pISSN 2005-8063
eISSN 2586-5854
2021. 6. 30.
Vol.13 No.2
pp. 9-17

말소리와 음성과학

Phonetics and Speech Sciences

한국음성학회지

https://doi.org/10.13064/KSSS.2021.13.2.009

Voice onset time in English and Korean stops with respect to a sound change*

Mi-Ryoung Kim**

Department of Practical Foreign Languages, Korea Soongsil Cyber University, Seoul, Korea

Abstract

Voice onset time (VOT) is known to be a primary acoustic cue that differentiates voiced from voiceless stops in the world's languages. While much attention has been given to the sound change of Korean stops, little attention has been given to that of English stops. This study examines VOT of stop consonants as produced by English speakers in comparison to Korean speakers to see whether there is any VOT change for English stops and how the effects of stop, place, gender, and individual on VOT differ cross-linguistically. A total of 24 native speakers (11 Americans and 13 Koreans) participated in this experiment. The results showed that, for Korean, the VOT merger of lax and aspirated stops was replicated, and, for English, voiced stops became initially devoiced and voiceless stops became heavily aspirated. English voiceless stops became longer in VOT than Korean counterparts. The results suggest that, similar to Korean stops, English stops may also undergo a sound change. Since it is the first study to be revealed, more convincing evidence is necessary.

Keywords: voice onset time (VOT), English stops, Korean stops, sound change, VOT change, three dimensions

1. Introduction

The languages of the world can vary depending on how many stops there are. Some languages have a two-stop category and others have a three- or four-stop category (Ladefoged & Maddieson, 1996). The former consists of English, French, and Polish and the latter Korean, Thai, and Hindi. Although the languages have the same number of categories, they differ in terms of voicing and aspiration. For example, although French and English have a voiced and voiceless contrast, French belongs to a true voicing language but English does not (Keating, 1984). Thai and Korean have a three-way contrast. Thai has a voiced-voiceless contrast, whereas Korean doesn't. To distinguish these stop categories, many phonetic parameters have been examined. One of the most popular acoustic parameters is voice onset time (the time interval stop release and onset of vocal cord vibration, henceafter VOT) (Lisker & Abramson, 1964).

Examining the eleven languages including two- to four-stop categories, Lisker & Abramson (1964) found that VOT served well to separate the stop categories of these languages in speech

* An earlier version of this paper was presented at Seoul International Conference on Speech Sciences (SICSS) on November 2017 in Seoul, Korea. I am very grateful to three anonymous reviewers for their thoughtful comments. I'd like to thank all participants. All the remaining errors are mine.

** kmrg@mail.kcu.ac, Corresponding author

Received 30 May 2021; Revised 18 June 2021; Accepted 18 June 2021

production. Abramson & Lisker (1970)'s perception study support that VOT serves as a primary cue for the distinction between voiced and voiceless pairs. Since then, VOT has come to be regarded as one of the best acoustic parameters for discriminating stops categories, especially in word-initial position. Stop categories across languages are categorized into the three VOT dimensions, voicing lead, short voicing lag, and long voicing lag (Keating, 1984; Lisker & Abramson, 1964). 'Voicing lead' refers to negative VOT values occurring before the release burst, whereas 'voicing lag' refers to positive VOT measurements occurring after the release burst. Voicing lag is further subdivided into 'short lag voicing' and 'long lag' voicing. They are well represented in Figure 1.

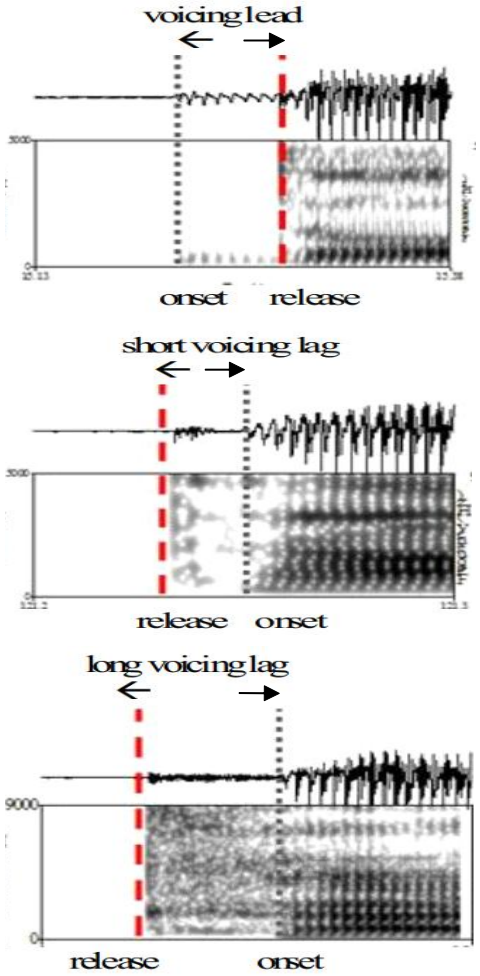

Figure 1. Wide-band spectrograms and waveforms showing three conditions of VOT for English /d/ and /t/ from top to bottom (Cited from Kim, 2011: 99).

It is well known that English stops are well divided into the three dimensions, while Korean stops are not (Lisker & Abramson, 1964 and among many others).

1.1. VOTs in English and Korean stops

In English, there are voiced and voiceless stops that voiced stops /b, d, g/ are realized as either voiced or devoiced and voiceless stops are realized as either aspirated or unaspirated. When voiced stops are fully voiced, they are realized with voicing lead (negative VOT). When they are devoiced, they are realized with short-lag VOT (Lisker & Abramson, 1964). When voiceless stops are aspirated, they are realized with long-lag VOT. When they are unaspirated (e.g., /ptk/ after /s/), they are realized with short-lag VOT. These phonetic realizations are well presented in early findings below.

Table 1. Mean VOTs and ranges (in parentheses) of English stops

Stop	Lisker & Abramson (1964)	Docherty (1992)	Chodroff & Wilson (2017)
/b/	1/–101	15	13 (11–20)
/d/	5/–102	21	21 (14–32)
/g/	21/–88	27	28 (19–42)
/p/	58 (20–120)	42	89 (46–139)
/t/	70 (30–105)	64	98 (57–156)
/k/	80 (50–135)	62	99 (67–137)

VOT, voice onset time.

