Linguistic Research 38(Special Edition): 77-100
DOI: 10.17250/khisli.38..202109.004

Perception-production relation in the cue use patterns of Korean stops by native Chinese speakers*

Eunjin Oh
(Ewha Womans University)

Oh, Eunjin. 2021. Perception-production relation in the cue use patterns of Korean stops by native Chinese speakers. *Linguistic Research* 38(Special Edition): 77-100. The purpose of this study was to investigate whether the perception and production patterns of non-native speakers are correlated in the context of using acoustic cues to differentiate between Korean lenis and aspirated stops. A perception test was conducted identifying tokens in which VOT and F0 values were systematically manipulated between /kan/ and /k^{h}an/. A production test was conducted in which words starting with a lenis or aspirated stop were read in a carrier sentence. Relative cue weightings were estimated via perceptual and production coefficients that were obtained from binary logistic regression analyses and discriminant function analyses of individual learner data (e.g., Shultz et al. 2012; Schertz et al. 2015). It was found that the correlation between perceptual and production coefficients was not significant for either VOT or F0. In general, individual learners used F0, which is the primary cue for the Korean stops, in perception, but did not use it in production, indicating a tendency for L2 perception to precede production. **(Ewha Womans University)**

Keywords perception, production, Korean, stops, VOT, F0, Chinese

1. Introduction

1.1 Relation between L2 speech perception and production

The relationship between speech perception (the process by which acoustic cues signaling phonetic contrast are interpreted) and speech production (the process by which the acoustic cues are realized into speech) has been a major interest in phonetics research. Several studies have reported that there is a close connection between the two processes. These studies have shown that the perception of a sound contrast shows a positive correlation to the production of the corresponding contrast. According to Perkell et al.

* I am grateful to anonymous reviewers for their useful comments and suggestions, and to Xing Fan for helping recruit experiment participants.

(2004), native speakers of American English who accurately perceived the vowel contrasts between /ɑ/ and /ʌ/ (as in “cod” - “cud”) and between /u/ and /ʊ/ (as in “who’d” - “hood”) produced the contrasts more distinctively, thus suggesting a close relationship between vowel perception and production. In contrast, other studies have reported no significant relationship between speech perception and production. Shultz et al. (2012) examined the patterns of native speakers of English using two voicing cues of English stops in perception and production. The study examined the voice onset time (VOT) as the primary cue and the fundamental frequency (F0) in the following segment as the secondary cue to differentiate voiced and voiceless stops in English. The correlation between the perceptual and production cue weights of individual participants was not significant for either VOT or F0.

The relationship between perception and production has also been an important research interest in second language (L2) speech, where there have been mixed results. Some studies have reported that the two competencies have a close relationship and exhibit similar stages in their development (e.g., Bohn and Flege 1992; Flege et al. 1997). Flege et al. (1997) reported that the perception and production of American English vowels /i/ - /ɪ/ and /ɛ/ - /æ/ by L2 learners were significantly correlated. That is, learners who accurately perceived those vowel contrasts tended to produce them well. Moreover, experienced L2 learners both perceived and produced the L2 vowel contrasts more accurately than inexperienced learners did. Further supporting this relationship are studies that have shown that L2 production was improved only by improvement in L2 perception. According to Bradlow et al. (1997), the English /l/ - /r/ production by native Japanese speakers could be improved only through perceptual identification training without production training, indicating a close connection between the two modalities. On the other hand, other studies have reported mismatched patterns between L2 speech perception and production. Bohn and Flege (1997) investigated the patterns of spectral and durational cue use of English vowels /ɛ/ - /æ/ by native German speakers. The use patterns of the two cues did not match the learners’ perception and production of the vowels, which differed according to their L2 experience. Inexperienced learners were able to perceive the English /ɛ/ - /æ/ contrast (using the durational cue) but could not produce the vowel contrast. Experienced learners were able to produce the corresponding vowel contrast but did not use the spectral cue in perception in a native-like way.

It has often been hypothesized that L2 perception precedes L2 production. In other words, it is generally believed that an L2 sound contrast must first be perceived before

the corresponding contrast can be produced (e.g., Flege 1993; Levy and Law 2010). Levy and Law (2010) reported that, for the French vowel contrast /y/ - /u/, English learners had a perception that was more native-like than their production, suggesting that perceptual learning appears to precede production learning. Laméris and Graham (2020) investigated the perception and production of an L2 suprasegmental structure. Native English advanced learners of Japanese perceived Japanese lexical pitch in a native-like way, but their production was significantly different from that of native Japanese speakers. On the other hand, although rarer, a few studies have provided evidence that the development of L2 production seems to precede L2 perception (e.g., Flege and Eefting 1987; Sheldon and Strange 1982). Sheldon and Strange (1982) examined the perception and production of the English /l/ - /r/ contrast by native Japanese speakers who were residing in the United States. Some participants produced the English /l/ - /r/ contrast more accurately than they perceived it, indicating that the L2 production development of the English /l/ - /r/ distinction preceded its perceptual development.

1.2 Production of Korean stops by native Chinese speakers

The stop system of Seoul Korean consists of three phonation types: fortis (/p'/, /t'/, /k'/), lenis (/p/, /t/, /k/), and aspirated stops (/p^h/, /t^h/, /k^h/). The main acoustic cues that distinguish the three stop categories are VOT and F0 of the following vowel. VOT refers to the time from the release burst of a stop to the beginning of the vocal cord vibration for the following segment. Fortis stops are cued with short-lag VOTs and higher F0s, lenis stops are cued with long-lag VOTs and low F0s, and aspirated stops are cued with long-lag VOTs and the highest F0s. In the present-day Seoul Korean, the VOT difference between the lenis and aspirated stops has been merged, and the stop contrast is cued mainly by F0 (Kang 2014 and references cited therein). Mandarin Chinese stops are contrasted as voiceless unaspirated (/p/, /t/, /k/) vs. voiceless aspirated (/p^h/, /t^h/, /k^h/) with three places of articulation. The main cue that distinguishes between the two stop categories is short-lag vs. long-lag VOTs, respectively.

Han and Kim (2014) published a longitudinal study that investigated the production of Korean stops by native Chinese speakers. Production experiments were conducted six times once every two months for one year, in order to examine the developmental process of the Korean stop system. For VOT, the fortis stop was the category showing the most

obvious development over the course of the year. The F0 value did not show a clear development by the end of the year. In particular, the F0 value of the following vowel was not used natively as the primary cue to distinguish between the lenis and aspirated stops.

Oh (2018) examined the effect of the L2 experience on the production of Korean stops by native Chinese speakers. The participants in this study were native speakers, beginning learners who had lived in Korea for less than three months, and advanced learners who had lived there for more than two years. The F0 values of the lenis and aspirated stops were significantly different between the three groups (i.e., the native speaker group and the two learner groups). Compared to the F0 values of the native group, the F0 values of the beginning learners were lower for aspirated stops and higher for lenis stops, resulting in smaller differences in the F0 values between the two stop categories. The advanced learner group showed F0 values closer to those of the native group but nonetheless showed a significant difference from the native group. Oh interpreted these data to indicate that native Chinese speakers may have found it difficult to use the F0 contrast for the two stops because in their first language (L1) F0 is not used as the primary cue to distinguish the stop categories (see also Holliday 2015).

