A PRODUCTION STUDY OF KOREAN CONSONANTS

Jeremy Perkins[a], Dahm Lee[b], Seunghun J. Lee[c]

University of Aizu[a], Seoul National University[b], International Christian University[c], IIT Guwahati[c]
jperkins@u-aizu.ac.jp, dahm1021@snu.ac.kr, seunghun@icu.ac.jp

ABSTRACT

A production study of the Korean three-way contrast of tense, lenis and aspirated obstruents was conducted that included all manners. In this paper, psychoacoustic roughness as a measure of laryngeal constriction was used to explore the phonetic acoustic properties of tense obstruents. Because it is a perceptual attribute, it may have stronger correlation with laryngeal constriction than spectral tilt. In addition, the behavior of the Korean laryngeal contrast among the fricatives and affricates was explored.

Tense obstruents correlated with increased roughness and reduced spectral tilt at the following vowel onset. In addition, tense fricatives had longer frication duration than aspirated fricatives; tense stops and affricates had longer closure duration than other obstruents; lenis obstruents had shorter closure duration. Closure and frication duration differences in stops and fricatives, respectively, were larger than in affricates, suggesting they play a more important role in producing the contrast in stops and fricatives.

Keywords: Korean obstruents, laryngeal setting, psychoacoustic roughness

1. INTRODUCTION

Korean has a three-way contrast (lenis vs. tense vs. aspirated) among stops and affricates. The inventory of Korean obstruents is shown in Table 1. Each cell in the table shows the possible laryngeal settings for each combination of place and manner.

Type	Labial	Coronal	Dorsal
Stops	p p’ pʰ	t t’ tʰ	k k’ kʰ
Affricates		tʃ tʃ’ tʃʰ	
Fricatives		s’ s	

Table 1: Korean obstruents by place & manner

This contrast shown in Table 1 has received much attention with most of that attention focused on the stop series [1, 2]. Previously, Seoul Korean speakers produced these obstruents with different voice-onset times (VOTs) [3, 4]. However, current speakers produce lenis stops with a large positive VOT, just like the aspirated series. Word-initially, a contrast in f0 of the following vowel has been observed, with lenis obstruents preceding low f0 and aspirated obstruents preceding high f0 [5, 6, 7, 8]. Despite this shift in cues from VOT to f0, the tense series still has a shorter VOT than lenis and aspirated [5]. Some research has found that VOT for lenis stops may have an intermediate value between tense and aspirated [3], and may perceptually not be distinguishable from the other two [9]. Kim et al. [10] showed that cues on the following vowel alone were necessary and sufficient in perception of the lenis category.

While this previous research has mostly focused on stops, there is less research on the status of affricates and fricatives. Some past research suggests affricates pattern with the stops [11]. Unlike stops and affricates, there is only a two-way contrast among the fricatives. In [12], frication duration was found to be longer in tense fricatives, but no f0 difference was found.

Laryngeal setting is also involved in the three-way contrast, and has been measured via spectral tilt. At higher frequencies, intensity drops relatively less in creaky voice than in modal voice [13], resulting in reduced values for spectral tilt measures like H1*−H2* with tense obstruents. On the other hand, lenis stops are often noted to involve breathiness, which are associated with significantly reduced intensity at higher frequencies [14, 12, 6].

The goals of this research are (1) to use psychoacoustic roughness to identify coarticulatory phonation types associated with the laryngeal contrast, in particular with (a) the tense series and (b) among fricatives and affricates; and (2) to assess which cues could be involved in the laryngeal contrast, based on the acoustic findings.

2. METHODS

2.1. Consultants

A production experiment was run with 24 Seoul Korean speakers between the ages of 20 and 27 (14 females). Most speakers spoke some English and all speakers except one reported no problems affecting their speech at the time of the recording.

2.2. Stimuli

A list of 66 CVCV bisyllables was created (most nonce), with all fourteen obstruents represented in the first consonant (C1), which was the focus of the main research question. All vowel positions were filled with the low vowel [a]. With labial and dorsal C1 (p, p', p^h, k, k', k^h), only the three lenis stops (p, t, k) were used in the second consonant (C2), yielding 18 words. With coronal C1 (t, t', t^h), the three aspirated (p^h, t^h, k^h) and the three lenis stops (p, t, k) were used in C2, yielding 48 more words, for a total of 66 words. More coronals were used because the analysis of fricatives is only done among coronals. A separate analysis of the tense series is done with all places though. As such, more coronals were needed to achieve the same relative statistical power among the places of articulation.

