http://dx.doi.org/10.17959/sppm.2020.26.3.461

Comparison of sensitivity to VOT and F0 cues in English stop voicing contrast between native English and Korean listeners*

Eunkyung Sung Sunhee Lee** Sehoon Jung
(Cyber Hankuk University of Foreign Studies) (Kyungsung University)

Sung, Eunkyung, Sunhee Lee and Sehoon Jung. 2020. Comparison of sensitivity to VOT and F0 cues in English stop voicing contrast between native English and Korean listeners. *Studies in Phonetics, Phonology and Morphology* 26.3. 461-485. This study compared sensitivity to VOT and F0 cues between native English and Korean listeners in detecting the voicing contrast of English stop consonants. Perceptual differences between the two listener groups for each step of VOT and F0 values were also examined. Furthermore, the effects of place of articulation on the perception of voicing contrast of word-initial stops were investigated. The stimuli were modified natural speech tokens varying along six steps of a VOT continuum intermixed with six steps of an F0 continuum. The results show that although the two listener groups utilized both VOT and F0 cues, the Korean listeners were more sensitive to gradual changes of F0 than the English listeners. In addition, the English listeners' judgments of voiceless stops were made at later steps of VOT than those of the Korean listeners. Specifically, the significant perceptual differences between the two listener groups were revealed at the fifth steps of F0 for alveolar and velar stops. In terms of the place of articulation, the bilabial and alveolar stops were identified as voiceless at earlier steps of VOT than the velar stops by both listener groups. Moreover, the asymmetry of gradual sensitivity to F0 across the two listener groups was more clearly shown for velar and alveolar stops than for bilabial stops. These results indicate that the second language (L2) listeners utilized acoustic cues in a language-specific way. **(Cyber Hankuk University of Foreign Studies, Professor and Associate Professor; Kyungsung University, Lecturer)**

Keywords: sensitivity, English stops, voicing contrast, VOT, F0, L2 listener, place of articulation

---

* This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2019S1A5A2A01047388). We would like to thank the three reviewers for their insightful comments and suggestions on this paper.

** Corresponding author

1. Introduction

Speech perception involves listeners' rapid decoding of highly variable acoustic signals. If listeners were to decipher sound signals differently from speakers' intended speech, communication would break down. Recent research has focused on decoding processes in which listeners used details of acoustic signals to determine speech categories (Kong and Yoon 2013, Kong and Edwards 2016, Kapnoula et al. 2017, Son 2017).

Voice onset time (VOT) and fundamental frequency (F0) transition characteristics are important phonetic cues for the voicing categorization of stop consonants. Voicing contrast in consonants is very common across the world's languages (Ladefoged and Maddieson 1996), and VOT is known as a primary cue for voicing contrast in word- and syllable-initial positions in many languages (Lisker and Abramson 1964, Ladefoged and Cho 2001, Kong et al. 2011). In general, the voiceless ([−voice]) category is represented by longer VOT values than the voiced ([+voice]) category. In English, word-initial voiced stop consonants are generally said to have a VOT of 15ms or less (short-lag VOT), and voiceless stop consonants 30ms or more (long-lag VOT) (Lieberman and Blumstein 1988, Nakai and Scobbie 2016). Regarding F0, it has been known that in English F0 is higher for voiceless than for voiced stops, but F0 values between these two categories are not significantly different (Haggard et al. 1970, Kingston and Diehl 1994). This means F0 values are different between voiced and voiceless stop consonants in English, but native English speakers do not use this phonetical cue to distinguish between the two stop categories. Thus, English stops have a two-way contrast, where VOT is a primary cue for voicing contrast, and F0 is used to a much smaller extent.

Korean, on the other hand, has a three-way contrast between tense, lax, and aspirated stops (e.g., p'ul (tense) 'horn', pul (lax) 'fire', p^hul (aspirated) 'grass'), which are all categorized as voiceless in the word-initial position. The three types of Korean stops (i.e., tense, lax, and aspirated stops) are phonetically characterized as having short lag (16.69ms), intermediate lag (70.82ms), and long lag VOT (80.08ms), respectively (Jeong 2010). Relative to lax and aspirated stops, tense stops have a shorter VOT. Although the VOT of tense stops is significantly different from that of the other two categories, there are no significant differences in VOT between the lax and aspirated stops (Silva 2006). In Korean, unlike English, F0 plays an important role in distinguishing between stop categories. Lax stops have a lower F0 than

aspirated and tense stops (Kim et al. 2002, Kim 2014, Jung and Kwon 2010, Kong et al. 2011, Lee and Jongman 2012, Kong and Lee 2018). Previous research has found that some dialects of Korean exhibit an emerging tonal system that aids the three-way stop contrast with the VOT overlap in word-initial position (Silva 2006, Kang and Guion 2008, Kang 2014). Kang (2014) indicated that this tonogenetic sound change was the case for younger female speakers of the Seoul dialect.

Phonetic cues such as VOT and F0 play a role in both Korean and English stop contrasts, but the degree to which these cues are used varies between these two languages. The purpose of this study was to investigate whether there were perceptual differences between native English and Korean listeners in terms of sensitivity to variability of VOT and F0 values of English stop consonants. Specifically, we examined the steps of VOT and F0 values where the two listener groups exhibited perceptual differences. Additionally, we explored the effects of place of articulation on the perception of voicing of word-initial stop consonants.

In order to address these issues, two groups of participants were asked to indicate which of the two images was represented by corresponding spoken stimuli. The stimuli were modified natural speech tokens that varied along six steps of the VOT continuum and six steps of the F0 continuum.

2. Literature review

2.1 Perception of English stop consonant voicing

While the majority of studies have found that perception of voicing in English word-initial stop consonants is mainly attributed to VOT, F0 also plays a role. Higher F0 onsets are characteristic of long VOT values and voiceless stops, whereas lower F0 onsets are related to short VOT values and voiced stops. F0 is not an important cue for stop consonants with canonical voiced and voiceless VOTs, but F0 exerts influence when VOTs are ambiguous and non-canonical (Abramson and Lisker 1985, Gordon et al. 1993, Whalen et al. 1993, Kingston and Diehl 1994, Francis et al. 2008, Winn et al. 2013). Gordon et al. (1993) and Francis et al. (2008) found individual differences in cue weighting in more demanding listening conditions. In their studies, listeners tended to rely on primary cues (such as VOT for the English stop voicing contrast) in ideal listening conditions, but changed their cue weighting toward redundant cues under less ideal conditions, such as listening to speech in noisy

environments, listening to multiple speakers, or listening to ambiguous speech lacking other resources. Whalen et al. (1993) found when F0 contour conflicted with VOT, reaction time was slower. The authors suggested that listeners were sensitive to F0 information even when identification curves were displayed differently.

