The current study investigated how tonal and non-tonal dialect speakers of Korean perceived the voicing contrast of English stops across different prosodic boundaries and accent conditions. Voice onset time and fundamental frequency of target segment were systematically manipulated across IP and Wd levels under accented and unaccented conditions. Both non-tonal Seoul and tonal Kyungsang listeners required a longer VOT for the perception of English voiceless stop in the IP level than in the Wd level. The ambiguous VOT range was also found to be different between the IP and Wd levels and a relatively higher onset F0 was required in the perception of voiceless stop for the stimuli in the ambiguous VOT range. The results from logistic regression analyses revealed that for both dialect groups, the effect of position was found to be significant, while the effect of accent was not found to be significant. The coefficient value of the VOT was greater than that of onset F0 for both dialect groups, meaning that when distinguishing the voicing contrast of English stop across different prosodic conditions, both dialect listeners used VOT as a primary and F0 as a secondary perceptual cue. However, the two dialect groups showed differences in the weight of each VOT and onset F0: Kyungsang listeners showed greater coefficient values of VOT and onset F0 than Seoul listeners. When the different prosodic conditions were given in the perception of the voicing contrast of L2, the preference and weightings of the perceptual cues were affected by the dialect of L2 learner's native language.

**Keywords:** voicing contrast, English stop, dialect difference, VOT, onset F0, perceptual weighting

---

* "This work was supported by the research fund of Hanyang University (HY-2015-G)"
1. Introduction

It is well-known that adult speakers are likely to have relative use of phonetic cues in their L1 when they produce or perceive phonological contrast in an L2 (McAllister et al., 2002; Iverson et al., 2003; Escudero et al., 2012). In recent studies of Korean stop categories, it was reported that non-tonal Seoul speakers and tonal Kyungsang speakers have distinct use of phonetic cues to signal the laryngeal contrast of Korean stops (Lee and Jongman, 2012; Lee et al., 2013). Given that the two dialect speakers have different weighting in the use of acoustic cues, the current study aims to examine how different dialects of native languages (L1) affect the perception of laryngeal contrast of a second language (L2). In particular, this study focused on the relative use of perceptual cues in distinguishing voicing contrast of English stops under different prosodic boundaries and accent conditions.

When L2 learners distinguish L2 phoneme contrast, it has been found that L2 learners are likely to make use of acoustic cues to produce and perceive L2 distinction based on their cue mapping system of their native language. As an example in Iverson et al. (2003), native English listeners put more weight on the use of the third formant values (F3), while Japanese listeners made more use of the second formant on the distinction of English /l/ and /r/. In the distinction of English vowel contrast for /i/ - /ɪ/, L2 learners relied on vowel duration rather than spectral difference, while native English speakers distinguished the vowel contrast based on the difference of the first formant and the second formant of target vowels (Cebrian, 2006 for Catalan; Flege et al., 1997 for Spanish and Mandarin; Kondaurova and Francis, 2008 for Russian; Morrison, 2002 for Japanese). Since the L2 learners had difficulty in perceiving the spectral differences of English vowels, they were found to rely on the durational difference to distinguish the contrast of the target vowels. In the perception and production of Dutch vowel contrast between /aː/ - /a/, Escudero et al. (2009) proved that native Dutch speakers and Spanish learners of Dutch were better in the perception of /aː/ - /a/ contrast than L1-German listeners without experience with the Dutch language. When it comes
to different perceptual cue weightings, the native Dutch and native German listeners were shown to have put more weight on vowel spectrum than vowel duration, whereas Spanish learners of Dutch relied more on the vowel duration than the spectral properties. The results suggested that the preference for one cue over the other was found to be different depending on the learners’ language background.