Table 1 shows that VOT changes are taking place. Previous studies have shown that some English speakers produce some or all of their initial voiced stops with prevoicing rather than short-lag VOT. For example, Lisker & Abramson (1964) provide two sets of VOT values for English voiced stops, one with a positive short lag, and the other with a negative voicing lead (Flege, 1982). In contrast, Docherty (1992) reports positive values only for both voiced /b, d, g/ and voiceless /p, t, k/ stops (see also Klatt, 1975). Keating (1984) discusses that English voiced stops are sometimes realized with some lead values but mainly realized with short lag and long lag values, corresponding to Chodroff & Wilson's (2017). In Table 1, mean VOTs for voiceless stops are getting longer. In Chodroff & Wilson (2017), their grand means for the voiceless stops were somewhat higher than previous findings. They speculate that it reflects an overall slow speaking rate in their experiment. Taking the results of previous studies, it is worth investigating the change in VOT of English stops.

With respect to the VOT boundary between voiced and voiceless stops in English, some studies report that the VOT of /b/ is less than 15 ms, whereas that of /p/ is more than 30 ms (Lieberman & Blumstein, 1988; Nakai & Scobbie, 2016). Other studies report that 25 ms was said to be the boundary between /p/ and /b/ (Abramson & Lisker, 1970). Based on previous reports, it is not easy to determine the ranges of VOT for English voiced and voiceless stops. Keating (1984) points out that, since English stops have a great deal of positional variation, it is hard to simply refer to VOT boundary with a specific number. In addition, there is some overlap in the VOT range of /b/ and /p/. This issue will be examined in this study.

In English, voiceless stops are unaspirated when they follow /s/. They are expected to have short-lag VOT (Keating, 1984). However, less attention was given to their VOT values on how voiceless unaspirated stops differ from devoiced stops as well as Korean tense stops. These questions are answered in the current study by examining them separately from voiceless aspirated stops in English.

In Korean, instead of a voiced-voiceless contrast, there is a three-stop voiceless category, called as 'tense', 'lax', and 'aspirated' stops and transcribed as [p*, t*, k*] (or [p', t', k']), [p, t, k], and [p^h, t^h, k^h], respectively (Han & Weitzman, 1970; Kim, 1965). Because of the lack of a voicing contrast, they are sometimes defined as 'unaspirated, slightly aspirated, and heavily aspirated' in terms of the amount of aspiration. The greater aspiration, the longer the VOT. Table 2 presents mean VOTs of Korean stops from previous

findings where it is clear that VOT changes are taking place. In Lisker & Abramson (1964), on average, VOTs are shortest (12 ms) for Korean tense stops, longer (30 ms) for lax, and longest (103 ms) for aspirated. There was a VOT overlap between tense and lax stops (Han & Weitzman, 1970; Kim, 1965). However, in the studies of Kim (1994) and Kim (2020), mean VOTs for lax stops are getting longer to 51 ms and 68 ms, while those for aspirated stops are getting shorter to 78 ms and 81 ms. That is, VOT merger between lax and aspirated stops have occurred, replicating previous findings such as Kim (2008), Silva (2006), Kang (2014) and among others.

Table 2. Mean VOTs and ranges (in parentheses) of Korean stops

Stop	Lisker & Abramson (1964)	Kim (1994)	Kim (2020)
Tense	12 (0–35)	9 (9–11)	14 (4–48)
Lax	30 (10–65)	51 (15–78)	68 (10–153)
Asp.	103 (65–200)	78 (75–87)	81 (28–145)

It is clear that Korean stops are hard to be differentiated in terms of short-lag and long-lag VOT. Many recent studies have already reported that VOT alone is not sufficient to contrast three voiceless stops in Korean and Korean stops are undergoing a sound change (Silva, 2006 and among many others). Since Kim (2000), fundamental frequency (f0 or tone) has arisen as a primary acoustic parameter in distinguishing lax from aspirated stops in Korean (e.g., '달' [tʰàl] and '탈' [tʰál]) (Kim et al., 2002 and among many others).

Unlike English stops, Korean stops have had a long-standing problem not to be explained well under the three VOT dimensions (Lisker & Abramson, 1964 and among many others). Much attention was given to each language, but not much attention was given to the cross-linguistic comparison between English and Korean stops. In addition, very few studies have examined English unaspirated stops. Including English voiceless unaspirated stops after /s/. separately, the research questions in this study are as follows:

1. Are there any VOT changes for English stops?
2. Are there any cross-linguistic differences on the effects of stop, place, gender, and individual on VOT?
3. Are English unaspirated and Korean tense stops similar or different in terms of VOT?
4. Are English voiceless (aspirated) stops and Korean aspirated (or lax) stops similar or different in terms of VOT?
5. Are the VOT boundary between short-lag and long-lag clear to differentiate stop types?

The current study is designed to answer these questions.

2. Method

2.1. Participants

Twenty four students (11 American and 13 Korean) at different universities participated in the experiment. The Korean data were replicated from Kim (2014). For the English data, eleven native speakers of American English (7 female and 4 male) participated. All were undergraduates or graduates at a university in the Western area of America. The mean age was 24.6 and the individuals ranged from 20 to 29. All but two spoke a western dialect. For the Korean data, thirteen native speakers of Seoul Korean (7 females, 6 males; mean age 25 years) participated. All were undergraduates at a university in Seoul, Korea. They grew up in Seoul and spoke the standard dialect of Seoul. All participants reported that they had no experience in living in an English speaking country.

All participants resided in their countries at the time of the recording. For participants employed in the current study, socio-indexical factors such as age, gender, and dialect were controlled to minimize any plausible effects. In addition, speaking rate was also controlled (see section 2.3). No speakers had any history of speaking pathology nor any phonetic training.