1.3 Research questions of the current study

While there have been several studies on the production of Korean stops by native Chinese speakers, studies on their perception are rare. Oh (2020) led one such study which investigated the perception of Korean lenis and aspirated stops by native and non-native listeners to examine the effect of base tokens used for manipulating perceptual stimuli. The study revealed a group pattern in which native Chinese listeners used the F0 cue significantly to differentiate the lenis and aspirated stops. While overall patterns appeared at the group level, individual learners' data showed considerable variability. This contrasted with the results for the individual native listeners who showed strikingly consistent patterns in stop cue use. In particular, some non-native listeners used F0 in the opposite pattern, identifying lower F0s as the aspirated stop and higher F0s as the lenis stop. This is in line with the production result of the study led by Holliday (2015) in which some native Chinese learners of Korean produced lenis stops with high F0s.

As such, previous studies have investigated either production or perception patterns

of Korean stops by native Chinese speakers. What is missing here is whether the cue use patterns are correlated between L2 perception and production of Korean stops. To investigate the relationship between perception and production, it is necessary to include the same participants for the two modalities and compare perception and production at the individual learners' level while also comparing groups (e.g., Levy and Law 2010). This study conducted both a perception experiment and a production experiment with the same participants to examine the relationship between L2 perception and production in the cue use patterns of Korean stops at the individual learners' level. It was also examined whether listeners who show the opposite perceptual pattern regarding the F0 cue use would reverse the F0 use in production, supporting the correlation between perception and production.

The research questions of this study are as follows: (1) Will perception and production in the cue use of L2 stops show similar stages of development or will one ability precede the other? In the former case, the two competencies will likely display correlation. If there is an indication of a developmental order, it is possible that the correlation between L2 perception and production does not appear. If this proves to be the case, a further question will be raised as to whether perception or production precedes the other in the acoustic cue use of L2 stops. (2) Will learners who use the F0 cue in the opposite pattern to native listeners in perception also show the opposite F0 use pattern in production?

Oh (2020) investigated whether the perceptual cue use patterns for Korean lenis and aspirated stops varied depending on base tokens (i.e., either /kan/ or /k^han/) used for stimuli manipulation. VOT and F0 were systematically manipulated with identical values for the two base tokens, and it was hypothesized that if there was a base token effect, the acoustic cues remaining in the base tokens (in addition to VOT and F0; e.g., H1−H2, F0 contour, vowel duration, aspiration noise, etc.) would have affected the stop perception. Native listeners did not show a base token effect, but non-native listeners did show an effect of the base token. In other words, there was no significant difference in the cue use patterns when either /kan/ or /k^han/ was used as the base token for native listeners, but the use pattern of the F0 cue was significantly different for non-native listeners as a function of the base token. When /k^han/ was used as the base token, the F0 cue was used more effectively to discriminate the lenis and aspirated stop contrast. Therefore, in this study, it was also examined whether the L2 perception-production relationship varies depending on the base token used for stimuli manipulation.

As input data to estimate the perception-production relation, coefficient values obtained from binary logistic regression analyses (perceptual β coefficients) and discriminant function analyses (production canonical coefficients) were calculated for each participant. The coefficient values from the two analyses provide measures of how reliably an individual learner uses the VOT and F0 cues to differentiate the lenis and aspirated stops in perception and production, respectively. Larger coefficient values indicate a larger relative importance of the corresponding acoustic cue in perceiving or producing the two stop categories for individual learners (e.g., Shultz et al. 2012; Schertz et al. 2015).

2. Perception experiment

2.1 Methods

2.1.1 Participants

Fifteen female native Mandarin Chinese learners of Korean participated in the perception and production experiments (mean age 22.9; age range 19-28). All of them were born and raised in mainland China and were undergraduate or graduate students at universities in Seoul at the time of this study. The average length of residence (LOR) in Korea among the participants was 2.3 years (LOR range 0.5-5 years). All the participants had lived exclusively in Seoul except for two who had the experience of living in other cities. The average TOPIK (Test of Proficiency in Korean; National Institute for International Education, Ministry of Education) level was 5.4 (level range 3-6, with level 6 corresponding to the highest linguistic competence). The languages that were being used by participants in daily life at the time of this study were Chinese for seven participants, Chinese/Korean for four participants, Korean for three participants, and Chinese/Korean/English for one participant. No participant had any experience living abroad outside of Korea. No participants reported speech or hearing impairments. They were paid for their participation in the experiments. Table 1 summarizes the participant information. The first column lists participant numbers.

Table 1. Participant information

	Age	Job	Major	LOR (year)	Area(s) of residence	TOPIK (level)	Daily language(s)
P1	21	Undergrad	Consumer Science	3.0	Seoul	4	Chi
P2	21	Undergrad	Multimedia	3.5	Seoul	6	Chi
P3	26	Grad	Business	4.5	Seoul	5	Chi
P4	21	Undergrad	Business	4.0	Seoul	5	Chi
P5	24	Undergrad	Business	5.0	Seoul/Daegu	5	Chi/Kor
P6	28	Grad	Korean	4.5	Seoul/Cheonan	6	Kor
P7	23	Grad	Economy	3.5	Seoul	6	Chi/Kor/Eng
P8	19	Undergrad	Business	0.5	Seoul	6	Chi
P9	24	Grad	Law	1.5	Seoul	6	Chi/Kor
P10	21	Undergrad	Business	4.0	Seoul	5	Chi/Kor
P11	23	Undergrad	Music	2.5	Seoul	3	Kor
P12	24	Undergrad	Business	1.0	Seoul	6	Chi
P13	22	Undergrad	Korean Education	1.5	Seoul	6	Kor
P14	22	Undergrad	Korean	2.0	Seoul	6	Chi
P15	24	Grad	Korean	3.5	Seoul	6	Chi/Kor

Notes: Undergrad = undergraduate student; Grad = graduate student; LOR = length of residence; Chi = Chinese; Kor = Korean; Eng = English

2.1.2 Perceptual stimuli

As base tokens for manipulating perceptual stimuli, the monosyllabic words /kan/ and /k^han/ were used (section 1). The base tokens /kan/ and /k^han/ were selected from a data pool that was collected in 2017. The selected base tokens were produced by a female native speaker of Korean who was born and raised in Seoul (age 22 at the time of recording; an undergraduate student at a university in Seoul). The words were read in a carrier sentence (tasi “CVn” poseyo, ‘see “CVn” again’). The VOT values of the base tokens were 75 ms for /kan/ and 76 ms for /k^han/. The F0 values were 208 Hz for /kan/ and 296 Hz for /k^han/.

