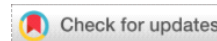




pISSN 2005-8063
eISSN 2586-5854
2026. 9. 30.
Vol.18 No.3
pp. 73-81



Acoustic realization and category split in L1 English speakers' L2 Korean stop production*

Eun Jong Kong¹ · Soyoung Kang^{2,**} · Misun Seo³

¹*School of Liberal Arts and Sciences, Korea Aerospace University, Goyang, Korea*

²*School of Linguistics and Language Studies, Carleton University, Ottawa, ON, Canada*

³*Department of English Language and Literature, Hannam University, Daejeon, Korea*

Abstract

This study investigated L2 Korean lenis and aspirated stops produced by L1 English adult learners in the early stages of L2 learning, examining how they utilize voice onset time (VOT) and F0 for category distinction over the course of L2 learning. Narrative speech samples (568 word-initial stops) were collected from 34 university students enrolled in first-year (25 instruction hours) and second-year (125 instruction hours) Korean courses in Canada. Bayesian mixed-effects logistic regression models were constructed to quantify the acoustic structures of the VOT and F0 for lenis-aspirated types across learner groups. The results revealed that the predictive roles of VOT and F0 were patterned differently by proficiency: first-year speakers solely utilized VOT, whereas second-year speakers exhibited increased use of F0 and reduced use of VOT. In addition, higher prediction accuracies based on the acoustic models were observed for second-year students than for first-year students, and for lenis stops than for aspirated stops in both groups. This prediction accuracy distribution suggests that the initial acoustic structure of the undifferentiated category is closer to the lenis target in its production representation. The current findings are significant in depicting the category split of L2 sounds assimilated into a single L1 category in terms of English speakers' acoustic realization of Korean stops.

Keywords: L1 English learners of L2 Korean, L2 Korean stop production, lenis-aspirated contrast, voice onset time, fundamental frequency, category split

1. Introduction

It is well-established that adult L2 sound learning is heavily mediated by the learner's ingrained L1 acoustic, phonetic and

phonological knowledge in production (Flege et al., 2021) and perception (Best & Tyler, 2007; Kuhl et al., 2007). New sounds in L2 are processed by assimilating or dissimilating with the learners' existing native categories. Consequently,

* This work was supported by the National Research Foundation of Korea (No. NRF-2024S1A5A2A01024518) awarded to the first author.

** kang139@gmail.com, Corresponding author

Received 15 August 2026; Revised 11 September 2026; Accepted 12 September 2026

© Copyright 2026 Korean Society of Speech Sciences. This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

adult L2 learners' relative difficulties in mastering the new sounds are predicted based on the discrepancy of acoustic-phonetic and phonological characteristics between L1 and L2 sounds. Specifically in the context of learning the Korean stops with a laryngeal contrast, L1 English learners of L2 Korean are to deal with the L1-L2 differences in terms of more complex phonological structure and multi-dimensional acoustic realization of the target contrast. The L2 Korean stops are contrastive in three ways among the fortis-lenis-aspirated types along multiple acoustic dimensions such as voice onset time (VOT: asp.=lenis>fortis) and F0 (fundamental frequency: asp.>fortis>lenis) in bilabial (ㅂ /p/, ㅃ /p^h/, ㅅ /p^h/), alveolar (ㄷ /t^{*}/, ㅌ /t/, ㅈ /t^h/) and velar (ㄱ /k^{*}/, ㅋ /k/, ㆁ /k^h/) places (Lee et al., 2020, among others). Compared to this, the L1 English stops are differentiated in two ways between voiced (phonetically fully voiced or unaspirated variants, /b d g/) and voiceless (phonetically aspirated /p t k/) types sufficiently along a single acoustic dimension of VOT (voiceless>voiced) with F0 redundantly utilized (voiceless>voiced) (Gordon et al., 1993; Llanos et al., 2013). Given the complexity of the stop laryngeal contrast in the target language, this study aims to investigate how L1 English speakers realize the acoustic characteristics of L2 Korean stops focusing on the lenis-aspirated stops produced by the beginner level classroom learners.

1.1. Predicted Difficulty in L2 Production

Reflecting the L1-L2 mismatch of the categories, several identification studies reported that English native speakers tended to map both the aspirated and lenis stops in Korean onto the English voiceless stops with the fortis stops onto the English voiced stops (Lee, 2024; Nagle et al., 2023; Schmidt, 2007). The Perceptual Assimilation Model-L2 (PAM-L2) by Best & Tyler (2007) defines this as a single category (SC) assimilation, i.e., two L2 categories mapped onto a single L1 category, and predicts a challenge for L2 learners because a category split needs to occur in the listeners' L2 representation. Such a perceptual challenge is demonstrated in Lee (2024) where L1 English listeners were poorer at discriminating the lenis-aspirated stop contrast compared to the lenis-fortis or aspirated-fortis contrast.

