Check for updates

The acoustic cue-weighting and the L2 production-perception link: A case of English-speaking adults' learning of Korean stops

Eun Jong Kong[1] · Soyoung Kang[2,*] · Misun Seo[3]

[1]Department of English, 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

The current study examined English-speaking adult learners' production and perception of L2 Korean stops (/t/ or /t'/ or /tʰ/) to investigate whether the two modalities are linked in utilizing voice onset time (VOT) and fundamental frequency (F0) for the L2 sound distinction and how the learners' L2 proficiency mediates the relationship. Twenty-two English-speaking learners of Korean living in Seoul participated in the word-reading task of producing stop-initial words and the identification task of labelling CV stimuli synthesized to vary VOT and F0. Using logistic mixed-effects regression models, we quantified group- and individual-level weights of the VOT and F0 cues in differentiating the tense-lax, lax-aspirated, and tense-aspirated stops in Korean. The results showed that the learners as a group relied on VOT more than F0 both in production and perception (except the tense-lax pair), reflecting the dominant role of VOT in their L1 stop distinction. Individual-level analyses further revealed that the learners' L2 proficiency was related to their use of F0 in L2 production and their use of VOT in L2 perception. With this effect of L2 proficiency controlled in the partial correlation tests, we found a significant correlation between production and perception in using VOT and F0 for the lax-aspirated stop contrast. However, the same correlation was absent for the other stop pairs. We discuss a contrast-specific role of acoustic cues to address the non-uniform patterns of the production-perception link in the L2 sound learning context.

Keywords: production-perception link, acoustic cue-weighting, L2 proficiency, voice onset time (VOT), fundamental frequency, stop laryngeal contrast, second language learning

1. Introduction

While theoretical frameworks of L2 speech learning (e.g., Speech Leaning Model: Flege, 1995; Flege, 2003; PAM-L2: Best & Tyler, 2007) postulate the perception-production link, empirical studies only showed null or weak evidence that perception and production are closely connected. On one hand, some studies supported the link, showing that L2 learners' gains in one modality were transferred to the other modality (Bradlow et al., 1997; Huensch & Tremblay, 2015; Kartushina et al., 2015; Nagle, 2018; Nagle, 2021;

Wang et al., 2003). Bradlow et al. (1997) found that Japanese speakers' identification training of English /l/-/r/ sounds facilitated their production accuracy. Reversely, feedback on French learners' Danish vowel production improved their ability to discriminate the vowels as well as production accuracy at the post-training test (Kartushina et al., 2015, but also see Baese-Berk, 2019). On the other hand, other studies reported contradictory findings that correlation was not found (Peperkamp & Bouchon, 2011) or there was only a moderate amount of interaction between the two modalities in L2 speech learning precisely conditioned by tasks, contrast pairs, levels of processing, or L2 experience (Hanulíková et al., 2012; Hattori & Iverson, 2010; Levy & Law, 2010; Melnik-Leroy et al., 2022). So far, it is inconclusive whether perception is closely linked to production in updating and processing L2 sounds.

Given the mixed picture of the perception-production relationship, the current study aims to add experimental evidence regarding whether the link between the two modalities exists in L2 users' utilizing multiple acoustic cues. We examine the case where English-speakers learn Korean as their second language. Since speech categories are defined along multiple acoustic dimensions, language users should learn to prioritize relevant dimensions for successful category distinctions in the target language. As the cue prioritization in L2 may not be identical to the order in L1, L2 learners have to re-adjust their cue weighting whereby they should enhance or ignore the use of specific cues (Francis & Nusbaum, 2002). Related to this L2 learning process, Schertz et al. (2015) investigated how Korean learners of English weight the voice onset time (VOT) and fundamental frequency (F0) cues in producing and identifying the English stops of /t/ and /d/. When L1 Koreans who use both VOT and F0 equally importantly for three-way stop contrast in L1 (tense /t'/ vs. lax /t/ vs. aspirated /tʰ/) learn a two-way contrast in L2 English stops (voiced /d/ vs. voiceless /t/, primarily differentiated by VOT alone: Shultz et al., 2012), cue weighting patterns in their L2 stop production and perception were not the same. The lack of connection came from the mismatch of the acoustic dimension they attend to in production and perception. In production, Korean learners consistently used VOT for the distinction, whereas in perception, various individual patterns were observed where some relied exclusively on VOT or F0 while others used both dimensions equally. Kong & Yoon (2013) further showed that L2 proficiency was not necessarily related to the same acoustic dimensions in production and perception. Korean learners' knowledge of L2 English was correlated with their better control of VOT in the /t/-/d/ production, whereas it was related to suppression of F0 in perception. That is, the two modalities did not necessarily manipulate the same acoustic dimension to reflect their knowledge of L2 sounds.

Given the results in Korean learners' cue weighting patterns discrepant in their perception and production of English stops, we are curious to see if the same incongruent relationship between the modalities would be observed for English learners of Korean. When English-speakers, who have a two-way stop contrast in their L1 are to learn L2 Korean stops of a three-way contrast, they need to utilize F0 as well as VOT for accurate production and perception of the stops. In the contemporary standard Korean, F0 is a primary acoustic dimension to differentiate the lax stop (lower F0) from the tense and aspirated stops (higher F0); without F0 differences, the lax stop VOT overlaps with the aspirated stop VOT because of sound change (Silva, 2006). For English-speaking L2 Korean learners,

whether and how much F0 is used is not just an option yielding individual variations but a crucial cue for *accurate* understanding of the L2 sounds. Considering this redefined role of F0 and VOT, we may expect that English-speaking learners of Korean may exhibit different patterns of link between production and perception along these acoustic dimensions, allowing for individual differences.