The base tokens were cut from 60 ms before the start of the VOT sections of the stops to 30 ms after /n/ to prepare tokens for stimuli manipulation. The stimuli were manipulated using the PSOLA function of Praat (Boersma and Weenink 2016). VOT and F0 were manipulated each in seven steps. As presented in Table 2, VOT was manipulated in 11 ms intervals from 32 ms to 98 ms, and F0 was manipulated in 28 Hz intervals from 170 Hz to 338 Hz. The ranges of the manipulated VOT and F0 values reflected the ranges of VOT and F0 values found in the female data pool that was collected in

2017.

Table 2. Manipulation steps of VOT and F0 cues

Steps	1	2	3	4	5	6	7
VOT (ms)	32	43	54	65	76	87	98
F0 (Hz)	170	198	226	254	282	310	338

The VOT values were reduced or increased from the original values to the intended values using the relative duration function. The F0 values were lowered or raised from the original values to the intended values at the midpoint of the following vowel duration. The original F0 contour shapes were maintained. First, the seven steps of VOT were manipulated, and then the F0 steps were manipulated so as to correspond each VOT step. For each base token, 49 stimuli were manipulated, and each stimulus was presented five times. Thus, each participant answered a total of 490 trials (7 VOTs*7 F0s*5 repetitions*2 base tokens).

2.1.3 Procedure

A forced-choice identification task was conducted using the MFC function of Praat (Boersma and Weenink 2016). The participants listened to the perceptual stimuli through Audio-Technica ATH-M40X headphones in a quiet room. They clicked on either /kan/ or /k^han/, which were presented in Korean spelling on a computer monitor. They were instructed to choose the word that sounded most similar to the audio example. They did not have a time limit to answer but were not allowed to change their answers after making a selection. The /kan/-based and /k^han/-based stimuli were blocked, and their presentation order was counterbalanced. The stimuli were randomized for each participant. Before the experiment, a practice session was conducted with eight stimuli selected from the full stimulus pool.

2.2 Results

2.2.1 Perceptual cue use patterns as a group

The L2 use patterns of the VOT and F0 cues to identify the Korean lenis and

aspirated stops were examined first at the group level. Figure 1 presents the percentage of aspirated (/k^{h}an/) responses for each step of VOT and F0. Native Korean data are presented on the right for comparison (adapted from Oh 2020; data from twenty native female listeners of Korean). F0 was used more reliably than VOT to identify the L2 stop categories for both /kan/-based (gray) and /k^{h}an/-based stimuli (black). The differences between the maximum and minimum percentages of the aspirated responses were larger for F0 (60% for /kan/-based stimuli; 51% for /k^{h}an/-based stimuli) than for VOT (9% for /kan/-based stimuli; 16% for /k^{h}an/-based stimuli). The differences between the maximum and minimum percentages of the aspirated responses for F0 were smaller for non-native listeners than for native listeners (99% for /kan/-based stimuli; 98% for /k^{h}an/-based stimuli).

The learners' response data were analyzed using mixed-effects logistic regression analysis to examine the effects of the two acoustic cues (VOT and F0) and the effects of the base tokens. The program R (R Development Core Team 2016) and the package "lme4" (Bates et al. 2015) were used for the analysis. The manipulated VOT and F0 values were coded as seven steps (−3, −2, −1, 0, 1, 2, 3). The response was the dependent variable; VOT, F0, and the base tokens were the fixed effects; and the subject was the random effect. The interaction terms of VOT*Base Token and F0*Base Token were also included to examine the effects of the base tokens on the use of the acoustic cues. The reference for the base token was /k^{h}an/. Table 3 presents the outputs of the analysis.

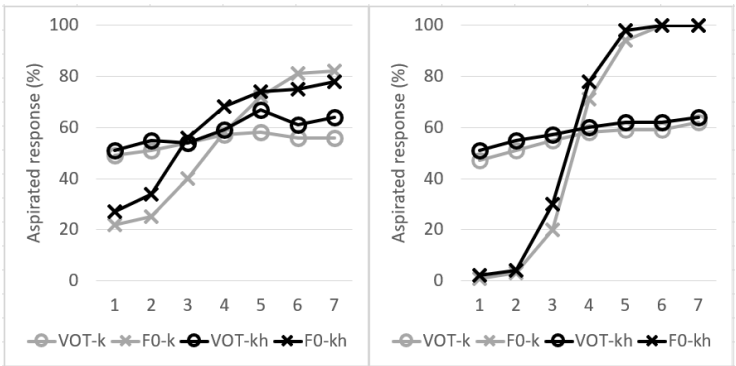

Figure 1. Percentages of aspirated responses (y-axis) along the manipulated seven steps of VOT and F0 (x-axis): /kan/-based (gray) and /k^{h}an/-based stimuli (black). Native Korean data on the right for comparison (adapted from Oh 2020)

Table 3. Outputs of the mixed-effects logistic regression analysis

	Estimate	Std. Error	z value	Pr (>|z|)
(Intercept)	0.23272	0.13501	1.724	0.08475
VOT	0.06986	0.01923	3.633	0.00028
F0	0.56904	0.02145	26.529	<0.0001
BaseToken	0.20977	0.05386	3.895	<0.0001
VOT*BaseToken	0.04475	0.02683	1.668	0.09529
F0*BaseToken	−0.13101	0.02906	−4.508	<0.0001

The effects of VOT (β=0.06986, *SE*=0.01923, z=3.633, p<0.0005) and F0 (β=0.56904, *SE*=0.02145, z=26.529, p<0.0001) were both statistically significant, indicating that both cues were used by non-native listeners to identify the lenis and aspirated stops. The estimated coefficient for F0 (0.56904) was larger than that for VOT (0.06986), indicating a greater reliance on F0 than on VOT in the stop identification. The effect of the base token was significant (β=0.20977, *SE*=0.05386, z=3.895, p<0.0001), which indicates that the use patterns of the acoustic cues differed depending on the base tokens selected for stimuli manipulation. The learners answered more as /k^{h}an/ for /k^{h}an/-based stimuli (2164 responses) than for /kan/-based stimuli (1997 responses), suggesting that the L2 learners used acoustic cues other than VOT and F0 when identifying the Korean lenis and aspirated stops (section 1.3).

While the interaction between VOT and the base token was not significant (β =0.04475, *SE*=0.02683, z=1.668, p=0.09529), the interaction between F0 and the base token was significant (β=−0.13101, *SE*=0.02906, z=−4.508, p<0.0001). In other words, the F0 use patterns differed depending on the base tokens used for stimuli manipulation. The difference between the maximum and minimum percentages of the aspirated responses was greater for /kan/-based stimuli (60%) than for /k^{h}an/-based stimuli (51%), indicating that the F0 cue was used more reliably when the base token was /kan/ (cf., Oh 2020). This interpretation (i.e., more reliable use of the F0 cue when the base token was /kan/) can also be confirmed by the fact that the estimated coefficient of the interaction effect showed a negative value (−0.13101).

2.2.2 Perceptual cue use patterns of individual learners

In terms of a group pattern, the learners used F0 more reliably than VOT to distinguish between the lenis and aspirated stops of Korean. This section examines the

perceptual cue use patterns at the individual learner level. Figure 2 shows individual learners' percentages of aspirated responses along the seven steps of VOT and F0. Individual learners demonstrated a considerable degree of variability in the patterns of stop cue use.