This perceptual deficit is highly likely to manifest as a corresponding challenge in L2 production. According to the Speech Learning Model-r (SLM-r; Flege et al., 2021), production of new L2 sounds is closely constrained by perceptual assimilation and dissimilation between L1-L2 categories. That is, L2 learners are predicted to assimilate new L2 sounds into familiar L1 sounds to develop a composite category, whereas a new, separate category is formed only when assimilation is blocked due to perceived phonetic dissimilarity. With this framework applied to the present context of interest, the inexperienced English-speaking learners of Korean would produce both lenis and aspirated stops as undifferentiated phonetic categories (phonetically similar to their L1 voiceless stops) with poor acoustic distinction due to their perceptual similarity to the single L1 category, a voiceless stop.

Given the challenge in both perception and production of the lenis and aspirated types, it is not known how the category split is realized in the inexperienced English-speaking learners' stop production. If the undifferentiated category is the aspirated type intended by the learners, native-like acoustic characteristics of lenis

stops would emerge gradually over the course of learning, and vice versa. English speakers' goodness ratings of the lenis and aspirated stops as L1 voiceless stops were comparably high (Lee, 2024; Nagle et al., 2023), indicating that neither one was a better exemplar than the other. This renders it difficult to predict in which pattern the category split would occur in the production of L2 Korean stops over time.

1.2. Learner's Acoustic Realizations of the Stops

Two useful acoustic studies are worth mentioning: They compared the acoustic realizations of the L2 Korean stops produced by beginning and advanced level English-speaking adult learners who were residing in Seoul (Kim et al., 2006), and Gwangju (Oh & Kim, 2011). Examining the beginning level learners' acoustic realizations of VOT and F0, the lenis stop VOT values were intermediate – instead of being comparable to that of the aspirated stops – between the fortis and the aspirated stops (asp.>lenis>fortis), and the F0 of the lenis stop were no lower than the fortis stops (asp.>fortis=lenis). Even for the advanced learners taking the highest level of the program, the acoustic properties of the fortis-lenis-aspirated stops were not native-like with the lenis type judged by native speakers as incorrect more than the aspirated and fortis types. Oh & Kim (2011)'s study echoed these patterns of VOT and F0 realizations by showing that the L1 English-speaking learners in their first year of instruction in Korea produced the lenis stops with a lag VOT significantly shorter than the aspirated stops (asp.>lenis), and with a F0 no lower than the aspirated stops (asp.=lenis). Another group of L1 English learners with two- and three-years of academic instruction realized the lenis stop F0 low enough to be differentiated from the aspirated stops (asp.>lenis), but still failed to realize the lag VOT long enough to be comparable to the aspirated stops (asp.>lenis).

Both studies demonstrated that the English-speaking learners, especially at the beginner level, employed VOT solely for a lenis-aspirated distinction, while F0 did not differentiate the categories. The new L2 category seems to emerge over the course of learning as F0 differences are observed in the lenis and aspirated stops in the advanced learners. Findings from Kong et al. (2022) supports this trend where a higher L2 Korean proficiency based on TOPIK scores was correlated with more use of F0 in the stop production. What is still unclear is whether it is the lenis that is split from the aspirated or the other way around, i.e., which is default category in the learners' production? Experimental evidence from neither study provided explicit answer to this. The present study aims to fill this gap.

1.3. The Current Study

The present study investigates acoustic realizations of L1 English-speaking adults' productions of the L2 stops to examine whether and how the new category split emerges in the production of the lenis-aspirated stops. Although English-speaking learners of Korean serve as a valuable testing ground for the category split proposed by SLM-r, acoustic research on this population is fairly limited, and to date, no experimental work has explored this question.

Specifically, the current study examined the stop productions of pseudo-spontaneous speech produced by L1 English speaking adult learners of L2 Korean in a classroom of their home country, Canada. Classroom settings in the learners' home country present

a distinct environment from naturalistic immersion as the existing studies show (Kim et al., 2006; Kong et al., 2022; Oh & Kim, 2011) in that L2 input is highly limited and L1 usage is still dominant especially for beginners. Although this context reflects the most common learning condition for the majority of global L2 learners, phonetic studies on these classroom learners are still rare within the field of L2 Korean. This adds a diversity of learner environments by investigating the production patterns of these learners during their initial learning stages free from the compounding variables of daily L2 exposure. Also, acoustic analysis of spontaneous speech allows us to capture the dynamic and natural phonetic variations that occur during real-time speech production (Labov, 1972).

By analyzing the L2 lenis-aspirated stops from spontaneous speech, we aim to pursue the two research goals as follows. First, we describe how VOT and F0 are utilized relatively in the L2 Korean stops focusing on the aspirated vs. lenis contrast. By examining relatively inexperienced learners of Korean, we can address the acoustic nature of their undifferentiated, composite category for the stops. Second, we compare the acoustic structure of the stops for the first year and the second-year students to observe how the beginning level speakers split a composite category into new categories. This study contributes to robust language-specific evidence needed to support L2 production models.