To explore the cue-weighting patterns of L1 English learners of L2 Korean in producing and perceiving the stops (i.e., /t'/ vs. /t/ vs. /tʰ/), we compare the perception data, a part of which was analyzed in Kong (2019), with the production results of the same participants. In Kong (2019), it was shown that L2 English learners of Korean relied more on VOT than on F0 in the lax-aspirated stops. But compared to relative cue-weights in their L1 voiced-voiceless perception, their use of F0 in L2 perception was greater. Related to proficiency, individuals with more L2 experience tended to rely on F0 more than others, becoming similar to native speakers of Korean. From this, it is possible to predict that the learners' increased use of F0 (and accordingly reduced role of VOT) for accurate Korean stop perception might be reflected in their production. The current study will further analyze the learners' relative use of VOT and F0 in the tense-lax, and tense-aspirated stop perception, two pairs not reported in Kong (2019), and examine the correlations of the acoustic cues between perception and production. Because the relative cue-weighting can change as proficiency improves (Flege et al., 1997; Kong & Yoon, 2013; Levy & Law, 2010), we consider proficiency as a control factor in assessing the link between the two modalities.

In summary, we aim to answer the following research questions. First, what acoustic cues would L2 English learners of Korean primarily manipulate in producing Korean 3-way stop contrasts? We will examine the cue-weighting patterns in production in comparison with those in perception reported in Kong (2019). Second, how would individual learners' acoustic patterns be related to proficiency in Korean? Finally, given the results in Kong (2019), is there a correlation between perception and production in terms of using acoustic cues in distinguishing Korean 3-way stop contrasts? The findings of the study will extend our understanding of cue-weighting strategies of L2 learners especially when they have to learn to use them for a distinction of more categories than ones in their L1 and of the relationship in production and perception in using multiple acoustic cues.

2. Methods

2.1. Participants and Tasks

Participants: 22 English-speaking learners of Korean (F=10, M=12 in their 20s through 40s) participated in the experiment consisting of production (wordlist reading task) and perception [3-alternative forced-choice (3AFC) identification task] sessions. The participants were living in Korea for their occupations or higher education at the time of recording. To assess their L2 proficiency, a small set of Korean language tests were administered, which included vocabulary, reading and listening comprehension questions adopted from previous Test of Proficiency in Korean (TOPIK). Correct answers scored 3 or 4 points as provided in the TOPIK answer keys, and the total scores were used to represent L2 proficiency in this study due to the participants' widely varying length of stay in Korea (18 months to 13 years) and different amount of formal Korean language instruction. Perfect scores of 178 points

consisted of 98 points of vocabulary, 14 points of reading, and 66 points of listening comprehension components. The total scores for the participants in the study ranged from 44 points to 175 points with 125 point mean scores.

Wordlist reading task: The participants read words presented in hangul and were recorded through a microphone (Shure SM81) in a quiet room. The recording was made using a digital recorder (Marantz PMD661) with 16-bit quantization and 44,000 Hz sampling rate. The wordlist consisted of two or three syllable obstruent-initial 127 words followed by seven different vowels /a ɛ i o u ʌ ɨ/. In this study, 64 stop-initial words are the targets (see Appendix).

3AFC Identification task: The L2 learners listened to CV syllables and identified each syllable as /t'a/ '따', /ta/ '다' or /tʰa/ '타' by mouse-clicking in the task. The stimulus syllables were synthesized from a Korean male speaker's natural productions of /ta/ by combining 6-step VOT (9, 13, 19, 28, 40, and 59 ms) and 5-step F0s (98, 106, 114, 122, and 130 Hz) [see stimuli preparation details in Kong (2019)]. A set of 30 different CV stimulus were presented three times using an E-prime 2 software. The production and perception tasks were given to the participants in a counter-balanced order.

2.2. Analysis

Acoustic analysis: Target words produced by the 22 learners of Korean were acoustically examined by measuring VOT, F0 (fundamental frequency at the following vowel onset), and H1-H2 (energy difference between the first two harmonics). For measurements, two acoustic events were marked manually by the first author based on waveform and spectrogram displays: (1) Burst (an energy spike before a vowel), (2) Voicing Onset (beginning of a voicing bar). Referring to these events, VOT was calculated by subtracting timestamp of Burst from that of Voicing Onset. F0 was the fundamental frequency at the 10 ms past the Voicing Onset, and H1-H2 was obtained by subtracting the amplitude of the second harmonic (H2) from that of the first one (H1) at the spectrum of 25 ms analysis window taken at Voicing Onset. The process of acoustic measurements was done using Praat (Boersma & Weenink, 2022).

Statistical consideration: The mixed-effects regression models were made to estimate the effect of acoustic variables (VOT, F0 and H1-H2: fixed-effect variables) in predicting the lax-tense stops,[1] the lax-aspirated stops, and the tense-aspirated stops in production and perception tasks (*lme4* package in R, Bates et al., 2015). The intercept and slope of acoustic variables varied at the subject level. To quantify individuals' weights on each acoustic variable, we summed the fixed-effect coefficients (group-averaged weights) and by-subject random slope coefficients (individuals' deviations from the group-average).