In recent studies, the relative use of acoustic cues in phonetic categorization was found to be affected by a speaker’s dialect. Korean has a three-way laryngeal contrast among voiceless lenis, aspirated, and fortis stops. The phonetic correlates of the contrast were reported as voice onset time (VOT) and fundamental frequency (F0) at vowel onset. The VOT was known to be a primary cue to distinguish the three-way contrast: aspirated stops have the longest VOT, lenis stops an intermediate VOT, and fortis ones the shortest VOT (Lisker and Abramson, 1964; Cho et al., 2002). The onset F0 was used as a supplementary cue to distinguish the three-way contrast such that F0 of the following vowel was higher for aspirated and fortis stops than for lenis stops (Abberton, 1972; Han and Weitzman, 1970; Cho et al., 2002). However, the use of phonetic correlates of Korean stops was found to be different among different dialect speakers and different age groups. In the Seoul dialect, older speakers made clear VOT distinction between lenis and aspirated stops, while younger speakers were likely to rely on the use of F0 to distinguish the two categories (Silva, 2006; Kang and Guion, 2008). Lee and Jongman (2012) also claimed that the non-tonal Seoul and tonal Kyungsang dialects used the acoustic cues in a different way when distinguishing the three-way contrast of Korean stops. While Seoul speakers relied on both VOT and F0, Kyungsang speakers mainly used VOT in contrasting three stops of Korean. In the perception of the three-way contrast of Korean stops by Kyungsang and Seoul dialect listeners in Lee et al. (2013), it was found that the listeners of the two dialects showed different phonetic trading relation between VOT and F0. For the perception of lenis and aspirated stops, Seoul listeners used F0 for the identification of the lenis stop and both VOT and F0 were used in the perception of the aspirated stop. Kyungsang listeners, on the contrary, did not use F0 in perception of
both the lenis and the aspirated stops as much as Seoul listeners did.

Additionally, the two dialect speakers of Korean have shown distinct use of the VOT and onset F0 in distinguishing voiced and voiceless stops of L2. In the production of English voicing contrast in the word-initial position, it was found that Seoul and Kyungsang speakers used VOT and F0 in a similar way, while in the perception of English voiced and voiceless stops, Kyungsang listeners relied more heavily on VOT than F0 relative to Seoul listeners (Lim and Han, 2014). In the production study by Choi et al. (2013), the results of discriminant function analysis presented that the mean canonical coefficient of VOT was similar between the two dialect speakers but the mean coefficient of onset F0 was greater for Seoul speakers than for Kyungsang dialect speakers. When perceiving English voicing contrast, the result of logistic regression revealed that the listeners from the two dialects showed higher $\beta$ weight of the VOT (0.73 for Kyungsang listeners and 0.71 for Seoul listeners) than that of the F0 onset (0.31 for both Kyungsang and Seoul listeners), but the different use of phonetic cues between the two dialect groups in the production was not found in the perception. The results of the previous studies showed a discrepancy in the use of the two phonetic cues in the production and perception of English voicing contrast. Furthermore, since Lim and Han (2014) analyzed the identification rates with a mixed ANOVA but they did not include discriminant function analysis and logistic regression analysis in their study, it was hard to compare the relative weightings of the two cues with the other production and perception studies. Therefore, although the speakers of the two dialect groups showed different uses of phonetic cues in the production of the Korean stop category, it is still questionable how the preference of one phonetic cue over the other found in their L1 dialect is reflected in the perception of L2.

In addition, the former studies by Choi et al. (2013) and Lim and Han (2014) were limited in the production and perception of English stops at the word level but it has not been studied how the listeners of the two dialectal groups perceive the laryngeal contrast of L2 across different prosodic conditions. Many previous studies indicated that prosodic conditions such as prosodic position and accent affect listeners’ phonetic
categorization. For example, English and Korean were found to show enhanced VOT durations in higher prosodic domain-initial positions than lower prosodic domain-initial positions (Fougeron and Keating, 1997; Cho and Keating, 2009; Jang, 2011). In a perception study of manipulated VOT for /b/ and /p/ contrast by native English listeners and non-native Korean listeners, Cho and Kim (2013) reported that both native English listeners and Korean listeners required longer VOT in identifying /p/ in the IP boundary than in the Wd boundary. They suggested that when the IP boundary cues such as long closure duration were provided, listeners expected a long VOT duration for English voiceless stop, and native and non-native English listeners showed a similar perceptual shift. When a short closure duration was provided under an unaccented condition, listeners needed even longer VOT for the IP-initial voiceless stop, showing that there was a trading relation between the closure duration and the VOT. However, the study by Cho and Kim (2013) focused on the use of VOT but they did not consider the relative relation between the two phonetic cues such as VOT and onset F0 when non-native Korean speakers signaled the voicing contrast of English stops across the different prosodic conditions.

