Instructional Effects on Voice Onset Time Perception in L2 Speech Segmentation

In Young Yang
(Seoul National University)

Yang, In Young. (2017). Instructional Effects on Voice Onset Time Perception in L2 Speech Segmentation. ***Language Research*****, 53.3, 561-585.**

This study examined the effects of instruction on the English word boundary perception of Korean learners of English. The focus of this study was the appropriate use of Voice Onset Time (VOT) of English voiceless stops (*p*, *t*, *k*) in word initial (e.g., *top*) and after word initial *s* position (e.g., *stop*) signaling the presence and absence of a word boundary before the stop consonant. There were 48 participants in the study, who were assessed three times: before instruction (pre-test), immediately after instruction (post-test), and five weeks after instruction (delayed post-test). Results revealed that participants' perceptual accuracy improved significantly in both VOT positions in segmentation after instruction. However, the improvement was more prominent in word initial than after word initial *s*. Among word initial stop consonants, velar stops were the most difficult to acquire, while after initial *s*, bilabial stops were the most difficult and alveolar stops were the easiest to acquire.

Keywords: VOT, Segmentation, Word boundary perception, English stop consonants, L2 perception, Pronunciation teaching

1. Introduction

The initial stage of acquiring a new language begins with listening (Rost 2002). Infants hear sounds around them and first acquire sound contrasts such as /p-b/, /t-d/, and /k-g/, and then their phonology develops to encompass stop consonant clusters (Wolfram and Johnson 1982). This process requires the correct segmentation of speech sounds in listening based on the acoustic information in speech signals. However, learning

segmentation through acoustic cues is not always straightforward for second language (L2) learners, especially in English as a Foreign Language (EFL) context. L2 learners frequently suffer interference from their native language (L1) processes in learning an L2 (Brown 2007). When it comes to the use of acoustic-phonetic cues, learners experience difficulty perceiving L2 cues accurately because their perception strategies are based on L1 perception (Guion and B. Lee 2006). When their target language and native language have different acoustic cue weighting strategies, it is difficult to acquire the L2 acoustic information that encodes sound contrasts or the appropriate segmentation strategies based on the respective acoustic cues, but L2 experience may help learners tune their acoustic cue weighting systems (Guion and B. Lee 2006).

This study investigates the learnability of segmentation strategies in terms of Voice Onset Time (VOT) cue salience and place of articulation of stop consonants. In a certain stream of speech sounds such as *loose pills* vs. *Lou spills*, voiceless stops are pronounced with quite long VOTs right after the word boundary that are more readily noticeable (more salient), but have very short VOTs (less salient) after word initial *s* (Altenberg 2005). Thus, learners must understand that the presence of the cue signals the initial position, and the absence[1)] of the cue signals the absence of the word boundary after initial *s*. Furthermore, different places of articulation cause different degrees of difficulty in acquiring the contrasts (Yavaş 2016). Thus, this study examines the effect of place of articulation (bilabial, alveolar, and velar) in each VOT type that may also entail different degrees of difficulty in L2 speech segmentation.

1) An anonymous reviewer commented that the term 'absence' could be misleading in that stops after word initial *s* do have (short) aspiration. However, here, 'absence' denotes the absence of long VOT cues in accordance with previous research by Altenberg (2005). In addition, in perception, VOT is categorically perceived (Clark and Yallop 1995), so this term was deemed appropriate in this context.

2. Background

2.1. Acoustics and functions of stop distinctions in English and Korean

English has two contrasting series of stop consonants — voiced and voiceless — distinguished by voicing (Ladefoged and Johnson 2014). However, phonetically there are three series of stops in initial position. Voiceless stops are aspirated at the beginning of a syllable and unaspirated after initial *s*. Essentially, if a voiceless stop is preceded by *s*, it is pronounced with no perceivable VOT, while if it is in syllable initial position, its VOT is rather long especially in a stressed syllable. Additionally, voiced stops are devoiced in utterance initial position that is phonetically unaspirated voiceless sounds, while they are fully voiced between voiced sounds. Therefore, English has three phonetically different stop sounds in syllable initial position: voiceless aspirated, voiceless unaspirated, and voiced (Ladefoged and Johnson 2014). In short, English VOT cues indeed constitute the primary acoustic cue for stop voicing distinction (Ladefoged and Johnson 2014). Furthermore, in some instances, VOT cues signal important information in speech segmentation (Altenberg 2005). For example, in phrases such as *loose pills* vs. *Lou spills*, longer VOT indicates that the voiceless stop is the beginning of a new word, thus signaling the presence of a word boundary right before the segment. However, the absence of VOT in voiceless stop consonants signals that the stop does not start a new word, indicating that the word should start elsewhere in the speech stream (Altenberg 2005).[2)]

According to H. Ahn (1999), Korean stop consonants have a three-way

2) As a reviewer indicates, cues other than VOT are present in the segmentation of such phrases, and this should not be disregarded since in natural speech the whole interaction of several cues signal the presence or absence of word boundary (Altenberg 2005). However, it has been proved that VOT plays a primary role comparing with the other cues (Christie 1974; Cohen 1987; Nakatani and Dukes 1977; cited from Altenberg 2005) such as durational difference of *s* (Klatt 1974), stop closure (Ladefoged 1975), and higher amplitude of word initial *s* than final *s* (Umeda and Coker 1974) (cited from Altenberg 2005). Please refer to Altenberg (2005) for a more detailed discussion on this subject. This study focuses on the primary cue VOT based on its importance in English segmentation and a more distinctive cross-linguistic difference between English and Korean.