2. Methods

2.1. Speech Materials: Oral Exam Recordings

Speech samples were produced by 34 native English-speaking university students enrolled in first-year (N=18: 2 males) and second-year (N=16: 3 males) Korean courses during the Fall 2023 semester at Carleton University, Ottawa, Canada.¹ At the time of recording, *first-year* students had 25 hours of instruction, and *second-year* students 125 hours.² While the samples were collected from the oral test, the speech is pseudo-spontaneous as students used pre-written scripts to talk about themselves. Students were provided with advance guidelines for the oral exam detailing length requirements, expected topics, and target structures. To ensure appropriateness for their proficiency level, the instructor edited the students' scripts for grammatical accuracy and language level. Students subsequently delivered their presentations using these instructor-corrected scripts. Topics were varied as students were asked to talk about themselves but common expressions were about where they lived, where they were from, which they liked to do, etc. All the recordings were done in class by one of the authors' laptop computer in an open field using Praat software with a sampling rate of 22,050 Hz and a 16-bit quantization depth in a single channel (mono). Only speech samples from native English speakers were used for the analyses.

The total duration of the speech materials was 2,213.7 seconds,

excluding silence intervals. The mean durations were 32.4 seconds (*SD*: 8.6) for the first-year students and 103 seconds (*SD*: 25.1) for the second-year students, exhibiting a considerable difference between the groups. The numbers of syllable-initial stops also differed, with second-year students having more tokens (Asp.: 299, Lenis: 1443, Fortis: 97) than first-year students (Asp.: 113, Lenis: 238, Fortis: 8). Because too few or no fortis tokens were found in the speech materials from the first-year students, we only included lenis and aspirated stops for the group comparison analysis in the current study. Regarding the scarcity or absence of fortis tokens in the students' oral test speech, we suspect that the input of fortis-initial words was too few during the classroom instruction. Based on the glosses of the textbooks (Jeon et al., 2020), only six and 26 fortis-initial words appeared for the first- and second-year textbook, respectively, whereas the lenis- or aspirated-initial word types more commonly appeared: 113 and 111 lenis initial words and 36 and 41 aspirated initial words for the first and second year, respectively. Although textbooks may not be the sole source for acquiring fortis-initial words, we suspect that this asymmetrical distribution of laryngeal types likely contributed to the infrequent production of fortis stops in the students' oral testing.

2.2. Acoustic Analysis

The recordings were initially segmented in vowels and consonants using Montreal Forced Aligner (McAuliffe et al., 2017) based on the transcripts submitted by the students. The authors manually inspected all the word-initial stops to identify their prosodic position in the accentual phrases (AP). Only the AP-initial tokens were included for further acoustic measurements, resulting in 568 tokens of the lenis and aspirated stops: 41 and 110 aspirated-initial stops, 75 and 342 lenis-initial stops for the first- and second-year speakers. Two acoustic indices—VOT and F0—were measured by the authors based on visually identified acoustic events on the spectrogram and waveform in Praat (Boersma & Weenink, 2026). VOT was calculated by subtracting timestamps of burst signal (sudden spike after silence in the spectrogram, 'burst' in Figure 1) from those of voicing onset (onset of voicing bar, 'vOnset' in Figure 1). F0 was the fundamental frequency measured at the 20ms into the following vowel, which is later converted into semitones relative to the respective group mean for each gender.

¹ This study was granted clearance by the Carleton University Research Ethics Board-A with the Ethics clearance ID (#122432).

² Unsurprisingly, a large majority—though not all—of the students had prior exposure to K-pop, K-dramas, and YouTube videos. Furthermore, several students noted that they had previously attempted to learn Korean using Duolingo. The instructor observed that students with self-reported exposure to K-dramas demonstrated superior pronunciation. However, the impact of alternative learning materials was not formally evaluated in this study.

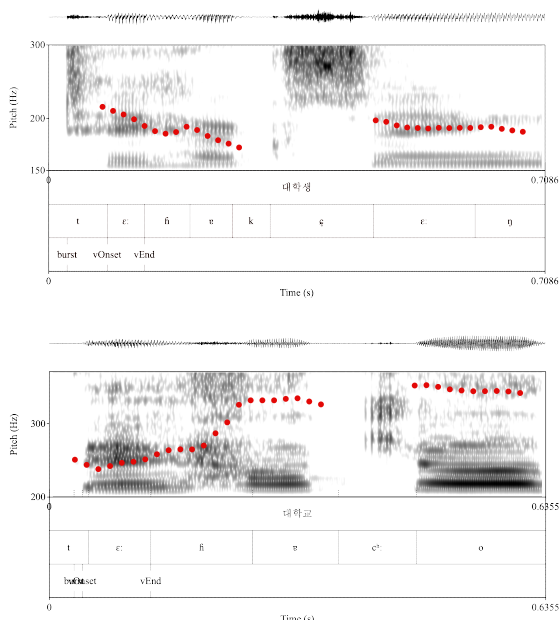


Figure 1. Two examples of word-initial lenis ‘ㄷ’ /t/ tokens realized in the words *대 학생* /te.hak.seŋ/ ‘an university student’ (top) and *대 학교* /te.hak.gjo/ ‘an university’ (bottom), illustrating acoustic events (‘burst’, ‘vOnset’, and ‘vEnd’) and F0 realizations (high vs. low F0 for top and bottom panels).