2.2. Speech materials

For the English data, eighteen monosyllabic words were constructed from a balanced list of the three stop types (voiced, voiceless aspirated, voiceless unaspirated after /s/) and the three places of articulations (labial, alveolar, and velar) followed by a vowel /a/ context. The syllable type was either CV(C) or sCV(C) where the final coda was an unreleased stop [t]. The words consisted of real and nonsense words (e.g., bot, pot, dot, tot, got, cot, ba, pa, da, ta, ga, ka, spa, sta, ska, spot, stot, scot). Voiceless unaspirated stops (s/ptk/) were intentionally added to see how they differ from English devoiced stops and Korean tense stops.

The same Korean data as in Kim (2014) were employed. Eighteen monosyllabic words (3 stops×3 places×2 syllable types) were constructed from a balanced list of the three stop types (lax, aspirated, and tense) and the three places of articulations (labial, alveolar, and velar) followed by a vowel /a/ context. The syllable type was either CV or CVC where the final coda was an unreleased stop [t] or [k]. The words consisted of real and nonsense words in Korean (e.g., [pat] 'farm', [tat] 'anchor', [gat] 'hat', [pa] 'method', [ta] 'all', [ka] 'the edge', [pʰat] 'red bean', [tʰat] 'blame', [kʰat] 'not applicable', NA', [pʰa] 'green onion', [tʰa] 'Get in!', [kʰa] 'NA', [p*ak] 'head'(slang), [t*ak] 'precisely', [k*ak] 'NA', [p*a] 'to grind', [t*a] 'to pick', and [k*a] 'to peel').

2.3. Procedure

Prior to recording, participants completed a brief language background questionnaire and has a practice session. They read target words in "Say _____ again" in English or [igə _____ *hasɛjo*] "Say this _____" in Korean where the words were located in phrase-initial position. In the position, the target words were expected to be fully emphasized. Sentences were presented either in English or *Hangul*, the writing system of Korean.

Recordings were done in either Korea or America. In Korea, recordings were done in a quiet office directly into a Samsung SENS NT900XC4C-A78 laptop computer using a Safa voice recorder (model SR-M190N). In the US, recordings were made in a sound-attenuated booth in the Phonetics Lab using a Shure (model SM 10A) head-mounted microphone. Recordings were saved on a flash card using a Marantz digital recorder (PMD 670) at a sampling rate of 44.1 kHz. Each speaker was asked to read the words on the monitor in a natural intonation. The monitor connected to the computer was inside the lab but the computer itself was outside the lab to minimize background noise. The words were automatically popped up at a 3-second interval. This can control speech rate by using the frame sentence and limiting the speakers' production time of sentences to a fixed 3 sec. There were 792 English tokens (18 words×11 speakers×4 repetitions) and 936 Korean tokens (18 words×13 speakers×4 repetitions), respectively. A total of 1,728

tokens excluding filler words were obtained for analysis. All utterances were analysed using Praat 6.0.37, a speech analysis program (Boersma & Weenink, 2018).

VOT was measured from the release burst of a stop to the onset of periodicity in the waveform (Lisker & Abramson, 1964). The onset of the vowel in the waveform was determined by the onset of the first full glottal pulse of the vowel as well as the pitch of the spectrogram. The onset of the voicing energy in the second formant shown in a time-locked spectrogram was used to help determine voicing onset in conjunction with the waveform. The onset of voicing (=vowel onset) was defined as the first and periodic pulse of a vocalic waveform that show features typical of a vowel. The measurements were semi-automatically done in the following steps: 1) all target words were labeled into consonants (C) and vowels (V) using waveforms and spectrograms, 2) the duration of C and V were automatically obtained using a textgrid.

The VOT data obtained were statistically tested using repeated measures analysis of variance (RM ANOVA) in the context of a univariate context of a general linear model (SPSS/PASW, 2017). The main goal of the present study was to examine whether there is any main effect of stop (i.e., voiceless, voiced 1 & 2, s/p, t, k/ for English and aspirated, lax, and tense for Korean), place (i.e., labial, alveolar, and velar), gender (i.e., male vs. female), and individual speakers on VOT and whether there was any interaction effect between factors. Repeated measures ANOVA includes both "between" subjects effects (language and gender) and "within" subject effects (i.e., stop and place). Their main and interaction effects were statistically analyzed at a 0.05 significant level. *Post hoc* Tukey HSD (Honest Significant Difference) multiple tests were also run to answer the following questions: (i) for each factor, whether any differences in pairs among types were significant and (ii) for any factor, whether any differences in pairs among the stop types in the individual data were significant. For statistical analysis, the VOT differences among stop types in the two languages were mainly focused because of most central interest in the current paper.

3. Results

3.1. The English results

For the group-normative data averaged across eleven native speakers of English, mean VOTs of English stops according to four phonetic types are illustrated in Figure 2. For the voiced category, fully voiced stops with negative VOT values were labeled as "voiced 1" and devoiced stops with positive VOT values were statistically labeled as "voiced 2." For the voiceless category, voiceless unaspirated stops were labeled as "s/ptk/" and voiceless aspirated stops were labeled as "voiceless" for short. Results of univariate RM ANOVA showed a main effect of stop on VOT [F(3, 691)=4,073.064, p<.0001], in that VOT was positively long for voiceless (104.4 ms), short for both s/ptk/ (20.9 ms) and voiced 2 (20.1 ms), and negatively long for voiced 1 (−115.2 ms). *Post hoc* Tukey HSD multiple comparisons revealed that differences in any pairs except for the voiced 2-s/ptk/ pair were statistically significant (p<0.5). When voiced stops were devoiced, their VOTs were statistically the same as those of s/ptk/, as clearly seen in Figure 2. This finding is the first to reveal that the VOT difference between devoiced and s/ptk/ stops were not statistically significant. However, this does not mean that they are phonetically the same because of other phonetic correlates.