In a recent production study by Jang (2015), it was found that Seoul and Kyungsang dialect speakers used the phonetic cues across prosodic conditions such as IP-accented, IP-unaccented, Wd-accented, and Wd-unaccented conditions in a distinctive way. Both dialect speakers showed an enhanced voicing contrast in the higher prosodic level and in the accented condition but the contrast between voiced and voiceless stops was found to be different between the two dialect groups. Kyungsang speakers showed relatively clear distinction of VOT as a function of the prosodic position and the accent condition, while Seoul speakers produced an enhanced voicing contrast only in the higher prosodic position and the contrast between English voiced and voiceless stops was not as clear as with the results from Kyungsang speakers. On the contrary, both dialect speakers used the onset F0 to signal the voicing contrast of English stops across the different prosodic conditions in a similar way. The distinction of onset F0 between the two categories was only found under the IP-accented
and the Wd-accented conditions, while under the unaccented condition in the IP and the Wd levels, the two dialect speakers did not show a significant difference in the use of onset F0 for /t/ and /d/.

Given that the two dialect speakers showed different use of VOT and F0 in the production of the laryngeal contrast of L2 under the different prosodic conditions, it will be interesting to examine how the difference between the two dialects is reflected in the perception of the voiced and the voiceless stops of L2. In addition to the relative use of the two perceptual cues, this study will also examine the effect of prosodic position and accent condition on the perception of the English stops by Seoul and Kyungsang listeners. In order to compare the results of production with those of perception, the prosodic conditions of this study will be based on the previous study by Jang (2015). Thus, the identification pattern of the English laryngeal contrast will be compared in the IP and the Wd boundaries under the accented and the unaccented conditions between the two dialect groups.

The organization of this paper is as follows: the research methods are provided in section 2, the results of the experiment are given in section 3, and in section 4, the findings of the experiment are discussed.

2. Methods

2.1. Original stimuli and recording

English words that contrast in voiced vs voiceless stops, tea/Dee were used as the base token. The experimental design for the carrier sentence and variation of accent was from Cho and Kim (2013). The target monosyllabic words were placed at the initial position of Intonational Phrase (IP) and Word (Wd) in a carrier sentence, “Let’s hear ______ again.” In order to make the target syllable placed in the IP-initial position, the carrier sentence was produced with two IPs, [Let’s hear]_IP [tea/Dee again]_IP. On the other hand, to place the target syllable in the Wd-initial position, the carrier sentence was produced with one IP [Let’s hear tea/Dee again]_IP. The target syllables were also placed in accented and unaccented positions.
In the accented position, a high pitch accent (H*) went to the target syllable, while in the unaccented position, H* was placed on again. In all conditions, a high accent was placed on the word, “let’s”.

A male native speaker of General American English participated in the recording. The talker was asked to produce the target words, tea and Dee in two different prosodic positions and two different accent conditions. The talker had prior training session before the recording and he was recorded using a digital recorder (Roland R-05) in a quiet room. The speech material was sampled at 44,100 Hz with a 16-bit quantization.

2.1. Participants

Thirty nine Korean learners of English (20 from Seoul and 19 from Kyungsang) were recruited from Hanyang University. The ages in Seoul dialect ranged from 20 to 26 years old (mean=21.6), and the ages in Kyungsang dialect ranged from 20 to 41 (mean= 22). All subjects from each dialect had lived and been educated in the target dialect area for at least 20 years and their parents spoke the same target dialect. The subjects from Kyungsang dialect resided in Seoul or Gyeonggi region for less than one year. The English proficiency level of the participants was determined by their average TOEIC scores. They were intermediate or advanced English learners and their TOEIC scores were between 750 and 960.