phonemic distinction (lenis, tense, and aspirated) that is phonetically and functionally different from English stops. In Korean, both lenis and aspirated stops are produced with aspiration — lenis stops have short aspiration and aspirated stops have long aspiration. However, Korean speakers demonstrate considerable overlaps in VOTs between lenis and aspirated, especially female speakers (E. Oh 2010). It is even doubtful that VOT primarily contributes to the three-way stop categorization in Korean (H. Ahn 1999; M.-R. Kim, Beddor, and Horrocks 2002) because vowels following lenis stops have much lower f0 than those following aspirated stops (Han and Weitzman 1970). English also displays a difference in the pitch of vowels following voiceless and voiced stops, but the difference is about 15%: on average the pitch of the vowel after voiceless stops is 15% higher than voiced stops (Lehiste and Peterson 1961). The difference in Korean is much greater than in English; the difference is about 40% — 165.01 Hz for aspirated stops, and 117.86 Hz for lenis stops (measured at voice onset) (H. Ahn 1999). Furthermore, Korean stop consonants are not subject to VOT-related segmentation strategies.

The acoustic differences and dissimilarities in phonological contrasts between the two languages shape native listeners' cue weighting strategies differently (Abramson and Lisker 1985; Guion and B. Lee 2006). Korean native listeners attend to vocalic cues (e.g., the pitch of the following vowel) primarily in categorizing the three classes of stops that are not primary cues in categorizing English stops, and sometimes ignore VOT cues (T. Cho 1996; M.-R. Kim et al. 2002; Y.-H. Kim 2007) that are crucial for English voicing distinction (Abramson and Lisker 1985).

2.2. L1 transfer in L2 segmentation

The differences in cue weighting strategy cause undetected problems in segmenting continuous speech streams. Although Korean learners of English must focus on the VOT of stop consonants when listening to English sounds, they fail to attend to it appropriately (I. Y. Yang 2014). It is well known that second or foreign language learners perceive English word boundaries more imperfectly than native speakers of English do

(Altenberg 2005), and Korean learners particularly are worse at word boundary perception when VOT cues are involved rather than glottal stops (H-Y Um 2006).

The difficulty that such segmentation presents to Korean learners of English could be due to Korean cue weighting strategies in stop categorization that weight relevant vocalic cues (e.g., pitch) more highly than consonantal cues (e.g., VOT) (I. Y. Yang 2014), implying that Korean learners of English are not as sensitive to English VOT cues as required for accurate perception.

I. Y. Yang (2014) demonstrated that this strategy was transferred to L2 English stop perception in EFL contexts, revealing that the vowel dependency of Korean native listeners played a negative role in determining word boundaries indicated by VOT cues. Korean L2 English listeners do not pay sufficient attention to consonantal cues that are crucial in categorizing and segmenting L2 English, but rather attend to vocalic cues that lack information imperative for English segmentation. This native language transfer might hinder the appropriate awareness of the function of VOT cues resulting in low performance in VOT-related L2 segmentation.

2.3. Learnability problems in L2 segmentation strategy

This study examines two further issues in the VOT-related perception of word boundaries. First, which cues are easier to acquire — VOT cues immediately after the word boundary e.g., *loose pills* (henceforth "long VOT") or after word initial *s* e.g., *Lou spills* (henceforth "short VOT")? In such English phrases, long VOT cues indicate the presence of a word boundary before a stop, but short VOT cues indicate that a stop is not the beginning of a word, and thus the boundary should be placed elsewhere. However, when learning such a perception strategy, knowledge of long VOT cues does not imply knowledge of short VOT cues and vice versa, especially for Korean learners who 'under-attend' (Guion and B. Lee 2006: 124) the VOT cues in their native language. H.-Y. Um (2006) and I. Y. Yang (2014) demonstrated that Korean college and high school students performed worse on short VOT cues than long VOT cues

in a speech segmentation task. This indicates that, in an uninstructed situation, short VOT cues after initial *s* position are not immediately noticed and learned. Therefore, this study investigated whether instructional effects differ depending on long and short VOT cues as cue salience may affect the acquisition or learning of a particular foreign language component. In this sense, the acquisition of short VOT cues that are not accompanied by audible aspiration, could be more difficult for L2 learners to learn (Altenberg 2005), hence implying lesser effect of instruction. In short, in second/foreign language learning, the degree of salience of a specific cue may play an important role in the ease and sequence of acquisition. The first research question is stated below:

Research Question 1: Does the presence or absence of the cue affect the learnability of the contribution of VOT cues to word boundary perception?

Essentially, can learners acquire both VOT cues to the same degree if the instructional effect is meaningful under both conditions?

Prediction 1: The instructional effect would be greater for long VOT than short VOT.

When learners are provided instructions on VOT related segmentation, they will more readily learn the long VOT because the cue is auditorily more noticeable and the relation between the cue and segmentation is more explicit.

Second is the effect of the place of articulation. According to Yavaş (2016), the acquisition of velar stop voicing contrast imposes more difficulty on learners than alveolars, and alveolars more than bilabials in terms of markedness. This allows us to expect that the instructional effect would be least for velars, greater for alveolars, and greatest for bilabials.