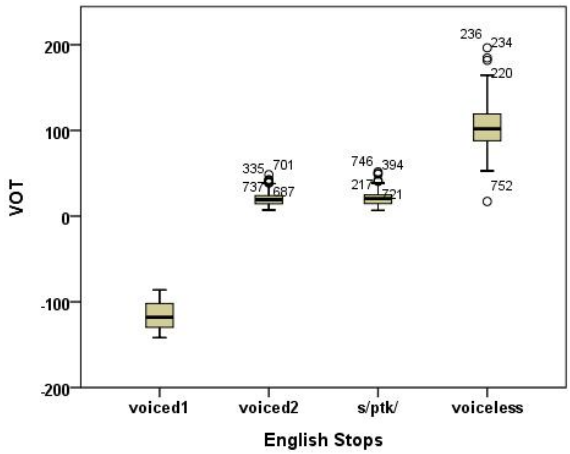

Figure 2. Mean VOTs (ms) of English stops according to voiced 1, voiced 2, s/ptk/, and voiceless (aspirated) stops aggregated from 11 American speakers (a total of 792 tokens). Error bars indicate ±1 standard deviation.

Table 3 presents the numeric VOT value of each phonetic type and their VOT dimensions along with means, ranges, and the number of tokens.

Table 3. Mean VOTs and ranges (in parentheses) for each stop for English stops

Stop	Type	Dimension	Mean (R)	N
/bdg/	Voiced 1	Voicing lead	−115 (−141- −86)	3
	Voiced 2	Short lag	20 (7-48)	261
/ptk/	s/ptk/	Short lag	21 (7-52)	264
	Voiceless	Long lag	104 (17-196)	264

The results in Table 3 tells that, for the voiced category, almost all speakers produced voiced stops with a short-lag VOT, especially in phrase-initial position. This correspond to Chodroff & Wilson (2017), Docherty (1992), Keating (1984), and Klatt (1975), but does not analogous to Flege (1982) and Lisker & Abramson (1964). Taken the findings together, it is safe to say that initial voiced stops in English became fully devoiced. For the voiceless category, mean VOTs for voiceless (aspirated) stops were even longer than those in Chodroff & Wilson (2017), suggesting that English voiceless stops became heavily aspirated. In this study, the devoiced and lengthening process can be considered a sound change.

Concerning the VOT boundary between voiced and voiceless category, as pointed out by Keating (1984), it is not easy to determine the ranges of VOT for English voiced and voiceless stops. The results in Table 3 tells that there is a remarkable overlap in the VOT range of /b/ and /p/. Thus, it is hard to simply refer to VOT boundary with a specific number as previous studies did (Abramson & Lisker, 1970; Lieberman & Blumstein, 1988; Nakai & Scobbie, 2016). However, since there was a huge VOT difference (84 ms) between voiced and voiceless stops, the three VOT dimensions seem to work well for the English data.

Next, there was a main effect of place of articulation on VOT [F(2, 691)=116.671, p<.0001] in that mean VOTs were shortest (40.8 ms) for labials, longer (50.1 ms) for alveolars, and longest (53.1 ms) for velars (labials<alveolars<velars). *Post hoc* Tukey HSD multiple comparisons revealed that differences among the

three places were statistically significant (p<.05). Table 4 shows that, although there is some overlap in the range, the effects of place of articulation on VOT hold well for across stop types. The place effect have been extensively documented in the literature for a variety of languages (Docherty, 1992; Klatt, 1975; Lisker & Abramson, 1964).

Table 4. Mean VOTs and ranges (in parentheses) of English stops for the places of articulation

Place	Labial	Alveolar	Velar
Voiced1	NA	−86 (1 token)	−129 (−141- −118)
Voiced2	14 (7-35)	23 (11-41)	27 (7-48)
s/ptk/	15 (7-34)	21 (9-42)	27 (13-52)
Voiceless	94 (17-160)	111 (62-185)	109 (61-196)
Mean	41 (7-160)	50 (−86-185)	53 (−141-196)

Next, consider the effects of gender on VOT. Figure 3 represents mean VOTs of stops according to gender and Table 5 presents their values. There was a main effect of gender on VOT [F(1, 691)= 105.231, p<.0001] in that mean VOTs were statistically shorter (45.7 ms) for female speakers than those (52.1 ms) for male speakers (female<male).

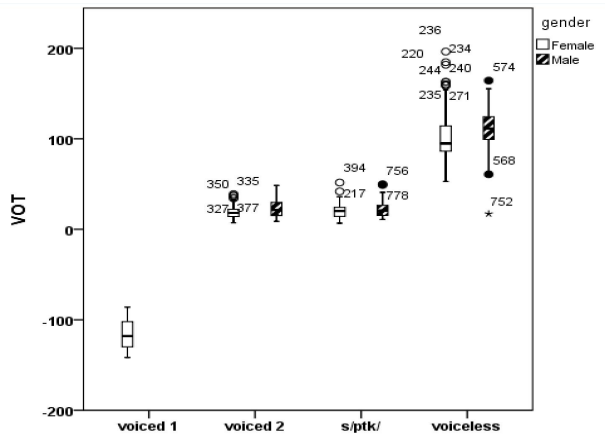

Figure 3. Mean VOTs (ms) of English stops according to voiced 1, voiced 2, s/ptk/, and voiceless stops by male and female speakers. Error bars indicate ±1 standard deviation.

Table 5. Mean VOTs and ranges (in parentheses) of English stops for female and male speakers

Place	Female	Male
Voiced 1	−115 (−141- −86)	NA
Voiced 2	18 (7-39)	23 (8-48)
s/ptk/	20 (6-51)	22 (10-49)
Voiceless	101 (53-196)	111 (17-164)
Mean	46 (−141-196)	52 (8-164)

Taken Figure 3 and Table 5 together, it is clear that the effects of gender on VOT were very small and linguistically not important. We can speculate that the difference may be mainly due to only one female speaker who produced both voiced and voiceless stops slightly different from other speakers, as discussed below.

Next, consider how individual speakers produced four phonetic types. Figure 4 illustrates individual speaker's VOT production and Table 7 presents their numeric values. Statistical analysis confirmed significant differences among individual speakers [F(9, 691)=25.667, p<.0001] and a weak interaction between stop and individual [F(18, 691)=20.983, p<.0001] in that individual speakers produced stops differently. However, the differences were not very big, as clearly seen in Figure 4. Out of eleven speakers, one female speaker AF4's production alone was remarkably different from others. Her data was intentionally included to capture individual speaker variations.