Furthermore, previous studies (Kong and Edwards 2016, Kapnoula et al. 2017) indicated that gradient manner of perception was closely linked to listeners' greater sensitivities to secondary cues (e.g., F0 in English /d/-/t/ contrast). Kong and Edwards (2016) examined whether there were individual differences in how categorically listeners perceived speech sounds, and whether individual differences were related to listeners' sensitivity to phonetic details, by using an eye-tracking paradigm. They administered two perception tasks, visual analogue scaling (VAS) and anticipatory eye movement (AEM), to 39 adult English speakers, using a /ta/-/da/ continuum. The authors found that the listeners who had a gradient response pattern on the VAS task also demonstrated more sensitivity to F0 in the AEM task. They suggested that the extent to which listeners were able to categorically perceive speech sounds was consistent within individuals.

The effect of place of articulation on the voicing perception of stop consonants has also been investigated (Kuhl and Miller 1975, 1978; Miller 1977; Benki 2001; Nakai and Scobbie 2016). These studies found that bilabial stops were classified as [−voice] more often than velar stops, and alveolars were in the intermediate position between bilabials and velars. However, another set of studies (Lisker 1975, Kluender 1991) did not confirm any interaction between perception of the two phonological properties (i.e., voicing and place of articulation).

Nakai and Scobbie (2016) examined whether listeners shifted the VOT category boundary of word-initial stop consonants with a change in articulation rate in spontaneous English speech. The authors found that perceptual VOT category boundaries did not shift with a change in articulation rate under normal circumstances. The results also showed that perceptual VOT category boundaries between voiced and voiceless stops were 16ms for bilabials, 24ms for alveolars, and 27ms for velars. These findings are comparable to the mean VOT values of corresponding stop consonants reported in previous perception studies (Summerfield 1975, 1981; Miller 1977).

Previous studies have verified the main role of VOT in English stop voicing contrast. Also, it has been shown that use of acoustic cues such as VOT and F0 in the perception of English stop voicing contrast was affected by listening conditions.

Furthermore, the majority of research presented a correlation between voicing perception and place of articulation. However, previous research failed to find a consistent correlation between voicing perception and place of articulation.

2.2 Perception of stop consonant voicing in L2 research

Perceptual studies using synthesized stimuli found that adult L2 listeners identified stop consonants mostly along a VOT continuum according to their L1 stop inventories (Lisker and Abramson 1964, 1967). Schmidt (2007) also demonstrated that L1 English listeners assimilated word-initial Korean lax and aspirated stops to L1 aspirated stops, and Korean tense stops were assimilated to L1 unaspirated stops. That is, native English L2 Korean learners equated two Korean stop categories with a single English category. Vautour (2012) provided Korean word-medial stops to American English listeners who were not familiar with the Korean language. The English listeners perceived Korean word-medial lenis stops as homorganic English voiced stops and Korean word-medial aspirated stops as homorganic English voiceless stops. Korean word-medial tense stops were perceived mostly as voiced, though considerable variation was found across participants.

Another set of perception studies on Korean stops indicated important acoustic information about a three-way contrast. Kwon (2013) investigated English speakers' discrimination of a three-way laryngeal distinction of word-initial Korean stops /p, t, k/ using disyllabic minimal pairs. The results showed a relatively low correct discrimination level on the lax-tense contrast. These findings confirmed the link between L1 acoustic patterns and L2 perceptual discrimination. The author indicated that F0 is as important as VOT for L2 listeners to fully perceive the three-way Korean stop contrast. Kwon (2019) observed Korean stop consonants identification among three different language groups. Each of the language groups has a different VOT feature to distinguish their own native stop consonants. English and Chinese have short lag and long lag VOT. Both French and Russian possess lead and short lag VOT. Swedish and Turkish instead have lead and long lag VOT. The participants who had lead-short and lead-long groups in their L1 demonstrated difficulty in perceiving tense and lenis, and the short-long group had trouble perceiving lenis compared to tense and aspirated. There seemed to be more explanatory power with VOT types in interpreting perceiving patterns of Korean stop consonants compared to that of phonological types.

Some previous studies have investigated Korean listeners' production and perception of English word-initial stops (Lim and Han 2014, Son 2017). Lim and Han (2014) explored whether the relative importance of specific acoustic properties in L1 dialects (Kyungsang and Seoul Korean) affects L2 production and perception. They found that in production, both dialect speakers showed a similar pattern for VOT and F0. However, in perception, Kyungsang listeners had greater reliance on VOT but less on F0 compared with Seoul listeners. The authors argued that Kyungsang listeners primarily employed VOT due to the use of F0 for lexical tone contrast in their dialect. Son (2017) examined how Korean speakers articulated and perceived English stops using their native acoustic cues, VOT and F0. She suggested that to the native Korean speakers, VOT played a primary role in voicing contrast in English while articulatory assimilation occurred in the F0 dimension. In addition, the native Korean speakers utilized F0 as a significant acoustic cue for phonetic distinction for English voicing, despite the fact that the native English speakers did not adopt the cue significantly.

A number of previous L2 studies on acoustic cues such as VOT and F0 have focused on analyses of L2 speakers' production and category identification. To our knowledge, however, no study has directly compared sensitivity to gradient acoustic cues between English and Korean listeners on English words, and it is the goal of this study to fill this gap in the current literature. More specifically, we compared contribution of various steps of VOT and F0 values in the perception of English word-initial stop voicing between native English listeners and native Korean listeners. With this goal, we also examined the effect of place of articulation on voicing perception.

3. Methods

3.1 Participants

For the native Korean listener group, twenty adults (seven male and thirteen female) were recruited from a university in Seoul. All of them were undergraduate or graduate students. The participants' ages ranged from 20 to 38 years old (the average being 28 years old). Their English proficiency levels were considered to be low-intermediate or low, based on a self-report and a short voice recording of English sentences (see the paragraph for voice recording in Appendix A). Their majors were

not related to English, and none of them had stayed in an English-speaking country for more than six months.