In contrast, in the acquisition of first language phonology, English native infants acquire English sounds by contrast, meaning they acquire /p/ and /b/ around the same age (1;6) and similarly for the /k-g/ and /t -d/ distinctions (Wolfram and Johnson 1982). Native English-language infants are known to acquire the bilabial distinction the earliest and the alveolar

stop distinction the latest, implying that bilabial stop distinctions are easiest to acquire and alveolar stops the most difficult (Ingram 1976, cited from Wolfram and Johnson 1982). As this concerns the acquisition of phonemic contrasts, we cannot explicitly state that this developmental sequence is valid for the allophonic distribution indicated by short VOT cues. However, we can still test whether the same pattern by place of articulation holds for the allophonic distribution in the L2 context.

In addition, it would not be equally easy to hear the VOT cues of stop sounds in different places of articulation. More specifically, the VOT cue of velar stops is produced further inside the oral cavity than those of alveolar and bilabial stops, in that order. Thus, the audibility of the VOT cues of stop consonants depends on the place of articulation: the long VOT cue of bilabial stops is most salient, that of alveolar stops the subsequent, and that of velar stops the least salient. However, this tendency is valid only for long VOT cues. The salience difference according to place of articulation is blurred concerning stop consonants with short VOT after *s* because they have virtually no audible aspiration. Therefore, it would be significant to examine how the perceptual salience difference among the three places of articulation interact with long and short VOT cues in the acquisition of L2 segmentation strategy.

Research Question 2: Which place of articulation is easiest to acquire with long and short VOT cues, respectively?

Prediction 2: For long VOT, learners would have the most difficulty in learning the VOT cues of velar stops for segmentation that will thus indicate the least instructional effect, but for short VOT, the difficulty may not vary across the three places.

Korean learners' difficulty in the production of English stop clusters has been well documented (e.g., H.-Y. Lee 2000; among others). Production is more directly comparable between native and non-native speakers to the finest aspects such as VOT or intensity, etc. However, perceptual difficulties faced by L2 learners are not easily testable and observable, and should be addressed more indirectly with a controlled

aspect. Although this study limits the focus to VOT instruction, future research with more cues can examine the exact nature of the problem.

3. Method

The experiment required Korean learners to locate the word boundary in English consonant sequences in pairs of words such as *Lou spills* vs. *loose pills*, for which segmentation is primarily triggered by the presence or absence of VOT (aspiration) cues. All the test materials were from Altenberg (2005).

3.1. Participants

Altogether 48 students participated. All the participants were taking English phonetics and phonology courses at two different universities during the data collection period. Most of them were majoring in English language education (mostly sophomores) or English language and literature (mostly juniors and seniors). None reported hearing disorders.

3.2. Test materials

The participants were tested on materials from Altenberg (2005). The stimuli are presented in the following table according to the presence and absence of VOT (long and short), the type of consonant sequence (VsC, CsC, and CsCC),[3] and place of articulation (bilabial, alveolar, and velar). The stop consonants for which the respective VOT measurements were made are underlined. A male native speaker of American English from Indiana recorded the experimental stimuli. He was 35 years old and had lived in Korea for six years at the time of recording. Altogether 36 stimuli (18 pairs) were employed, and the same number of stimuli was included as distracters. The students were presented the 72 test stimuli randomly and provided a trial question before taking the test. Each ques-

3) *V* = vowel, *C* = consonant.

tion offered two answer options. The test took approximately 10 minutes to complete.

Table 1. Test Items with Mean Values of VOT (ms)

Place of Articulation	Short VOT stimuli	Long VOT stimuli
Bilabial	Lou spills	loose pills
	lay speech	lace peach
	keep sparking	keeps parking
	chief sport	chief's port
	cook sprints	cook's prints
	top spry	tops pry
Mean of VOT (S.D.) (ms)	11.0 (2.5)	60.3 (8.0)
Alveolar	Lou stops	loose tops
	lay stable	lace table
	keep stalking	keeps talking
	chief star	chief's tar
	cook struck	cook's truck
	top strains	tops trains
Mean of VOT (S.D.) (ms)	23.9 (14.1)	59.5 (7.0)
Velar	Lou skis	loose keys
	lay scar	lace car
	keep scanning	keeps canning
	chief school	chief's cool
	cook screams	cook's creams
	top scrawled	tops crawled
Mean of VOT (S.D.) (ms)	21.4 (17.1)	73.0 (19.6)
Total	18.8 (13.4)	64.3 (13.7)

3.3. Procedures

The experiment employed the following processes: (1) Pre-Test → (2) Instruction → (3) Post-Test → (4) Delayed Post-Test. The course content included instruction on aspiration and VOT cues of English stop consonants and their contribution to English word boundary perception. The instruction was composed of two phases. In the first phase, students were

provided with an explanation on the realization of English stop consonants in syllable initial position based on Ladefoged and Johnson (2014). Then they were presented a spectrogram of aspirated and unaspirated stop realizations employing Praat speech analysis software. They heard the aspiration of voiceless stop consonants in word initial position singled out, compared the aspirated voiceless stop, unaspirated voiceless stop after *s*, and voiced stop (partially devoiced in the utterance initial position), and understood the perceptual nature of aspiration and unaspirated stop after *s*. In the second phase, students were provided with the opportunity to detect the presence and absence of VOT cues in stop consonant perception, utilizing 20 phrases whose segmentation was triggered primarily by VOT cues from the same material employed for the test but randomly presented. They were instructed to pay special attention to the VOT length of the stops, and indicate whether they heard word boundary before the voiceless stop or not.