2.3. Statistical Analysis

Bayesian generalized linear mixed-effects model (Bayesian GLMM) was constructed using the *brm* package (Bürkner, 2017) in R & RStudio (Posit team, 2025) to predict the stop laryngeal types (Lenis vs. Aspirated) based on the fixed variables of acoustic parameters (VOT and F0 in semitone, scaled) and their interaction with classroom instruction (Year: *first-year* and *second-year*).³ Random intercepts and random slopes of VOT and F0 by speaker were entered to estimate differential effects of acoustic variables at the speaker-level as a random effect.

Based on the acoustic structure of this Bayesian regression model, we further performed 10-fold cross-validation using a *kfold* function of the *brms* package.⁴ This cross-validation task produces predicted categories (Lenis or Aspirated) based on individual tokens’ VOT and F0 values in a way that 10% of the total data is predicted by the model based on the remaining 90% of the data iteratively 10 times. While the primary purpose of this subsequent analysis is to validate the robustness of the Bayesian model prediction, we used this to compare the predicted types based on the model acoustic structure to the target types intended by the L1 English learners, so that we can examine the match/mismatch rates of each level of the Year and Type variables. This would reveal how the statistical distinctiveness and

structural organization of the VOT-F0 space develop for the lenis-aspirated contrast across the different learning years.

3. Results

3.1. Descriptive Patterns: VOT and F0 by Year

When visually inspecting the distributions of VOT and F0 of the tokens in Figure 2, the lenis and aspirated stops produced by *first-year* speakers were poorly differentiated compared to those by *second-year* speakers. While the first-year speakers’ individual tokens (smaller plot characters) as well as their mean values (bigger plot characters) largely overlapped for the lenis and aspirated stops in the left panel, the ellipse of the aspirated stops was located at a slightly longer VOT range with a substantially overlapping F0 range. Notably, the histogram of F0 in the figure forms two peaks, where the aspirated stops have a lower F0 peak than the lenis stops do, although the F0 values of both types overlap extensively.

In contrast, the differentiation between the lenis and aspirated stops was clearer for the second-year speakers in the right panel. Overall, more speakers in the second-year class produced aspirated stops with longer VOT and higher F0 than their lenis stops. When focusing on the single acoustic dimension of VOT, the peak of the VOT histogram for the lenis stops still aligns with that of the aspirated stops. In contrast, the F0 peak of the lenis stops was overall lower than that of the aspirated stops. The means and standard deviations of VOT and F0 were summarized in Appendix (Table A).

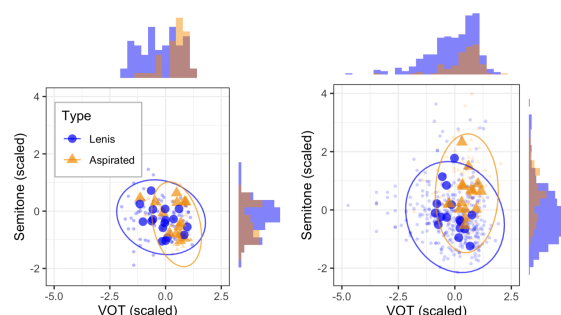


Figure 2. Scatterplots and histograms of voice onset time (VOT) and F0 in semitone separated by Year (left panel: *first-year*, right panel: *second-year*). Bigger plot characters represent individual speakers’ mean values, whereas smaller plot characters show individual tokens.

3.2. Statistical Model Results

Statistical results of Bayesian logistic mixed-effects model confirmed the observations that the first- and second-year students produced the lenis and aspirated stops with differential utilizations of VOT and F0. Figure 3 displays the differential effects of VOT

```
3 >brm(formula=Type ~ (VOT.scale + f0.scale) +
      VOT.scale:Year + f0.scale:Year +
      (VOT.scale + f0.scale | Speaker), data=data,
      family=bernoulli(link="logit"), control=list(adapt_delta=0.95), chains=4, iter=2000, cores=parallel::detectCores(), seed=1234)
Default priors provided by the brms package were used.

4 >kfold(model, K=10, save_fits=TRUE).
```

and F0 in predicting the aspirated stops over the lenis stops, and Table 1 presents the model summary.

Table 1 shows that VOT but not F0 was used by the *first-year* students (as a reference level) for the distinction between the lenis and aspirated stops. The posterior means of the group estimate yielded strong evidence that an increase of VOT ($\beta=1.67$) predicted the aspirated stops over the lenis stops in the first-year students' productions, as the entire 95% credible interval (CI) was well above zero for VOT (95% CI [0.85, 2.62]). However, there was no reliable evidence from the model that an increase of F0 ($\beta=0.2$) predicted the aspirated stops over the lenis stops produced by the *first-year* students, as the 95% CI of F0 estimate widely contained zero (95% CI [-0.92, -1.51]). These model estimates of the VOT and F0 variables support that the *first-year* students realized the aspirated stops with distinctively longer VOT compared to the lenis stops, while F0 did not differ between aspirated and lenis types. The logistic curves in Figure 3 visualize this result by displaying a steep curve of predicted probability along the VOT dimension (left panel) and an almost flat curve with fairly wide 95% confidence intervals along the F0 dimension (right panel).

