Language and Linguistics 95, 121-150 (2022)
DOI http://dx.doi.org/10.20865/20229505

Perception of Fortis and Lenis Stops of English in Word-Initial Position in Terms of VOT

Yongjin Sung* · Hyunsong Chung**
(Korea National University of Education)

Sung, Yongjin · Chung, Hyunsong. 2022. Perception of Fortis and Lenis Stops of English in Word-Initial Position in Terms of VOT. *Language and Linguistics* 95, 121-150. The purpose of this study was to investigate the differences between the responses of native Korean and English listeners to voice onset time (VOT) cues when distinguishing word-initial fortis and lenis stops in English. Additionally, this study examined the effects of the place of articulation and the height and backness of the following vowels. The first finding of the study is that the native Korean listeners needed a longer VOT to differentiate fortis from lenis stops. Second, the VOT boundary was the lowest for bilabial stops, while there was no statistically significant difference between alveolar and velar stops. Third, the boundary was found to be the highest for high vowels, the second highest for mid vowels, and lowest for low vowels. Finally, the boundary of high vowels was formed at a greater value than that of low vowels.

Keywords: speech perception, voicing contrast, word-initial stop, VOT boundary, English

---

* First author
** Corresponding author

1. Introduction

The perceptual difference of a speech sound between one speaker and another may function as an obstacle in verbal communication. For example, Korean stop consonants /p*/, /p/, and /pʰ/ as in 뿔 [p*ul], 불 [pul], and 풀 [pʰul] are virtually undistinguishable for native speakers of English, while voicing contrast in English is not readily identified by native speakers of Spanish. These perceptual differences in stop consonants have been researched by comparing various acoustic cues, such as voice onset time (VOT), fundamental frequency (F_0), and closure duration. Particularly, the production and perception of native and non-native stop contrast have been widely discussed with the VOT cue, which is known as the primary acoustic measure of phonetic differences in stop consonants (Cho & Ladefoged, 1999; Docherty, 1992; Klatt, 1975; Lisker & Abramson, 1964, 1967; Morris et al., 2008). VOT refers to the time between the burst of a stop and the start of vocal cord vibration for the following vowel. The VOT values have been reported to vary depending on various phonetic or socio-phonetic factors, such as vowel contexts, places of articulation of stop consonants, speaking rate, utterance types (e.g., spoken in isolated syllables or in sentences), speaker age, and speaker gender (Oh, 2019). Several studies have examined the VOT boundaries of stop phonemes of English by adult native Korean listeners in terms of VOT (Lim & Han, 2014; Schertz et al., 2015; Sung et al., 2020), but few have considered all three stop places of bilabial, alveolar, and velar contrast (/b/-/p/, /d/-/t/, and /g/-/k/). Furthermore, little attention has been given to the cross-linguistic comparisons between native English and Korean listeners within the context of various vowel qualities. Thus, it is necessary to shed more light on the perceptual L2 (English) stop contrast by native Korean listeners in the comparison with the perceptual L1 stop contrast of native English listeners in various points of view. This study makes an attempt to fill the gap of the current literature.

The main purpose of this study is to investigate how different native English and Korean listeners respond to the VOT cue when distinguishing between lenis and fortis stops of English (/b, d, g/ vs. /p, t, k/), specifically in the word-initial positions. In addition, this study examines the effects of the stop place of articulation (bilabial, alveolar, and velar), the heights of the following vowels (high, mid, and low), and the backness of the following vowels (front and back) with the use of VOT cue to perceptually differentiate fortis stops from lenis stops.

2. Literature Review

A number of studies explored the perceptual VOT boundaries between the lenis and fortis stops of English in word-initial position, focusing on either or both of the native English and native Korean listeners (Choi, 2012; Lim & Han, 2014; Masip, 2020; McCarthy et al., 2014; Oh, 2019; Park & Kang, 2006; Schertz et al., 2015; Sung et al., 2020; Yun, 2019). The majority of these studies designed multiple steps of VOT continua for the sound stimuli of their lenis vs. fortis forced-choice identification experiments. While some studies reported the specific mean values for the perceptual VOT boundary between lenis and fortis (Masip, 2020; McCarthy et al., 2014), other studies offered the rates of responses of the fortis stops instead (Lim & Han, 2014; Oh, 2019; Schertz et al. 2015; Sung et al. 2020; Yun, 2019).

Some studies presented the native English listeners' VOT boundaries between the lenis and fortis stops of English in word-initial position (Masip, 2020; McCarthy et al., 2014; Oh, 2019). Masip (2020) demonstrated the effects of the L2 experience on both the perception and production of L1 and L2 bilabial and velar stops. Regarding the perception, the study found most of the Spanish learners of English differentiated the L2 stops from the L1 stops, whereas in the case of the English learners of Spanish, they heard both L1 and L2 stops with L1-like (English-like) VOT values.

There were some other studies that reported native Korean listeners' L2 (English) stop phoneme contrasts as well (Lim & Han, 2014; Schertz et al., 2015; Sung et al., 2020; Yun, 2019). Lim and Han (2014) investigated whether there were differences between adult Kyungsang and Seoul speakers when it came to the production and perception of their L2 word-initial bilabial stops, particularly by using the pair pea-bee (/pi/-/bi/). The study revealed that there was no significant difference between the two dialect speakers' VOT and F_0 patterns in production, whereas in terms of perception, the Kyungsang listeners had a greater reliance on the VOT cue and less on the F_0 cue, contrary to the Seoul listeners.