The native English listener group was comprised of nineteen adults (eleven male and eight female), who were a mixture of undergraduate and graduate students, professors, and English instructors. Their ages ranged from 21 to 52 years old (the average of which was 37 years old). The participants were from the United States, Australia, Canada, and the United Kingdom. The subjects were all paid for their participation in the identification experiment, and none reported any hearing problems.

3.2 Stimuli

The stimuli were pseudo-synthetic CVC(C) syllables that were created from six minimal pairs containing a stop in the initial position followed by a /a/ or /o/ vowel. The target stops were bilabial, alveolar, or velar (i.e., palm-bomb, pole-bowl, tart-dart, toe-dough, card-guard, coat-goat). The target words were spoken in a carrier sentence 'Look at the _____' by a male native speaker of American English three times. The speaker was from California and in his twenties. The stimuli were recorded using an Olympus LS-P4 voice recorder in a soundproof studio. The second or third one was used among the three tokens.

The stimuli were constructed by manipulating VOT in 6 steps and F0 in 6 steps for /b/-/p/, /d/-/t/, and /g/-/k/, respectively, using Praat 6.1.07 (Boersma and Weenink 2019). The VOT (in milliseconds) of word-initial stops was measured as the time from the onset of the noise burst to the onset of the following vowel.

Following the method adopted in previous studies (Andruski et al. 1994, Kong and Yoon 2013, Winn et al. 2013, Kong and Edwards 2016), short onset VOT portions of /b/, /d/, and /g/ were progressively replaced with long portions of onset VOT from /p/, /t/, and /k/, respectively. The vowel from each item was from the /b/, /d/, or /g/-initial tokens. The range of VOT manipulation was set in each place of articulation, so the VOT ranges were different in the three places of articulation. For the /b/-/p/ continuum, the VOT range spanned from 10ms (original VOT of balm) to 70ms (original VOT of palm) in six steps (i.e., 10ms, 22ms, 34ms, 46ms, 58ms, 70ms). For the /d/-/t/ continuum the VOT range was from 8ms to 101ms (i.e., 8ms, 26ms, 44ms, 63ms, 82ms, 101ms), and the VOT range for the /g/-/k/ continuum spanned from

15ms to 87ms (i.e., 15ms, 29ms, 43ms, 58ms, 72ms, 87ms). The interval between two steps was 11–19ms for all three pairs (i.e., /b/-/p/ /d/-/t/ /g/-/k/).

At each VOT step, the original F0 value during the vowel was replaced with one of six different sustained F0 values (i.e., 120Hz, 130Hz, 140Hz, 150Hz, 160Hz, and 170Hz). The F0 contour of each token was manipulated using the pitch synchronous overlap-add method (PSOLA) in Praat. The pitch tiers were extracted and stylized by reducing the number of dots. After manipulating pitch tiers, the resulting pitch tier object and the other manipulation object were selected together and resynthesized.

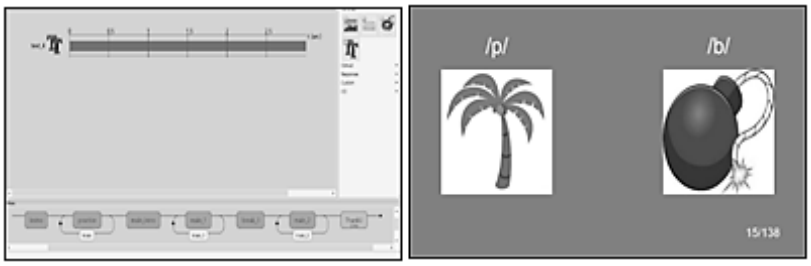
Thus, 36 target tokens (a 6-step VOT x 6-step F0 continuum) were created. Each item was created by preposing a phrase 'look at the' to the CVC(C) target tokens, and the phrase was 510ms. The fully-crossed design of place of articulation, VOT, F0, and vowel contexts (i.e., /a/, /o/) yielded 216 tokens (3 places × 6 VOT steps × 6 F0 steps × 2 vowels). In addition, 60 filler items were included in the stimuli. Thus, in total, there were 276 items (216 target items including /p/-/b/, /t/-/d/, and /k/-/g/ continua, and 60 filler items including /s/-/ʃ/, /l/-/r/, and /m/-/h/ pairs). All the items in the stimuli were presented in random order.

3.3 Procedure

In order to test the effects of VOT, F0, and place of articulation on voicing classification, picture identification data were collected using the PsychoPy program ver. 3.1.1 (Peirce 2007). The participants were tested with 276 items split into two blocks. The Stimuli were presented binaurally over headphones, and the participants adjusted the volume to a comfortable level. The stimulus items were randomly presented to each participant.

Before the main experiment began, they were given the list of images used in the experiment with related words and auditory instructions for the experiment. Then they had a training session to familiarize them with the task. For all items including ten training items, an auditory stimulus item was presented with two images on a computer screen. The participants were asked to choose which of the two images matched with the auditory stimuli by pressing one of the designated buttons on the keyboard. One image was accompanied by a voiceless sound symbol (e.g., /p/, /t/, or /k/), and the other with a voiced sound symbol counterpart (e.g., /b/, /d/, or /g/). The image examples used in the experiment are shown in Appendix B. Figure 1 manifests

(a) the PsychoPy builder screen and (b) the screen of the experiment prompted by PsychoPy.



(a) **(b)**

Figure 1. (a) PsychoPy builder screen, (b) a screenshot of the experiment (palm vs. bomb)

4. Results

The results are first presented using grayscale intensity grids. Averaged group responses of [−voice] identification to the continua of both VOT and F0 are given in the tiled grids in three places of articulation separately in Figure 2. Next, the results are presented by plots of [−voice] identification in Figure 3. In addition, the logistic regression analysis results regarding sensitivity differences between the two listener groups are presented in Table 1 and Table 2. The statistical analyses of [−voice] identification for each step F0 values are also presented in Table 3.

The following grids display how the details of acoustic cues like VOT and F0 affected listeners' perception of voicing category. Grids with rows presenting differing grayscale intensity relate to listeners who utilized VOT cues to distinguish between voiceless and voiced stops. On the other hand, grids with columns showing varied grayscale intensity represent listeners who used F0 cues to differentiate between the two stop categories. Each grid illustrates the use of both cues in varying proportions. The darker the color is, the more voiceless stops were selected by the listeners.