In terms of interaction terms in this model, the estimate of VOT $\times$ Year (the difference in the effect of VOT between the *first* and *second* years) showed that the effect of VOT decreased for the distinction of the aspirated-lenis stops in the second-year speaker group relative to the first-year group. This evidence of interaction ($\beta_{\text{VOT}\times\text{Year}}=-0.87$) was not strong but moderate by having a 95% CI slightly covering zero (95% CI [1.84, 0.04]). Regarding the role of F0, the estimate of F0 $\times$ Year interaction (the difference in the effect of F0 between the *first* and *second* years) was positive indicating that the effect of F0 increased for the aspirated-lenis distinction produced by the second-year speakers. Evidence of this interaction ($\beta_{\text{F0}\times\text{Year}}=-0.87$) was not strong, however, with 95% CI slightly including zero (95% CI [-0.29, 2.49]). These results of the interaction terms indicate that the second-year speakers utilized VOT less and F0 more than the first-year speakers for the distinction between aspirated and lenis stops. In Figure 3, these patterns are illustrated by the second-year speakers' shallower VOT slope (left panel) and steeper F0 slope (right panel) than those of the first-year speakers.

To assure that the effects of VOT and F0 were credible for the second-year speakers, the same logistic mixed-effects model was fitted with the *second-year* group as the reference level. The posterior means of the group estimates were positive for both VOT and F0, and the 95% CIs were above zero ($\beta_{\text{VOT}}=0.81$, 95% CI [0.39, 1.33]; $\beta_{\text{F0}}=1.34$, 95% CI [0.62, 2.23]), indicating that the second-year speakers used F0 as well as VOT to distinguish aspirated and lenis stops. Based on the beta coefficients of the two acoustic parameters, the second-year speakers utilized F0 more than VOT for the contrast (F0>VOT), which is opposite to the first-year group (VOT>F0). Compared to the beta coefficients for the first-year speakers, VOT played a smaller role for the second-year group, whereas F0 played a greater role in the second-year speakers.

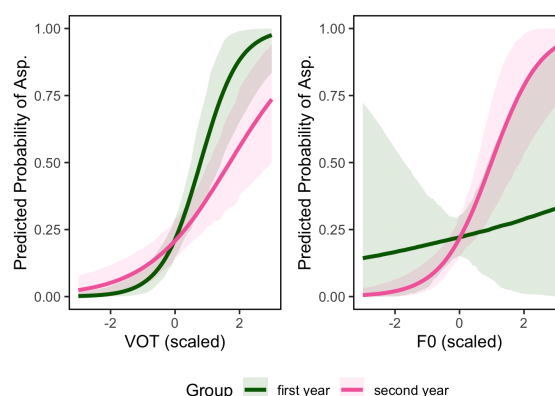


Figure 3. Predicted logistic curves of voice onset time (VOT; left) and F0 (right) for stop classification (aspirated over lenis type) produced by *first-year* and *second-year* L2 Korean learners.

Table 1. Posterior estimates of Bayesian mixed-effects regression model for lenis-aspirated stop differentiation (reference level for Year: *first-year*)

	Estimate	Est.Error	l%-95% CI	u-95% CI	Rhat
Intercept	-1.31	0.23	-1.77	-0.87	1
VOT	1.67	0.45	0.85	2.62	1
F0	0.2	0.62	-0.92	1.51	1
VOT: Year (second-year)	-0.87	0.48	-1.84	0.04	1
F0: Year (second-year)	1.13	0.7	-0.29	2.49	1

VOT, voice onset time; CI, credible interval.

With this Bayesian mixed effects regression model, we further conducted a 10-fold cross-validation to predict the lenis-aspirated types of individual tokens based on their VOT and F0. When utilizing the `kfold()` function in the `brms` package, the predicted categories (Lenis or Asp.) matched the speakers' target categories with an accuracy of 68.97% for the first-year speakers and 80.09% for the second-year speakers (Table 2). This accuracy difference suggests that the second-year speakers produce lenis and aspirated stops with clearer acoustic boundaries in the VOT-F0 space, indicating that the acoustic distributions of VOT and F0 become statistically more distinct and structured as the learning year increases. More interestingly, the prediction accuracy was higher for the lenis type than for the aspirated type in both groups of the learners: 81%>46% for the first-year group, and 94%>36% for the second-year group. This suggests that the acoustic structure of VOT and F0 in the model is likely to predict the lenis stops with more reliability and less ambiguity than the aspirated stops. As the accuracy distribution reflects the internal category mapping within the learner groups' acoustic space, the current patterns imply that the learners' lenis stops form a more stable acoustic target in the learners' VOT-F0 space than their aspirated stops, characterizing the latter as the acoustically more variable type.