Although some previous studies have examined the phoneme boundaries of English lenis and fortis stops by adult native Korean listeners in terms of VOT (Lim & Han, 2014; Schertz et al., 2015; Sung et al., 2020), few have dealt with all bilabial, alveolar, and velar stop contrasts. Also of note, little research has examined the effect of the following vowels on the auditory distinction of L2 (English) stop phonemes by native Korean listeners. Making an attempt to fill in the pieces missing from the current literature, the research questions of this study were set as follows:

(1) Do native English listeners and native Korean listeners differ from each other in their perception of fortis and lenis stops of English in the word-initial position in terms of VOT?
(2) Does the place of articulation of a word-initial stop affect the perception of fortis and lenis stops of English in the word-initial position in terms of VOT?
(3) Does vowel height affect the perception of fortis and lenis stops of English in the word-initial position in terms of VOT?
(4) Does vowel backness affect the perception of fortis and lenis stops of English in the word-initial position in terms of VOT?

3. Method

3.1 Participants

Twenty-two participants (10 native Korean and 12 native English speakers) participated in the experiment who had no speech, language, or hearing problems. The native Korean listener group (hereafter NKL group) included 10 adult listeners (six females and four males; mean age = 39.2). Their majors were not related to English, and none of them had stayed in an English-speaking country for more than six months according to their self-reports. The Korean listeners varied in their accent-backgrounds, ranging from Busan, Daegu, and Seoul, and also spanned in age from 20s, 30s, and 50s. On the other hand, the native English listener group (hereafter NEL group) included

12 native adult listeners (12 females; mean age = 27.3). The English listeners were a mix of diverse nationalities including Canada, South Africa, the U.K. and the U.S. Their lengths of residence in Korea spanned from seven to 84 months (mean length = 35). All the participants were compensated for their participation.

3.2 Stimuli

Stimuli were created from 42 CVC(C)-syllable English words (21 minimal pairs) containing a fortis or lenis stop in the word-initial position which was followed by an /i/, /ɪ/, /ɛ/, /æ/, /ɑ/, /ɔ/, or /u/ vowel. The word-initial stop was bilabial, alveolar, or velar (/p/-/b/, /t/-/d/, or /k/-/g/). All words were real words of English except keld and geld which were pseudo-pairs that include a word-initial velar stop followed by a mid-front vowel. 〈Table 1〉 shows the target words. Each target word in 〈Table 1〉 was spoken in a carrier sentence 'Let's say __ again' by a male native speaker of American English three times, in a row. The second or third one was used for stimuli among the three tokens.

	/p/-/b/	/t/-/d/	/k/-/g/
/i/	peach-beach	teen-dean	keek-geek
/ɪ/	pig-big	tip-dip	kill-gill
/ɛ/	pest-best	tech-deck	keld-geld
/æ/	pack-back	tab-dab	cap-gap
/ɑ/	palm-bomb	tart-dart	card-guard
/ɔ/	pour-bore	tore-door	core-gore
/u/	pooh-boo	tomb-doom	coup-goo

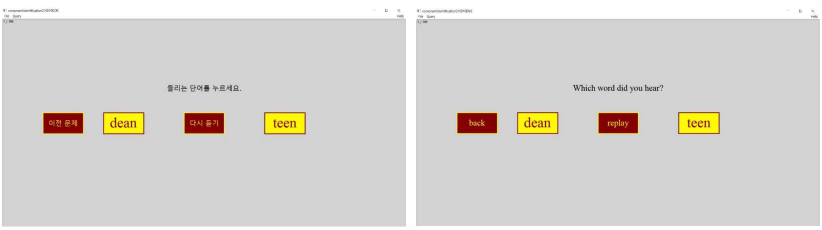
〈Table 1〉 Target words

Stimuli varied by VOT in 18 steps for each /b/-/p/, /d/-/t/, and /g/-/k/. Following the method used by the previous studies (Andruski et al., 1994; Kong & Edwards, 2016; Kong & Yoon, 2013; Winn et al., 2013), onset portions of words with the word-initial lenis stops (/b/, /d/, and /g/) were progressively replaced with the equivalent portion (5 ms) of onset aspiration from their fortis stop counterparts (/p/, /t/, and /k/). Accordingly, the VOT range spanned from 5 ms to 90 ms in the 18 steps (i.e., 5, 10, 15, ⋯, 85, and 90 ms) for each /b/-/p/, /d/-/t/, and /g/-/k/. In other words, an 18-step VOT continuum, containing 18 target tokens, was created for each minimal pair. Additionally, the original F_0 value during the vowel was automatically adapted to each VOT step using Praat 6.1.39 (Boersma & Weenink, 2019).

The fully designed stimuli with all the places of articulation, VOT, and the places of vowel equated to 378 tokens (3 places × 18 VOT steps × 7 vowels). In addition to the target items, 18 filler items with the word-initial /s/-/ʃ/, /l/-/r/, and /m/-/h/ were also added, for example, seat-sheet, lead-read, and meat-heat. Thus, in total, 396 tokens were created for the experiment.