2.2. Manipulated stimuli

In the production by a male native speaker of English, the mean VOT duration for /t/ was 95 ms in the IP-accented condition and 77 ms in the IP-unaccented condition. In the Wd-initial position, the mean value of the VOT for /t/ was 76 ms and 72 ms in the accented and the unaccented positions, respectively. For /d/, the VOT values were not greatly influenced by the prosodic positions and the accent conditions (8 ms in the accented IP condition, 11 ms in the unaccented IP condition, 6 ms in the accented Wd condition, and 7 ms in the unaccented Wd condition). Accordingly, the VOT values from the base token were manipulated rang-
ing from 0 ms to 70 ms in 10 ms steps. In addition to the 10 ms difference of the VOT values, 5 ms VOT value was also included in the manipulation, resulting in 9 steps of VOT values.

The F0 onset also varied depending on the prosodic position and the accent condition. In the results of production by the male English speaker, the mean F0 onset for /t/ was 198 Hz in the IP accented condition, 166 Hz in the IP unaccented condition, 222 Hz in the Wd accented condition, and 168 Hz in the Wd unaccented condition. On the other hand, the mean onset of F0 for /d/ was 178 Hz in the IP accented condition, 113 Hz in the IP unaccented condition, 191 Hz in the Wd accented condition, and 114 Hz in the Wd unaccented condition. Based on the results in the production study, the onset F0 values of the token ranged in 10 steps from 110 Hz to 200 Hz.

The original token with the voiceless stop, *tea*, was used for manipulation of the stimuli. When selecting the best token and carrier sentence from each prosodic condition, the manipulation was conducted by using the PSOLA (Pitch Synchronous Overlap and Add) algorithm. To manipulate VOT duration, the interval between the release of the target stop and the onset of voicing was modified using “To manipulation…” function of Praat. For the F0 manipulation, since the minimum and maximum F0 values varied depending on each prosodic condition, the average range of the F0 from the onset to the offset of the target vowel was calculated based on the results of the production by the above male English speaker. For example, in the IP accented condition, the F0 was raised to 200 Hz at the midpoint of the vowel and fell to 120 Hz at the end of the vowel. In order to make the test sound natural, the F0 points across the onset, midpoint, and offset of the vowel were interpolated using F0 contours in Pitch manipulation tier of Praat.

To sum up, 90 stimuli were generated (9 steps of VOT × 10 steps of F0) by synthesizing a new token with modified VOT and F0. The manipulated stimuli were included in the frame sentence “Let’s hear /Xi/ again” for the IP accented, IP unaccented, Wd accented, and Wd unaccented conditions. Overall, 360 stimuli were created in total. Depending on prosodic phrasing, the frame sentences had different duration. The duration
of "Let's hear" was longer in the IP context than in the Wd context. In addition, the following vowel /i/ after target English stops had higher F0 and duration in the accented condition than in the unaccented condition.

2.3. Procedure

The stimuli were repeated three times and each subject was given randomized blocks of the stimulus set. They were asked to listen to test sentences and to identify what they had heard by clicking one of the response options, tea or Dee on the computer screen. Prior to the actual test, test trials were given to the subjects to familiarize them with the test. The experiment was conducted with headphones connected to a computer in a quiet room on campus, and the task took approximately 30 to 35 min. Subjects were paid for their participation.

For the statistical analysis, binary logistic regression was conducted with an identification response as the dependent variable, and position, accent, VOT, and F0 as covariates.

3. Results

Figure 1 presents the perceptual pattern of English voicing contrast as a function of VOT across the IP-accented, IP-unaccented, Wd-accented, and Wd-unaccented conditions by Seoul and Kyungsang dialect listeners. In this result, the response rates for /t/ were averaged across the 10 different onset F0 values. Thus, only the function of VOT was shown in the figure. Both dialect groups showed a perceptual shift depending on the VOT variation and the perceptual shift was different between the IP and Wd boundaries. When identifying English voiceless stop, both dialect listeners needed longer VOT in the IP boundary than in the Wd boundary. At a 50% response rate, both group listeners also showed greater VOT values in the IP boundary than in the Wd boundary. Furthermore, it was shown that the estimated 50% crossover point was slightly greater in the unaccented condition than in the accented condition in each prosodic level (35 ms in the IP-accented, 38 ms in the IP-unaccented, 28 ms
in the Wd-accented, and 30 ms in the Wd-unaccented for Kyungsang listeners; 34.5 ms in the IP-accented, 37 ms in the IP-unaccented, 26 ms in the Wd-accented, and 28 ms in the Wd-unaccented for Seoul listeners). It seems that relatively longer VOT values are required for the identification of voiceless stop when the target-bearing syllables are not prominent for both dialect groups.