4. Discussion

The present study investigated the spontaneous speech of L2 Korean produced by the first and second year L1 English-speaking learners who were taking classroom instruction at a university in

Table 2. 10-fold cross-validation accuracy matrix by Year and Type

Year	Type (counts)	Predicted lenis	Predicted aspirated	Type accuracy (%)	Overall accuracy [95% CI] (%)
First year	Lenis (75)	61	14	81.33	68.97 [59.71, 77.23]
	Aspirated (41)	28	19	46.34	
Second year	Lenis (342)	322	20	94.15	80.09 [76.10, 83.67]
	Aspirated (110)	70	40	36.36	

CI, credible interval.

Canada. Specifically, productions of the Korean stops were examined in terms of VOT and F0 domains, focusing on the lenis-aspirated contrast with a goal to investigate a category split from the composite one by describing the initial-stage acoustic properties in realizing the new L2 sounds. Echoing previous findings, the results from the current dataset showed that lenis and aspirated stops were poorly differentiated in the VOT-F0 space during the initial stage of L2 Korean learning. However, the learner group with higher proficiency exhibited shifts in acoustic properties toward L2 Korean, ultimately establishing a laryngeal contrast in an acoustically native-like manner.

One feature of the study is that the target consonants were naturally elicited in the pseudo-spontaneous speech. Due to that, the current data differed from those based on a word-list reading task and yielded unbalanced distributions of the word-initial target categories across the fortis-lenis-aspirated stops: The lenis stops were most frequently elicited by the first- and second-year learners, and the fortis type least frequently observed. We tentatively reason that this skewed distribution is most likely influenced by the input frequency from the textbooks in the classroom. Although the fortis stops were not included in the current analysis due to the narrow scope of the study, it seems necessary to investigate whether L2 Korean learners' higher production frequency in the spontaneous speech correlates with more native-like acoustic realizations of the L2 stops (fortis, lenis, and aspirated). If so, this would bear pedagogical implications regarding the importance of providing a balanced amount of segmental input for L2 sound acquisition.

In terms of acoustic realizations of the L2 Korean stops, the group of the first-year English-speaking learners tended to realize the L2 lenis-aspirated stops distinctively with each other utilizing the acoustic dimensions *not* in a native-like manner. This inexperienced learner group solely utilized VOT (relatively shorter VOT for the lenis), which is an acoustic parameter important in L1 stop contrast but not in L2 Korean stops as they overlap (e.g., Bang et al., 2018). Also, the F0 parameter was not employed by the beginner group, although the lenis-aspirated stops in Korean are differentiated primarily along F0 dimension. Notably, the F0 values for the learners' lenis stops were even higher than those for their aspirated stops. This is an opposite pattern of using F0 for the contrast by Korean native speakers (Aspirated > Lenis), and by the L2 learners' production elicited in a word-reading task (Oh & Kim, 2011, Aspirated > Lenis, descriptively but not statistically significant). Together, this means that for the English-speaking learners, the native-like use of VOT and F0 for the lenis-aspirated stop contrast does not emerge in their spontaneous speech until after one year of L2 learning. Importantly, F0 was not involved

in the English-speaking learners' composite category of the lenis-aspirated stops.

Different from the first-years' stops, the group of the second-year learners of L2 Korean showed slightly more distinct acoustic realizations of the lenis-aspirated stops. That is, the second-year speakers used F0 for a lenis-aspirated differentiation by realizing the aspirated type with higher F0 values. Also, their utilization of F0 was influential than that of VOT, with both dimensions involved for the contrast. Although the second-year learners' use of VOT for the contrast also differs from the native Korean speakers' sole use of F0 (Bang et al., 2018; Lee et al., 2024), this result shows that the English-speaking learners of L2 Korean were able to employ L2 specific acoustic characteristics in their spontaneous speech production within around two years of instruction. This evidence of the second-year learners' meaningful use of F0 is consistent with existing studies based on word-reading tasks over resident learners in Korea (Kim et al., 2006; Oh & Kim, 2011) in terms of relative role of acoustic parameters.

The differences between the learner groups indicate that for the lenis-aspirated type contrast, a category split from the composite category acoustically occurs with the learners' native-like use of F0, L2 specific primary acoustic dimension. More interestingly, when we depicted how the acoustic realizations of F0 and VOT changed as L2 learning continued, the acoustic splits of the lenis-aspirated stops along VOT and F0 occurred by producing higher F0 for the aspirated stops rather than by producing lower F0 for the lenis stops. This direction of F0 change associated with L2 Korean proficiency is consistent with findings in Oh & Kim (2011, figure 7 therein) and Kim et al. (2006, figure 6 therein) where more advanced speakers showed greater F0 for the aspirated stops than beginning learners, with little F0 change for the lenis stops across proficiency groups. One interpretation of these consistent results is that for the English-speaking adult learners of Korean, the acoustic structure of the lenis type is more similar to that of the default or undifferentiated category for the lenis-aspirated contrast. The multi-fold cross-validation task based on the acoustic models supports this interpretation, showing that the predicted categories by the acoustic models better matched with the lenis target than with the aspirated target. This suggests that the learners have more robust acoustic structure for the lenis stops and more variable tokens for the aspirated stops at least at the initial stage of L2 Korean learning.