3.3 Procedure

Each participant was tested with the forced-choice identification experiment that consisted of the 396 tokens split into two sessions (198 tokens for each session). A break time was offered in between the two sessions in order for the participants to maintain their focus throughout the experiment. In auditory stimulus of a single word was presented binaurally over headphones with two words on the computer screen. The participants were asked to choose which of the two words on the screen was pronounced by clicking one of the two words with the mouse. The word on the left side of the screen included a word-initial lenis stop (/b/, /d/, or /g/), and its minimal pair counterpart with the word-initial fortis stop (/p/, /t/, or /k/) was presented on the right side. A replay button was provided in the middle of the screen in order for the participants to be able to repeat the stimuli to clarify what they had heard. Also, a back button was offered on the very left of the screen to let the participants change their previously marked answers. 〔Figure 1〕 shows the screenshots of the experiment implemented using Praat.



[Figure 1] Screenshots of the experiment (Dean vs. Teen for each listener group)

3.4 Data analysis

This current study defined the perceptual VOT boundary as the first fortis-stop response on the VOT continuum. However, there were the cases that the lenis response(s) occurred in between the fortis responses. It should be noted that the VOT boundaries for this kind of lenis-fortis traversed data were judged by the fortis-stop response right after the final lenis-stop response on the VOT continuum. The speech perception data from the 22 participants were statistically analyzed using a two-way analysis of variance (ANOVA) with a general linear model of *SPSS* 12.0. The effects of listeners' L1, the place of articulation, vowel height, and vowel backness (and their interactions) were measured at a .05 significant level.

4. Results

4.1 The effects of L1 and place of articulation

Dependent Variable	L1	Place	*Mean*	*SD*	*n*
Boundary	Korean	Bilabial	44.93	13.39	70
		Alveolar	57.57	14.54	70
		Velar	60.43	12.21	70
		Total	54.31	14.96	210
	English	Bilabial	37.74	10.16	84
		Alveolar	57.02	14.42	84
		Velar	55.71	9.76	84
		Total	50.16	14.56	252
	Total	Bilabial	41.01	12.24	154
		Alveolar	57.27	14.43	154
		Velar	57.86	11.16	154
		Total	52.04	14.87	462

〈Table 2〉 Descriptive statistics of L1 and place of articulation (in milliseconds)

The effects of L1 and place of word-initial stops (and their interaction) on the VOT boundaries within 18 steps of VOT continua (5, 10, 15, ..., 85, and 90 ms) were investigated. 〈Table 2〉 presents the descriptive statistics of L1 and place of articulation. Ten native Korean listeners were grouped in the NKL group (M = 54.31), and twelve native English listeners made up the NEL group (M = 50.16). The VOT boundary was the lowest for the bilabial stops (M = 41.01), while the mean gap between the alveolar stops (M = 57.27) and the velar stops (M = 57.27) was noticeably short, equaling .59.

In addition to the descriptive statistics, a 2 × 3 factorial ANOVA was performed. Levene's test indicated that the assumption of homogeneity of variances had not been met, F(5, 456) = 8.23, p 〈 .05. 〈Table 3〉 presents the summarization of the 2 × 3 factorial ANOVA of L1 and place of articulation.

Source	*Type III Sum of Squares*	*df*	*Mean Square*	*F*	*Sig.*	*Eta Squared*
Corrected Model	31009.78	5	6201.96	39.84	.00	.30
Intercept	1250105.11	1	1250105.11	8030.85	.00	.95
L1	1973.51	1	1973.51	12.68	.00	.03
Place	27267.98	2	13633.99	87.59	.00	.28
L1 * Place	860.62	2	430.31	2.76	.06	.01
Error	70982.26	456	155.66			
Total	1353425.00	462				
Corrected Total	101992.04	461				

〈Table 3〉 2 × 3 factorial ANOVA of L1 and place of articulation

The results revealed that L1 had a statistically significant main effect on the VOT boundary, F(1, 460) = 12.68, p 〈 .05, with a small partial eta^2 value, .03. Also, the place of the word-initial stops had a statistically significant main effect on the VOT boundary, F(2, 459) = 87.59, p 〈 .05, with a medium partial eta^2 value, .28. In contrast, the interaction of L1 and the place of the word-initial stops on the VOT boundary was not statistically significant, F(5, 456) = 2.76, p = .06.

	Mean Difference	*Std. Error*	*Sig.*	*95% Confidence Interval*		*d*
				Lower Bound	*Upper Bound*	
Bilabial-Alveolar	-16.27	1.52	.00	-19.93	-12.60	1.22
Bilabial-Velar	-16.85	1.33	.00	-20.05	-13.65	1.44
Alveolar-Velar	-0.58	1.47	.97	-4.11	2.94	.05

〈Table 4〉 Multiple comparisons between the three groups of place of articulation

In order to further examine the relationships between the bilabial, alveolar, and velar groups, the post hoc multiple comparisons of Dunnett's T3 were administered. 〈Table 4〉 presents the summarization of the multiple comparisons between the three groups of place of articulation. As shown from the 95% CIs for each group-comparison, there were two separate homogeneous subsets: alveolar/velar and bilabial (bilabial 〈 alveolar = velar).