Before the instructional session began, students were tested on their awareness of the VOT length of English stop consonants and its contribution to word boundary. After being taught the VOT realization of stop consonants in word initial position and after *s* within a word, they were tested again on the same material. Five weeks later, they were retested on the same material to examine if their awareness of VOT and word boundaries could be sustained. This study utilized the same material in the three tests. As a reviewer indicates, the discussion on instructional effect ought to be cautious considering that employing the same test material may allow learners to grow accustomed to it, weakening the instructional effect. Regarding this point, further research should develop a more refined method. In this study, the discussion on instructional effect will primarily focus on the difference between pre-test and delayed post-test; the five-week term between post-test and the delayed post-test could partly complement this methodological issue.

4. Results and Discussion

This section presents and discusses the results of the experiments with reference to the research questions and predictions. First, participants' overall performance before the relevant instruction is presented in 4.1. Second, the effects of instruction on the perception of VOT cues regarding L2 speech segmentation are submitted in 4.2. Finally, a general discussion on the findings is provided in 4.3.

4.1. Pre-Instruction

4.1.1. Presence or absence of the cue in the perception of English word boundaries

The participants in this research demonstrated approximately 73% accuracy in determining the word boundary (approximately 26 correct responses out of 36). The participants performed relatively better than those in previous studies (H.-Y. Um 2006; I. Y. Yang 2014), especially in terms of word boundaries with short VOT cues that we attributed to the fact that they were college students, predominantly English majors, who had greater exposure to spoken English than high school or college students with other majors. They performed equally well on long and short VOT-related word boundary perception before instruction.

Table 2. Descriptive Statistics of Overall Performance before Instruction

	N	Mean (Correct %)	S.D.
Long VOT	48	13.1042 (72.8)	2.8971
Short VOT	48	13.2917 (73.8)	3.36413
Total	48	26.3958 (73.3)	5.32253

The accuracy in the perception of long and short VOT cues was almost identical (approximately 13 correct responses out of 18 responses). Statistical analyses were performed using SPSS. The results of a paired *t*-test indicated that the difference between the long VOT and short VOT was insignificant (*Mean difference* = -.18750, t = -.390, p = .698). This indicates that participants could perceive word boundaries by the two

types of VOT cues equally well before instruction. However, the standard deviation for short VOT cues was higher than long VOT cues, demonstrating that the differences among the students were greater for short VOT, thus implying that short cues are more difficult to acquire than long ones.

4.1.2. The effect of place of articulation on the perception of English word boundaries

The following table displays the accuracy of word boundary perception among Korean college students in terms of three places of articulation of the stop consonants (bilabial, alveolar, and velar). The participants performed best on bilabial stops and worst on velar stops when long VOT cues were involved. However, the results for the short VOT category did not indicate the tendencies displayed for long VOT as the participants demonstrated the highest accuracy for alveolar stops, then bilabial, and the lowest for velar stops. Nonetheless, they had the most difficulty in perceiving word boundaries with velar stops regardless of the presence or absence of the cue. Additionally, it is noteworthy that the standard deviation was considerable among the participants before the instruction, implying that some learners have more difficulty in segmentation than others. Later we will examine how instruction affects this large deviation.

Table 3. Descriptive Statistics of Place of Articulation Effect before Instruction

		Long VOT		Short VOT	
	N	Mean[4] (Correct %)	S.D.	Mean (Correct %)	S.D.
Bilabial	48	4.71 (78.5)	0.988	4.42 (73.7)	1.285
Alveolar	48	4.37 (72.8)	1.409	4.77 (79.5)	1.448
Velar	48	4.02 (67.0)	1.28	4.1 (68.3)	1.325

Repeated Measures ANOVA was conducted for long and short VOT with the three places of articulation as within subject variables, respectively, and demonstrated that the differences among the three places of articulation were statistically meaningful for both long (F = 7.640, p = .001)

4) Averages of three places of articulation targets — six items for each place, altogether 18 items.

and short VOT (F = 6.317, p = .004). Pairwise comparisons indicated that the difference between bilabial and velar stops (p < .001) primarily contributed to the significance of the within-variable difference for long VOT (p = .103 for the difference between bilabial and alveolar stops, p = .091 for the difference between alveolar and velar stops). For short VOT, only the difference between alveolar and velar stops was statistically meaningful (p = .001; p = .088 for the difference between bilabial and alveolar stops, p = .087 for the difference between bilabial and velar stops).

In short, before instruction, when there was an audible acoustic cue to distinguish word boundaries, it was easiest to acquire aspiration of the bilabial stops. Interestingly, when the stop consonants are after *s* in *s* + stop sequences, the participants displayed the highest performance on the stimuli containing alveolar stops.

4.2. Post-Instruction

This section presents the effect of instruction in terms of the presence or absence of VOT cues and place of articulation of stop consonants. First, overall performance differences before and after instruction will be compared, then the differences among the three places of articulation in relation to the long and short VOT will be presented.

4.2.1. Overall performance

The following table presents participants' scores before and after the instruction. As stated above, the participants were tested with 36 stimuli with either long or short VOT (18 stimuli respectively).

Table 4. Descriptive Statistics of Comparison between Before and After Instruction

Test Time	Long VOT			Short VOT		
	Mean (Correct %)	S.D.	N	Mean (Correct %)	S.D.	N
Pre-Test	4.37 (72.8)	1.26	144	4.43 (73.8)	1.37	144
Post-Test	5.31 (88.5)	1.09	144	5.17 (86.2)	1.1	144
Delayed Post-Test	5.4 (90.0)	0.99	144	5.17 (86.2)	1.18	144

In placing the word boundary before stops with audible aspiration (long VOT), the participants demonstrated a significant improvement after instruction. They ultimately obtained 90% accuracy in perceiving word boundaries with audible VOT cues on the delayed post-test, for an increase in their listening performance of 17 percentage points. Regarding short VOT, where the participants need to be aware that there is no boundary before the target stop consonants, they also demonstrated improvement after instruction, but not as significant as the long VOT. Before instruction, the participants displayed 73.8% accuracy in placing word boundaries before the *s* of *s* + stop consonant sequences. With awareness of short VOT cues after word initial *s*, their accuracy increased to 86.2%. The standard deviation of the mean scores of the participants also decreased, implying more consistent performance.