Figure 1. Identification rate for the voicing contrast of English stops as a function of VOT for Seoul and Kyungsang listeners.

Figure 2 illustrates the identification rates of English stops as a function of onset F0 for Seoul and Kyungsang listeners. In Fig. 2, the response for /t/ was averaged across the 9 different VOT values and only the function of F0 was presented in the figure. As the onset F0 values of target stops get higher, the perception of the voiceless stop increases for
all prosodic conditions. The identification rates for English voiceless stops are greater in the Wd conditions than in the IP conditions across all F0 values for both dialect groups.

For both dialect listeners, a greater onset F0 value was required for the identification of English voiceless stop for the IP level than for the Wd level, and in the accented condition than in the unaccented condition. For Seoul listeners, the estimated 50% crossover point was 180 Hz for the IP-accented, 175 Hz for the IP-unaccented, 145 Hz for the Wd-accented, and 140 Hz for the Wd-unaccented condition. Kyungsang listeners showed 175 Hz for the IP-accented condition, 170 Hz for the IP-unaccented condition, 155 Hz for the Wd-accented condition, and 140 Hz for the Wd-unaccented condition for the estimated 50% crossover point. When the onset F0 was given below 150 Hz, Seoul listeners showed greater /t/ responses than Kyungsang listeners in the Wd-accented condition. However, consistent results were not found in the other prosodic conditions.

Compared to Seoul listeners, Kyungsang listeners showed a sharper perceptual curve as the onset F0 increased. Kyungsang listeners seemed to be more heavily affected by the increase of the onset F0 in the identification of English voiceless stop under the different prosodic conditions. This result, in part, conflicted with the findings in the previous study by Lim and Han (2014) such that Seoul listeners showed a sharper perceptual curve than Kyungsang listeners when distinguishing /pi/ and /bi/ in the word-initial position.
Figure 2. Identification rate for the voicing contrast of English stops as a function of F0 for Seoul and Kyungsang listeners.

Figure 1 and Figure 2 presented the identification pattern as a function of VOT and F0, respectively but in each figure, one of the two perceptual cues was averaged across the four different prosodic conditions. Consequently, they did not reveal the relative effect of the two perceptual cues in the identification of English stops across the different prosodic positions and the accent conditions. In order to examine the relational cues of VOT and onset F0 in the perception of voicing contrast of English stops, gradations of color were used to present the responses for the stimuli varying the VOT and the onset F0 in different combination. This heat plot was previously used in Lee et al. (2013). In the following heat plots in Fig. 4, 10 levels of F0 and 9 levels of VOT are shown on the horizontal and the vertical axes, respectively. The darkness of each cell represents
the proportion of identification rate. As the proportion of responses for the stimuli increase, the target cell becomes darker. The specific proportion of response rates according to the darkness of the cells is shown in Fig. 3.

<table>
<thead>
<tr>
<th>Identification %</th>
<th>1-19</th>
<th>20-39</th>
<th>40-59</th>
<th>60-79</th>
<th>80-100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Color</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 3. The range of identification rate for different color of cells.

Figure 4 presents the identification response by Seoul and Kyungsang listeners in the IP-accented, IP-unaccented, Wd-accented and Wd-unaccented conditions, respectively. As found in the results of Fig. 1, the two dialect groups were likely to require a longer VOT in the identification of /t/ in the IP boundary than in the Wd boundary. For both dialect listeners, the stimuli with VOTs above 50 ms were perceived as a voiceless stop in the IP boundary with a response rate more than 60% as in (a) and (b) in Fig. 4. In the Wd boundary, the stimuli with VOTs above 40 ms were identified as voiceless stop by both dialect groups as in (b) and (c) in Fig. 4.