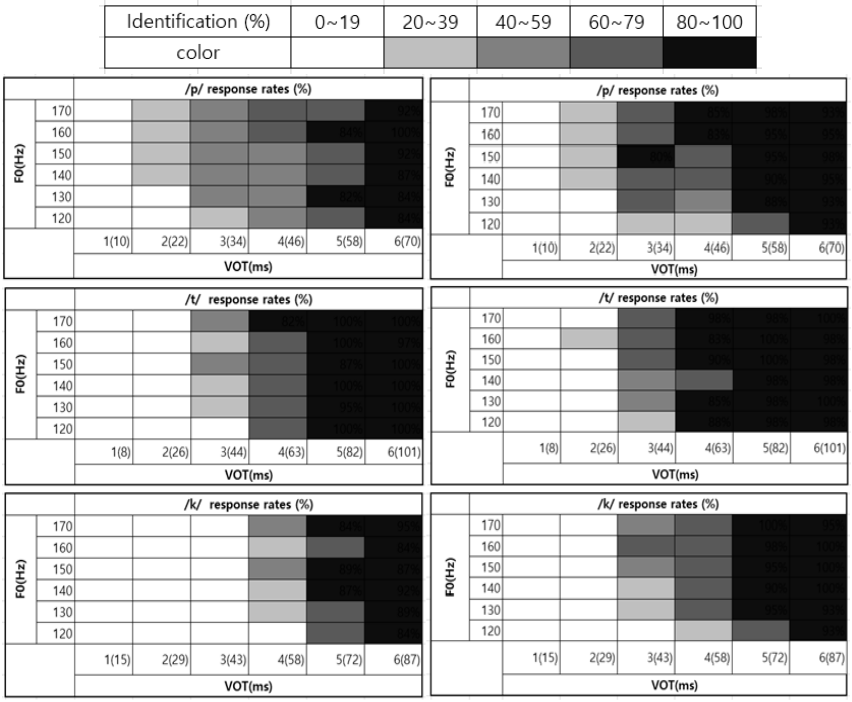


Figure 2. Tiled grids showing the response rates of [−voice] responses (/p/, /t/, and /k/) at each step of the VOT and F0 continua by native English listeners (left side) and Korean listeners (right side)

The grids in Figure 2 demonstrate that the two listener groups utilized both VOT and F0 cues to distinguish between voiced and voiceless stops. Comparing perceptual patterns between the two listener groups, the English listeners identified [−voice] of stop consonants with longer VOT values than the Korean listeners in all places of articulation. Also, in terms of F0 cues, the English listeners' responses of [−voice] were shown with higher F0 values than the Korean listeners.

Furthermore, the Korean listeners showed more gradient patterns than the English listeners in relation to F0 cues. For example, at the third VOT step of alveolar stops, the Korean listeners' response rates of [−voice] increased as F0 values became higher. However, the gradual patterns depending on F0 values were not clearly exhibited by the English listeners. The same tendency was shown at the third VOT step of velar

stops. The English listeners' response rates of [−voice] were less than 20%, even at the highest F0 value.

When it comes to places of articulation, both the Korean and English listeners showed the response of [−voice] with higher values of VOT in velar stops than in bilabial or alveolar stops. That is, bilabial and alveolar stops were more likely to be identified as voiceless than velar stops. In addition, bilabial stops induced more voiceless responses than alveolar stops for both listener groups.

Averaged group responses are displayed with the continua of VOT and F0 separately in Figure 3. On the left side, the results are presented by plots of [−voice] responses as a function of synthesizer VOT at the second step (130Hz) of F0. The second step of F0 was chosen based on F0 normalization. Also, in previous studies (Takefuta et al. 1972, Graddol 1986, Johns-Lewis 1986), native male English speakers' average F0 in a reading task was 130Hz.

On the right side, the results of [−voice] responses as a function of synthesized F0 continuum with the third step of VOT (34ms for bilabials, 44ms for alveolars, and 43ms for velars) are displayed. The third step of VOT was selected since the values at this step were close to the boundaries between voiced and voiceless stops presented in previous studies (Lieberman and Blumstein 1988, Nakai and Scobbie 2016). The solid lines represent the proportions of voiceless sound selection by the English listeners, while the dotted lines illustrate the Korean listeners' response proportions.

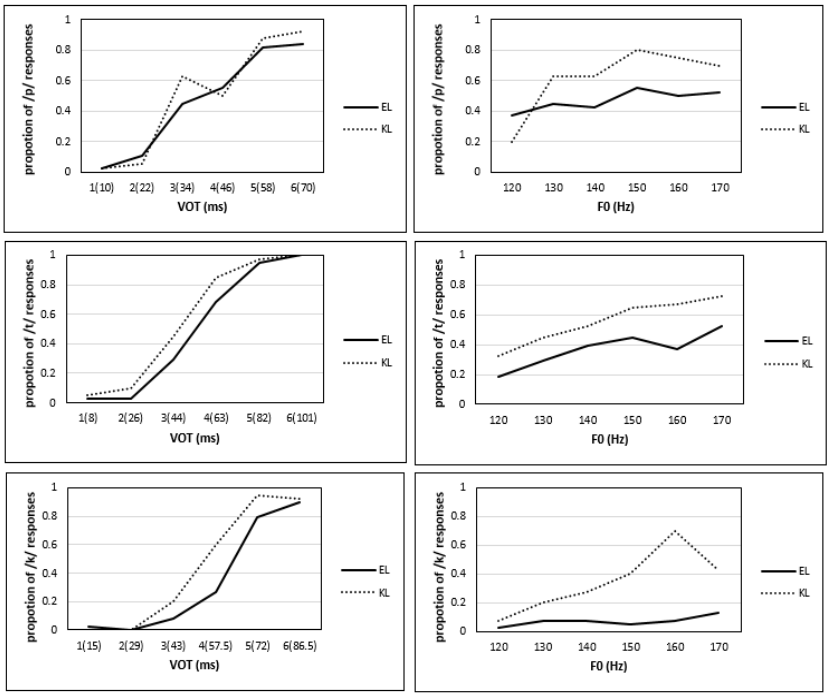


Figure 3. Proportions of voiceless stop (/p/, /t/, or /k/) responses as a function of VOT at the second step (130Hz) of F0 (left side), and as a function of F0 at the third step of VOT (right side) by English listeners (EL, solid lines) and Korean listeners (KL, dotted lines)

As shown on the left side in Figure 3, the response proportions of [−voice] responses rose as the VOT values increased for both English and Korean listeners. This means both groups utilized VOT as a major acoustic cue to identify voiceless stop categories. Some differences, however, emerged across different places of articulation. Particularly, the Korean listeners produced slightly more [−voice] responses than the English listeners. This perceptual difference was more clearly shown for alveolar and velar stops than bilabial stops. Also, when the response proportions in the three places of articulation were compared, bilabials and alveolars were categorized as more [−voice] sounds at the earlier VOT steps than velars in the corresponding steps. For both listener groups, more than half of the responses were [−voice] at around 34-46ms of VOT in bilabial and alveolar stops. On the other hand,

velars were identified as [−voice] at much higher VOT values than bilabials and alveolars. For velar stops, approximately half of the Korean listeners' responses were [−voice] at 58ms or higher VOT values while the native English listeners' response proportions of [−voice] at around 58ms VOT were lower than 30%.

