tongue imaging [7], this study examined the articulation of the tongue dorsum (TD) in Seoul Korean (SK) bilabial onset obstruents (BOs) (fortis $/p^*/$, aspirated $/p^h/$, lenis /p/) [4], which lack clearly distinguishing acoustic and articulatory properties. TD has been argued to move during the acoustic closure to indicate the aerodynamic characteristics of BOs due to the linkage to the laryngeal structure. Lowering TD lowers the laryngeal structure and may increase subglottal air pressure, signaling the consonantal voicing in English, which correlates to its lower *f0* [8, 9]. Raising TD was assumed to have the opposite effect, increasing *f0* [6]. This study argues that the TD movement during the closure and the duration of the following vowels (Vs) should be related to the phonetic properties in SK BOs, employing the laryngeal contrasts, instead of voicing. Following [8, 9], this study argues that SK speakers should produce $/p^*/$ with longer and larger TD movement by raising TD to indicate its fortis-ness, along with higher *f0* and longer duration of V, compared to /p/ and $/p^h/$. $/p^h/$ should also be distinguished from /p/ in TD movement, indicating its articulatory strength associated with its longer voice onset time (VOT) and higher *f0* [5]. The difference may be more apparent with varying focus prominence conditions [2].

Audio-synchronized ultrasound tongue images (frame rates = 60 - 66Hz) were collected from seven female native SK speakers. Speakers recorded sentence-medial words with three Vs (/i, u, a/) and three BOs under broad and contrastive focus prominence seven times (**Table 1**). Tongue contours were estimated using DeepLabCut™ [10]. TD movements were annotated as the distance between the hyoid bone and the TD surface under the velum when it starts (ONSET), reaches its target (TARGET), and ends its movement (OFFSET) in timing. Linear mixed-effects models [1] with pairwise post-hoc tests [3] found some significant conditional differences in articulatory distance and duration between TD events and intergestural timing between acoustic (BO closure and V beginning) and TD articulation for BOs and Vs, which gauges laryngeal distinctions.

The models (**Figure 1; Table 2**) estimated that $/p^h/$ and $/p^*/$ were produced with longer ONSET-to-TARGET distance than /p/ when lowering ($/p/ > /p^h/ = /p^*/$) (a), while /p/ was articulated with shorter TARGET-to-OFFSET distance after lowering ($/p/ > /p^h/ = /p*/$) (b). $/p^*/$ was articulated with longer ONSET-to-TARGET duration than /p/ and $/p^h/$ ($/p/ = /p^h/ < /p^*/$) (c). /p/ was articulated with shorter TARGET-to-OFFSET duration than $/p^h/$ and /p*/ ($/p/ < /p^h/ = /p^*/$) (d). Altogether, $/p^*/$ was produced with *longer*, while /p/ was articulated with *shorter* TD ONSET-to-OFFSET duration than $/p^h/$ ($\mathbf{/p/ < /p^h/ < /p^*/}$) (e). Vs with $/p^h/$ and $/p^*/$ were produced with *shorter* TD ONSET-to-OFFSET distance and duration than those with /p/ ($\mathbf{/pV/ > /p^hV/ = /p^*V/}$) (f)-(g). Regarding intergestural timing (**Figure 2**; **Table 3**), TD starts to move earlier with $/p^*/$ from the acoustic closure than with $/p^h/$ and /p/ (a). All BOs reach the target similarly from the V target in time (b). TD movement ends later with $/p^*/$ than with /p/ and $/p^h/$ (c). $/p^h/$ and $/p^*/$ had a *later* offset than /p/ from the beginning of C1 (d). V with $/p^*/$ has all *later* TD events, while $/p^h/$ has the *earlier* V target and offset from the acoustic V beginning than /p/ and $/p^h/$ ($\mathbf{/p^hV/ < /pV/ < /p^*V/}$) (e). TD indicates clear phonetic contrast on C1s and Vs. Correlations were insignificant between *f0*, focus types, and TD movement.