4.2 The effects of L1 and vowel height

The effects of L1 and vowel height (and their interaction) on the VOT boundaries within 18 steps of VOT continua were examined. 〈Table 5〉 shows the descriptive statistics of L1 and vowel height. The VOT boundary was the highest for the high vowel group (M = 60.48), the second-highest for the mid vowel group (M = 48.18), and the lowest for the low vowel group (M = 43.26). Moreover, each mean VOT boundary in the high, mid, and low vowel of the NKL group (63.44 ms, 50.25 ms, and 44.67 ms) was observed at the later VOT steps than each one of the NEL group (58.01 ms, 46.46 ms, and 42.08 ms).

Dependent Variable	L1	Vowel Height	*Mean*	*SD*	*n*
Boundary	Korean	High	63.44	15.46	90
		Mid	50.25	10.10	60
		Low	44.67	9.56	60
		Total	54.31	14.96	210
	English	High	58.01	15.16	108
		Mid	46.46	9.91	72
		Low	42.08	11.47	72
		Total	50.16	14.56	252
	Total	High	60.48	15.50	198
		Mid	48.18	10.14	132
		Low	43.26	10.68	132
		Total	52.04	14.87	462

〈Table 5〉 Descriptive statistics of L1 and vowel height (in milliseconds)

In addition to the descriptive statistics, a 2 × 3 factorial ANOVA was conducted. Levene's test indicated that the assumption of homogeneity of variances had not been met, $F(5, 456) = 7.91$, $p < .05$. 〈Table 6〉 shows the summarization of the 2 × 3 factorial ANOVA of L1 and vowel height. The outcome indicated that L1 had a statistically significant main effect on the VOT boundary, $F(1, 460) = 10.60$, $p < .05$, with a small partial eta^2 value, .02. Additionally, vowel height had a statistically significant main effect on the VOT boundary, $F(2, 459) = 81.80$, $p < .05$, with a medium partial eta^2 value, .26. On the other hand, the interaction of L1 and vowel height on the VOT boundary was not statistically significant, $F(5, 456) = .51$, $p = .60$.

Source	*Type Ⅲ Sum of Squares*	*df*	*Mean Square*	*F*	*Sig.*	*Eta Squared*
Corrected Model	28388.87	5	5677.78	35.18	.00	.28
Intercept	1141012.03	1	1141012.03	7069.01	.00	.94
L1	1711.81	1	1711.81	10.60	.00	.02
Vowel Height	26408.00	2	13204.00	81.80	.00	.26
L1 * Vowel Height	165.61	2	82.81	.51	.60	.00
Error	73603.17	456	161.41			
Total	1353425.00	462				
Corrected Total	101992.04	461				

〈Table 6〉 2 × 3 factorial ANOVA of L1 and vowel height

In order to further understand the relationships between the high, mid, and low vowel groups, the post hoc multiple comparisons of Dunnett's T3 were performed. 〈Table 7〉 presents the summarization of the multiple comparisons of the three groups of vowel height. As 95% CIs for each group-difference show, there were three separate homogeneous subsets: high, mid, and low (low 〈 mid 〈 high).

	Mean Difference	*Std. Error*	*Sig.*	*95% Confidence Interval*		*d*
				Lower Bound	*Upper Bound*	
High-Mid	12.30	1.41	.00	8.91	15.68	.91
High-Low	17.22	1.44	.00	13.76	20.68	1.27
Mid-Low	4.92	1.28	.00	1.84	8.00	.03

〈Table 7〉 Multiple comparisons between the three groups of vowel height

4.3 The effects of L1 and vowel backness

The effects of L1 and vowel backness (and their interaction) on the VOT boundaries within 18 steps of VOT continua were investigated. 〈Table 8〉 shows the descriptive statistics of L1 and vowel backness. The mean VOT boundary of the back vowels (*M* = 53.56) was higher than that of the front vowels (*M* = 50.90), the mean difference equating to 2.66. Furthermore, the mean VOT boundary for each front and back vowels of the NKL group (52.96 ms and 56.11 ms) was found at the later VOT steps than those of the NEL group (49.20 ms and 51.43 ms).

Dependent Variable	L1	Vowel Backness	*Mean*	*SD*	*n*
Boundary	Korean	Front	52.96	16.08	120
		Back	56.11	13.19	90
		Total	54.31	14.96	210
	English	Front	49.20	15.98	144
		Back	51.43	12.38	108
		Total	50.16	14.56	252
	Total	Front	50.91	16.11	264
		Back	53.56	12.93	198
		Total	52.04	14.87	462

〈Table 8〉 Descriptive statistics of L1 and vowel backness (in milliseconds)

In addition to the descriptive statistics, a 2 × 2 factorial ANOVA was performed. Levene's test indicated that the assumption of homogeneity of variances had not been met, $F(3, 458) = 5.16$, p 〈 .05. 〈Table 9〉 presents the summarization of the 2 × 2 factorial ANOVA of L1 and vowel backness.

Source	*Type III Sum of Squares*	*df*	*Mean Square*	*F*	*Sig.*	*Eta Squared*
Corrected Model	2792.66	3	930.89	4.30	.00	.03
Intercept	1233629.70	1	1233629.70	5695.62	.00	.93
L1	1944.87	1	1944.87	9.21	.00	.02
Vowel Backness	813.93	1	813.93	3.76	.05	.01
L1 * Vowel Backness	23.69	1	23.69	.11	.74	.00
Error	99199.39	458	216.59			
Total	1353425.00	462				
Corrected Total	101992.04	461				

〈Table 9〉 2 × 2 factorial ANOVA of L1 and vowel backness

The results showed that L1 had a statistically significant main effect on the VOT boundary, $F(1, 460) = 9.21$, $p < .05$, with a small partial eta^2 value of .02. However, the main effect of vowel backness was not statistically significant, $F(1, 460) = 3.76$, $p > .05$. Also of note, the interaction of L1 and vowel backness on the VOT boundary was not statistically significant, $F(3, 458) = .11$, $p = .74$.