Binary logistic regression analyses were conducted to estimate the perceptual weights (i.e., relative importance) of VOT and F0 in discriminating the stops as β-coefficient values. The response was the dependent variable and the VOT/F0 steps were the covariates. Larger β coefficients indicate larger contributions of the corresponding cues for the stop categorization (e.g., Shultz et al. 2012; Schertz et al. 2015). The significance alpha level was set at 0.05. Table 4 presents the perceptual β-coefficient values of individual learners obtained from the analyses. The mean β-coefficient value of the VOT was 0.189 for /kan/-based stimuli and 0.235 for /k^han/-based stimuli, and that of the F0 was 1.483 for /kan/-based stimuli and 1.471 for /k^han/-based stimuli. For both types of stimuli, the mean β-coefficient value of F0 was larger than that of VOT, indicating that the use of F0 was more important for the categorization of the two Korean stops.

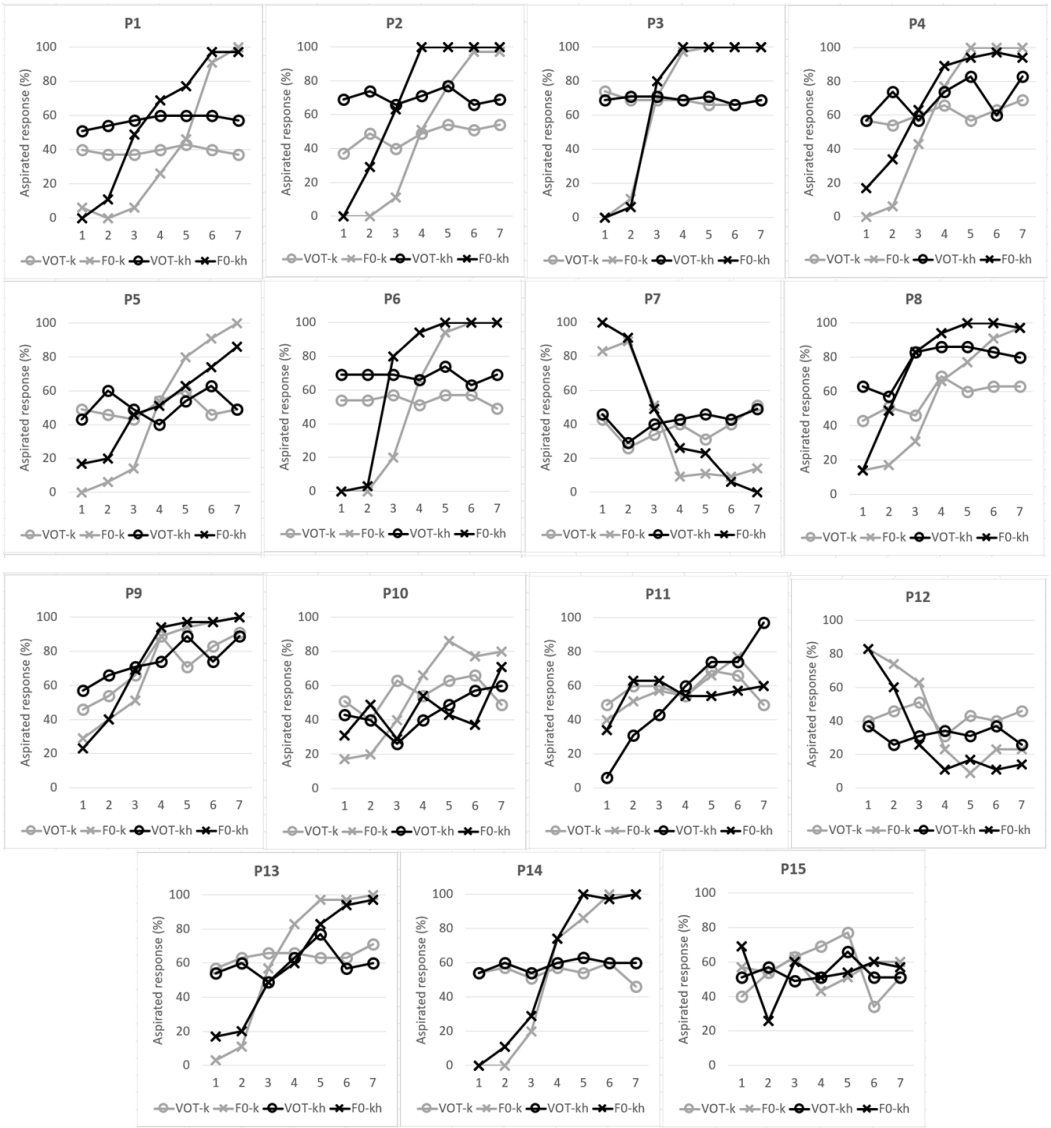

Figure 2. Individual learners' percentages of aspirated responses (y-axis) along seven steps of VOT and F0 (x-axis): /kan/-based (gray) and /k^{h}an/-based stimuli (black)

For /kan/-based stimuli, 12 learners (P1, P2, P3, P4, P5, P6, P8, P9, P10, P11, P13, and P14) used F0 significantly to distinguish the lenis and aspirated stops. Nine of them (P1, P2, P3, P4, P5, P6, P9, P13, and P14) showed β-coefficient values for F0 larger than 1. Four of them (P2, P4, P8, and P9) used not only F0 but also VOT significantly to distinguish the stops. P7 (a negative β-coefficient value of −0.852) and P12 (−0.586) used F0 significantly but in the opposite pattern. In other words, they tended to identify

larger F0 values in the following vowel as /kan/ and smaller F0 values as /k^han/, displaying perceptual patterns opposite to those of native listeners. For both learners, the absolute β-coefficient values were greater for F0 than for VOT (0.125 for P7; −0.005 for P12), indicating that they placed greater importance on the F0 cue, even if reversed, to distinguish the two stops. P15 did not use either VOT or F0 significantly to differentiate the stops. In particular, this learner showed the lowest absolute β coefficient value for F0 (0.012) among the participants.