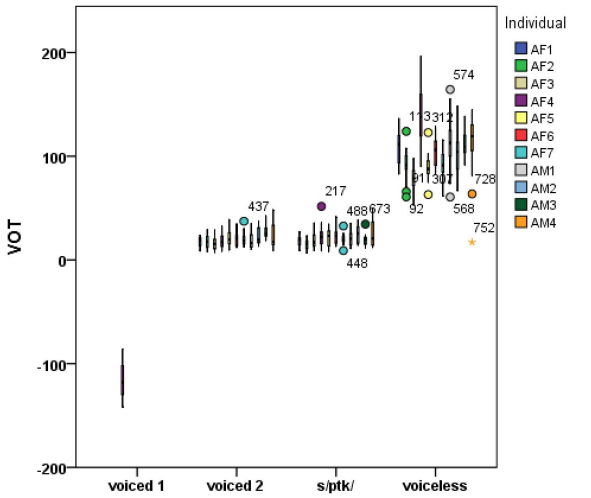

Figure 4. Mean VOTs (ms) of English stops according to voiced 1, voiced 2, s/ptk/, and voiceless aggregated from each individual speaker. Error bars indicate ±1 standard deviation.

Table 6. Mean VOTs and ranges (in parentheses) of English stops for each individual speaker

Spk.	Voiced 1	Voiced 2	s/pt/	Voiceless
AF1	−115 (−142- −86)	17 (9–24)	17 (9–27)	110 (83–136)
AF2		17 (8–29)	15 (7–23)	91 (61–124)
AF3		15 (7–29)	19 (9–36)	77 (53–98)
AF4		18 (8–33)	23 (9–52)	141 (90–196)
AF5		21 (9–39)	22 (8–35)	89 (63–123)
AF6		20 (12–35)	23 (13–42)	104 (83–129)
AF7		20 (12–37)	20 (9–33)	93 (62–116)
Mean		18 (7–39)	20 (7–52)	101 (53–196)
AM1		18 (10–35)	21 (11–35)	114 (61–164)
AM2		23 (13–37)	25 (14–38)	104 (67–148)
AM3		28 (19–43)	19 (12–34)	112 (92–138)
AM4		23 (9–48)	25 (12–50)	113 (17–145)
Mean		23 (9–48)	23 (11–50)	111 (17–164)

Spk, speaker; AF, American female; AM, American male.

Taking Figure 4 and Table 6 into a consideration together, we can conclude that there are little speaker variations on VOT for the production of the English stop types. This can be comparable to those of the Korean stops next section.

In summary, the results of English stops show the main effects of stop, place, gender, and speaker on VOT. Voiced stops were rarely produced with voicing lead but mostly with short voicing lag. Devoiced stops were statistically the same as voiceless unaspirated stops after /s/. Mean VOTs for voiceless aspirated stops were relatively longer than previous findings. The results in the current study suggest that English stops undergo a sound change in VOT.

3.2. The Korean results

Figure 5 shows mean VOT of Korean stops. For the pool data, results of univariate RM ANOVA showed a main effect of stop on

VOT [F(2, 933)=1,188.657, p<0.001] in that mean VOT values were significantly longer (83 ms) for the aspirated and lax stop (71 ms) than for the tense stop (15 ms). Although mean VOTs for aspirated stops were higher than those for lax stops, the VOT differences between them were not statistically significant.

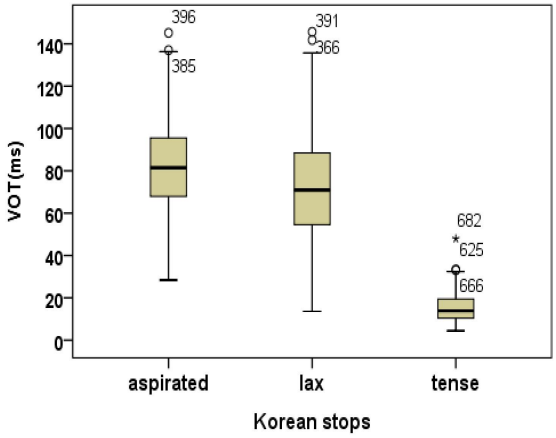

Figure 5. Mean VOTs (ms) of Korean stops according to aspirated, lax, and tense aggregated from 13 Seoul Korean speakers (a total of 936 tokens). Error bars indicate ±1 standard deviation.

Post hoc Tukey HSD multiple comparisons revealed that there were two significant subsets: aspirated/lax and tense (tense< lax=aspirated). The present results correspond to previous findings on the VOT merger of lax and aspirated stops (Kang, 2014; Kim, 2020; Silva, 2006). The results suggest that Korean stops truly undergo a sound change.

There was a main effect of place of articulation on VOT [F(2, 933)=11.500, p<.0001] in that mean VOTs were shortest (51 ms) for labials, longer (54 ms) for alveolars, and longest (64 ms) for velars (labials<alveolars<velars). The effect of place of articulation on VOT held well across the three types of stops, as seen in Figure 6(a). Pairwise comparisons revealed that differences in each pair among the three places were statistically significant (p<.005).

There was also a main effect of gender on VOT [F(1, 930)= 40.634, p<.0001] and an interaction between phonation and gender F(2, 930)=28.429, p<.0001] in that VOT differences between the aspirated and lax stop were significantly smaller for female speakers than male speakers. Mean VOTs for both aspirated and lax stops were significantly longer for female speakers than male speakers, as can be seen in Figure 6(b).

Table 7 summarizes individual mean VOT values for Korean stops. Compared to the English individual data, inter- and intra-speaker variations on VOT in Korean were much greater.

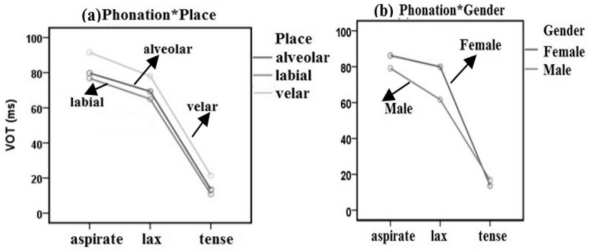

Figure 6. Mean VOTs (ms) of (a) Phonation*Place and (b) Phonation *Gender according to aspirated, lax, and tense targets in initial position. Error bars indicate ±1 standard deviation.