4.4 The effects of place of articulation and vowel height

The effects of the place of the word-initial stops and vowel height (and their interaction) on the VOT boundaries within 18 steps of VOT continua were examined. 〈Table 10〉 shows the descriptive statistics of stop place of articulation and vowel height. The mean VOT boundaries of the bilabial stops followed by the high, mid, and low vowels were 44.85, 42.84, and 33.41 ms. Those of the alveolar stops read as 71.44, 50.23, and 43.07 ms, while those of the velar stops equated to 65.15, 51.48, and 53.29 ms.

Dependent Variable	Place	Height	*Mean*	*SD*	*n*
Boundary	Bilabial	High	44.85	12.18	66
		Mid	42.84	12.96	44
		Low	33.41	7.45	44
		Total	41.01	12.24	154
	Alveolar	High	71.44	9.27	66
		Mid	50.23	5.60	44
		Low	43.07	3.77	44
		Total	57.27	14.43	154
	Velar	High	65.15	10.07	66
		Mid	51.48	8.32	44
		Low	53.29	8.69	44
		Total	57.86	11.16	154
	Total	High	60.48	15.50	198
		Mid	48.18	10.14	132
		Low	43.26	10.68	132
		Total	52.04	14.87	462

〈Table 10〉 Descriptive statistics of stop place and vowel height (in milliseconds)

In addition to the descriptive statistics, a 3 × 3 factorial ANOVA was conducted. Levene's test indicated that the assumption of homogeneity of variances had not been met following, $F(8, 453) = 8.82$, $p < .05$. 〈Table 11〉 shows the summarization of the 3 × 3 factorial ANOVA of place of articulation and vowel height. The results revealed that the place of word-initial stops had a statistically significant main effect on the VOT boundary, $F(2, 459) = 135.65$, $p < .05$, with a large partial eta^2 value, .37. Vowel height also had a statistically significant main effect on the VOT boundary, $F(2, 459) = 150.03$, $p < .05$, with a large partial eta^2 value, .40. Additionally, the interaction of the place of word-initial stops and vowel height on the VOT boundary was statistically significant, $F(8, 453) =$

22.69, $p < .05$, with a small partial eta² value, .17. This interaction will be further discussed in Section 5.2.

Source	*Type III Sum of Squares*	*df*	*Mean Square*	*F*	*Sig.*	*Eta Squared*
Corrected Model	62363.64	8	7795.45	89.11	.00	.61
Intercept	1142432.32	1	1142432.32	13059.36	.00	.97
Place	23732.59	2	11866.30	135.65	.00	.37
Height	26249.75	2	13124.87	150.03	.00	.40
Place * Height	7938.24	4	1984.56	22.69	.00	.17
Error	39628.41	453	87.48			
Total	1353425.00	462				
Corrected Total	101992.04	461				

〈Table 11〉 3 × 3 factorial ANOVA of place of articulation and vowel height

4.5 The effects of place of articulation and vowel backness

Dependent Variable	Place	Backness	*Mean*	*SD*	*n*
Boundary	Bilabial	Front	36.93	10.60	88
		Back	46.44	12.24	66
		Total	41.01	12.24	154
	Alveolar	Front	59.26	15.00	88
		Back	54.62	13.28	66
		Total	57.27	14.43	154
	Velar	Front	56.53	12.11	88
		Back	59.62	9.54	66
		Total	57.86	11.16	154
	Total	Front	50.91	16.11	264
		Back	53.56	12.93	198
		Total	52.04	14.87	462

〈Table 12〉 Descriptive statistics of stop place and vowel backness (in milliseconds)

The effects of the place of word-initial stops and vowel backness (as well as their interaction) on the VOT boundaries within 18 steps of VOT continua were examined. 〈Table 12〉 shows the descriptive statistics of stop place of articulation and vowel backness.

The mean VOT boundaries of bilabial stops followed by the front and back vowels were 36.93 ms and 46.44 ms, while those of the alveolar stops read as 59.26 ms and 54.62 ms. Those of the velar stops equated to 56.53 ms and 59.62 ms.

In addition to the descriptive statistics, a 3 × 2 factorial ANOVA was performed. Levene's test indicated that the assumption of homogeneity of variances had not been met, $F(5, 456) = 7.22$, $p < .05$. 〈Table 13〉 presents the summarization of the 3 × 2 factorial ANOVA of place of articulation and vowel backness.