2.3. Procedure

Each word was embedded into a Korean carrier sentence, 단어 X 는 무슨 뜻인가요?, meaning "What does the word X mean?" The 66 sentences were arranged in a random order and presented twice via slides displayed on a screen. Five practice items were included. Participants wore a head mounted Shure WH-30 microphone connected to a Tascam MK-2 with an XLR cable. A native Korean-speaking experimenter advanced the slides, noting any pronunciation errors. The resulting audio files were saved as mono wav files at a sampling rate of 44.1 kHz.

All files were automatically segmented and labeled using a Korean forced alignment tool [15]. Further annotation of the audio files was done using Praat [16]. Additional segments were labeled for consonant closure, frication, and aspiration. All boundaries were automatically moved to the nearest zero-crossing.

2.4. Data extraction

Laryngeal constriction (here, spectral tilt and psychoacoustic roughness) at the onset of a following vowel, VOT, closure duration, frication duration, and f0 in the vowel following C1 were measured. VoiceSauce [17] was used to measure spectral tilt, f0 and duration. Formant-normalized spectral tilt was measured via H1*−H2*, H1*−A1*, and H1*−A2*. F0 was normalized within each speaker relative to that speaker's median f0, to account for higher- and lower-pitched voices. F0 and spectral tilt measurements were made during the entire vowel, at 10 ms intervals. F0 was converted to cents (hundredths of a semitone). Duration measures included closure duration for stops and affricates, frication and aspiration duration for affricates and fricatives, and VOT for stops. Release duration (frication plus aspiration) was also included for affricates and fricatives.

Psychoacoustic roughness is defined as a sensory attribute related to rapid changes on the amplitude envelope of a sound (15–300 Hz). It was found to correlate positively with creaky tones in Burmese [18]. As such, it can provide a psychoacoustic measure of creakiness, like loudness or pitch for intensity and f0, respectively. Since it relates to perception, it may relate more closely to phonological differences, so if laryngeal constriction is used as a cue to distinguish tense from lax, roughness may correlate more strongly with this difference than spectral tilt. Roughness was measured in Aspers in the vowel following the first consonant using the Matlab routine from Villegas et al. [18].

2.5. Analysis

F0, spectral tilt and roughness were plotted via smooth scatterplots using geom_smooth with either the gam or loess method automatically chosen in R's ggplot2 library [19, 20].

Linear mixed models were constructed for all duration measures with random intercepts for speaker. Fixed effects for obstruent type, place of articulation and manner were included with mutual interactions, when relevant. Type III ANOVAs with Satterthwaite's method were used to confirm fixed effect significance, with any insignificant effects removed in the final models. The model intercepts are lenis alveolar stops, unless otherwise noted.

3. RESULTS

The combined results are plotted in the following figures by laryngeal setting (tense is light grey or with long dashed lines, aspirated is dark grey or with short dashed lines, and lenis is solid black). Results for VOT are not illustrated as they conformed with previous findings [5] (effect sizes in linear mixed model: lenis (intercept): 73.0 ms; tense: −64.0 ms; aspirated: −7.5 ms, all differences significant with $p < 0.001$).

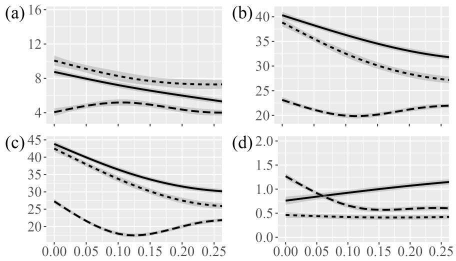

Figure 1: Stops. Spectral tilt in dB, (a) H1*−H2*, (b) H1*−A1*, (c) H1*−A2* and (d) psychoacoustic roughness in Aspers by laryngeal setting in the initial 25% of the vowel (normalized time on the x-axes)