Table 4. Perceptual β coefficients (coeff.) of individual learners obtained from binary logistic regression analyses (an asterisk indicating the significant use of the acoustic cue in the same pattern as native listeners)

	/kan/-based stimuli					/k^han/-based stimuli			
	VOT		F0			VOT		F0	
	coeff.	*p*-value	coeff.	*p*-value		coeff.	*p*-value	coeff.	*p*-value
P1	0.011	p=0.918	1.443	p<0.001*	P1	0.097	p=0.304	1.195	p<0.001*
P2	0.325	p<0.01*	1.811	p<0.001*	P2	−0.034	p=0.793	2.379	p<0.001*
P3	−0.199	p=0.191	3.066	p<0.001*	P3	−0.138	p=0.463	4.378	p<0.001*
P4	0.279	p<0.05*	2.294	p<0.001*	P4	0.216	p<0.05*	1.021	p<0.001*
P5	0.065	p=0.534	1.496	p<0.001*	P5	0.053	p=0.468	0.554	p<0.001*
P6	−0.068	p=0.604	2.361	p<0.001*	P6	−0.051	p=0.751	3.520	p<0.001*
P7	0.125	p=0.138	−0.852	p<0.001	P7	0.140	p=0.153	−1.266	p<0.001
P8	0.251	p<0.01*	0.955	p<0.001*	P8	0.467	p<0.001*	1.623	p<0.001*
P9	0.730	p<0.001*	1.283	p<0.001*	P9	0.488	p<0.001*	1.360	p<0.001*
P10	0.083	p=0.266	0.606	p<0.001*	P10	0.163	p<0.05*	0.168	p<0.05*
P11	0.030	p=0.646	0.179	p<0.01*	P11	0.729	p<0.001*	0.122	p=0.121
P12	−0.005	p=0.941	−0.586	p<0.001	P12	−0.025	p=0.754	−0.626	p<0.001
P13	0.171	p=0.128	1.690	p<0.001*	P13	0.099	p=0.243	0.884	p<0.001*
P14	−0.089	p=0.466	2.079	p<0.001*	P14	0.121	p=0.300	1.878	p<0.001*
P15	0.012	p=0.847	0.012	p=0.847	P15	0.008	p=0.898	0.041	p=0.522

For /k^han/-based stimuli, 11 learners (P1, P2, P3, P4, P5, P6, P8, P9, P10, P13, and P14) used F0 significantly to distinguish the L2 stops. Eight of them (P1, P2, P3, P4, P6, P8, P9, and P14) showed β-coefficient values for F0 larger than 1. Four of them (P4, P8, P9, and P10) used not only F0 but also VOT significantly to distinguish the stops. P11 did not use F0 (β coefficient 0.122) but used VOT (0.729) significantly for stop identification. P7 (a negative β-coefficient value of −1.266) and P12 (−0.626) used F0 significantly but in the opposite pattern to native listeners for /k^han/-based stimuli. The absolute β-coefficient values for these two learners were greater for F0 than

VOT (0.140 for P7; −0.025 for P12), indicating that they placed more importance on the F0 cue, even if reversed in use, to distinguish the two stops for /k^{h}an/-based stimuli. P15 did not use either VOT or F0 significantly for the stop categorization for /k^{h}an/-based stimuli. The learner showed the lowest absolute β-coefficient value for F0 (0.041) among the participants.

3. Production experiment

3.1 Methods

The 15 native Chinese learners of Korean who participated in the perception experiment also participated in the production experiment (section 2.1). Each participant completed the production test first and then completed the perception test. If the perception test had been performed first, the task focusing on the changes in the stop cue values may have affected the production performance (Shultz et al. 2012).

Monosyllabic words with the /CVn/ structure were used as test materials wherein C was a stop with one of the three places of articulation (bilabial, denti-alveolar, or velar) and V was /i/ or /a/. Twelve target words (/pin/ ‘empty’, /tin/ (non-word), /kin/ ‘long’, /pan/ ‘half’, /tan/ ‘sweet’, /kan/ ‘liver’, /p^{h}in/ ‘blooming’, /t^{h}in/ (non-word), /k^{h}in/ ‘on’, /p^{h}an/ ‘board’, /t^{h}an/ ‘burnt’, and /k^{h}an/ ‘compartment’) and 10 filler words were presented in random order. The words were read in a carrier sentence ‘tasi “CVn” poseyo’ (‘see “CVn” again’). The target lenis stops were produced in a phrase-initial position and mostly as voiceless, but there were four times when the lenis stops were produced as pre-voicing (see below).

The test materials were presented on a computer monitor one sentence at a time every two seconds in order to control individual differences in speaking rate to some extent (e.g., Kessinger and Blumstein 1998). Recordings were made with Praat (Boersma and Weenink 2016) using a TAKSTAR GX6 USB condenser microphone. The recordings were digitized at a sampling rate of 44,100 Hz and stored as WAV files. The VOT values of the stops and the F0 values in the following vowels were measured by hand using Praat (Boersma and Weenink 2016). The VOTs were measured from the release burst of the stops to the beginning of periodic cycles in sound waveforms, and the measurements were confirmed both aurally and through wideband spectrograms. The

F0s were measured at the vowel midpoint following the stop (Kang and Guion 2008; Kang 2014). The F0 values of each participant were converted from hertz (Hz) to semitones (St) using the St conversion equation '12[ln(Hz/100)/ln2]' in order to normalize differences across speakers in speech physiology. The participants read the entire set of test materials three times. The second and third utterances were analyzed, and the first utterance was analyzed when a production problem took place (four times with producing lenis stops as pre-voicing, and eight times with articulation mistakes).

The relative weightings of the VOT and F0 cues in the stop categorization were estimated using discriminant function analyses of the production data of each participant. The stop served as the group variable, and VOT and F0 served as the predictor variables. A larger canonical discriminant function coefficient value of an acoustic cue indicates the cue's contribution in producing the stop contrast was more important (e.g., Shultz et al. 2012; Schertz et al. 2015).

3.2 Results

This section reports the experimental results regarding the patterns of VOT and F0 values of the Korean stops produced by native Chinese speakers. At the group level, the average value of VOT was 65.4 ms (SD 30.0 ms) for lenis stops and 84.5 ms (SD 22.8 ms) for aspirated stops. The difference in the VOT values between the two stops was 19.1 ms. In the study by Oh et al. (2018), the average VOT values of 10 native female speakers of Korean were 58.5 ms (SD 20.9 ms) for lenis stops and 59.4 ms (SD 18.8 ms) for aspirated stops, and the VOT difference between the two stops was 0.9 ms, suggesting that the learners participating in this study made larger VOT distinctions between the two stops than native speakers did. The average F0 value was 229.5 Hz (SD 30.1 Hz) for lenis stops and 251.6 Hz (SD 42.7 Hz) for aspirated stops. The F0 values in St were 14.2 St (SD 2.3 St) for lenis stops and 15.7 St (SD 2.8 St) for aspirated stops. The difference in F0 between the two stops was 1.5 St. In Oh et al. (2018), the average F0 values of native female speakers of Korean were 13.3 St (SD 1.7 St) for lenis stops and 18.0 St (SD 1.9 St) for aspirated stops, and the F0 difference between the two stops was 4.7 St, suggesting that the learners in this study made smaller F0 distinctions between the two stops than native speakers did.

Figure 3 presents the individual learners' means and the standard deviations of the

VOT (above) and F0 (below) values for the lenis and aspirated stops. Table 5 presents the production coefficient values obtained from the discriminant function analyses that were conducted for individual learners (section 3.1). The significance alpha level was set at 0.05. Three learners (P2, P9, and P11) differentiated the lenis and aspirated stops with both VOT and F0 values. Three learners (P1, P4, and P8) differentiated the two stops only with VOT values, and three learners (P13, P14, and P15) differentiated the stops only with F0 values. Six learners (P3, P5, P6, P7, P10, and P12) did not differentiate either VOT or F0 values in the production of the two stops. Thus, six out of 15 learners significantly used F0, the primary cue that distinguishes the lenis and aspirated stops.