The relationship between individuals' weights of acoustic variables and L2 Korean proficiency, and the relationship between each acoustic variable between production and perception were assessed by conducting a series of partial correlation tests (*ppcor* package in R, Kim, 2015). This test is useful because the correlation of the two test variables is assessed with the effect of control variables removed: e.g., ppcor (variable *x*, variable *y*, a set of control variable *z*). For example, when L2 proficiency is a control variable in testing correlation between production VOT and perception VOT, this test partials out the amount of correlation that L2 proficiency might have with VOT coefficients either in production or perception. This way, the partial correlation coefficients can represent the relationship solely between the two test variables independent of the control variable.

3. Results

3.1. Descriptive Patterns: Acoustic Variables (VOT and F0) and Proficiency

Figure 1 presents strip-charts of individual learners' acoustic realizations of /t'/, /t/, and /tʰ/ in terms of VOT, F0, and H1-H2. Ordered by L2 proficiency, each acoustic dimension exhibited distribution patterns differentiating higher proficiency learners from lower ones. In the VOT panels (top), higher proficiency learners had longer lax stop VOT than those of the tense stops (leftmost) and had a greater VOT overlap between the lax and aspirated stops (center). In contrast, lower proficiency learners at the bottom of each panel had their lax stops whose VOT values were as short as tense stop VOT and shorter than aspirated stop VOT. The H1-H2 distribution as a function of proficiency (bottom) patterned similar to the VOT distribution. In the Figure 1-middle panels, the high proficiency learners' lax stop F0s were lower than the tense and aspirated stop F0s, whereas the low proficiency learners' lax stop F0s overlapped with the tense and aspirated stops. The acoustic realizations of VOT, F0, and H1-H2 were least relevant to proficiency in differentiating the tense from the aspirated stops (rightmost panels): the aspirated stops had longer VOT, higher F0s, and greater H1-H2 than the tense stops regardless of proficiency.

In Figure 2, we selected six participants of low, intermediate, and high level L2 proficiency to show bivariate distributions of F0 and VOT in each type of the stops (top panels). Regardless of proficiency level differences, all six learners' aspirated stops were well separated from the other types by having relatively longer VOT. It was the lax type that the learners realized differently across the L2 proficiency. The high proficiency learners' lax stops (ID210 and ID216, right panels) had longer VOT than the tense stops and lower F0s than the aspirated stops, differentiating the three types from one after another in the F0 by VOT dimension. Lower and intermediate level learners had almost a complete overlap of the tense and lax types in production, where both had shorter VOT than the aspirated stops.

These production patterns were similar to the perception patterns of F0 and VOT (Figure 2-bottom three panels). The same all six listeners identified tokens with longer VOT as the aspirated type (bottom panels), and the two high proficiency learners perceived the tokens with both longer VOT and lower F0 as the lax type.

---

[1] Lax-tense model: glmer(/t/ or /t'/~VOT+F0+(VOT+F0|learners), binomial, data).
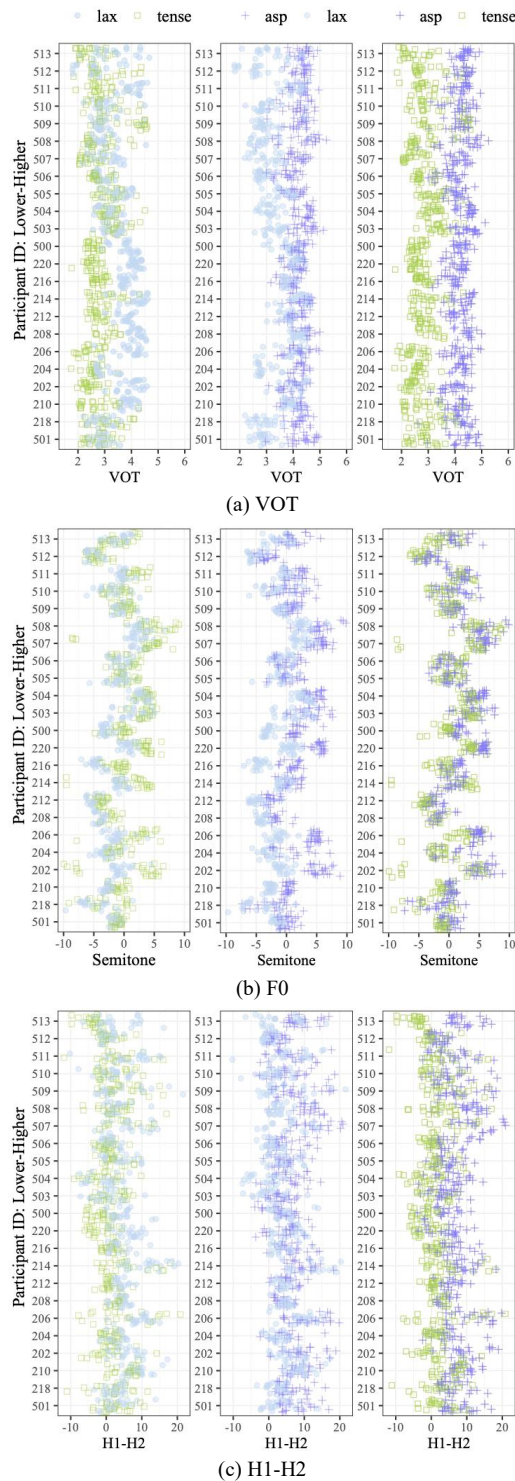
(a) VOT