In examining sensitivity to F0 in voicing identification, the response proportions of [−voice] rose as the F0 values increased for both the English and Korean listeners for bilabial and alveolar stops. However, for velar stops the English listeners did not produce the similar rising curves. Even at the highest F0 values (160~170Hz), the native English listeners' response proportion was less than 20%. In other words, the English listeners were not sensitive to F0 changes for velar stops at all.

Comparing response curves between VOT and F0, the VOT curves presented more abrupt changes than the F0 curves for both the English and Korean listeners. Although [−voice] stop category responses were associated with longer VOT and higher F0 values, the listeners' gradient sensitivity to the two acoustic cues was different.

In order to examine how the two different phonetic cues (i.e., VOT, F0) affected the voicing identification of stop consonants in the three places of articulation separately, we carried out a series of generalized logistic mixed-effects models (GLMMs). First, we compared the responses of the two groups with the items varying with VOT in six steps at the second step of F0 (130Hz). For each place of articulation separately, a mixed effects logistic regression analysis with the option of the robust covariance matrix estimation was performed. The raw binary participant responses (1= [+voice], 2= [−voice]) were modeled, with group (English listeners, and Korean listeners) and VOT (6 levels) as the fixed effects, and subjects as random effects. The following table summarizes the statistical results.

Table 1. Summary of listener group and VOT effects

POA (= place of articulation)	Group			VOT			Group × VOT		
	df1, df2	*f*	*p*	*df1, df2*	*f*	*p*	*df1, df2*	*f*	*p*
Bilabial	1, 456	.397	.529	5, 456	25.796	.0001	5, 456	.566	.726
Alveolar	1, 456	2.004	.158	5, 456	29.239	.0001	5, 456	.150	.980
Velar	1, 456	.001	.996	5, 456	18.723	.0001	5, 456	.358	.877

The results showed that the effect of VOT was significant in all three places of articulation ($p < .001$ for all three places of articulation). However, neither a significant group effect nor a significant group by VOT interactions were found in any of the analysis. These results could suggest that while both groups displayed comparable levels of sensitivities to VOT cues in voicing perception of stop consonants, the patterns across different degrees of VOT levels also formed similar patterns. In other words, although the English listeners identified [−voice] at higher VOT values than the Korean listeners, the sensitivity patterns regarding VOT cues were not significantly different between the two listener groups.

Second, we compared the responses of the two groups with the items varying with F0 levels (6 steps) at the third step of VOT. For each place of articulation, the same mixed effects logistic regression analysis was performed, with group and F0 as the fixed effects and subject as random effects. The following table summarizes the statistical results.

Table 2. Summary of listener group and F0 effects

POA	Group			F0			Group × F0		
	df1, *df2*	*f*	*p*	*df1*, *df2*	*f*	*p*	*df1*, *df2*	*f*	*p*
Bilabial	1, 456	2.473	.117	5, 456	6.749	.0001	5, 456	2.115	.063
Alveolar	1, 456	5.880	.016	5, 456	6.110	.0001	5, 456	.357	.878
Velar	1, 456	15.735	.000	5, 456	3.459	.004	5, 456	1.753	.121

As shown in Table 2, the effect of F0 was also statistically significant in all three places of articulation. Thus, F0 was used as an important acoustic cue in detecting [−voice] of stop consonants in all places of articulation. However, the effect of language group varied across places of articulation. There was no effect of listener group regarding the bilabial stops, whereas the responses to the alveolar and velar stops yielded significantly different results between the two groups. The perceptual differences were more clearly shown with velar stops than alveolar stops ($p = .016$ for alveolar, $p < .001$ for velar stops). As shown in the plots in Figure 3, the English listeners were less sensitive to F0 increase in perceiving [−voice] of alveolar and velar stops.

In order to get a clearer picture of perceptual differences that were found significant between the two groups, we carried out a series of post-hoc comparisons on each steps of F0 separately. The summary of the analysis is provided in Table 3.

Table 3. Summary of comparisons between the two listener groups for each step of F0

F0 Step (Hz)	Bilabial		Alveolar		Velar	
	f	*p*	*f*	*p*	*f*	*p*
1(120)	0.762	0.383	0.662	0.416	0.34	0.563
2(130)	1.739	0.188	1.165	0.281	0.629	0.428
3(140)	2.449	0.118	1.030	0.311	0.971	0.325
4(150)	1.371	0.242	2.635	0.105	1.634	0.202
5(160)	2.034	0.155	6.718	0.010**	11.648	0.001***
6(170)	1.227	0.269	1.833	0.176	1.999	0.158

'***' ≤ .001, '**' ≤. 01

The Korean listeners, compared to the English listeners, displayed greater sensitivity to F0 cues in most cases (see Figure 3). However, as shown in Table 3, the follow-up analysis revealed that such higher sensitivity patterns were statistically significant on the two occasions only, at the fifth step (160 Hz) of alveolar (p = .010) and velar (p = .001) stops, respectively.

Taken together, the descriptive and statistical analysis reported above demonstrate that while the two groups patterned similarly to one another in the use of VOT cues, they presented some perceptual differences in their use of F0 cues. Sensitivity differences, however, were found only at the fifth step of alveolar and velar stops. There were no differences between the two listener groups for F0 values at any steps of bilabial stops.

5. Discussion

5.1 Differences of sensitivity to VOT and F0 cues between English and Korean listeners

The current study compared gradient sensitivity to acoustic cues in identifying [−voice] of word-initial stop consonants between native English and Korean listeners. The results showed that response patterns in the [−voice] judgment of stop consonants were similar between the English and Korean listeners in that group-averaged response curves rose as a function of increase in the VOT and F0 values. Also, their response curves switching from [+voice] to [−voice] were more abrupt along a VOT dimension than along a F0 dimension. However, the two listener groups also revealed some different perceptual sensitivity to two acoustic cues.