Table 7. Mean VOTs and ranges (in parentheses) of Korean stops for individual speakers

Spk.	Aspirated	Lax	Tense	Asp-Lax
SF1	94 (68-131)	72 (41-100)	14 (7-30)	23
SF2	119 (89-145)	111 (87-146)	14 (7-28)	9
SF3	87 (69-103)	81 (57-102)	14 (9-21)	0.4
SF4	70 (44-102)	60 (34-86)	15 (7-30)	9
SF5	92 (71-129)	96 (84-110)	11 (6-21)	–4.1
SF6	65 (28-94)	65 (36-136)	14 (6-23)	–0.2
SF7	77 (46-111)	74 (50-96)	14 (6-27)	3
Mean	86 (28-145)	80 (34-146)	14 (6-30)	
SM1	63 (45-82)	49 (33-61)	17 (10-26)	14
SM2	71 (49-124)	33 (14-57)	14 (5-29)	39
SM3	82 (62-116)	61 (37-92)	20 (9-48)	22
SM4	94 (64-124)	73 (44-105)	20 (11-33)	21
SM5	97 (79-128)	86 (57-109)	16 (10-31)	12
SM6	68 (38-90)	70 (35-90)	14 (8-24)	–1.8
Mean	79 (38-128)	62 (14-109)	17 (5-33)	

SF, Seoul female; SM, Seoul male.

As already discussed, differences are apparent between male and female speakers in Table 7. There are also intra-speaker differences even in the same gender. For example, all female speakers but SF1, aspirated and lax stops were undergoing a merger as evidence of a sound change. In contrast, all male speakers but SM6 showed a three-way contrast of stops in terms of VOT. Thus, there were inter- and intra-speaker variations on VOT. And the VOT ranges among individuals were largely overlapped.

3.3. A cross-linguistic comparison

Figure 7 shows mean VOTs of English and Korean stops together.

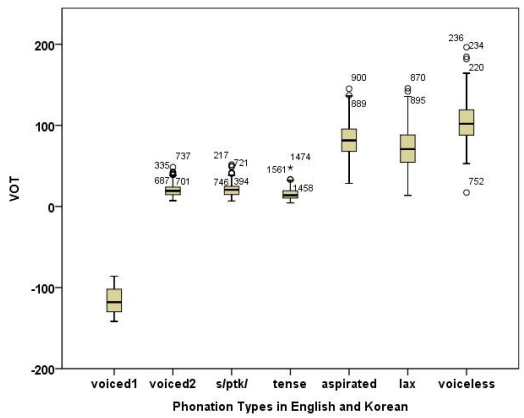

Figure 7. Mean VOTs (ms) of English and Korean stops averaged across a total of 1,728 tokens. Error bars indicate +1 standard deviation.

There was a main effect of language on VOT (p<.05) in that mean VOTs were significantly shorter in English than in Korean. *Post hoc* Tukey HSD multiple comparisons revealed that there were four significant subsets: Group 1 for voiced 1, group 2 for the voiced 2, s/ptk/ and tense categories, group 3 for lax and aspirated, and group 4 for voiceless category, as presented in Table 8.

Table 8. Four statistical subsets on VOT across English and Korean stops

Stop	N	1	2	3	4
Voiced 1	3	−115			
Tense	312		15		
Voiced 2	261		20		
s/ptk/	264		21		
Lax	312			71	
Aspirated	312			83	
Voiceless	264				104

English has a voiced-voiceless contrast, whereas Korean doesn't. However, the results in Figure 7 and Table 8 show that the two languages are very similar in that their stops were mainly produced with either short-lag or long-lag VOT. Under the long-lag VOT dimension, there are two statistical groups: lax and aspirated stops in Korean form one group and voiceless aspirated stops in English form the other group in Table 8. One questions arises: are they belong to the same dimension or not? Either way is problematic. Table 9 summarizes cross-linguistic differences on VOT.

Table 9. Summary of cross-linguistic comparisons on VOT

Aspects	English	Korean
VOT shift	Yes	Yes
VOT dimension	Short-long	Short-long
Individual variation	Small	Big
Gender variation	Small	Big
Voiceless aspirated	Longer lag	Long lag
VOT Merger	Devoiced-unasp.	Lax-aspirated
	Devoiced/unasp.	Tense
Phonemic contrast	Easy	Difficult
VOT boundary	Clear	Not clear

unasp, unaspirated.

In Table 9, it is common that both English and Korean stops undergo a VOT shift and the stops in the two languages were mainly produced with either short-lag or long-lag VOT. Except for these, English stops are slightly different from Korean stops in several aspects, as summarized in Table 9.

4. Discussion

While many studies have shown VOT in Korean stops, there has not been any investigation into changes in English stops. This study investigated VOT shift of both English and Korean stops produced by 11 English and 13 Korean native speakers and provided a cross-linguistic comparison between the two languages. The new findings of the current study are as follows: First, there was a VOT change for both Korean and English stops. The aspirated-lax merger on VOT in Korean stops was apparently replicated in this study, as reported in previous studies (Kang, 2014; Kim, 2000; Silva, 2006). Initial voiced stops in English were always or nearly always produced a single kind of /b, d, g/ as having short-lag VOT. Voiceless aspirated stops in English were produced with longer-lag VOT and voiceless unaspirated stops in English was produced with short-lag VOT. The devoicing process for the voiced stop as well as the lengthening process for the voiceless (aspirated) stop can be considered as a VOT shift. Compared with early findings (60 ms to 90 ms), VOTs for English voiceless aspirated stops in this study were much longer (104 ms). However, since it is the first study, more convincing evidence is necessary to confirm whether English stops truly undergo a sound change. Second, English voiceless aspirated stops were much longer than Korean aspirated stops (104 ms vs. 83 ms). This is comparable to Lisker & Abramson's (1964) where the opposite pattern was reported. This may be due to the fact that both languages undergo a sound change. Third, mean VOTs of English devoiced and voiceless unaspirated stops (s/ptk/) were the same as those of Korean tense stops (21 ms vs. 15 ms). This fact can be a proof of Kim's (2000) assumption that Korean tense stops could be an ordinary stop rather than an unusual one.