(b) F0

(c) H1-H2

Figure 1. Distributions of (a) VOT, (b) F0 in semitone and (c) H1-H2 values separated by the contrast type. The speakers are ordered by their L2 proficiency.
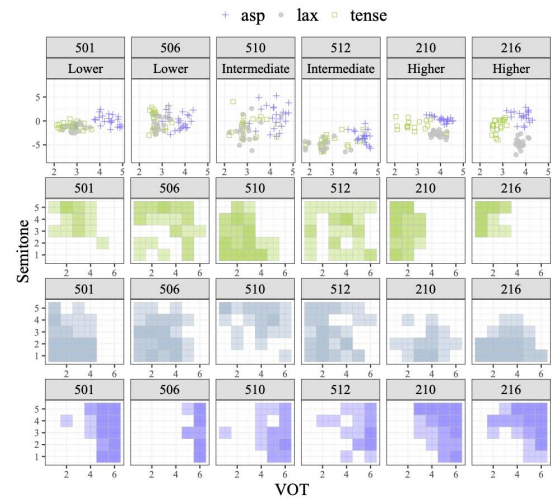

Figure 2. [Top panels] Scatter plots of F0 against VOT from six low, intermediate, and high proficiency learners. [Bottom three panels] Heat maps of the perception responses by the same six learners (reanalyzed based on data from Kong, 2019): Tense (green), and lax (grey) and aspirated (purple) responses, respectively.

3.2. Relationship between Acoustic Cue Weighting and Proficiency

Table 1 presents a summary of mixed-effects regression model results: tense-lax, lax-aspirated, and tense-aspirated contrast models. Across the three models, the coefficients of VOT, F0, and H1-H2 variables were significant (except H1-H2 for the lax-aspirated contrast model), suggesting that at the group level, each acoustic variable was useful in distinguishing one type from the other in the L2 learners' stop production. The cue-weighting patterns of greater VOT coefficients than F0 coefficients indicate that on average, English-speaking L2 learners differentiated the Korean stops primarily by VOT and secondarily by F0 in production.

These patterns in production differed from the perception results in that the primary role of VOT over F0 was not always observed across the contrast pairs. The tense-lax model of the perception revealed a reversed order of cue primacy between F0 and VOT where the absolute value of F0 coefficient was greater than the VOT coefficient (Table 1b). In summary, at the group level, the cue-weighing patterns were not necessarily identical between production and perception across all the contrast pairs.

Table 1. Summary of the mixed effects regression models: (a) production (b) perception

(a)	Tense-Lax		Lax-Asp.		Tense-Asp.	
	Estimate	p-value	Estimate	p-value	Estimate	p-value
VOT	4.48 (1.08)	<.0001	6.75 (1.01)	<.0001	9.17 (1.26)	<.0001
F0	−0.99 (.21)	<.0001	4.43 (.81)	<.0001	0.72 (.26)	.005
H1-H2	0.61 (.19)	.001	0.52 (.40)	.19	0.60 (.24)	.011
(b)	Tense-Lax		Lax-Asp.		Tense-Asp.	
	Estimate	p-value	Estimate	p-value	Estimate	p-value
VOT	0.539 (.26)	.0452	3.189 (.49)	<.001	4.292 (.58)	<.001
F0	−1.233 (.25)	<.001	1.069 (.20)	<.001	0.099 (.19)	.61

In Figure 3, the individual learners' F0 coefficients were plotted as a function of their VOT coefficients to confirm the group patterns of cue primacy. In production and perception, all or most datapoints were below the diagonal line, indicating greater VOT coefficients than F0 values. This conforms to the group averaged pattern of

cue-weighting, utilizing VOT more than F0 in production and perception (except the perception tense-lax model). Deviated from this general trend, however, the two contrast pairs (i.e., the lax-aspirated pair in production and tense-lax pair in perception) had datapoints almost evenly distributed above or below the line. While group averaged coefficients were estimated to be greater in VOT than F0 (the lax-asp. production) and greater in F0 than VOT (the tense-lax perception), not all individuals had cue-weighting patterns represented as a group tendency.
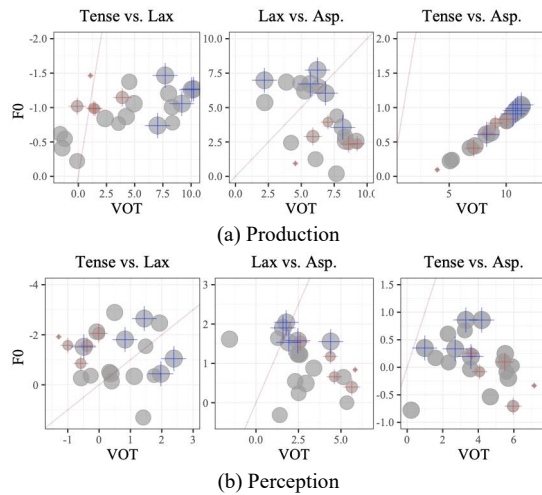

Figure 3. Scatterplots of individual listeners' F0 coefficients against VOT coefficients: (a) production, (b) perception. Blue cross indicates learners with highest top six test scores (higher proficiency learners), and red cross indicates those with bottom six scores (lower proficiency learners).