4.2.2. Instructional effect on long VOT cues

Examining the results comprehensively indicated that the amount of improvement among the three places did not greatly differ. The following table presents the improvements at each place of articulation of the stop consonants.

Table 5. Descriptive Statistics of Long VOT Before and After Instruction

Place of Articulation		Bilabial Stop		Alveolar Stop		Velar Stop	
Test Time	N	Mean (Correct %)	S.D.	Mean (Correct %)	S.D.	Mean (Correct %)	S.D.
Pre-Test	48	4.71 (78.5)	0.99	4.38 (73)	1.41	4.02 (67)	1.28
Post-Test	48	5.5 (91.7)	1.09	5.4 (90.0)	1.03	5.02 (83.7)	1.12
Delayed Post-Test	48	5.67 (94.5)	0.69	5.5 (91.7)	0.95	5.04 (84.0)	1.18

In placing the word boundary before the stops with audible aspiration, the participants demonstrated consistent improvement after instruction across the three places of stop consonants. The participants scored highest on bilabial stop consonants after instruction, which is expected since the cue was considered the most easily acquired for this stop before instruction

as well. Participants also improved in perceiving word boundaries before velar stops, but ultimately the scores were the lowest for velar stops after instruction. The amount of improvement was greatest for alveolar, then velar, and least for bilabial stops. Comparison of the standard deviations' values indicated that velar stops demonstrated the least difference between the pre-test and delayed post-test, implying that among the three places of articulation, velar stops seemed to be the most difficult to internalize with learners displaying great variability in employing the cue. Repeated-measures ANOVA was conducted to investigate whether instructional effects were statistically significant. The following table presents the results of the multivariate tests.

Table 6. Multivariate Tests of Long VOT Before and After Instruction

Effect		Value	F	Hypothesis df	Error df	Sig.
Test Time	Wilks' Lambda	.615	43.887[b]	2.000	140.000	.000
Test Time * Place of Articulation	Wilks' Lambda	.991	.315[b]	4.000	280.000	.868

a. Design: Intercept + Place of Articulation
Within Subjects Design: Test Time

b. Exact statistic

c. The statistic is an upper bound on F that yields a lower bound on the significance level.

As presented in the table above, the instructional effect reached statistical significance (F = 43.887, p < .001). Pairwise comparisons by test time revealed that the difference was primarily due to that between the pre-test and the post-test (p < .000)/delayed post-test (p < .000). No significant change was reported between the post-test and delayed post-test (p = .193). The instructional effect was consistent across the three places of articulation, as indicated by the insignificance of the interaction of test time and place of articulation (F = .315, p = .868).

4.2.3. Instructional effect on short VOT cues

The following table indicates the improvements at each place of articulation of the stop consonants after word initial *s*.

Table 7. Descriptive Statistics of Short VOT Before and After Instruction

Place of Articulation		Bilabial Stop		Alveolar Stop		Velar Stop	
Test Time	N	Mean (Correct %)	S.D.	Mean (Correct %)	S.D.	Mean (Correct %)	S.D.
Pre-Test	48	4.42 (73.7)	1.29	4.77 (79.5)	1.45	4.1 (68.3)	1.32
Post-Test	48	5.13 (85.5)	1.21	5.29 (88.2)	0.97	5.1 (85)	1.12
Delayed Post-Test	48	4.9 (81.7)	1.48	5.33 (88.8)	1.02	5.27 (87.8)	0.96

The participants also displayed considerable improvement in perceiving *s* + stop sequences accurately. The amount of improvement was largest for velar stops. Ultimately, the participants performed best in alveolar stop perception. It is noteworthy that they had the most difficulty in segmenting bilabial stops after *s*. The accuracy obtained after instruction was not sustained on the delayed post-test, whereas the segmentation with alveolar and velar short VOT rather improved. They had an accuracy of 81.7% on bilabial stops after word initial *s*, well below the 94.5% of long VOT. Additionally, standard deviation increased on the delayed post-test.

Table 8. Multivariate Tests of Short VOT Before and After Instruction

Effect		Value	F	Hypothesis df	Error df	Sig.
Test Time	Wilks' Lambda	.748	23.645[b]	2.000	140.000	.000
Test Time * Place of Articulation	Wilks' Lambda	.935	2.390[b]	4.000	280.000	.051

a. Design: Intercept + Place of Articulation
Within Subjects Design: Test Time

b. Exact statistic

c. The statistic is an upper bound on F that yields a lower bound on the significance level.

Multivariate test results indicate that the instructional effect was also significant for short VOT (F = 23.645, p < .001) as well as long VOT. Pairwise comparisons indicated that this difference was primarily due to the differences between the pre-test and the post-test (p < .000)/delayed post-test (p < .000) (p = .933 for the difference between post-test and delayed post-test).

The interaction of test time and place of articulation should be noted here. As mentioned above, participants' performance on bilabial stops deteriorated on the delayed post-test, leading to an interesting interaction pattern between these two variables as demonstrated in Figure 1.