Fourth, compared to the Korean data, the English data had smaller gender and speaker variations on VOT. For the English data, out of eleven speakers, only one female speaker showed deviant VOT patterns for both voiced and voiceless stops. Finally, the VOT boundary among stop categories varies in both English and Korean. It is hard to fix it in a specific number. In addition, the VOT dimension between voicing lead and voicing lag was reasonable, while that between short- and long-lag was not. For the long-lag VOT dimension, Korean stops had long-lag VOT, while English stops had longer-lag VOT. As a result, the three VOT dimensions were not sufficient to differentiate both Korean and English stops. This suggests that it may require another dimension to separate Korean from English.

Table 10. Relative VOT distributions in English and Korean stops

VOT	Aspiration	English	Korean
Longer lag (>100 ms)	Strong aspiration	Yes	Yes
Long lag (>60 ms)	Moderate aspiration	(Yes)	Yes
Intermediate lag (>35 ms)	Mild aspiration		
Short lag (>15 ms)	Unaspirated	Yes	Yes
	Partially voiced	(Yes)	(Yes)
Voicing lead (< 0 ms)	Fully voiced	Yes	

Table 10 presents relative VOT distributions for English and Korean stops. According to the amount of aspiration, three VOT boundaries-35 ms, 60 ms, and 100 ms-for the same voiceless aspirated stops are possible. Regardless of different VOT values, there is only one VOT dimension — long-lag VOT. Among the three VOT dimensions, voicing lead and short voicing lag would be sufficient, whereas long voicing lag would be problematic. The current data shows that, in English, voiceless unaspirated stops could produce a 52 ms long-lag VOT, while voiceless aspirated stops could produce a 17 ms long-lag VOT. In addition, in the case of VOT and the voicing distinction, a first difficulty is that the boundary between voiced and voiceless categories varies with place of articulation. Other things being equal, VOT will be greatest (42 ms) for velar stops and smallest (25 ms) for labial stops, with alveolar stops occupying an intermediate position (35 ms) (Lisker & Abramson, 1967). This mean that, for English listeners, a VOT of

35 ms is likely to be interpreted as voiceless (pa) if the stop in question is labial but as voiced (ga) if it is velar; it could be ambiguous (ta/da) if the stop is coronal.

In English, initial voiced stop phonemes are generally said to have a VOT of 15 ms or less (short-lag VOT or prevoiced), and voiceless stop phonemes some 30 ms or longer (long-lag VOT) (Lieberman & Blumstein, 1988: 215). For native speakers of English, it is supposed that if a stop consonant has a VOT greater than 25 ms, listeners would report a voiceless percept while if the VOT was less than 25 ms they would report a voiced percept. Since the boundary between voiced and voiceless categories varies, fixing boundaries in that way may result in false consequences. For the voiceless aspirated stops, some languages may have a 35 ms long lag, while others have a 110 ms long lag. Kong et al. (2012: 742) states, "Japanese voiceless stops realized in an intermediate lag VOT range differing from the long lag VOT in the English voiceless stops. The intermediate lag VOT might be understood as a fourth type of VOT type – a "partially" aspirated stop – which was previously noted in Canadian French (Caramazza et al., 1973) and in Korean as described in older studies... Greek voiced stops differ from Japanese and English voicing stops in having a partially nasalized variant rather than a short lag variant."

The situation becomes more complicated when there is no voicing but aspiration distinctions as in the Korean data. Lisker & Abramson's (1964) states, "VOT will certainly suffice to distinguish the aspirated set from the other two and it may still be the single most important measure for separating the latter." The Korean data in this study replicate that VOT is not sufficient to separate lax from aspirated stops. Since long-lag VOT alone cannot separate voiceless lax from voiceless aspirated and voiceless aspirated in Korean from voiceless aspirated in English, it is questionable whether there needs to be another dimension to separate them.

The present results have a couple of implications. First, in addition to short and long voicing lag, there may need another VOT dimension to differentiate stop categories across languages. Second, except for VOT, other phonetic correlates may play a role in differentiating them. For example, consonants do not have much information to represent themselves. Instead, vowels contain more information about consonants. The representative acoustic correlate is f0 which is usually known to be a secondary cue to contrast consonants. In fact, f0 arises as the primary cue to contrast lax and aspirated stops in Korean. Besides, other acoustic correlates such as F1 (cutback and transition), H1-H2, burst intensity, vowel duration carry consonantal information (Haggard et al., 1970; Kohler, 1982; Whalen et al., 1993). There have also been studies about articulatory and aerodynamic cues (Liberman et al., 1958; Rothenberg, 2009; Winn et al., 2013).

Despite many limitations and problems, VOT has long been considered the most important cue in classifying stop categories in the world's languages and is still used in many acoustic and perceptual studies (Abramson & Whalen, 2017; Fowler et al. 2008). VOT is widely used because it is easy to measure and to capture, compared with articulatory and aerodynamic cues. The question is, are we missing any important cues due to the excessive use of VOT? A number of problems arose in defining VOT in some languages, and there is a call for reconsidering whether this speech synthesis parameter should be used to replace articulatory or aerodynamic model parameters which do not have these problems, and which have a strong explanatory significance. Rothenberg (2009) states, "..any explication of VOT variations will in variably lead back to such aerodynamic and articulatory concepts, and there is no reason presented why VOT adds to an analysis, other than that, as an acoustic parameter, it may sometimes be easier to measure than an aerodynamic parameter (pressure or airflow) or an articulatory parameter (closure interval or the duration, extent and timing of a vocal fold abductory gesture)."

What this study reveals is that we have to face the practical problems of VOT and do not rely on it too much. And we should try to find an alternative cue to replace it if any. Future research will be needed to do this.

5. Conclusion

This paper examined the VOT change of English stops and compared it with that of Korean stops that are currently undergoing a VOT merger between lax and aspirated stops. Based on a large and systematic quantitative data, the results showed that voiced stops are exclusively pronounced with voicing lag and voiceless aspirated stops are pronounced with relatively longer-lag VOTs than previous findings. The results suggest that, similar to Korean stops, English stops may also undergo a sound change in terms of VOT. However, since it is the first study to be revealed, more convincing evidence is necessary.