Compared to /p/, $/p^*/$ was articulated with larger and longer TD movement, resulting in maximal temporal expansion of TD. The differences between /p/ and $/p^h/$ and $/p^h/$ and $/p^*/$ in TD moving distance still seem unclear. SK Speakers, instead, contrast three BOs more clearly with Vs by shifting TD timing from the acoustic V beginning. $/p^h/$ is subject to TD expansion on both C1 and V and is further distinguished by having the *earlier* TD events of the following V from the acoustic beginning of V from /p/ and $/p^*/$. It is coordinated with the earlier glottal [5] and bilabial opening, evidenced in the longer VOT. The constricted TD articulation of $/p^*V/$ is due to the expanded TD articulation of $/p^*/$, but its *later* TD events from acoustic V beginning must be related to speakers' intricate timing control of the laryngeal articulation. $/p^*/$ is articulated with the minimal VOT and longest V duration, involving early and immediate (i.e., longer) vocal folds vibration after release. This laryngeal and supralaryngeal coordination make TD events for V to appear *later* even with gestural reduction. Taken together, we suggest that this "secondary" articulation of TD and its spatiotemporal coordination with the laryngeal articulation are the key components of the clear three-way phonological contrast of SK BOs in speech production.

Table 1 Prime and target sentences with the target word '빠' /p*a/ in broad (BF) and contrastive focus (CF) prominence conditions.

BF	Q: 어제 무엇을 했어요? (What did (you) do {today, yesterday}?) /ʌ.dʑɛ. mu.ʌ.sɯɾ hɛ.s*ʌ.jo/ A: 카드에 빠하라고 적었어요. ((I) wrote do /p*a/ on the card.) /kʰɑ.dɯ.ɛ p*a.ha.ɾɑ.go dʑʌg.ʌ.s*ʌ.jo/
CF	Q: 카드에 키하라고 적었어요? (Did (you) write do /kʰi/ on the card?) /kʰɑ.dɯ.ɛ **kʰi**.ha.ɾɑ.go dʑʌ.gʌ.s*ʌ.jo/ A: 아니요, 카드에 **빠**하라고 적었어요. (No, (I) wrote do /**p*a**/ on the card.) /ɑ.ni.jo kʰɑ.dɯ.ɛ **p*a**.ha.ɾɑ.go dʑʌ.gʌ.s*ʌ.jo/

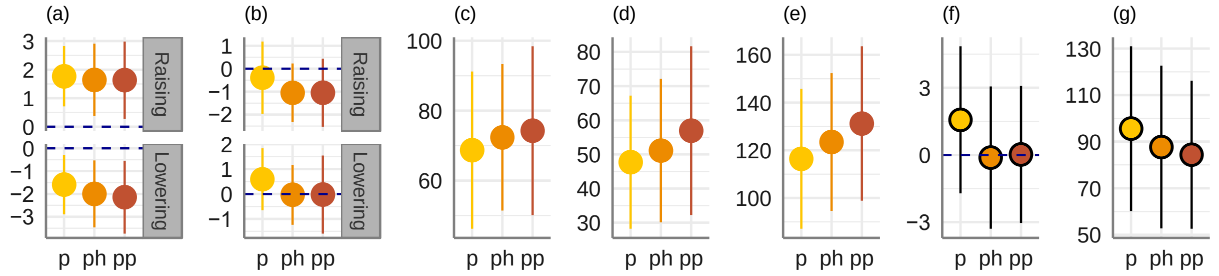

Figure 1 (a) ONSET-to-TARGET, (b) TARGET-to-OFFSET distance (.mm), (c) ONSET-to-TARGET, (d) TARGET-to-OFFSET distance duration (.ms), (e) ONSET-to-OFFSET duration of BOs, (f) ONSET-to-TARGET distance and (g) duration of V ('p' = /p/, 'ph' = /pʰ/. 'pp' = /p*/).

Table 2 Post-hoc test results of the models estimating the conditional differences shown in Figure 1.