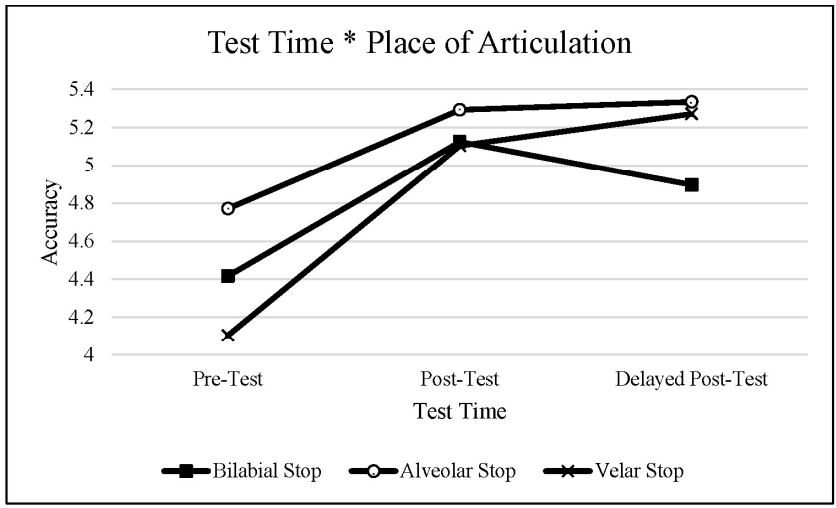

Figure 1. Interaction between Test Time and Place of Articulation in Short VOT.

As shown in the figure above, the accuracy results in bilabial stops deteriorated while alveolar and velar stop perception improved slightly in delayed post-test. Although the interaction did not reach significance (F = 2.390, p = .051), the participants seemed to struggle to learn short VOT compared to long VOT, especially for bilabial stops.

4.3. General Discussion

This study addressed two research questions regarding the instructional effect on VOT perception and L2 segmentation strategy. The first inquired whether the instructional effect would be the same for long and short VOT that signal the placement of the word boundary differently. The author hypothesized that long VOT related segmentation would display more improvement after instruction than short VOT related ones based on the influence of perceptual salience claimed by previous research (Altenberg 2005), which was supported by the experimental results of this study. However, examining the results thoroughly regarding the place of articulation of English stop consonants — which is the second research question — presented mixed results for long and short VOT cues, respectively. More specifically, the findings reveal rather contradicting patterns in terms of markedness and perceptual salience. In general, long VOT cues displayed patterns conforming to predictions corresponding to markedness theory. Conversely, short VOT cues seemed to conform to predictions corresponding to perceptual salience. The results are discussed comprehensively from the two perspectives, and the learnability of this type of segmentation cue is addressed.

4.3.1. Markedness Effect

According to Yavaş (2016), markedness theory predicts that unmarked structures are acquired more easily than marked ones. Therefore, it was hypothesized that the instructional effect would be greatest for bilabial stops, smaller for alveolar stops, and smallest for velar stops as the voicing distinction of velar stops is considered to be most marked among the three places of articulation (Yavaş 2016).

Before being provided with instruction on segmentation strategy regarding VOT cues, the participants performed best on bilabial stops and worst on velar stops when long VOT cues were involved, compliant with the prediction based on markedness theory (Yavaş 2016). In uninstructed settings without explicit attention to VOT cues at word boundary, it was observed that the more marked it was, the more difficult it was to acquire.

This study examined whether this tendency is observed in the instructional effect with explicit attention to VOT cues. Comparing the results between pre-test and delayed post-test indicated that the instructional effect was largest for alveolar stops, smaller for velar stops, and smallest for bilabial stops. If the relationship between markedness and ease or difficulty of learning was sustained due to the instructional effect, bilabial stops should have displayed the most increase in the accuracy of the responses.

However, we cannot assert that markedness effect was absent at this point because the difference among the three places is insignificant. In addition, the accuracy in perceiving word boundary with bilabial long VOT was higher than the other two places, and the instructional effect on bilabial stops might indicate that the participants' response accuracy cannot exceed this level (the ceiling effect). On delayed post-test, the participants scored 94.5% accuracy on perceiving bilabial long VOT at the word boundary. This is consistent with the correct response percentage of native English speakers in Altenberg (2005) displayed, which is 95% accuracy for long VOT cues in general. Among the three places of articulation of aspirated stops, the bilabial stop was the only place where learners attained native-like accuracy. Therefore, it can be argued that bilabial stops with long VOT are easiest in relevant L2 segmentation cues. In addition, the participants attained the lowest score for velar stops on delayed post-test, essentially 84% of correct responses. This indicates that velar stops, the most marked among the three, may take longer, and require more exposure/attention to the relevant cues to attain native-like accuracy in perception.

4.3.2. Perceptual Salience Effect

Altenberg (2005) mentions that the more perceptually salient they are, the more easily they are acquired, citing Suomi (1985) on Finnish word boundary perception. According to Altenberg (2005), this perceptual salience effect provides two structures — segmentation with long VOT and alveolar stops — with the ease of perception that leads to ease of acquisition. First, the presence of long VOT cues leads to the presence of word boundary and absence of cues to absence of word boundary.

The relationship between the cue and segmentation is more explicit for long VOT than short VOT. Moreover, long VOT cues are more perceptually salient. This predicts that the instructional effect would be more significant for segmentation with long VOT.