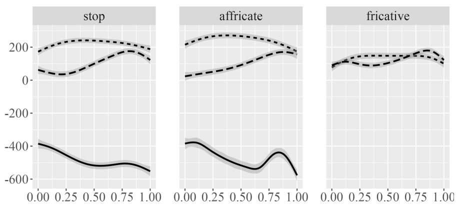

Figure 2: Coronals. f0 in cents relative to speaker median by laryngeal setting and manner (normalized time on the x-axes)

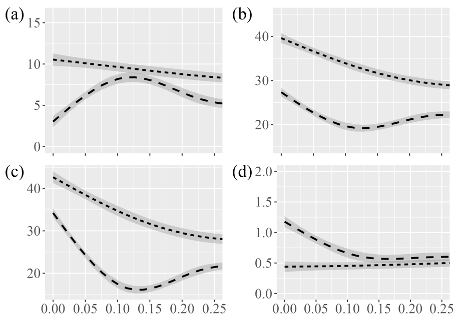

Figure 3: Fricatives. Spectral tilt in dB, (a) H1*−H2*, (b) H1*−A1*, and (c) H1*−A2*, and (d) psychoacoustic roughness in Aspers by laryngeal setting in the initial 25% of the vowel (normalized time on the x-axes)

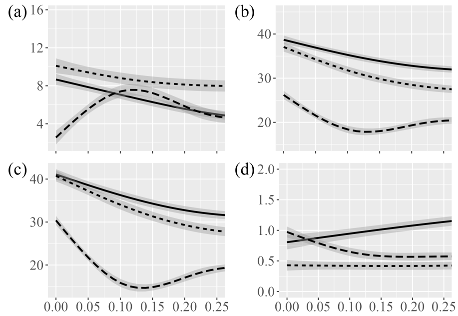

Figure 4: Affricates. Spectral tilt in dB, (a) H1*−H2*, (b) H1*−A1*, and (c) H1*−A2*, and (d) psychoacoustic roughness in Aspers by laryngeal setting in the initial 25% of the vowel (normalized time on the x-axes)

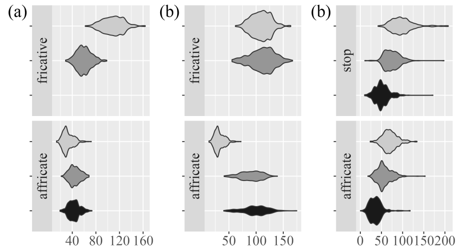

Figure 5: (a) Frication duration, (b) release duration, and (c) closure duration, all in ms, by manner and laryngeal setting

3.1. Laryngeal constriction and f0 in tense obstruents

Laryngeal setting was measured among the stop series via spectral tilt and psychoacoustic roughness, as shown in Fig. 1. In tense stops, reduced spectral tilt and raised roughness was seen at vowel onset, indicating the presence of coarticulatory creaky phonation. This effect extended throughout the initial 25% of the vowel for H1*−A1*, and H1*−A2*, but only for about the first 5% to 10% for H1*−H2* and roughness. On the other hand, aspirated and lenis stops induced raised spectral tilt, indicative of breathiness. Additionally, roughness increased over the initial 25% of the vowel following a lenis stop. This is likely due to the lowered f0 of the lenis stop (see Fig. 2); roughness has been found to correlate inversely with f0 during modal phonation in Thai for example [21]. Despite this, roughness is actually reduced at vowel onset where coarticulatory effects of the lenis obstruent are expected. Like spectral tilt, roughness results also indicate the presence of laryngeal constriction following tense stops.

Fricatives and affricates showed the same pattern as the stops: The tense series had reduced spectral tilt difference and increased roughness at the following vowel onset (see Fig. 3 for fricatives and Fig. 4 for affricates). This effect extended throughout the initial 25% of the vowel for H1*−A1*, and H1*−A2*, but only the initial 10% for H1*−H2* and roughness, like the stops. Regarding f0, there was almost no difference between tense and aspirated fricatives (see Fig. 2). However, the affricates showed the same pattern as the stops, where f0 was lowest following lenis affricates and higher following aspirated and tense affricates, with the tense series having a slightly lower f0 than the aspirated series.