To find out the relationship with the learner's L2 proficiency, individual coefficients of each acoustic variable were considered in the partial correlation tests. As in Table 2, the learners' L2 Korean proficiency was significantly correlated with their VOT coefficients from the perception models of the three contrast pairs and was correlated with their F0 coefficients from the production models (except the tense-asp. contrast pair). In the lax-aspirated pair models, higher proficiency was associated with less use of VOT in perception and with more use of F0 in production. Contrastively, in the tense-lax pair models, higher proficiency was associated with more use of VOT in perception and with less use of F0 in production (note that F0 coefficients for the tense-lax model were negative values). These association patterns suggest that the English-speaking learners of Korean manipulate multiple dimensions to effectively highlight a single cue, a primary one, and to suppress less important cues. Given the understanding, it is the F0 dimension that the learners targeted to highlight for the lax-aspirated contrast, and the VOT dimension for the tense-lax contrast.

Table 2. Summary of the partial correlation tests where L2 proficiency was correlated with VOT or F0 coefficients. Bold indicates correlation coefficients greater than 0.4

		Tense-Lax		Lax-Asp.		Tense-Asp.	
		Estimate	*p*-value	Estimate	*p*-value	Estimate	*p*-value
L2 proficiency (=x) correlated with [y] with [z] controlled							
(a) Production							
y	z						
VOT	F0 & H1-H2	0.212	0.368	−0.12	0.613	0.075	0.745
F0	VOT& H1-H2	**0.425**	**0.061**	**0.502**	**0.023**	−0.039	0.865
H1H2	VOT & F0	0.15	0.526	0.14	0.538	−0.039	0.865
(b) Perception							
y	z						
VOT	F0	**0.625**	**0.002**	**−0.485**	**0.025**	**−0.42**	**0.057**
F0	VOT	−0.052	0.819	0.207	0.365	0.31	0.17

3.3. Relationship between Production and Perception

The VOT and F0 coefficients were further used as test variables in the partial correlation tests to examine the relationship between production and perception. Since L2 proficiency was correlated with acoustic coefficients as reported in Section 3.2, we set the L2 proficiency as one of control variables in the partial correlation tests. As in Table 3, for the lax-aspirated contrast pair, both VOT and F0 coefficients from production were significantly correlated with those from perception. Specifically, the learners who use VOT and F0 more in production (greater coefficients) relied on VOT and F0 more in perception. For the other contrast pairs, correlation coefficients were not statistically meaningful, suggesting a lack of consistent association between production and perception.

Since the correlation test is limited in assessing linear relationship between the variables, we supplemented the analysis by visually inspecting the distributions of perception coefficients against production coefficients as in Figure 4. Top six proficiency learners (blue) and bottom six learners (red) were indicated with cross symbols. On the whole, the coefficients were distributed at any quadrant domains. Although it appeared more general to have most datapoints located at the 1st and 3rd quadrants, there *were* some datapoints at the 2nd and 4th quadrants, suggesting that the learners could be good at perception or production but poor at the other modality.

Table 3. Summary of the partial correlation tests where VOT and F0 coefficients of production were correlated with those of perception. Bold indicates correlation coefficients greater than 0.4

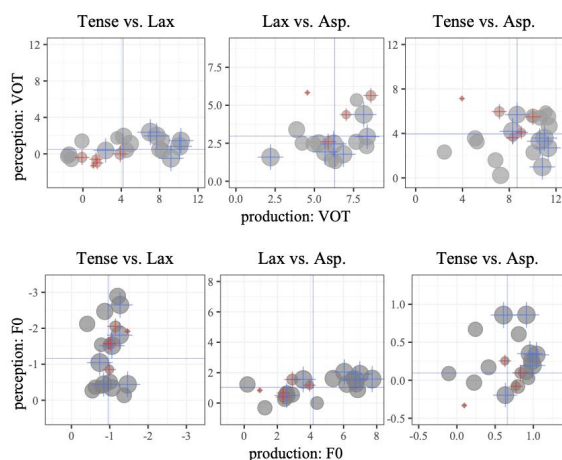|		Tense-Lax		Lax-Asp.		Tense-Asp.	
		Estimate	*p*-value	Estimate	*p*-value	Estimate	*p*-value
Correlation between [x] and [y] with [z] controlled:							
[z] includes L2 proficiency & the coefficients not used in [x] and [y]							
prod.	perc.						
x	y						
VOT	VOT	0.305	0.218	**0.682**	**0.001**	0.199	0.411
VOT	F0	0.204	0.415	−0.397	0.102	−0.315	0.187
F0	F0	0.237	0.343	**0.585**	**0.010**	0.323	0.176
F0	VOT	0.308	0.212	−0.263	0.289	−0.169	0.487

Figure 4. Scatterplots of individual listeners' VOT; top and F0; bottom coefficients in perception against those in production. Blue cross indicates learners with highest top six test scores (higher proficiency learners), and red cross indicates those with bottom six scores (lower proficiency learners).