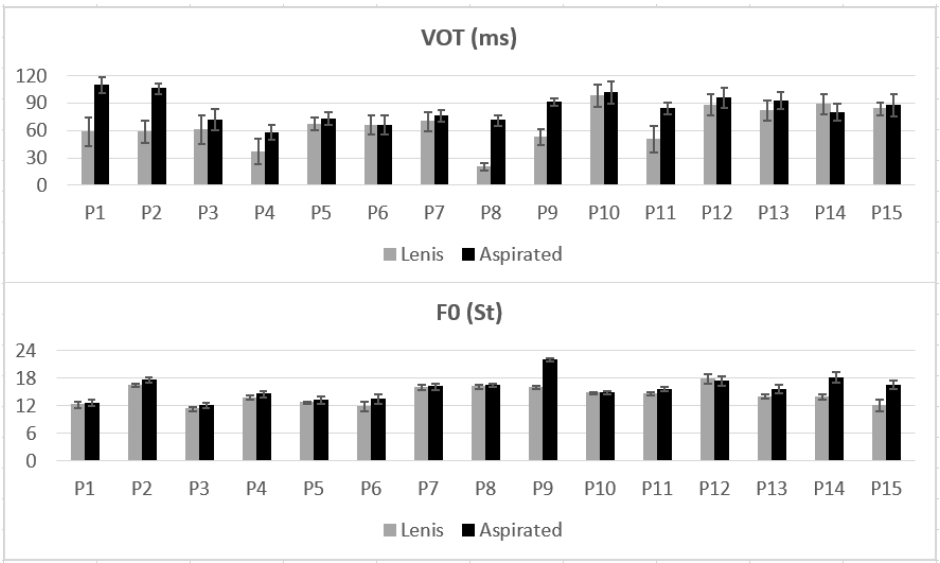

Figure 3. Individual learners' means and standard deviations of VOT (ms) and F0 (St) values for Korean lenis (gray) and aspirated (black) stops

Table 5. Standardized canonical coefficient values obtained from discriminant function analyses for individual learners (an asterisk indicating the significant use of the acoustic cue in the same pattern as native speakers)

	VOT	F0		VOT	F0		VOT	F0
P1	1.005 (p<0.001)*	0.268 (p=0.540)	P6	0.932 (p=0.961)	1.353 (p=0.082)	P11	0.851 (p<0.005)*	0.768 (p<0.01)*
P2	0.930 (p<0.001)*	0.548 (p<0.01)*	P7	0.837 (p=0.459)	0.367 (p=0.660)	P12	0.921 (p=0.400)	−0.116 (p=0.552)
P3	0.629 (p=0.335)	0.835 (p=0.174)	P8	0.997 (p<0.001)*	0.056 (p=0.276)	P13	0.450 (p=0.201)	0.907 (p<0.01)*
P4	0.748 (p<0.05)*	0.433 (p=0.088)	P9	0.204 (p<0.001)*	0.952 (p<0.001)*	P14	0.274 (p=0.329)	1.079 (p<0.001)*
P5	0.737 (p=0.272)	0.805 (p=0.214)	P10	0.892 (p=0.680)	0.948 (p=0.663)	P15	−0.130 (p=0.579)	1.022 (p<0.001)*

4. Perception-production relation in the stop cue use by individual learners

This section examines whether the patterns of the acoustic cue use by individual learners in the Korean stop categorization show a correlation between perception and production. The variables used for the correlation analyses were the perceptual β coefficients (Table 4) and the production canonical coefficients (Table 5) of individual learners, which were obtained from the binary logistic regression analyses and the discriminant function analyses, respectively (sections 2.2.2 and 3.1). The larger the coefficient values of an acoustic cue, the more reliably the corresponding cue is assumed to be used for the stop categorization.

Figure 4 displays scatterplots of the perceptual β coefficients (x-axis) and the production canonical coefficients (y-axis) for VOT (left) and F0 (right). The scatterplots are presented separately for /kan/-based stimuli (above) and /k^han/-based stimuli (below). A linear regression line is indicated on each graph. Pearson product-moment correlation analyses were performed using the perceptual β coefficients and the production canonical coefficients for individual learners. The correlation between the perceptual and the production coefficients was not significant for either VOT (Pearson correlation coefficient r = −0.141, p = 0.617 for /kan/-based stimuli; r = 0.032, p = 0.910 for /k^han/-based stimuli) or F0 (r = 0.418, p = 0.121 for /kan/-based stimuli; r = 0.387, p = 0.154 for /k^han/-based stimuli) (cf., Shultz et al. 2012; Schertz et al. 2015). Positive trends were found in the F0 use between perception and production (the slope of the

linear regression line = 0.1516, r = 0.418 for /kan/-based stimuli; the slope = 0.1064, r = 0.387 for /k^{h}an/-based stimuli) in which learners who used F0 with importance in perception tended to use F0 with importance in production. However, the trends were not found to be statistically significant (p = 0.121 for /kan/-based stimuli; p = 0.154 for /k^{h}an/-based stimuli).

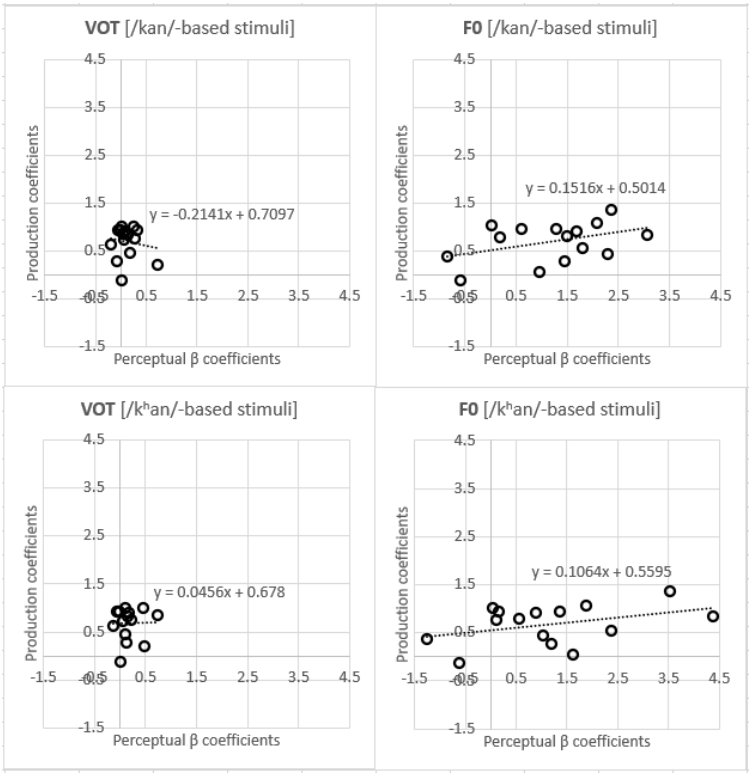

Figure 4. Scatterplots of perceptual β coefficients (x-axis) and production canonical coefficients (y-axis) for VOT (left) and F0 (right): /kan/-based (above) and /k^{h}an/-based stimuli (below)