Before instruction, the students in this study demonstrated higher performance on long VOT cues than short VOT cues. This indicates that in uninstructed settings, long VOT was easier to acquire. Comparing before- and after-instruction indicated that the improvement after instruction was significant for both long and short VOT, but the amount of improvement for short VOT was not as prominent as long VOT. More exposure would be required to learn the 'negative signals' (Altenberg 2005: 328) linking absence of the cue to absence of the boundary.

Second, among the three places of articulation, alveolar stops have a perceptual advantage as they have higher intensity energy than bilabial and velar stops (Ferrand 2001). Before instruction, the results for the short VOT category indicated the highest accuracy for alveolar stops, then bilabial, and the lowest for velar stops. This partially conforms to Altenberg's (2005) results where Spanish learners of English perceived junctures best when alveolar stops were involved in segmentation. Altenberg (2005) did not provide separate analysis on long and short VOT, but in general interpreted the results to imply that the higher intensity energy of alveolar stops affected the Spanish participants' performance.

The Korean participants in this study displayed mixed results for the two VOT types. For long VOT, they ultimately scored highest for bilabial, then alveolar, and lowest for velar stops. However, the amount of improvement after instruction between pre-test and delayed post-test was greatest for alveolar stops. As for short VOT, they performed best on alveolar stops before instruction, ultimately achieving highest accuracy in the delayed post-test as well.

This result confirms that the students were assisted by the high intensity produced in alveolars during the segmentation process. In addition, how this intensity salience of alveolar stops interacts with long VOT cues in general is noteworthy. The results of this study imply that prior audible aspiration cues grant learners more ease in acquisition of L2 segmentation;

when these audible cues are unavailable, intensity cues adopt the role of primary salience.

This discussion seems to accommodate an interesting cross-linguistic observation regarding L2 segmentation cues by comparing this study's Korean learners and Altenberg's (2005) Spanish learners. As mentioned in Section 2, aspiration primarily contributes to the distinction of English stops. In Korean, long aspiration cues are present in aspirated stops, but the contribution of aspiration to stop distinction is limited. Spanish has no aspirated stops. This cross-linguistic difference seems to play an interesting role in L2 learners with different linguistic backgrounds acquiring VOT related English segmentation strategies. Essentially, this may allow us to establish positions among perceptual cues. Korean learners seemed to perceive both aspiration and intensity cues without much difficulty. Moreover, between the two, aspiration cues were more readily accessible by Korean learners as long VOT related segmentation was easier than short VOT related ones. Spanish learners of English might have been unable to perceive English aspiration cues as easily as Korean learners due to the lack of corresponding cues in their native language. Therefore, they naturally focused on the subsequent available cue that was most salient; in the segmentation of phrases in this study, it was high intensity spectral energy. The results of this study suggest further exploration of the relative perceptual salience of acoustic-phonetic cues involved in L2 segmentation from a cross-linguistic perspective.

4.3.3. Learnability

Finally, this section will conclude with a discussion on learnability of L2 segmentation cues. The linguistic features handled in this study are relatively less perceptually salient when compared to other linguistic features such as segments, intonation, vocabulary, grammar, etc. Thus, the acoustic cues not utilized in L1 are likely to be unnoticed during L2 acquisition. Therefore, in natural (uninstructed) settings, the question arises whether L2 learners can attain native-like proficiency in segmentation. Altenberg (2005) claims that although her participants acquired certain information required for L2 segmentation above the chance level they

barely attained native-like employment of segmentation cues. This may be the situation in uninstructed contexts where less salient cues (i.e., acoustic cues) are frequently unnoticed due to differences between L1 and L2 acoustic weighting systems. However, explicit instruction on acoustic cues and appropriate training on attention to acoustic cues may facilitate learners' acquisition of acoustic cues that are important in L2 speech streams. The findings of this study corroborate this, as participants were able to attain native-like perceptual proficiency in bilabial long VOT. Therefore, improving acoustic awareness in L2 pronunciation and listening teaching is advantageous. In EFL contexts with limited input and exposure time, attentional focus may be a significant factor in successful acquisition. The more attention learners allot to a form, the more they acquire. In this respect, a final remark can be made regarding the participants' low performance on bilabial short VOT and greatest improvement for velar short VOT.

The participants displayed considerable improvement in perceiving *s* + stop sequences correctly. Ultimately, the participants performed best in alveolar stop perception. However, the improvement was greatest for velar stops. Furthermore, they had the most difficulty in segmenting bilabial stops after *s*.

Lack of attention may provide an answer to this issue in part. Before instruction, the participants were already proficient at segmentation with long VOT cues of bilabial stops. As mentioned in Section 2, long VOT cues of bilabial stops may be easier to perceive. This may have rendered their short VOT counterparts to be perceived more easily, creating a sharp contrast between long and short VOT in bilabial stops. Moreover, standard deviations of VOT length were very short for bilabial stops, indicating a clearer distinction of VOT length in two positions. This, in turn, implies that learners found it relatively easy to learn the short VOT counterparts, reducing cognitive load. In other word, it is likely that learners did not have to concentrate on it as much as the other two places.

Instead, we may conclude that the reverse of this bilabial pattern was what occurred for velar short VOT. Velar stop voicing distinction is more difficult to acquire because it is most marked among the three places

of articulation. In addition, the length of aspiration of velar stops is not as audible as bilabials. The VOT length of the stimuli employed in this study exhibited significant standard deviations compared to bilabial stimuli. Therefore, they may have concentrated more on segmentation with short VOT of velar stops. Further research with a refined design is required to clarify this result.