First, in terms of VOT, the English listeners' [−voice] judgments were made at later steps of VOT than those by the Korean listeners. Specifically, the mean English listeners' [−voice] response that exceeded 50 % emerged from the fourth step of VOT continua, whereas the Korean listeners displayed such trend from the third VOT step. These results were not consistent with those of previous research where the boundary between voiced and voiceless stop consonants was around 30ms (Lisker and Abramson 1964, Lieberman and Blumstein 1988, Nakai and Scobbie 2016). Note that in contrast to previous studies that used natural speech, the present study made use of manipulated speech tokens to create gradient acoustic cues. Taking this into account, some properties of the manipulated speech may have affected the English listeners' voicing perception. The English listeners may have required more solid evidence such as longer VOT in order to determine the [−voice] category resulting in some more delays relative to the Korean listeners. In addition, the discrepancy of [−voice] judgment between the two listener groups was more obviously shown for alveolar or velar stops than for bilabial stops.

Second, the two listener groups' perceptual asymmetry was relatively more apparent in respect to F0. As shown in Figures 2 and 3 in the result section, the Korean listeners were more sensitive to differences in F0 cues, and their reliance on F0 was relatively greater in identifying [−voice] than the English listeners. Such different levels of sensitivities to F0 cues were more significant on alveolar and velar stops. These results are in line with those of previous studies (Haggard et al. 1970, Lieberman and Blumstein 1988, Kingston and Diehl 1994, Nakai and Scobbie 2016).

It has been known that in English VOT is a primary cue, and F0 is a secondary cue. F0 is higher for voiceless than for voiced stops, but F0 values between these two categories are not significantly different. Conversely, F0 cues play an important role in Korean for distinguishing between lax stops and aspirated stops (Kim 2002, Silva 2006, Jeong 2010, Kong et al. 2011, Lee and Jongman 2012, Kim 2014, Kong and Lee 2018). These acoustic properties were reflected in the perceptual patterns in the present study. Also, the current results partially support the view from previous research. Kim (2012) and Kong and Yoon (2013) pointed out that Korean L2 learners of English, in general, consistently differentiated English voiced stops from voiceless stops in the F0 dimension, which is the primary acoustic dimension in their L1 perception. Also, Kong and Yoon (2013) indicated that more proficient L2 learners were able to hinder the unimportant acoustic cue in L2 (i.e., F0) in identifying L2 voicing contrast. In the present study, the Korean listeners with low-intermediate or low English proficiency level judged voicing contrast based on both VOT and F0. In particular, it was clear that they utilized F0 cues consistently when processing the English stops, arguably due to their application of a language-specific use of acoustic cues available in the L1. Thus, the present results verified a language-specific use of acoustic cues by L2 listeners.

Third, the results of the present study showed that perceptual disparities were manifested to a certain step of F0 cues. That is, although the English listeners were less sensitive, by and large, to F0 cues than the Korean listeners, statistically significant discrepancies between the two listener groups were revealed only at the fifth step of alveolar and velar stops. At the fifth step of F0 the Korean listeners' identification of [−voice] for alveolar or velar stops were around 70%, whereas the English listeners' [−voice] identification was 37% for alveolar stops and 8% for velar stops. The highest step of F0 (170Hz) did not induce as high rates of [−voice] identification as the fifth step for the Korean listeners. The manipulated tokens with 170Hz of F0 sounded somewhat unnatural and might reduce the [−voice] identification rates. If the natural stimuli are used in the experiment, perceptual discrepancies between the two listener groups will be more clearly revealed with respect to F0 values.

5.2 Effects of place of articulation

The present study found different perceptual patterns of stop voicing contrast depending on the place of articulation. As shown in Figures 2 and 3, bilabial and alveolar stops were identified as the [−voice] category at the earlier VOT steps than velar stops. These perceptual patterns were clearly shown by both native English listeners and Korean listeners. As displayed in Figure 3, the English listeners judged velar stops as [−voice] at the third step of VOT in less than 20% of the responses throughout all F0 steps. In addition, the Korean listeners identified [−voice] of velar stops at the second step of VOT in less than 20% of the responses in all F0 steps. These results for velar stop were different from those for bilabial and alveolar stops.

These results associated with place of articulation were partially consistent with those of previous studies. Kuhl and Miller (1975, 1978), Miller (1977), Nakai and Scobbie (2016) found that English bilabial stops were classified as [−voice] more often than velar stops by native English listeners, and alveolars were in the intermediate position between bilabials and velars. The current results, however, did not confirm the perceptual VOT differences between bilabial and alveolar stops. Furthermore, according to Jang (2012), Korean velar stops' VOT values were 23ms for tense, 77ms for lax, and 89ms for aspirated stops. These VOT values were 10ms or longer than those of bilabial and alveolar stops. It seems that these Korean stops' acoustic properties affected the Korean listeners' [−voice] judgment of English stops in three places of articulation.

Moreover, the aerodynamic or articulatory-based explanations suggested that the smaller air cavity behind more posterior constrictions and the larger cavity in front of the constriction suspended the initiation of vocal fold vibration. Thus, stop consonants that had the constriction at the posterior places such as velars had a shorter duration with longer VOTs (Lisker and Abramson 1967, Goldstein and Browman 1986, Ladefoged and Cho 2001). Therefore, the present results regarding the interaction between voicing identification and place of articulation support previous perception and production studies.

6. Conclusion

The purpose of the present study was to compare a sensitivity to VOT and F0 cues in the perception of English stop voicing contrast between native English and Korean

listeners. The steps of VOT and F0 values where the two listener groups showed perceptual discrepancies were also examined. In addition, the effects of place of articulation on the perception of voicing in word-initial stop consonants were explored. To these ends, two listener groups listened to modified natural speech tokens that varied along six steps of VOT continuum and six steps of F0 continuum. The results showed that while the Korean listeners utilized both VOT and F0 cues, the English listeners judged the voicing contrast mainly based on the VOT cue. The Korean listeners were significantly more sensitive to the gradual change of F0 than the English listeners. The English listeners' identification of the stop's voicing specification in synthesized speech was not aided by F0 cues, but mainly based on VOT cues.