As another method to estimate the correlation between perception and production, differences in the cue values between the lenis and aspirated stops were used as input data for the analyses. For perception, maximum-minus-minimum percentage values of /k^{h}an/ responses (i.e., perceptual differential values) for individual learners were calculated for each acoustic cue (e.g., 97% for /k^{h}an/-based stimuli of P1 [= 97% − 0%] in Figure 2). For the two learners who used the F0 cue in an opposite manner to that of native listeners (i.e., identifying lower F0s with the aspirated stop and higher F0s with the lenis stop), the corresponding negative values were used as input data for the differences between the maximum and the minimum percentage values of /k^{h}an/

responses. For production, [VOT mean of aspirated stops − VOT mean of lenis stops] and [F0 mean of aspirated stops − F0 mean of lenis stops] (i.e., production differential values) of individual learners were used as input data (e.g., 51 ms for VOT [= 110 ms of aspirated stops − 59 ms of lenis stops] and 0.34 St for F0 [= 12.61 St of aspirated stops − 12.27 St of lenis stops] as for P1 in Figure 3). P14 showed a negative production differential value for VOT (−9 ms), and P12 showed a negative production differential value for F0 (−0.51 St), which indicate that the learners used the VOT or F0 cue in the opposite pattern to native speakers (i.e., producing lower mean VOT/F0 with aspirated stops and higher mean VOT/F0 with lenis stops). In these cases, the corresponding negative values were used as the production differential values. The correlation between the perceptual and the production differential values was not significant for either VOT (r = 0.062, p = 0.827 for /kan/-based stimuli; r = 0.238, p = 0.392 for /k^han/-based stimuli) or F0 (r = 0.183, p = 0.515 for /kan/-based stimuli; r = 0.292, p = 0.291 for /k^han/-based stimuli).

In sum, the correlations between L2 perception and production in the patterns of the stop cue use by Korean learners were found to be non-significant by means of both the perceptual and the production coefficient values obtained from the binary logistic regression and the discriminant function analyses, and the perceptual and the production differential values of the VOT/F0 cues to distinguish the lenis and aspirated stops.

Two learners (P7 and P12) used the F0 cue in a manner opposite to native listeners in that they identified lower F0s with the aspirated stop and higher F0s with the lenis stop (Figure 2 and Table 4). The opposite use of F0 in perception was statistically significant (Table 4). In production, the differences in the F0 values between the two stops (i.e., F0 mean of aspirated stops − F0 mean of lenis stops) were 3.67 Hz (0.23 St) for P7 and −8.33 Hz (−0.51 St) for P12 (Figure 3 and Table 5). That is, P7 produced lenis stops with a lower mean F0 and aspirated stops with a higher mean F0, although the difference was not statistically significant. P12 tended to also use F0 in an opposite pattern in production, producing lenis stops with a higher mean F0 and aspirated stops with a lower mean F0, but this difference in production was not statistically significant.

Except for the above two learners who demonstrated the opposite pattern in the use of F0, two general types of perception-production relationships were found for the use of the F0 cue; one that used F0 significantly in both perception and production, and the other that used F0 in perception but not in production. Four learners (P2, P9, P13, P14)

showed the former pattern, demonstrating the significant use of F0 in both perception and production. Seven learners (P1, P3, P4, P5, P6, P8, P10) displayed the latter pattern, displaying the significant use of F0 in perception but not in production.[1] Two learners deviated from the above two types of perception-production relationships of F0 use. One learner (P11) distinguished the two stops with F0 in production but showed different patterns of F0 use in perception as a function of the base tokens. F0 was used significantly in /kan/-based stimuli but not in /k^han/-based stimuli. Except for P11, no learner showed the base token effect on the use of the F0 cue. P15 was the only learner who used F0 significantly in production but not in perception.

In sum, for the use of F0, which is the primary cue that categorizes the Korean lenis and aspirated stops, more learners demonstrated that their perception learning of the L2 stop cue preceded their production learning, while one learner seemingly demonstrated a pattern of L2 production of the stop cue preceding its perception.

5. Summary and discussion

The purpose of this study was to investigate whether the perceptual and production patterns of non-native speakers are correlated through the use of acoustic cues differentiating Korean lenis and aspirated stops. A perception test was conducted to identify tokens in which VOT and F0 values were systematically manipulated between /kan/ and /k^han/. A production test was administered in which words starting with lenis or aspirated stops were read in a carrier sentence. Relative cue weightings were estimated via perceptual and production coefficients that were obtained from binary logistic regression analyses and discriminant function analyses for individual learner data (Shultz et al. 2012; Schertz et al. 2015).

The first research question was whether perception and production in the cue use of L2 stops show similar stages of development or whether one ability precedes the other. In production, while six learners used F0 (the primary cue) significantly, the other nine did not use F0 significantly. Although the participants of this study were living in Korea as undergraduate or graduate students and their Korean proficiencies were at intermediate or higher levels, it seemed to be difficult for them to produce the L2 stop contrast using the primary cue. In perception, 12 learners used F0 significantly for the

1 The fact that the F0 cue was used significantly to distinguish the two stops in perception or production does not imply the L2 acquisition of native-like F0 values.

stop identification (for /kan/-based stimuli; eleven learners used F0 significantly for /k^han/-based stimuli). The other learners did not use F0 significantly or used it in an opposite pattern to native listeners. That is, more learners used F0 reliably for the stop categorization in perception than production, suggesting that their perception of the primary cue for the L2 stops preceded their production.

The correlation between the perceptual and production coefficients was found to be non-significant for either VOT or F0. For the use of F0, five out of 15 learners used it both in perception and production, and seven learners used it in perception but not in production. Two learners did not use F0 in production but perceived it in an opposite pattern to native listeners. The remaining one learner did not use F0 significantly in either perception or production.[2] Overall, the results did not differ between /kan/-based and /k^han/-based stimuli, indicating that the perception-production relationship in the cue use patterns was not different as a function of the base tokens used for the manipulation of perceptual stimuli.

As stated in the Introduction, Holliday's production study (2015) and Oh's perception study (2020) reported that some Chinese participants used F0 in the opposite pattern to native values, producing or identifying lower F0 values with aspirated stops and higher F0 values with lenis stops, respectively. The second research question of this study was whether, given the perception-production link, individual learners who use F0 in the opposite pattern to native listeners in perception would also use it reversely in production; the answer to this question would be found by conducting both a perception experiment and a production experiment for the same participants. Two learners reversed the perception of F0, and this opposite use was statistically significant. One of them did not show opposite cue use patterns in production, as they produced lenis stops with a lower mean F0 than aspirated stops, but the F0 difference in production was not statistically significant. The other learner showed the opposite pattern with F0 use in production, producing lenis stops with a higher mean F0 than aspirated stops, but the F0 difference was not statistically significant here, either. That is to say that the two learners used F0 significantly in reverse in perception, but their F0 difference between

2 Schertz et al. (2015) investigated the perception and production of L2 English voiced and voiceless stops by native Korean speakers. The production-perception link was examined at the individual learner level. The patterns using acoustic cues were examined (e.g., VOT as the primary cue and F0 in the following vowel as the secondary cue). Their results are in line with those in the current study in that both demonstrated no significant L2 perception-production correlation and considerable individual learner variability for the patterns using the L2 stop cues.

the two stops was not significant in production.