Source	*Type III Sum of Squares*	*df*	*Mean Square*	*F*	*Sig.*	*Eta Squared*
Corrected Model	32756.25	5	6551.25	43.15	.00	.32
Intercept	1234831.82	1	1234831.82	8132.83	.00	.95
Place	25279.95	2	12639.98	83.25	.00	.27
Backness	795.45	1	795.45	5.24	.02	.01
Place * Backness	3785.15	2	1892.57	12.46	.00	.05
Error	69235.79	456	151.83			
Total	1353425.00	462				
Corrected Total	101992.04	461				

〈Table 13〉 3 × 2 factorial ANOVA of place of articulation and vowel backness

The results showed that the place of word-initial stops had a statistically significant main effect on the VOT boundary, $F(2, 459) = 83.25$, $p < .05$, with a medium partial eta^2 value, .27. Also of note, vowel backness had a statistically significant main effect on the VOT boundary, $F(1, 460) = 5.24$, $p < .05$, with a small partial eta^2 value, .01. Following this, the interaction of the place of word-initial stops and vowel backness on the VOT boundary was statistically significant, $F(5, 456) = 12.46$, $p < .05$, with a small partial eta^2 value, .05. This interaction will be further discussed in Section 5.2.

4.6 The effects of vowel backness and vowel height

The effects of vowel backness and vowel height (as well as their interaction) on the VOT boundary within the 18 steps of VOT continua were examined. 〈Table 14〉 shows the descriptive statistics of vowel backness and vowel height. The mean VOT boundaries of the high, mid, and low front vowels were 58.94 ms, 44.32 ms, and 41.44 ms. Those back-vowel counterparts read as 63.56 ms, 52.04 ms, and 45.07 ms.

Dependent Variable	Backness	Height	*Mean*	*SD*	*n*
Boundary	Front	High	58.94	17.11	132
		Mid	44.32	10.91	66
		Low	41.44	8.63	66
		Total	50.91	16.11	264
	Back	High	63.56	11.12	66
		Mid	52.04	7.59	66
		Low	45.07	12.20	66
		Total	53.56	12.93	198
	Total	High	60.48	15.50	198
		Mid	48.18	10.14	132
		Low	43.26	10.68	132
		Total	52.04	14.87	462

〈Table 14〉 Descriptive statistics of vowel backness and vowel height (in milliseconds)

In addition to the descriptive statistics, a 2 × 3 factorial ANOVA was conducted. Levene's test indicated that the assumption of homogeneity of variances had this time also not been met, $F(4, 456) = 13.45$, $p < .05$. 〈Table 15〉 presents the summarization of the 2 × 3 factorial ANOVA of vowel backness and vowel height.

The results revealed that vowel backness had a statistically significant main effect on the VOT boundary, $F(1, 460) = 19.31$, $p < .05$, with a small partial eta^2 value, .04. Additionally, vowel height also had a statistically significant main effect on the VOT boundary, $F(2, 459) = 85.25$, $p < .05$, with a medium partial eta^2 value, .27. On the other hand, the interaction of vowel backness and vowel height on the VOT boundary was not statistically significant, $F(5, 456) = .96$, $p = .38$.

Source	*Type III Sum of Squares*	*df*	*Mean Square*	*F*	*Sig.*	*Eta Squared*
Corrected Model	29596.21	5	5919.24	37.28	.00	.29
Intercept	1119074.45	1	1119074.45	7048.72	.00	.94
Backness	3066.18	1	3066.18	19.31	.00	.04
Height	27070.08	2	13535.04	85.25	.00	.27
Backness * Height	305.83	2	152.92	.96	.38	.00
Error	72395.83	456	158.76			
Total	1353425.00	462				
Corrected Total	101992.04	461				

〈Table 15〉 2 × 3 factorial ANOVA of vowel backness and vowel height

5. Discussion

5.1 The differences between native Korean and English listeners

In respect to the perception of the native English listeners, this present study obtained the VOT boundaries of the bilabial, alveolar, and velar stops equating to 37.74, 57.02, and 55.71 ms of VOT. These results are partially consistent with those of Sung et al. (2020) which reported the VOT boundaries (50% cross-over points on the gradients of the fortis response rates) in between 34-46, 44-63, and 57.5-72 ms of VOT for the bilabial, alveolar, and velar stops. Conversely, Masip (2020) proposed that the mean VOT boundary between /b/ and /p/ by monolingual English listeners was 14 ms and the one between /g/ and /k/ was 29.7 ms of VOT. These VOT boundary values are noticeably lower than those found in this current study. In contrast to Masip's (2020) English monolingual listeners, the native English listeners in the present study have lived in Korea for 35 months in average. This implies that the greater VOT boundary values for the NEL group in the current study than those in Masip (2020) may have been affected by their L2 experience in Korea. On the other hand, Oh (2019) suggested that the VOT boundary between /d/ and /t/ (50% cross-over points on the gradients of the rates of the fortis-stop response) was formed at less than 28 ms of VOT in male stimuli and at less than 18 ms of VOT in female stimuli. Considering that this current study used a male native speaker's recording as its stimuli, the native English speakers in the NEL group needed 29.02 ms longer VOT to differentiate /t/ from /d/ than those in Oh's (2019).

When it comes to the perception of native Korean listeners, this present study found the VOT boundaries for the bilabial, alveolar, and velar stops as 44.93, 57.57, and 60.43 ms of VOT. These results were partially consistent with Sung et al. (2020) which obtained the VOT boundaries (50% cross-over points on the gradients of the fortis-stop response rates) in between 22-34, 44-63, and 43-57.5 ms of VOT. Additionally, Lim and Han (2014) reported the VOT boundary between /b/ and /p/ by adult native Korean listeners as less than 30 ms, and Yun (2019) reported the boundary between /b/ and /p/ by Korean female high school students to be in between 44-60 ms of VOT (50% cross-over points on the gradients of the fortis-stop response rates). The /b/-/p/ boundary of the NKL group in the current study (44.93 ms) corresponds to Yun's (2019).