In the IP boundary, VOTs which ranged from 20 ms to 40 ms were too ambiguous to be perceived as voiceless stop but listeners were likely to perceive voiceless stop in this ambiguous VOT range when higher F0 was provided. When short VOT values were given, a relatively higher F0 was required in the identification of English voiceless stop. For example, when 40 ms of VOT was given, the stimuli with the onset F0 above 160 Hz were perceived as a voiceless stop. At 30 ms of VOT, listeners identified the stimuli as a voiceless stop when the stimuli had the onset F0 above 190 Hz in the IP accented condition and 200 Hz in the IP unaccented condition.

In the Wd level, the intermediate VOT values ranged from 20 ms to 30 ms, and the stimuli with higher F0 were likely to be perceived as voiceless stop by both dialect groups. In the Wd-accented condition, both dialect listeners showed a relatively greater difference in identifying the laryngeal contrast of English as a function of onset F0. When the stimuli with 40 ms VOT was given in the Wd accented condition, Kyungsang listeners perceived the stimuli with the onset F0 above 140 Hz as voiceless stop but Seoul listeners identified the same stimulus as voiceless stop.
across all F0 values. At 30 ms of VOT in the Wd accented condition, the onset F0 above 140 Hz was required for the perception of voiceless stop by Seoul listeners, while Kyungsang listeners perceived the stimuli with the onset F0 above 170 Hz as a voiceless stop. In short, when the ambiguous VOT values were given in the Wd accented condition, relatively higher onset F0 values were required for the perception of voiceless stop by Kyungsang listeners than by Seoul listeners. However, the listeners of the two dialectal groups did not show different perceptual pattern in the Wd unaccented condition.

(a) Identification rates in the IP accented condition

(b) Identification rates in the IP unaccented condition
The results from the heat plots confirmed that both dialect groups used VOT as a primary perceptual cue, and F0 as a secondary cue. When ambiguous VOT was provided, the dependence on onset F0 increased. However, the relative use of the two perceptual cues was found to be slightly different between the two dialect groups and across the different prosodic conditions. In order to find out the relative use of perceptual cues across different prosodic conditions, binary logistic regression was
performed using SPSS 1.8. Both dialect groups showed significant main effects for position ($\beta=0.449$, $p =.005$ for Seoul listeners, $\beta=0.694$, $p <.000$ for Kyungsang listeners), VOT ($\beta=0.081$, $p <.000$ for Seoul listeners, $\beta=0.157$, $p <.000$ for Kyungsang listeners), and F0 ($\beta=0.019$, $p <.000$ for Seoul listeners, $\beta=0.039$, $p <.000$ for Kyungsang listeners). On the contrary, the effect of accent was not found to be significant for both dialect groups ($\beta=-0.031$, $p =0.805$ for Seoul listeners, $\beta=0.141$, $p =0.290$ for Kyungsang listeners). For Seoul listeners, the interaction effect was found to be significant between VOT and F0 ($p<.000$), and across VOT, F0 and accent ($p<.000$), and across VOT, F0, and position ($p<.000$). The significant interaction of VOT by F0 found by Seoul listeners indicated that they used the two perceptual cues complementarily. For Kyungsang listeners, on the contrary, the interaction effect was only found to be significant across VOT, F0, and accent ($p<.000$).

Figure 5. Beta-coefficients from individual logistic regression analyses for Seoul and Kyungsang listeners

Figure 5 displays the distribution of beta-coefficients for VOT and F0 from the individual listeners of the two dialects. For Kyungsang listeners, the difference of beta-coefficients between VOT and F0 were greater compared to the results from Seoul listeners. Seoul listeners showed relatively greater range of beta-coefficient values for VOT compared Kyungsang listeners. In the results from the individual logistic regression analyses, the individual difference was not found in the beta-coefficient values for
VOT and F0. All subjects from the two dialect groups showed greater beta-coefficient values for VOT than those for F0 except for one subject from Seoul dialect. The subject from Seoul dialect showed nearly similar beta-coefficient values for both VOT and F0.