Moreover, the two listener groups revealed a different perceptual pattern in the use of VOT cue. The English listeners' [−voice] judgments were made at later steps of VOT than the Korean listeners. Also, the further examination of each step of F0 revealed that perceptual discrepancies between the two listener groups were revealed at the fifth step (170 Hz) of alveolar and velar stops.

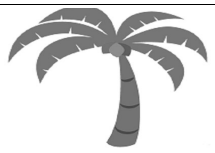
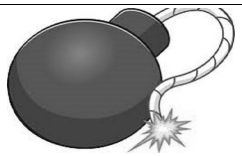
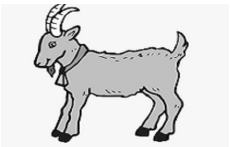
The results also suggested that there were different levels of listener sensitivity across different places of articulation. Bilabial and alveolar stops were identified as [−voice] at earlier steps of VOT than velar stops by both listener groups. Also, the asymmetry of gradual sensitivity to F0 between the two listener groups was more clearly shown for velar stops than for alveolar stops, and for bilabial stops no significant differences were found between the two groups. These results indicated that the L2 listeners utilized acoustic cues in a language-specific way, and their perception strategies are based on L1.

The experimental results reported here have shed some additional light on the perception of English stop consonants by L1 and L2 listeners. In order to investigate more concrete characteristics of L2 listeners' stop perception, other phonetic properties such as closure duration, and F1 and H1-H2 (dB) at the vowel onset need to be considered. In order to further explore the interaction of L2 proficiency and cue-weighting change between VOT and F0, native Korean listeners with multiple proficiency levels of English need to be included in perceptual experiments. Moreover, future research may be warranted with substantially larger samples of listeners to better understand cross-linguistic perceptual differences involving voicing contrast.

Appendix A. A diagnostic paragraph (Celce-Murcia et al. 2010: 481)

Is English your native language? If not, your foreign accent may show people that you come from another country. Why is it difficult to speak a foreign language without an accent? There are a couple of answers to this question. First, age is an important factor in learning to pronounce. Another factor that influences your pronunciation is your first language. You also need accurate information about English sounds. Will you make progress, or will you give up? It's your decision. You can improve!

Appendix B. Target item examples

	
palm	**bomb**
	
tart	**dart**
	
coat	**goat**

REFERENCES

ABRAMSON, ARTHUR S. and LEIGH LISKER. 1985. Relative power of cues: F0 shift versus voice timing. In Victoria A. Fromkin (ed.). *Phonetic Linguistics: Essays in Honor of Peter Ladefoged*, 25-33. San Diego: Academic Press.

ANDRUSKI, JEAN E., SHEILA E. BLUMSTEIN and MARTHA BURTON. 1994. The effect of subphonetic differences on lexical access. *Cognition* 52, 163-187.

BENKI, JOSE R. 2001. Place of articulation and first formant transition pattern both affect perception of voicing in English. *Journal of Phonetic* 29.1, 1-22.

BOERSMA, PAUL and DAVID WEENINK. 2019. Praat: doing phonetics by computer (Version 6.1.07) [Computer program]. Retrieved December 12, 2019 from https://www.fon.hum.uva.nl/praat/.

CELCE-MURCIA, MARIANNE, DONNA M. BRINTON, JANET M. GOODWIN and BARRY GRINER. 2010. *Teaching Pronunciation: A course book and reference guide.* Cambridge: Cambridge University Press.

FRANCIS, ALEXANDER L., NATALYA KAGANOVICH and COURTNEY DRISCOLL-HUBER. 2008. Cue-specific effects of categorization training on the relative weighting of acoustic cues to consonant voicing in English. *The Journal of the Acoustical Society of America* 124.2, 1234-1251.

GOLDSTEIN, LOUIS and CATHERINE P. BROWMAN. 1986. Representation of voicing contrasts using articulatory gestures. *Journal of Phonetics* 14.2, 339-342.

GORDON, PETER C., JENNFER L. EBERHARDT, and JAY G. RUECKL. 1993. Attentional modulation of the phonetic significance of acoustic cues. *Cognitive Psychology* 25.1, 1-42.

GRADDOL, DAVID. 1986. Discourse specific pitch behavior. In Catherine Johns-Lewis (ed.). *Intonation in Discourse*, 221-237. London: Croom Helm.

HAGGARD, MARK., STEPHEN AMBLER and MO CALLOW. 1970. Pitch as a voicing cue. *The Journal of the Acoustical Society of America* 47.2, 613-617.

JANG, HYEJIN. 2012. *Acoustic Properties and Perceptual Cues of Korean Word-initial Obstruents*. PhD Dissertation. Korea University.

JEONG, YUN JA. 2010. Speech perception of Korean plain, aspirated and tense considering pitch of the following vowel. *Urimaryeongu* 27, 73-94. Urimalhakoe.

JOHNS-LEWIS, CATHERINE. 1986. Prosodic differentiation of discourse modes. In Catherine Johns-Lewis (ed.). *Intonation in Discourse*, 199-219. London: Croom Helm.

JUNG, MIJI and SUNGMI KWON. 2010. A study on the phonetic discrimination and acquisition ability of Korean language learners. *Phonetics and Speech Sciences* 2.1, 23-32. The Korean Society of Speech Sciences.

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.

KANG, YOONJUNG and SUSAN G. GUION. 2008. Clear speech production of Korean stops: changing phonetic targets and enhancement strategies. *The Journal of the Acoustical Society of America* 124, 3909-3917.

KAPNOULA, EFTHYMIA C., MATTHEW B. WINN, EUN JONG KONG, JAN EDWARDS and BOB MCMURRAY. 2017. Evaluating the sources and function of gradiency in phoneme categorization: An individual differences approach. *Journal of Experimental Psychology: Human Perception and Performance* 43.9, 1594-1611.

KIM, MI-RYOUNG. 2012. L1-L2 transfer in VOT and f0 production by Korean English listeners: L1 sound change and L2 stop production. *Phonetics and Speech Sciences* 4.3, 81-90. The Korean Society of Speech Sciences.

_____. 2014. Ongoing sound change in the stop system of Korean: A three- to two-way categorization. *Studies in Phonetics, Phonology and Morphology* 20.1, 51-82. The Phonology-Morphology Circle of Korea.

KIM, MI-RYOUNG, PATRICE SPEETER BEDDOR and JULIE HORROCKS. 2002. The contribution of consonantal and vocalic information to the perception of Korean initial stops. *Journal of Phonetics* 30.1, 77-100.