4. Discussion and Conclusion

The current study compared the cue-weighting patterns of VOT and F0 in English-speaking learners' production and perception of L2 Korean stops. Our goal was to investigate whether there is consistent relationship between the two modalities when the learners deal with multiple acoustic dimensions for L2 sound contrasts. Results showed that the English-speaking learners were able to manipulate both VOT (a primary cue in L1) and F0 (a new primary cue in L2) in producing and perceiving the Korean stops. To understand the cue-weighting patterns, examining both group and individual trends was important. As a group, the production results showed that for English learners of Korean, VOT, F0 and H1-H2 were all useful cues to distinguish Korean stops. In terms of a relative cue weighting pattern, the results also showed that VOT was the primary cue and F0 was the secondary cue in all three stop contrast pairs, demonstrating a clear L1 effect. These results are slightly different from the perception results where the listeners relied more on F0 than VOT for at least one contrast pair, the tense-lax stop pair. Individual patterns further showed that their L2 proficiency was consistently associated with F0 in production, whereas it was associated with VOT in perception. When these individual variabilities and their L2 proficiency carefully controlled, partial correlation analyses revealed that production and perception were connected in both VOT and F0 dimensions when the English-speaking learners deal with the lax-aspirated stops in Korean. The findings provide a partial support for the production- perception link since evidence was confined to a single contrast pair.

In investigating the link between production and perception, it was important to consider individual learners' L2 proficiency. Individual variations associated with L2 proficiency were evident; descriptively in production, speakers with a lower proficiency tended to have an overlapping VOT value between the lax and tense stops while the lax and aspirated stops were clearly distinguished from each other on the same acoustic dimension. Reversely, the speakers with a higher proficiency seemed to be able to separate the lax from the tense stops while the lax and aspirated stop VOT values

tended to merge. Given that the VOT values of the lax and aspirated stops in Seoul Korean are often overlapped (Silva, 2006; Lee & Jongman, 2012), the merging VOT values for the two stops by English speakers demonstrated a learning effect in their L2 phonology. In terms of F0 dimension, the proficiency difference was also found. For the two pairs involving the lax stop (i.e., tense-lax, and lax-aspirated stop pairs), which is realized as a low F0 by native Korean speakers, speakers with a higher proficiency tended to separate the lax from the tense and aspirated stops by a lower F0, another illustration of their phonological development in L2 Korean. A similar pattern of proficiency-related differences was found in individuals' perception results as well. Statistically, the learners' proficiency was correlated with VOT in perception (all three stop contrast pairs) but with F0 in production (tense-lax, and lax-aspirated pairs). For the lax-aspirated pair, high proficiency learners made less use of VOT in perception but more use of F0 in production. For the tense-lax pair, high proficiency learners relied more on VOT in perception but less on F0 in production. The findings are congruent with those in Kong & Yoon (2013) where Korean learners' L2 English proficiency was correlated with multiple acoustic dimensions in L2 production and perception. The pattern of association between acoustic dimension and modality was language-specific, however, showing that any acoustic cues are flexible enough to efficiently control the target language. Given that F0 is a new primary cue in L2 for the L1 English learners, it appears natural that experienced learners are better at utilizing F0 in L2 stop production. Confusingly enough, the same pattern of F0 association with L2 proficiency was not observed in perception. We may speculate this inconsistency comes from different nature of the two modalities, that is, perception is more flexible than production (Pinget et al., 2020). Pinget et al. (2020) argued that perception precedes production in the course of sound change, demonstrating that language users' perception adapted to innovative form at the initial stage of sound change and then flexibly adjust to conservative forms at the later stage as their changes in production are in progress. The same mechanism may be working in learning new L2 sounds. To test the speculation, we need more deliberate experimental study in the future.

Most importantly, there was evidence that production and perception was correlated when these individual differences related to the L2 proficiency were taken into consideration in the statistical assessment. For the lax-aspirated stop contrast, the English-speaking learners that use VOT more in production also used VOT more in perception and the same was true for F0 use. This is an important finding supporting that production and perception are linked at multiple acoustic dimensions for the L2 sound contrast. What is puzzling in the results, however, is that this link was not consistently observed in the other two stop pairs, i.e., the tense-lax and tense-aspirated stops. We speculate that for these two contrast pairs, the absence of correlation might be attributable to a differing degree of contribution of the VOT and F0 cues. For one, accurate differentiations of the tense-lax stops can be sufficiently made along a single acoustic dimension: either VOT alone or F0 alone is *equally* useful. As such, different individuals may opt for a different strategy in producing and perceiving the stops. Some learners may rely more on VOT in production while they may rely more on F0 in perception. Conversely, others may rely more on F0 in production, while they may rely more on VOT in perception. Finally, still others may rely equally on both VOT and F0 in production and perception. For the