In the logistic regression analysis, the $\beta$ coefficients represented participants’ reliance on each perceptual cue. The results of the logistic regression analysis suggested that the mean $\beta$ weights of VOT, and F0 were 0.081, and 0.019 for Seoul listeners and 0.157, and 0.039 for Kyungsang listeners, indicating that the coefficients of VOT and F0 were greater for Kyungsang listeners than for Seoul listeners. The comparison of coefficients for each participant was conducted between the two dialects using an independent-samples t-test. The coefficients for position and the interaction across accent, VOT and F0 were not found to be significantly different between the two dialect groups. On the contrary, the results of the t-test for the averaged coefficients for VOT and F0 were found to be nearly significant between the listeners of the two dialect groups ($t=-1.967$, $p=.057$ for VOT and $t=-1.922$, $p=.062$ for F0). When perceiving the voicing contrast of English stops across different prosodic conditions, both dialect groups were more heavily dependent on the use of VOT than on the use of F0, but the reliance on the use of VOT was much greater for Kyungsang listeners than for Seoul listeners. In addition, the results revealed that compared to Seoul listeners, Kyungsang listeners were more greatly affected by the difference of F0 onset values for the perception of English voicing contrast across the different prosodic conditions.

4. Discussion

The current study investigated the perception of English stops across the different prosodic conditions by the listeners of the two dialectal groups of Korean. In the production study by Jang (2015), it was found that when Seoul and Kyungsang dialect speakers produced English stops across different prosodic levels and the accent conditions, they showed relative use of phonetic cues to signal English voiced and voiceless stops:
Kyungsang speakers put more weight on the use of VOT to signal the laryngeal contrast of L2, while Seoul speakers made less distinctive use of the VOT compared to Kyungsang speakers. With regard to the use of onset F0, both dialect speakers showed the significantly different onset F0 for the voiceless and voiced stops of English only in the accented condition of the IP and Wd levels.

Drawing on the different use of each phonetic cue by the two dialect speakers of Korean, this study sought to find out how the listeners of the tonal and non-tonal Korean dialects perceived the voicing contrast of L2 across the different prosodic contexts. The results from the present study demonstrated that both dialect listeners required a longer VOT duration to identify an English voiceless stop in the IP level than in the Wd level. The results paralleled the findings in Cho & Kim (2013) such that native and non-native listeners of English require a longer VOT for the perception of voiceless stop in the IP boundary than in the Wd boundary. As Cho & Kim (2013) noted, when IP boundary cues were given, listeners were likely to expect a longer VOT for the identification of voiceless stop. On the contrary, the listeners of the two dialectal groups did not show systematic perceptual differences between the accented and un-accented conditions. Although the VOT and onset F0 of the voiceless and voiced stop of English were affected by the accented condition, as shown in the production study in Jang (2015), the perceptual response pattern of the voicing contrast was not influenced by the accent condition. The results were also supported by the findings in the logistic regression analysis such that only the effect of position was found to be significant.

The results of the heat plot confirmed that VOT was a robust perceptual cue in the identification of English voicing contrast and that when the stimuli had ambiguous VOT ranges, higher onset F0 was required for the identification of the voiceless stop. The ambiguous VOT ranges were found to be different between the IP level and the Wd level (from 20 ms to 40 ms in the IP level and from 20 ms to 30 ms in the Wd level). It also revealed that the dependency on the onset F0 was different depending on how ambiguous VOT was, and the relative use of VOT and onset F0 varied across the different prosodic conditions. For example, when
short VOT values were provided, a relatively higher F0 was required in the perception of English voiceless stop. Although the required VOT duration for the perception of the English voiceless stop and the ambiguous range of VOT were different between IP and Wd level, the results manifested that the VOT and onset F0 were in a trading relation under the different prosodic conditions.

The results of the logistic regression presented that for the listeners of the two dialectal groups, the coefficient of VOT was greater than that of onset F0. Furthermore, it demonstrated that when distinguishing the voicing contrast of the English stops under the different prosodic conditions, the listeners of the two dialects used VOT as a primary and F0 as a secondary perceptual cue. However, the coefficient value for VOT by F0 interaction was found to be significant for Seoul listeners, but not for Kyungsang listeners. This means that Seoul listeners used the two perceptual cues in a complementary way compared to Kyungsang listeners.