3.2. Duration measures

Fig. 5a shows frication duration by manner and laryngeal setting. The linear mixed model had significant effects for laryngeal setting ($F = 4170$, $p < 0.001$), manner (fricative vs. affricate) ($F = 49,108$, $p < 0.001$) and their interaction ($F = 21,814$, $p < 0.001$). The intercept condition, here [$tʃ^h$], had a frication duration of 44.0 ms ($t = 29.4$, $p < 0.001$). The tense series had the shortest frication duration (-11.8 ms, $t = -39.9$, $p < 0.001$) and lenis affricates were slightly shorter than aspirated affricates (-1.3 ms, $t = -4.2$, $p < 0.001$). Fricatives had longer frication duration than affricates (+15.3 ms, $t = 47.8$, $p < 0.001$). A highly significant interaction effect was seen, with tense fricatives having very long frication durations (+61.2 ms, $t = 147.7$, $p < 0.001$).

Fig. 5b shows the results for release duration (frication plus aspiration). Tense obstruents do not include periods of aspiration, unlike the aspirated and lenis series. As a result, the tense series had significantly shorter release durations, which may play an important role in distinguishing the tense series. Aspirated fricatives, meanwhile had very similar release durations to tense fricatives and so release duration is likely less important than frication duration among fricatives.

The results for closure duration are shown in Fig. 5c. A linear mixed model was run only on coronals to test for differences between laryngeal setting and between affricates and stops (manner). The intercept condition, [t^h], had a closure duration of 74.3 ms ($t = 30.0$, $p < 0.001$). Affricates were found to have shorter closure duration than stops (-17.9 ms, $t = -35.7$, $p < 0.001$). The lenis series had the shortest closure duration (-23.2 ms, $t = -48.1$, $p < 0.001$) and the tense series had the longest (+23.2 ms, $t = 51.5$, $p < 0.001$). A significant interaction between laryngeal setting and manner indicated that the relative differences in closure duration for each laryngeal setting were shorter among the affricates ($F = 256.0$, $p < 0.001$). Lenis affricates had relatively longer closure duration (+2.6 ms, $t = 3.8$, $p < 0.001$), and tense affricates had relatively shorter closure duration (-10.3 ms, $t = 3.8$, $p < 0.001$).

4. DISCUSSION

Word-initial tense obstruents saw increased roughness and decreased spectral tilt, indicative of laryngeal constriction across all manners and places of articulation. In addition, duration differences were found; tense affricates had longer closure duration, shorter frication duration, and significantly shorter release duration. Meanwhile, tense fricatives had longer frication duration (confirming [12]) and tense stops had longer closure duration.

Regarding the lenis/aspirated distinction, this was traditionally made via VOT, with shorter VOT for lenis obstruents. We explored other cues, including those related to laryngeal setting. Small differences were seen in spectral tilt and roughness, but individual speaker variation was considerable and so these cues are unlikely to be useful. However, regarding closure duration, coronal lenis stops and affricates had consistently shorter closure duration for all speakers (notably, dorsals and labials were less consistent across speakers). Here, future work would be needed to determine if these closure duration differences are perceptually relevant in Korean.

5. CONCLUSION

A production study of the three-way contrast in Korean between tense, lenis and aspirated obstruents confirmed previous research showing an emerging tonal contrast with lenis obstruents preceding low tone and aspirated and tense obstruents preceding high tone. This tonal contrast was seen here for fricatives and affricates as well. In addition to spectral tilt, psychoacoustic roughness also confirmed previous findings that tense obstruents occur with laryngeal constriction, which was seen in tense fricatives and affricates in addition to stops. Among the fricatives, phonation cues in the following vowel and differences in frication duration were found.

VOT is unavailable as a cue among affricates and fricatives, but our results suggested frication duration and release duration (frication plus aspiration) may fill its role among the fricatives and affricates, respectively. Among the stops and affricates, closure duration was shorter for lenis and longer for tense; however, the difference between aspirated and lenis was small enough that its status as a perceptual cue is unclear. Future research on closure duration as a perceptual cue is needed to address this question.