Estimates	**(a) Rai. / Low.**		**(b) Rai. / Low.**		**(c)**	**(d)**	**(e)**	**(f)**	**(g)**
p – ph	0.12	**0.41***	**0.63****	**0.67****	**-3.98***	-3.47	**-7.51***	**1.37*****	**7.18***
p – pp	0.13	**0.55****	**0.62****	**0.66****	**-6.09*****	**-9.43*****	**-15.67*****	**1.20*****	**11.26****
ph – pp	0.01	0.14	-0.01	-0.02	-2.11	**-5.96****	**-8.16****	-0.16	4.08

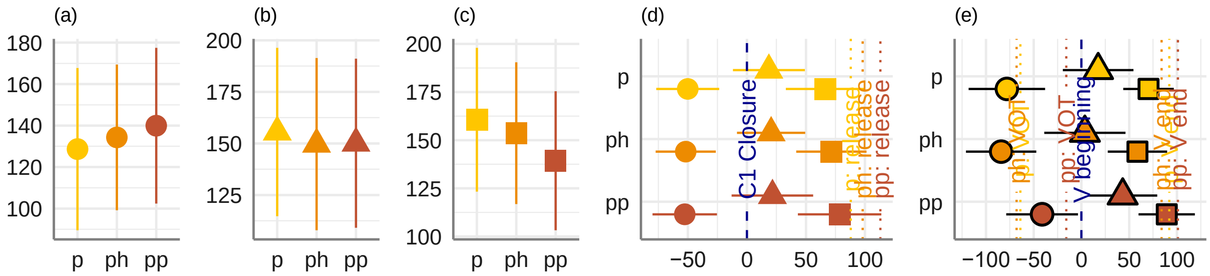

Figure 2 Intergestural timing (.ms) between (a) onsets (circles), (b) targets (triangles), and (c) offsets (squares) of bilabial C1 and V. Relative timing (d) between onsets, targets, and offsets of C1 and acoustic C1 closure (blue dashed line) and (e) between those of following Vs from the acoustic V beginning (blue dashed line).

Table 3 Post-hoc test results of the models estimating the conditional differences in Figure 2.

Estimates	**(a)**	**(b)**	**(c)**	**(d) C1 clo. – C1 on.**	**C1 clo. – C1 p.**	**C1 clo. – C1 off.**	**(e) V begin. – C1 clo.**	**V begin. – C1 clo.**	**V begin. – V clo.**
p – ph	-6.16	5.98	7.25	1.73	-2.24	-5.78	6.05	**14.29*****	**12.04*****
p – pp	**-12.06****	5.56	**21.84*****	2.67	-3.36	**-12.95*****	**-36.85*****	**-25.26*****	**-18.49*****
ph – pp	-5.90	-0.43	**14.59*****	0.94	-1.11	**-7.17***	**-42.91*****	**-39.54*****	**-30.54*****

References

[1] Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using *lme4*. *Journal of Statistical Software*, *67*(1).
[2] 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.
[3] Halekoh, U., & Højsgaard, S. (2014). A Kenward-Roger Approximation and Parametric Bootstrap Methods for Tests in Linear Mixed Models—The *R* Package *pbkrtest*. *Journal of Statistical Software*, *59*(9).
[4] Kang, Y., Schertz, J., & Han, S. (2022). The Phonology and Phonetics of Korean Stop Laryngeal Contrasts. In S. Cho & J. Whitman (Eds.), *The Cambridge Handbook of Korean Linguistics* (pp. 215–247).
[5] Kim, H., Honda, K., & Maeda, S. (2005). Stroboscopic-cine MRI study of the phasing between the tongue and the larynx in the Korean three-way phonation contrast. *Journal of Phonetics*, *33*(1), 1–26.
[6] Ohala, J. J. (1978). Production of Tone. In V. A. Fromkin (Ed.), *Tone: A Linguistic Survey* (pp. 5–39). Elsevier.
[7] Stone, M. (2005). A guide to analysing tongue motion from ultrasound images. Clinical Linguistics & Phonetics, 19(6–7), 455–501.
[8] Svirsky, M. A., Stevens, K. N., Matthies, M. L., Manzella, J., Perkell, J. S., & Wilhelms-Tricarico, R. (1997). Tongue surface displacement during bilabial stops. *The Journal of the Acoustical Society of America*, *102*(1), 562–571.
[9] Vazquez-Alvarez, Y., & Hewlett, N. (2007). The 'Trough Effect': An Ultrasound Study. *Phonetica*, 64(2–3), 105–121.
[10] Wrench, A., & Balch-Tomes, J. (2022). Beyond the Edge: Markerless Pose Estimation of Speech Articulators from Ultrasound and Camera Images Using DeepLabCut. *Sensors*, *22*(3), 1133.