One possible interpretation of this result is that the two learners who perceived F0 in reverse were in a learning stage just before that in which perception would be learned but where production was yet to be learned. It is probable that they were in the stage of learning such that they knew F0 is an important cue for the distinction of the two Korean stops but displayed the error of using the cue in the opposite pattern. It can be further hypothesized that in the next learning stage, the learners would come to use F0 in the perception of the Korean stops in a native-like pattern.

Holliday (2015) led a longitudinal study on the production of Korean stops by native Chinese speakers. Production experiments were conducted three times over the course of one year, and various production patterns were found among learners. Only one participant used F0 in a native-like pattern from Time 1, and by Time 3, two participants had learned the F0 difference but three had not. It was interpreted that Chinese learners of Korean did not use F0 to distinguish the lenis and aspirated stops because F0 is not the primary cue for the stop categorization in their L1 (see also Oh 2018). In connection with this, the result of the current study indicates that more L2 learners learned F0 as the primary cue for the distinction of Korean stops in perception, but the cue learning had not yet been reflected in their L2 production. This is in line with the result of Flege (1993) who investigated the perception and production of the English /t/ - /d/ contrast in word-final position by native Chinese speakers. Experienced learners were more native-like in the perception of the L2 /t/ - /d/ contrast than in its production.

In a cross-sectional study on the perception-production relationship in L2 speech, two scenarios are possible, as described in section 1.3. If the perception and production of L2 speech show similar stages of development, the two modalities are likely to be correlated. If L2 perception and production develop sequentially, it is probable that correlation would not be found due to a time-lapse between the development of the two modalities. A longitudinal study for individual learners will provide more accurate evidence for the developmental processes between L2 perception and production learning. If longitudinal evidence is found suggesting that perception and production develop in a sequential way, a clearer understanding of the general developmental patterns of L2 perception and production may be possible. Through a longitudinal study, it will also be possible to discover a process for approaching native-like cue use patterns in perception and production in the cases of learners who displayed developmental stages that deviated from the general patterns in a cross-sectional study (e.g., production preceding perception,

an error pattern that reverses the F0 use in perception, etc.).

References

Bates, Douglas, Martin Maechler, Ben Bolker, and Steven Walker. 2015. *lme4: Linear mixed-effects models using Eigen and S4*. R package version 1.1-9.

Boersma, Paul and David Weenink. 2016. Praat: Doing phonetics by computer. Software retrieved from <http://www.fon.hum.uva.nl/praat/>.

Bohn, Ocke-Schwen and James E. Flege. 1992. The production of new and similar vowels by adult German learners of English. *Studies in Second Language Acquisition* 14: 131-158.

Bohn, Ocke-Schwen and James E. Flege. 1997. Perception and production of a new vowel category by adult second language learners. In Allan James and Jonathan Leather (eds.), *Second-Language Speech*, 53-74. New York: De Gruyter Mouton.

Bradlow, Ann R., David B. Pisoni, Reiko Akahane-Yamada, and Yoh'ichi Tohkura. 1997. Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America* 101: 2299-2310.

Flege, James E. 1993. Production and perception of a novel, second-language phonetic contrast. *Journal of the Acoustical Society of America* 93: 1589-1608.

Flege, James E., Ocke-Schwen Bohn, and Sunyoung Jang. 1997. Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics* 25: 437-470.

Flege, James E. and Wieke Eefting. 1987. Cross-language switching in stop consonant perception and production by Dutch speakers of English. *Speech Communication* 6: 185-202.

Han, Jeong-Im and Joo-Yeon Kim. 2014. Acquisition of Korean stops by native speakers of Mandarin Chinese: A one-year longitudinal study. *Korean Journal of Linguistics* 39: 377-403.

Holliday, Jeffrey J. 2015. A longitudinal study of the second language acquisition of a three-way stop contrast. *Journal of Phonetics* 50: 1-14.

Kang, Kyoung-Ho and Susan G. Guion. 2008. Clear speech production of Korean stops: Changing phonetic targets and enhancement strategies. *Journal of the Acoustical Society of America* 124: 3909-3917.

Kang, Yoonjung. 2014. Voice onset time merger and development of tonal contrast in Seoul Korean stops: A corpus study. *Journal of Phonetics* 45: 76-90.

Kessinger, Rachel H. and Sheila E. Blumstein. 1998. Effects of speaking rate on voice-onset time and vowel production: Some implications for perception studies. *Journal of Phonetics* 26: 117-128.

Laméris, Tim J. and Calbert Graham. 2020. L2 perception and production of Japanese lexical pitch: A suprasegmental similarity account. *Journal of Monolingual and Bilingual Speech* 2: 106-136.

Levy, Erika S. and Franzo F. Law. 2010. Production of French vowels by American-English learners

of French: Language experience, consonantal context, and the perception-production relationship. *Journal of the Acoustical Society of America* 128: 1290-1305.

Oh, Eunhae. 2018. Effects of L2 experience on the production of Korean stop contrasts by Mandarin Chinese learners. *Linguistic Research* 35: 233-251.

Oh, Eunjin, Kaori Idemaru, and Boram Kim. 2018. The use of a voice onset time cue in the perception of Seoul Korean stops as a function of listener gender. *Korean Journal of Linguistics* 43: 761-780.

Oh, Eunjin. 2020. Effects of base token for stimuli manipulation on the perception of Korean stops among native and non-native listeners. *Phonetics and Speech Sciences* 12: 43-50.

Perkell, Joseph S., Frank H. Guenther, Harlan Lane, Melanie L. Matthies, Ellen Stockmann, Mark Tiede, and Majid Zandipour. 2004. The distinctness of speakers' productions of vowel contrasts is related to their discrimination of the contrasts. *Journal of the Acoustical Society of America* 116: 2338-2344.

R Development Core Team. 2016. *The R project for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Schertz, Jessamyn, Taehong Cho, Andrew Lotto, and Natasha Warner. 2015. Individual differences in phonetic cue use in production and perception of a non-native sound contrast. *Journal of Phonetics* 52: 183-204.

Sheldon, Amy and Winifred Strange. 1982. The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics* 3: 243-261.

Shultz, Amanda A, Alexander L. Francis, and Fernando Llanos. 2012. Differential cue weighting in perception and production of consonant voicing. *Journal of the Acoustical Society of America* 132: EL95-EL101.

Eunjin Oh
Professor
Department of English Language and Literature
Ewha Womans University
52 Ewhayeodae-gil, Seodaemun-gu
Seoul 03760, Korea
E-mail: ejoh@ewha.ac.kr

Received: 2021. 06. 08.
Revised: 2021. 08. 05.
Accepted: 2021. 09. 03.