6. ACKNOWLEDGEMENTS

This research was supported by JSPS KAKENHI Grant-in-Aid for Early Career Scientists Number JP19K13162 and also by the ILCAA joint research project “Phonetic typology from cross-linguistic perspectives (PhonTyp)” (2021-2023).

7. REFERENCES

[1] M. R. Kim and S. Duanmu, “"tense" and "lax" stops in korean,” *Journal of East Asian Linguistics*, vol. 13, no. 1, pp. 59–104, 2004.

[2] H. Lee, J. J. Holliday, and E. J. Kong, “Diachronic change and synchronic variation in the korean stop laryngeal contrast,” *Language and Linguistics Compass*, vol. 14, no. 7, 2020.

[3] L. Lisker and A. S. Abramson, “Cross-language study of voicing in initial stops: acoustical measurements,” *Word*, vol. 20, pp. 384–422, 1964.

[4] C. W. Kim, “On the autonomy of the tensity feature in stop classification,” *Word*, vol. 21, pp. 339–359, 1965.

[5] D. J. Silva, “Acoustic evidence for the emergence of tonal contrast in contemporary korean,” *Phonology*, vol. 23, pp. 287–308, 2006.

[6] K.-H. Kang and S. G. Guion, “Clear speech production of korean stops: Changing phonetic targets and enhancement strategies,” *Journal of the Acoustical Society of America*, vol. 124, no. 6, pp. 3909–3917, December 2008.

[7] J. Kirby, *Origins of sound change: Approaches to phonologization*. Oxford University Press, 2013, ch. The role of probabilistic enhancement in phonologization, pp. 228–246.

[8] H.-Y. Bang, M. Sonderegger, Y. Kang, M. Clayards, and T.-J. Yoon, “The emergence, progress, and impact of sound change in progress in seoul korean: Implications for mechanisms of tonogenesis,” *Journal of Phonetics*, vol. 66, pp. 120–144, 2018.

[9] H. Lee and A. Jongman, “Effects of tone on the three-way laryngeal distinction in korean: An acoustic and aerodynamic comparison of the seoul and south kyungsang dialects,” *Journal of the Acoustical Society of America*, vol. 42, no. 2, pp. 145–169, August 2012.

[10] M. R. Kim, P. S. Beddor, and J. Horrocks, “The contribution of consonantal and vocalic information to the perception of korean initial stops,” *Journal of Phonetics*, vol. 30, pp. 77–100, 2002.

[11] M. Kim, “Correlation between vot and f0 in the perception of korean stops and affricates,” in *Eighth International Conference on Spoken Language Processing*, 2004.

[12] T. Cho, S.-A. Jun, and P. Ladefoged, “Acoustic and aerodynamic correlates of korean stops and fricatives,” *Journal of Phonetics*, vol. 30, pp. 193–228, 2002.

[13] E. Abberton, “Some laryngographic data for korean stops,” *Journal of International Phonetic Association*, vol. 2, pp. 67–78, 1972.

[14] H. Ahn, “Post-release phonatory processes in english and korean: Acoustic correlates and implications for korean phonology,” Ph.D. dissertation, The University of Texas at Austin, 1999.

[15] T.-J. Yoon, “Korean forced alignment system,” Oct. 2022. [Online]. Available: https://tutorial.tyoon.net

[16] P. Boersma and D. Weenink, “Praat: doing phonetics by computer (version xxx),” 2022.

[17] Y.-L. Shue, “The voice source in speech production: Data, analysis and models.” Ph.D. dissertation, UCLA, 2010.

[18] J. Villegas, K. Markov, J. Perkins, and S. J. Lee, “Prediction of creaky speech by recurrent neural networks using psychoacoustic roughness,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 2, pp. 355–366, 2020.

[19] H. Wickham, *ggplot2: Elegant Graphics for Data Analysis*. New York: Spring-Verlag, 2016.

[20] R Core Team, *R: A language and environment for statistical computing*. URL: https://www.R-project.org/: Vienna, Austria: R Foundation for Statistical Computing, 2019.

[21] J. Perkins, “Acoustic measurement of laryngeal constriction in thai consonants,” in *35th General Meeting of the Phonetics Society of Japan*, September 25-26 2021.