In previous studies, it was proved that the VOT was used as a robust perceptual cue in distinguishing English voicing contrast and the listeners of the two dialectal groups showed similar patterns in perception. Choi et al. (2013) noted that the listeners of the two dialect groups made more use of the VOT than the onset F0 and the perceptual weightings of VOT and F0 were similar between the listeners of the two dialectal groups in the identification of voiceless English stops in the Wd level. In the current study, on the other hand, the results from the independent-samples t-test revealed that the coefficient values of onset F0 and VOT were significantly greater for Kyungsang listeners than for Seoul listeners. The difference of the coefficient values between the two dialects was even greater for the VOT than for the onset F0. This showed that the effect of the VOT was more salient for Kyungsang listeners when distinguishing the voicing contrast of English stops compared to Seoul listeners. For the perception of voicing contrast in the Wd-initial position, the relative use of the two acoustic cues found in the production of the two dialects did not seem to be reflected in the perception. However, the perceptual pattern seemed to be affected by their L1 dialect when the different prosodic conditions were involved in the identification of laryngeal contrast.
As found in the results of Jang (2015), Kyungsang speakers put more weight on the use of VOT to enhance the voicing contrast of English stops in the higher prosodic position and in the accented condition, while Seoul speakers showed less distinctive production of the voicing contrast across the different prosodic conditions. Thus, the different use of phonetic cues by the two dialect speakers was reflected in the perception of the laryngeal contrast of L2. In the production and perception of Korean stop categories, it was also claimed that Kyungsang speakers were heavily dependent on the use of VOT compared to Seoul speakers. In the perception of a two-way contrast of English stops, the VOT alone was a sufficient perceptual cue to distinguish the voicing contrast. However, when the relatively complex prosodic condition was given, the non-native listeners seemed to show their own preference in the use of perceptual cues, which was based on their dialect in their native language.

In the production and perception of Korean stop categories, it was claimed that Kyungsang speakers were less dependent on the use of onset F0 to distinguish the laryngeal contrast of Korean stops since the onset F0 was used for the distinction of lexical tonal contrast of Kyungsang dialect (Lee and Jongman, 2012; Lee et al., 2013). However, this study found that Kyungsang listeners had greater weighting of onset F0 than Seoul listeners although the coefficient value of the onset F0 were relatively small compared to that of VOT. This result conflicted with the findings in the production study by Jang (2015). In the production study by Jang (2015), the two dialect speakers produced significant distinction of onset F0 between English voiced stop and voiceless stop in the IP-accented and the Wd-accented conditions but the distinction was not shown under the unaccented condition. Although both dialect groups showed similar patterns in the use of onset F0 when producing the laryngeal contrast of English stops under the different prosodic conditions, Kyungsang listeners had greater weighting in the use of onset F0 than Seoul listeners in the perception of the laryngeal contrast of English stops. It seemed that when the various prosodic conditions were involved in distinguishing the laryngeal contrast of L2, the presence of pitch accent in Kyungsang dialect made the listeners of the two dialect groups more sensitive to
detecting the differences in the vocalic properties of the target syllable.

In conclusion, the present study revealed that the perception of the voicing contrast of L2 was affected by the different dialect of L1. In the perception of voicing contrast of English stops across the different prosodic conditions, Seoul and Kyungsang dialect listeners used the perceptual cues in a similar way but their relative weightings of the perceptual cues were differently manifested.

References


Han, M. S. and Weitzman, R. (1970). Acoustic features of Korean /P, T, /, /p,
t, k/ and /pʰ, tʰ, kʰ/. *Phonetica* 22, 112-128.


Mi Jang
Hanyang University
55, Hanyangdae-hak-ro, Sangnok-gu, Ansan-si, Gyeonggi-do, 15588, Korea
E-mail: jj9094@hanyang.ac.kr

Received: October 31, 2016
Revised version received: December 1, 2016
Accepted: December 19, 2016