5.2 The differences between bilabial, alveolar, and velar stops

As shown in Section 4.1, the current study revealed that the bilabial stops were understood as the fortis stop (/p/) in shorter VOTs than the alveolar and velar stops had by both the NKL and NEL groups. This result was consistent with the previous studies (Kim, 2021; Miller, 1977; Nakai & Scobbie, 2016; Sung et al., 2020), where bilabial stops were classified as fortis during the shorter VOT steps than the alveolar and velar stops had. Also of note, this current study showed the alveolar and velar stops were not statistically different from each other in terms of VOT perception, as shown in the post hoc multiple comparisons located in Section 4.1. This outcome opposes the previous reports where the alveolar stops had the clear intermediate position between bilabial and velar stops.

There were statistically significant two-way interactions between the place of word-initial stops and the other two vowel-related variables. First, the interaction of the place of word-initial stops and vowel height proposes that, in the case of the velar stops, the lower the following vowel, the higher the VOT boundary, whereas for the bilabial and alveolar stops, the higher the following vowel, the higher the VOT boundary. Generally speaking, the velar stops have a higher point of articulation than the bilabial and alveolar stops. As a result of the increased space in the oral cavity, it is likely that more air pressure builds up between the velar stops and the low vowel. On the other hand, based on the data alone, it is difficult to generalize the patterns for the bilabial and alveolar stops.

In addition, the statistically significant two-way interaction between the place of word-initial stops and vowel backness demonstrates that, for the alveolar stops, the farther front the vowel, the higher the boundary, whereas for the bilabial and velar stops, the farther back the vowel, the higher the boundary. However, there was no obvious pattern in this interaction in terms of the distance between the word-initial stop and the following vowel. The distance could be precisely measured by ultrasound technology, which is now being used to analyze the articulatory gestures of the vocal organs, particularly the movement of the tongue (Davidson, 2006; Mielke, 2015; Tabain et al., 2020).

5.3 The differences between high, mid, and low vowels

The results of ANOVA showed that the perceptual separation between the fortis and lenis stops varied with the height of the following vowels. This was consistent with the previous studies

(Higgins et al., 1998; Klatt, 1975; Nakai & Scobbie, 2016; Nearey & Rochet, 1994; Summerfield, 1975, 1981) in that stop consonants tend to be accompanied by a longer VOT value when they were followed by a phonologically high vowel than they would have been followed by a non-high vowel. As Higgins et al. (1998) discussed, greater VOT values in the context of high vowels may have resulted from the greater oral and glottal constrictions of air flow produced when high vowels are formed. Additionally, the result was also partially consistent with Nakai and Scobbie's (2016) where a large VOT boundary value was observed for the high vowels while a small value was found for the low vowels in regards to the alveolar and velar stops. As shown in Section 4.4, the VOT boundary was the highest for the alveolar-high vowels (M = 71.44) and the second-highest for the velar-high vowels (M = 65.15), which corresponds with Nakai and Scobbie's (2016). However, the average VOT boundary of the velar-low vowels (M = 53.29) was higher than that of the velar-mid vowels (M = 51.48) in this current study, opposing Nakai and Scobbie's (2016) result.

5.4 The differences between front and back vowels

There was a statistically significant two-way interaction between vowel backness and the place of word-initial stops. The average VOT boundary of the alveolar stops was higher than that of the velar stops in regard of the front vowels, whereas inversely with the back vowels, the boundaries of the velar stops were higher than those of the alveolar stops. These opposite results between the front and the back vowels caused this particular interaction. This VOT change may have caused by the different contact points of the velar stops in accordance with the back/front nature of the following vowels. Yavaş (2020) stated that the velar stops are more front before a front vowel than before a back vowel. That is, it can be said that the word-initial velar stops (/k, g/) followed by the back vowels (/u, ɔ, ɑ/) as in coup-goo, core-gore, and card-guard were made at a significantly more back point in velum than those that contained the front vowels (/i, ɪ, ɛ, æ/) as in keek-geek, kill-gill, keld-geld, and cap-gap.

6. Conclusion

The main goal of this study was to investigate how different the native Korean and English listeners react to VOT cue when it comes to differentiating lenis and fortis stops of English (/b, d, g/ and /p, t, k/), specifically in word-initial position. Additionally, this study explored the effects of the place of articulation (bilabial, alveolar, and velar), the height of the following vowel (high, mid, and low), and the backness of the following vowel (front and back) on using VOT cue to distinguish the fortis stops from the lenis stops.

To summarize the findings of the present study, there were in fact largely remarkable differences between the two listener groups who have varying L1, the three groups of the stop places of articulation, the three groups of vowel height, and the two groups of vowel backness in the perceptual VOT boundaries when it came to the fortis and lenis stops. Additionally, two important additional interactions were also observed for this study: 1) the place of articulation of word-initial stops and vowel height and 2) the place of articulation of word-initial stops and vowel backness.

Despite this study's findings and its unique distinctions with previous literature, there were some limitations that were faced which makes for a few suggestions when it comes to future research. The major restriction with this study was the theoretical and technical resources were not sufficient enough to discuss the interactions in depth. There were two statistically significant interactions: 1) between the stop places of articulation and vowel height and 2) between the stop places of articulation and vowel backness. Although both interactions were discussed within available reasoning, it would be more reliable if ultrasound technology was utilized in future research to address the exact points of each place of articulation so that we could further observe the physical distance between the consonant and the following vowel.