tense-lax pair, different approaches to using VOT and F0 in individuals' L2 perception and production might yield a loose connection between the two modalities in terms of acoustic cue uses. This case is comparable to the production and perception of Korean learners' English stops. Schertz et al. (2015) also pointed out L1 Korean learners' variable uses of F0 for the voiced-voiceless stop perception in L2 English, while the learners used VOT primarily for the stop productions, resulting in the absence of the production-perception link. On the other hand, for the tense-aspirated pair, the *redundant* role of F0 may also account for the absence of the perception-production link. While individual learners' reliance on VOT was dominant for the tense-aspirated stops, they could also use F0 or not since F0 do not affect accurate production or perception of the pair. This redundant role of the F0 cue for the tense-aspirated stop contrast may result in a loose connection between production and perception. On a more general note, such an absence of correlation between production and perception may also be due to a difference in tasks used in production and perception, or asynchronous nature of production and perception learning across individuals. For the former, Melnik-Leroy et al. (2022) suggested that the link between production and perception may be found within but not across processing levels. In their study on French vowels, /u/ and /y/, the production task used 3 syllable pseudo-words (pre-lexical) and naming real words (lexical) and the accuracy of the target vowels was measured. The perception task was conducted with an ABX discrimination task using the same 3 syllable pseudo-words. The link between production and perception was found only for the same processing level, that is, pseudo-word reading (pre-lexical production) and ABX task (pre-lexical perception). High vowel production scores predicted higher perception accuracy, but the same relation was not found between naming of real words (lexical) and ABX task (pre-lexical). The production study here used real words (lexical level), whereas the perception study used non-words (pre-lexical level), which may account for the absence of such a link. For the latter, Nagle (2018) showed that L1 English speakers learning L2 Spanish could have individual differences in temporal dynamics of production and perception of the Spanish stops. Some learners showed perceptual sensitivity to L2 categories immediately accompanied by production accuracy, while others showed different degrees of time-lag between perception and production. For the future direction, conducting both production and perception tasks within the same processing level and recruiting a large number of learners to define learning trajectories should be considered.

To conclude, the current findings demonstrated that some English learners of Korean produced and perceived the L2 stop sounds in a connected way in using multiple acoustic dimensions. As previously reported (Hattori & Iverson, 2010; Levy & Law, 2010; Melnik-Leroy et al., 2022), however, the L2 learners' production-perception link was observed in only a limited context: the correlation was evident for the lax-aspirated stop pair (but not for the tense-lax or the tense-aspirated pairs) when individual variability due to their L2 proficiency was carefully controlled. Individual learners' freedom to employ a redundant cue might make it hard to find a consistent link between production and perception.

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.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1-48.

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O. S. Bohn, & M. J. Munro (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. (2022). Praat: Doing phonetics by computer (version 6.2.14) [Computer program]. Retrieved from http://www.praat.org/

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *The Journal of the Acoustical Society of America, 101*(4), 2299-2310.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233-277). Timonium, MD: York Press.

Flege, J. E. (2003). Assessing constraints on second-language segmental production and perception. In N. O. Schiller, & A. S. Meyer (Eds.), *Phonetics and phonology in language comprehension and production: Differences and similarities* (pp. 319-355). Berlin, Germany: De Gruyter Mouton.

Flege, J. E., Bohn, O. S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics, 25*(4), 437-470.

Francis, A. L., & Nusbaum, H. C. (2002). Selective attention and the acquisition of new phonetic categories. *Journal of Experimental Psychology: Human Perception and Performance, 28*(2), 349-366.

Hanulíková, A., Dediu, D., Fang, Z., Bašnaková, J., & Huettig, F. (2012). Individual differences in the acquisition of a complex L2 phonology: A training study. *Language Learning, 62*(s2), 79-109.

Hattori, K., & Iverson, P. (2010, September). Examination of the relationship between L2 perception and production: An investigation of English /r/-/l/ perception and production by adult Japanese speakers. *Proceedings of the Second Language Studies: Acquisition, Learning, Education and Technology*. Tokyo, Japan.

Huensch, A., & Tremblay, A. (2015). Effects of perceptual phonetic training on the perception and production of second language syllable structure. *Journal of Phonetics, 52*, 105-120.

Kartushina, N., Hervais-Adelman, A., Frauenfelder, U. H., & Golestani, N. (2015). The effect of phonetic production training with visual feedback on the perception and production of foreign speech sounds. *The Journal of the Acoustical Society of America, 138*(2), 817-832.

Kim, S. (2015). ppcor: an R package for a fast calculation to semi-partial correlation coefficients. *Communications for Statistical Applications and Methods, 22*(6), 665-674.

Kong, E. J. (2019). Individual differences in categorical perception: L1 English learners' L2 perception of Korean stops. *Phonetics and Speech Sciences, 11*(4), 63-70.

Kong, E. J., & Yoon, I. H. (2013). L2 proficiency effect on the acoustic cue-weighting pattern by Korean L2 learners of English: Production and perception of English stops. *Phonetics and Speech Sciences, 5*(4), 81-90.

Lee, H., & Jongman, A. (2012). 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 International Phonetic Association*, *42*(2), 145-169.

Levy, E. S., & Law, F. F. II. (2010). Production of French vowels by American-English learners of French: Language experience, consonantal context, and the perception-production relationship. *The Journal of the Acoustical Society of America*, *128*(3), 1290-1305.

Melnik-Leroy, G. A., Turnbull, R., & Peperkamp, S. (2022). On the relationship between perception and production of L2 sounds: Evidence from Anglophones' processing of the French /u/-/y/ contrast. *Second Language Research*, *38*(3), 581-605.

Nagle, C. L. (2018). Examining the temporal structure of the perception-production link in second language acquisition: A longitudinal study. *Language Learning*, *68*(1), 234-270.

Nagle, C. L. (2021). Revisiting perception-production relationships: Exploring a new approach to investigate perception as a time-varying predictor. *Language Learning*, *71*(1), 243-279.