KINGSTON, JOHN and RANDY L. DIEHL. 1994. Phonetic knowledge. *Language* 70.3, 419-454.

KLUENDER, KEIT R. 1991. Effects of first formant onset properties on voicing judgments result from processes not specific to humans. *The Journal of the Acoustical Society of America* 90.1, 83-96.

KONG, EUN JONG, MARY E. BECKMAN and JAN EDWARDS. 2011. Why are Korean tense stops acquired so early: The role of acoustic properties. *Journal of Phonetics* 39.2, 196-211.

KONG, EUN JONG and JAN EDWARDS. 2016. Individual differences in categorical perception of speech: Cue weighting and executive function. *Journal of Phonetics* 59, 40-57.

KONG, EUN JONG and HYUNJUNG LEE. 2018. Attentional Modulation and individual differences in explaining the changing role of fundamental frequency in Korean laryngeal stop perception. *Language and Speech* 61.3, 384-408.

KONG, EUN JONG and IN HEE YOON. 2013. L2 proficiency effect of 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. The Korean Society of Speech Sciences.

KUHL, PATRICIA K. and JAMES. D. MILLER. 1975. Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants. *Science* 190.4209, 69-72.

_____. 1978. Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli. *The Journal of the Acoustical Society of America* 63.3, 905-917.

KWON, NAHYUN. 2013. Acoustic observation for English speakers' perception of a three-way laryngeal contrast of Korean stops. In Lauren Gawne and Jill Vaughan (eds.). *Selected Papers from the 44th Conference of the Australian Linguistic Society*, 58-76.

KWON, SUNGMI. 2019. A Study on the Perception of Korean Plosives by Phonological and Phonetic Types. *Korean Education* 118, 223-254. The Association of Korean Education.

LADEFOGED, PETER and TAE HONG CHO. 2001. Linking linguistic contrasts to reality: The case of VOT. *UCLA Working Papers in Phonetics*, 1-9.

LADEFOGED, PETER and IAN MADDIESON. 1996. *The Sounds of the World's Languages*. Oxford: Blackwell.

LEE, HYUNJUNG and ALLARD JONGMAN. 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.

LIEBERMAN, PHILIP and SHEILA E. BLUMSTEIN. 1988. *Speech Physiology, Speech Perception, and Acoustic Phonetics.* Cambridge: Cambridge University Press.

LIM, SOOA and JEONG-IM HAN. 2014. Effects of dialectal differences in the use of native-language acoustic cues on the production and perception of second

language stops. *Studies in Phonetic, Phonology and Morphology* 20.3, 403-426. The Phonology-Morphology Circle of Korea.

LISKER, LEIGH. 1975. Is it VOT or a first formant transition detector? *The Journal of the Acoustical Society of America* 57, 1547-1551.

LISKER, LEIGH and ARTHUR S. ABRAMSON. 1964. A cross-language study of voicing in initial stops: Acoustical measurements. *WORD* 20.3, 384-422.

_____. 1967. Some effects of context on voice onset time in English stops. *Language and Speech* 10, 1-28.

MILLER, JOANNE L. 1977. Properties of feature detectors for VOT: The voiceless channel of analysis. *The Journal of the Acoustical Society of America* 62.3, 641-648.

NAKAI, SATSUKI and JAMES SCOBBIE. 2016. The VOT Category Boundary in Word-initial Stops: Counter-Evidence Against Rate Normalization in English Spontaneous Speech. *Laboratory Phonology* 7.1, 13.

PEIRCE, JONATHAN W. 2007. PsychoPy - Psychophysics software in Python. *Journal of Neuroscience Methods* 162.1-2, 8-13.

SCHMIDT, ANNA MARIE. 2007. Cross-language consonant identification: English and Korean. In Ocke-Schwen Bohn and Murray J. Munro (eds.). *Language Experience in Second language Speech Learning: In Honor of James Emil Flege*, 185-200. Amsterdam, the Netherlands: John Benjamins.

SILVA, DAVID J. 2006. Acoustic evidence for the emergence of tonal contrast in contemporary Korean. *Phonology* 23.2, 287-308.

SON, GAYEON. 2017. Phonemic categorization of English stops in the VOT-F0 dimensions for Korean stop contrast. *Language Research* 33.3, 363-389. The Modern Linguistic Society of Korea.

SUMMERFIELD, QUENTIN. 1975. Acoustic and phonetic components of the influence of voice changes and identification times for CVC syllables. *Report on Research in Progress in Speech Perception* 2, 73-98.

_____. 1981. On articulatory rate and perceptual constancy in phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance* 7, 1074-1095.

TAKEFUTA, YUKIKO, ELIZABETH JANCOSEK, MICHAEL BRUNT ANDRE RIGAULT and RENE CHABONNEAU. 1972. A statistical analysis of melody curves in the intonation of American English. *Proceedings of the 7th International Congress of Phonetic Sciences*, 1035-1039. IPA Montreal, Canada

VAUTOUR, DOUGLAS. 2012. *English Speakers' Perception of Korean Stops*. MA Thesis. Seoul National Univeristy.

WHALEN, DOUGLAS H., ARTHUR S. ABRAMSON, LEIGH LISKER and MARIA MODY. 1993. F0 gives voicing information even with unambiguous voice onset times. *The Journal of the Acoustical Society of America* 93.4, 36-49.

WINN, MATTHEW B., MONITA CHATTERJEE and WILLIAM J. IDSARDI. 2013. The roles of voice onset time and F0 in stop consonant voicing perception: Effects of masking noise and low-pass filtering. *Journal of Speech, Language, and Hearing Research* 56.4, 1097-1107.

Eunkyung Sung (Professor)
Department of English
Cyber Hankuk University of Foreign Studies
107 Imun-ro, Dongdaemun-gu
Seoul 02450, Republic of Korea
e-mail: eks@cufs.ac.kr

Sunhee Lee (Associate Professor)
Department of Chinese
Cyber Hankuk University of Foreign Studies
107 Imun-ro, Dongdaemun-gu
Seoul 02450, Republic of Korea
e-mail: lishanxi@cufs.ac.kr

Sehoon Jung (Lecturer)
Department of English Language and Literature
Kyungsung University
309 Sooyoung-ro, Nam-gu
Busan 48434, Republic of Korea
e-mail: sejung@ks.ac.kr

Received: November 23, 2020
Revised: December 19, 2020
Accepted: December 24, 2020