One might be concerned that the robustness of the lenis category is a byproduct of its predominant token frequency in the data. However, this asymmetry itself reflects the natural distributional properties of L2 Korean. The distinct tightly clustered distribution of the lenis tokens, contrasted with the highly variable and dispersed nature of the aspirated tokens, suggests a systematic structural difference rather than a simple frequency artifact. To further solidify this conclusion, it would be worthwhile to validate the current analysis in a separate study with more balanced frequency distributions. Since VOT is a primary acoustic cue that most distinctly separates fortis stops from lenis and aspirated ones, omitting this category in the current study might underestimate the role of VOT in L2 phonetic learning. The inclusion of the fortis stops in the future study would enable us to observe how the utilizing pattern of VOT shifts across different proficiency levels, or how VOT functions

within each specific proficiency group.

Another potential caveat to this interpretation is that the current acoustic analysis treated F0 as a segmental value not considering its dual role in suprasegmental intonation patterns (i.e., L-beginning AP tonal patterns in Seoul Korean unless it begins with aspirated or fortis obstruents, which yields H-beginning AP, Jun, 2000). The category split by heightening F0 for the aspirated target may also reflect the learner's realization of L2 Korean intonation. Given that the lenis stop is associated with the L-initial AP with no additional acoustic manipulation to be supra-segmentally correct, its acoustic properties may be viewed as closer to an undifferentiated category with less change over learning. The current study primarily centers on segmental features, without fully accounting for suprasegmental variations. Future research integrating both dimensions will undoubtedly yield a more complete understanding.

It is noted that the current view of lenis acoustics as a composite category in English-speaking learners' production is independent of learners' perception performance (Kim et al., 2006; Lee, 2024) as well as Korean natives' impressionistic judgements of those productions (Kim et al., 2006). Apart from the inconclusive debate regarding the perception-production link (e.g., Baese-Berk, 2019), there is no guarantee that learners' native-like perceptual representations will directly translate into actual production (Isbell, 2016 for a brief review). Schertz et al. (2015) has demonstrated that even when L1 Korean learners of L2 English establish robust perceptual categories of the stops, their articulatory control of acoustic cues such as VOT and F0 in actual speech production lagged. As such, we do not make any prediction about the learners' perceptual distinction of the lenis-aspirated stops based on this production data. Similarly, the production accuracy judged by native speakers reveals how native-like the learners' productions were, going beyond the mere differentiation of contrastive categories. Independent of these measures, the accurate prediction of the target categories in this study would indicate how effectively the current acoustic structures distinguish one category from the other and assess which category it fits better than the other. This measure from the multi-fold cross-validation task could be a useful tool to reveal which phonetic category is acoustically more robust based on L2 learners' initial production models.

5. Conclusion

This study investigates how English-speaking L2 Korean learners establish VOT and F0 acoustic structures for the lenis-aspirated contrast. Examining first- and second-year classroom learners, we found that category split emerges through the distinctive use of F0 -- a primary domain for Korean lenis-aspirated contrast but merely redundant for English voiced-voiceless contrast. Furthermore, this acoustic structure strengthens in second-year students and in lenis stops, yielding more accurate category predictions. This implies that lenis acoustics remain closer to the undifferentiated category during early L2 Korean learning. The current study contributes to the L2 production research by providing acoustic evidence of a category split of the L2 categories corresponding to a single category in L1.

Acknowledgement

This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (No. NRF-2024S1A5A2A01024518). An earlier version of this paper was presented orally at the 2025 Canadian Association of Teachers of Korean (CATK) Conference.