References

Andruski, J. E., Blumstein, S. E., and M. Burton. (1994). The effect of subphonetic differences on lexical access. *Cognition* 52: 163-187.

Boersma, P. and D. Weenink. (2021). Praat: Doing phonetics by computer (Version 6.1.39). Retrieved July 22, 2021 - https://www.fon.hum.uva.nl/praat

Choi, H. (2012). Effects of acoustic cues in the perception of English stop laryngeal consonants by English and Korean listeners. *Korea Journal of English Language and Linguistics* 12(2): 357-378.

Cho, T. and P. Ladefoged. (1999). Variation and universality in VOT: Evidence from 18 languages. *Journal of Phonetics* 27: 207-229.

Davidson, L. (2006). Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance. *The Journal of the Acoustical Society of America* 120: 407-415.

Docherty, G. J. (1992). *The timing of voicing in British English obstruents*. Foris Publication.

Higgins, M. B., Netsell, R., and Schulte, L. (1998). Vowel-related differences in laryngeal articulatory and phonatory function. *Journal of Speech, Language, & Hearing Research* 41: 712-724.

Kim, M-R. (2021). Voice onset time in English and Korean stops with respect to a sound change. *Phonetics and Speech Sciences* 13(2): 9-17.

Klatt, D. (1975). Voice onset time, frication, and aspiration in word-initial consonant clusters. *Journal of Speech & Hearing Research* 18: 686-706.

Kong, E. and J. Edwards. (2016). Individual differences in categorical perception of speech: Cue weighting and executive function. *Journal of Phonetics* 59: 40-57.

Kong, E. J. and I. H. 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.

Lim, S. and J.-I. 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 Phonetics, Phonology, and Morphology* 20(3): 403-426.

Lisker, L. and A. Abramson. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word* 20: 384-422.

Lisker, L. and A. Abramson. (1967). Some effects of context on voice onset time in English stops. *Language & Speech* 10: 1-28.

Masip, C. G. (2020). "Crosslinguistic influence in L2 speakers: The effect of language experience on L1 and L2 VOT production and perception". Doctoral Dissertation. Universitat Autònoma de Barcelona.

McCarthy, K. M., Mahon, M., Rosen, S., and B. G. Evans. (2014). Speech perception and production by sequential bilingual children: A longitudinal study of voice onset time acquisition. *Child Development* 85(5): 1965-1980.

Mielke, J. (2015). An ultrasound study of Canadian French rhotic vowels with polar smoothing spline comparisons. *The Journal of the Acoustical Society of America* 137: 2858-2869.

Miller, J. 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.

Morris, R. J., McCrea, C. R., and K. D. Herring. (2008). Voice onset time differences between adult males and females: Isolated syllables. *Journal of Phonetics* 36: 308-317.

Nakai, S. and J. M. Scobbie. (2016). The VOT category boundary in word-initial stops: Counter-evidence against rate normalization in English spontaneous speech. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 7(1): 1-31.

Oh, E. (2019). Effects of gender on the use of voice onset time and fundamental frequency cues in perception and production of English stops. *Linguistic Research* 36(1): 67-89.

Park, H. and S.-H. Kang. (2006). A cross-linguistic study of the perception of the voicing contrast in English plosives. *Journal of the Linguistic Society of Korea* 45: 3-37.

Schertz, J., Cho, T., Lotto, A., and N. Warner. (2015). Individual differences in phonetic cue use in production and perception of a non-native sound contrast. *Journal of Phonetics* 52: 183-204.

SPSS statistics. (2003). SPSS 12.0 for Windows (Version 12.0.0). IBM.

Summerfield, Q. (1975). How a full account of segmental perception depends on prosody and vice versa. In Cohen, A., & Nooteboom, S. G. (eds.), *Structure and process in speech perception*. Springer-Verlag.

Summerfield, Q. (1981). Articulatory rate and perceptual constancy in phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance* 7(5): 1074-1095.

Sung, E., Lee, S., and S. 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.

Tabain, M., Kochetov, A., and R. Beare. (2020). An ultrasound and formant study of manner contrast at four coronal places of articulation. *The Journal of the Acoustical Society of America* 148(5): 3195-3217.

Winn, M. B., Chatterjee, M., and W. J. Idsardi. (2013). 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.

Yavaş, M. (2020). *Applied English Phonology* (4th ed.). Wiley-Blackwell.

Yun, Y. (2019). L1 transfer in perception of English stops by Korean female teenagers. *The Journal of Linguistics Science* 90: 279-299.

Yongjin Sung
Korea National University of Education
Department of English Education
Graduate Student
250, Taeseongtabyeon-ro, Gangnae-myeon, Heungdeok-gu, Cheongju-si, Chungbuk 28173, Korea
e-mail: ysungkorea@gmail.com

Hyunsong Chung
Korea National University of Education
Department of English Education
Professor
250, Taeseongtabyeon-ro, Gangnae-myeon, Heungdeok-gu, Cheongju-si, Chungbuk 28173, Korea
e-mail: hchung@knue.ac.kr

Received: Jan. 10, 2022
Revised: Feb. 14, 2022
Accepted: Feb. 21, 2022