Peperkamp, S., & Bouchon, C. (2011, August). The relation between perception and production in L2 phonological processing. *Proceedings of the Twelfth Annual Conference of the International Speech Communication Association*. Florence, Italy.

Pinget, A. F., Kager, R., & Van de Velde, H. (2020). Linking variation in perception and production in sound change: Evidence from Dutch obstruent devoicing. *Language and Speech*, *63*(3), 660-685.

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.

Shultz, A. A., Francis, A. L., & Llanos, F. (2012). Differential cue weighting in perception and production of consonant voicing. *The Journal of the Acoustical Society of America*, *132*(2), EL95-EL101.

Silva, D. J. (2006). Acoustic evidence for the emergence of tonal contrast in contemporary Korean. *Phonology*, *23*(2), 287-308.

Wang, Y., Jongman, A., & Sereno, J. A. (2003). Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *The Journal of the Acoustical Society of America*, *113*(2), 1033-1043.

• **Eun Jong Kong**
Professor, Dept. of English, Korea Aerospace University
76 Hanggongdaehang-ro, Deogyang-gu, Goyang 10540, Korea
Tel: +82-2-300-0338
Email: ekong@kau.ac.kr
Fields of interest: Acoustic phonetics, Speech production & perception, First & second language acquisition

• **Soyoung Kang,** Corresponding author
Instructor, School of Linguistics and Language Studies
Carleton University
1125 Colonel By Dr, Ottawa, ON K1S 5B6, Canada
Tel: +1-613-520-2600
Email: kang139@gmail.com
Fields 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 1

Hangul	IPA	Hangul	IPA
Tense-initial words			
떠돌이 'a wonderer'	/t'ʌtoli/	띵동 'dingdong'	/t'iŋtoŋ/
떡집 'rice-cake shop'	/t'ʌktɕip/	뜨네기 'wondering one'	/t'inɛki/
떨이 'frozen stock'	/t'ʌli/	뜨다 'to float'	/t'ɨta/
따르릉 'ting-a-ling'	/t'aliliŋ/	뜰채 'a net'	/t'ɨltɕʰɛ/
딸기 'strawberry'	/t'alki/	또랑 'ditch'	/t'olaŋ/
땅콩 'peanut'	/t'aŋkʰoŋ/	똑똑 'knock knock'	/t'okt'ok/
때문에 'because'	/t'ɛmunɛ/	똘똘 'smart'	/t'olt'ol/
때리다 'to hit'	/t'ɛlita/	뚜껑 'a lid'	/t'uk'ʌŋ/
땡땡이 'playing hooky'	/t'ɛŋt'ɛŋi/	뚝배기 'earthen pot'	/t'ukpɛki/
띄엄띄엄 'sparsely'	/t'ɨʌmt'ɨʌm/	뚝뚝 'dripping sound'	/t'uk'tuk/
띠띠빵빵 'beep beep'	/t'it'ip'aŋp'aŋ/		
Lax-initial words			
더미 'a heap'	/tʌmi/	딩동댕 'ding-dong-dang'	/tiŋtoŋtɛŋ/
더덕 'Deodeok'	/tʌtʌk/	딩굴딩굴 'in the manner of rolling about'	/tiŋkultiŋkul/
덥석 'suddenly'	/tʌpsʌk/	드림 'offering'	/titim/
달팽이 'a snail'	/talpɛŋi/	든든 'safe'	/tɨntɨn/
당근 'a carrot'	/taŋkɨn/	도마 'cutting board'	/toma/
단추 'a button'	/tantɕʰu/	도라지 'balloon flower root'	/toratɕi/
대리 'substitute'	/tɛli/	동생 'younger sibling'	/toŋsɛŋ/
대신 'instead'	/tɛsin/	두부 'tofu'	/tupu/
댕기 'ribbon'	/tɛŋki/	두통 'headache'	/tutʰoŋ/
디스코 'disco'	/tiskʰo/	둥글게 'round'	/tuŋkilkɛ/
득녀 'having a daughter'	/tiknjʌ/		
Aspirated-initial words			
터미널 'a terminal'	/tʰʌminʌl/	팀장 'team leader'	/tʰimtɕaŋ/
텁텁 'unpleasant taste in one's mouth'	/tʰʌptʰʌp/	통로 'a path'	/tʰoŋro/
털이 'a robber'	/tʰʌli/	트집 'finding fault'	/tʰɨtɕip/
타조 'a turkey'	/tʰatɕo/	특집 'special issue'	/tʰɨktɕip/
타이어 'a tire'	/tʰaiʌ/	튼튼 'healthy'	/tʰɨntʰɨn/
탁자 'a table'	/tʰaktɕa/	토마토 'tomato'	/tʰomatʰo/
태극기 'national flag'	/tʰɛkikki/	토실 'chubby'	/tʰosil/
택배 'delivery service'	/tʰɛkpɛ/	투덜투덜 'grumblingly'	/tʰutʌltʰutʌl/
택시 'taxi'	/tʰɛksi/	투수 'a pitcher'	/tʰusu/
티켓 'a ticket'	/tʰikʰɛt/	툭툭 'hitting sound'	/tʰuktʰuk/
티셔츠 'a T-shirt'	/tʰiʃʌtɕʰɨ/	퉁퉁 'stamping sound'	/tʰuŋtʰuŋ/

IPA, International Phonetic Alphabet.