5. Conclusion

This study examined whether instructional effect would differ between two conditions — long and short VOT cues at the beginning of a word and after initial *s*, respectively, in segmentation. As predicted, the instructional effect was greater for word initial long VOT in word boundary perception. However, the second research question's prediction was not supported entirely. It was predicted that the velar stops with long VOT would be most difficult and bilabial stops with long VOT the easiest to learn, as was confirmed by the experimental results. However, interestingly, the VOT of bilabial stops after word initial *s* transpired to be the most difficult to learn and alveolar and velar stops were easier than bilabials.

The instructional effect was significant, and participants demonstrated that explicit instruction and attention to acoustic awareness contributed significantly to their segmentation performance. This study therefore supports explicit instruction in the pronunciation of L2 English. Currently, pronunciation is not considered as important as other content in the L2 classroom, in that phonetic factors do not contribute as much to communication. However, certain instances where phonetic or acoustic details play important roles in communication do exist as dealt with in this study. Thus, more instructional consideration should be provided to L2 phonetics.

References

Abramson, Arthur S. and Lisker, Leigh. (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. Academic Press.

Ahn, Hyunkee. (1999). *Post-Release Phonatory Processes in English and Korean: Acoustic Correlates and Implications for Korean Phonology*. Unpublished doctoral dissertation, University of Texas, Austin.

Altenberg, Evelyn P. (2005). The perception of word boundaries in a second language. *Second Language Research* 21.4, 325-358.

Brown, H. Douglas. (2007). *Principles of Language Learning and Teaching* (5^{th} edition). Pearson Education.

Cho, Taehong. (1996). *Vowel Correlates to Consonant Phonation: An Acoustic-perceptual Study of Korean Obstruents*. Unpublished master's thesis, University of Texas at Arlington.

Christie, William M. (1974). Some cues for syllable juncture perception in English. *Journal of the Acoustical Society of America* 55, 819-821.

Clark, John and Yallop, Colin. (1995). *An Introduction to Phonetics and Phonology*. Blackwell.

Cohen, Anthony. (1987). Juncture revisited. In Chanon, R and Shockey, L. ed., *In honor of Ilse Lehiste*, 7-18. Foris Publications.

Ferrand, Carole T. (2001). *Speech Science: An Integrated Approach to Theory and Clinical Practice*. Allyn and Bacon.

Guion, Susan G. and Lee, Borim. (2006). The role of phonetic processing in second language acquisition. *English Language and Linguistics* 21, 123-148.

Han, Mieko S. and Weitzman Raymond S. (1970). Acoustic Features of Korean /P, T, K/, /p, t, k/ and /p^h, t^h, k^h/. *Phonetica* 22, 112-128.

Ingram, David. (1976). *Phonological Disability in Children*. Edward Arnold.

Kim, Mi-Ryoung, Beddor, Patrice Speeter, and Horrocks, Julie. (2002). The contribution of consonantal and vocalic information to the perception of Korean initial stops. *Journal of Phonetics* 30.1, 77-100.

Kim, Yoon-Hyun. (2007). *The Effect of Selective Attention on Discrimination between Lenis /t/ and Aspirated /th/ of Korean Alveolar Stops as shown by the Korean and the Japanese*. Unpublished doctoral dissertation, Seoul National University.

Klatt, Dennis H. (1974). The duration of [s] in English words. *Journal of Speech and Hearing Research* 17, 51-63.

Ladefoged, Peter. (1975). *A Course in Phonetics*. Harcourt Brace Jovanovich.

Ladefoged, Peter and Johnson, Keith. (2014). *A Course in Phonetics* (7^{th} edition).

Cengage Learning.
Lee, Ho-Young. (2000). The pronunciation of English consonant clusters by Koreans. *Journal of the Phonetic Society of Korea* 40, 79-89.
Lehiste, Ilse and Peterson, Gordon. E. (1961). Some basic considerations in the analysis of intonation. *Journal of Acoustical Society of America* 33.4, 419-423.
Nakatani, Lloyd H. and Dukes, Kathleen D. (1977). Locus of segmental cues for word juncture. *Journal of the Acoustical Society of America* 62, 714-719.
Oh, Eunjin. (2010). Effects of speaker gender on voice onset time in Korean stops. *Journal of Phonetics* 39, 59-67.
Rost, Michael. (2002). *Teaching and Researching Listening*. Longman.
Suomi, Kari. (1985). On detecting words and word boundaries in Finnish: A survey of potential word boundary signals. *Nordic Journal of Linguistics* 8, 211-231.
Um, Hye-Young. (2006). The perception of word boundaries by Korean college EFL learners. *The Linguistic Association of Korea Journal* 14.3, 51-70.
Umeda, Noriko and Coker, Cecil H. (1974). Allophonic variation in American English. *Journal of Phonetic*s 2, 1-5.
Wolfram, Walt and Johnson, Robert. (1982). *Phonological Analysis: Focus on American English.* Prentice Hall.
Yang, In Young. (2014). Acoustic Cues in Korean High School Students' L2 Phoneme Perception and Speech Segmentation. *Korean Journal of Applied Linguistics* 30.4, 235-264.
Yavaş, Mehmet. (2016). *Applied English Phonology* (3rd edition). Willey Blackwell.

In Young Yang
English Language Education Department,
Seoul National University,
Gwanakro 1 Gwanakgu, Seoul, Korea
E-mail: inyoung@snu.ac.kr

Received: October 31, 2017
Revised version received: December 11, 2017
Accepted: December 22, 2017