References

- Baese-Berk, M. M. (2019). Interactions between speech perception and production during learning of novel phonemic categories. *Attention, Perception, & Psychophysics*, 81(4), 981-1005.
- Bang, H. Y., Sonderegger, M., Kang, Y., Clayards, M., & Yoon, T. J. (2018). The emergence, progress, and impact of sound change in progress in Seoul Korean: Implications for mechanisms of tonogenesis. *Journal of Phonetics*, 66, 120-144.
- Best, C. T., & Tyler, M. (2007). Nonnative and second-language speech perception. In M. J. Munro & O. S. Bohn (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 13-34). Amsterdam, Netherlands: John Benjamins.
- Boersma, P., & Weenink, D. (2026). Praat: Doing phonetics by computer (version 6.4.58) [Computer program]. Retrieved from <https://www.praat.org/>
- Bürkner, P. C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, 80(1), 1-28.
- Flege, J. E., Aoyama, K., & Bohn, O. S. (2021). The Revised Speech Learning Model (SLM-r) applied. In R. Wayland (Ed.), *Second language speech learning: Theoretical and empirical progress* (pp. 84-118). CambridgeUK: Cambridge University Press.
- Gordon, P. C., Eberhardt, J. L., & Rueckl, J. G. (1993). Attentional modulation of the phonetic significance of acoustic cues. *Cognitive Psychology*, 25(1), 1-42.
- Isbell, D. (2016). The perception-production link in L2 phonology. *MSU Working Papers in Second Language Studies*, 7(1), 57-67.
- Jeon, M., Ko, K., Kim, D., Choi, Y., & Lee, A. (2020). *New generation Korean 1 & 2: Beginner level* (2nd ed.). Toronto, ON: University of Toronto Press.
- Jun, S. A. (2000). K-ToBI (Korean ToBI) labelling conventions (version 3.1, October 2000). *UCLA Working Papers in Phonetics*, 99, 149-173.
- Kim, K. H., Park, Y., & Chun, Y. (2006). The production and perception of the Korean stops by English learners. *Speech Science*, 13(4), 51-67.
- Kong, E. J., Kang, S., & Seo, M. (2022). The acoustic cue-weighting and the L2 production-perception link: A case of English-speaking adults' learning of Korean stops. *Phonetics and Speech Sciences*, 14(3), 1-9.
- Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2007). Phonetic learning as a pathway to language: New data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society B*, 363(1493), 979-1000.

- Labov, W. (1972). Some principles of linguistic methodology. *Language in Society*, 1(1), 97-120.
- Lee, E. (2024, July). English listeners' perception of Korean laryngeal contrasts among word-initial stops. *Proceedings of 14th Pronunciation in Second Language Learning and Teaching Conference* (pp. 1-12). West Lafayette, IN.
- Lee, H., Holliday, J. J., & Kong, E. J. (2020). Diachronic change and synchronic variation in the Korean stop laryngeal contrast. *Language and Linguistics Compass*, 14(7), e12374.
- Lee, H., Kong, E. J., & Holliday, J. J. (2024). The perception and production of Korean stops in second dialect acquisition. *JASA Express Letters*, 4(6), 065202.
- Llanos, F., Dmitrieva, O., Shultz, A., & Francis, A. L. (2013). Auditory enhancement and second language experience in Spanish and English weighting of secondary voicing cues. *The Journal of the Acoustical Society of America*, 134(3), 2213-2224.
- McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. (2017, August). Montreal forced aligner: Trainable text-speech alignment using kaldi. *Interspeech 2017* (pp. 498-502). Stockholm, Sweden.
- Nagle, C., Bruun, S., & Baese-Berk, M. (2023, April). Perceptual assimilation of Spanish, Thai, and Korean stop consonants by native English speakers. *20th International Congress of Phonetic Sciences*. Prague, Czech.
- Oh, M., & Kim, Y. J. (2011). The production of Korean stops and intonation by English speakers. *Korean Language Education as a Foreign Language*, 36, 139-158.
- Posit team. (2025). RStudio: Integrated development environment for R. Posit Software, PBC. Retrieved from <http://www.posit.co/>
- Schertz, J., Cho, T., Lotto, A., & Warner, N. (2015). Individual differences in phonetic cue use in production and perception of a non-native sound contrast. *Journal of Phonetics*, 52, 183-204.
- Schmidt, A. M. (2007). Cross-language consonant identification. In O. S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 185-200). Netherlands, Amsterdam: John Benjamins.

• **Eun Jong Kong**

Professor, School of Liberal Arts and Sciences, Korea Aerospace University
76 Hanggongdaechang-ro, Deogyang-gu, Goyang-si 10540, Korea
Tel: +82-2-300-0338
Email: ekong@kau.ac.kr
Areas of interest: Acoustic phonetics, Speech perception & production, Phonological development in L1 & L2

• **Soyoung Kang**, Corresponding author

Instructor, School of Linguistics and Language Studies, Carleton University
1125 Coloney By Dr. Ottawa, ON K1S 5B6, Canada

Tel: +1-613-724-1570

Email: kang139@gmail.com

Areas of interest: Psycholinguistics, Second language acquisition, Syntax, Phonetics

• **Misun Seo**

Professor, Dept. of English Language and Literature, Hannam University
70 Hannamro, Daedeok-gu, Daejeon 34430, Korea
Tel: +82-42-629-8461
Email: misunseo@hnu.kr
Fields of interest: L2 perception, L2 production, Phonology

Appendix

Table A. Mean VOT (ms) and F0 (Hz) by Year and Type

Type	Year	VOT (SD)	F0:Female (SD)	F0:Male (SD)
Aspirated	First year	56.4 (20.2)	223 (23.3)	117 (11.4)
	Second year	59.8 (28.8)	261 (40.5)	143 (27.8)
Lax	First year	37.2 (24.8)	225 (22.7)	125 (12.4)
	Second year	45.4 (30.7)	231 (38)	124 (24)

Standard deviations are shown in parentheses.
VOT, voice onset time.