References

Abramson, A. S., & Lisker, L. (1970, January). Discriminability along the voicing continuum: Cross-language tests. *Proceedings of the Sixth International Congress of Phonetic Sciences* (pp. 569-573). Prague, Czech.

Abramson, A. S., & Whalen, D. H. (2017). Voice Onset Time (VOT) at 50: Theoretical and practical issues in measuring voicing distinctions. *Journal of Phonetics*, *63*, 75-86.

Boersma, P., & Weenink, D. (2018). *Praat: Doing phonetics by computer* (version 6.0.37) [Computer program]. Retrieved from http://www.praat.org/

Caramazza, A., Yeni-Komshian, G. H., Zurif, E. B., & Carbone, E. (1973). The acquisition of a new phonological contrast: The case of stop consonants in French-English bilinguals. *The Journal of the Acoustical Society of America*, *54*(2), 421-428.

Chodroff, E., & Wilson, C. (2017). Structure in talker-specific phonetic realization: Covariation of stop consonant VOT in American English. *Journal of Phonetics*, *61*, 30-47.

Docherty, G. L. (1992). *The timing of voicing in British English obstruents.* New York, NY: Foris.

Flege, J. E. (1982). Laryngeal timing and phonation onset in utterance-initial English stops. *Journal of Phonetics*, *10*(2), 177-192.

Fowler, C. A., Sramko V., Ostry, D. J., Rowland, S. A., & Hallé, P. (2008). Cross language phonetic influences on the speech of French – English bilinguals. *Journal of Phonetics*, *36*(4), 649-663.

Haggard, M., Ambler, S., & Callow, M. (1970). Pitch as a voicing cue. *The Journal of the Acoustical Society of America*, *47*(2B), 613-617.

Han, M. S., & Weitzman, R. S. (1970). Acoustic features of Korean /P, T, K/, /p, t, k/ and /p^h, t^h, k^h/. *Phonetica*, *22*(2), 112-128.

Kang, Y. (2014). Voice onset time merger and development of tonal contrast in Seoul Korean stops: A corpus study. *Journal of*

Phonetics, 45, 76-90.
Keating, P. A. (1984). Phonetic and phonological representation of stop consonant voicing. *Language*, *60*(2), 286-319.
Kim, C. W. (1965). On the autonomy of the tensity feature in stop classification (with special reference to Korean stops). *Word*, *21*(3), 339-359.
Kim, M. R. C. (1994). *Acoustic characteristics of Korean stops and perception of English stop consonants* (Doctoral dissertation). University of Wisconsin, Madison, WI.
Kim, M. -R. (2000). *Segmental and tonal interactions in English and Korean: A phonetic and phonological study* (Doctoral dissertation). The University of Michigan, Ann Arbor, MI.
Kim, M. -R. (2008). Lax stops in Korean revisited: VOT neutralization. *Studies in Phonetics*, *Phonology and Morphology*, *14*(2), 3-20.
Kim, M.-R. (2011). Native and non-native English speakers' VOT productions of stops. *The Linguistic Association of Korea Journal, 19*(1), 97-116.
Kim, M. -R. (2014). Ongoing sound change in the stop system of Korean: A three- to two-way categorization. *Studies in Phonetics*, *Phonology*, *and Morphology*, *20*(1), 51-82.
Kim, M. -R. (2020). The effects of length of residence (LOR) on voice onset time (VOT). *Phonetics and Speech Sciences*, *12*(4), 9-17.
Kim, M. -R., Beddor, P. S., & Horrocks, J. (2002). The contribution of consonantal and vocalic information to the perception of Korean initial stops. *Journal of Phonetics*, *30*(1), 77-100.
Klatt, D. H. (1975). Voice onset time, frication, and aspiration in word-initial consonant clusters. *Journal of Speech and Hearing Research*, *18*(4), 686-706.
Kohler, K. J. (1982). *F*0 in the production of fortis and lenis plosives. *Phonetica*, *39*(4-5), 199-218.
Kong, E. J., Beckman, M. E., & Edwards, J. (2012). Voice onset time is necessary but not always sufficient to describe acquisition of voiced stops: The cases of Greek and Japanese. *Journal of Phonetics*, *40*(6), 725-744.
Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages*. Oxford, UK: Blackwell.
Liberman, A. M., Delattre, P. C., & Cooper, F. S. (1958). Some cues for the distinction between voiced and voiceless stops in initial position. *Language and Speech*, *1*(3), 153-167.
Lieberman, P., & Blumstein, S. E. (1988). *Speech physiology*, *speech perception*, *and acoustic phonetics*. Cambridge, UK: Cambridge University Press.
Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, *20*(3), 384-422.
Lisker, L., & Abramson, A. S. (1967). Some effects of context on voice onset time in English stops. *Language and Speech*, *10*(1), 1-28.
Nakai, S., & Scobbie, J. M. (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.
Rothenberg, M. (2009). Voice onset time versus articulatory modeling for stop consonants. *Logopedics Phoniatrics Vocology*, *34*(4), 171-180.
Silva, D. J. (2006). Acoustic evidence for the emergence of tonal contrast in contemporary Korean. *Phonology*, *23*(2), 287-308.
SPSS/PASW statistics. (2017). *IBM SPSS statistics for Windows*. Version 25.0. Armonk, NY: IBM.
Whalen, D. H., Abramson, A. S., Lisker, L., & Mody, M. (1993). *F*0 gives voicing information even with unambiguous voice onset times. *The Journal of the Acoustical Society of America*, *93*(4), 2152-2159.
Winn, M. B., Chatterjee, M., & Idsardi, W. J. (2013). Roles of voice onset time and *F*0 in stop consonant voicing perception: Effects of masking noise and low-pass filtering. *Journal of Speech*, *Language*, *and Hearing Research*, *56*(4), 1097-1107.

• **Mi-Ryoung Kim,** Corresponding author
Professor, Department of Practical Foreign Languages
Korea Soongsil Cyber University
23 Samil-daero 30-gil, Jongno-gu, Seoul 03132, Korea
Tel: +82-2-708-7845
Email: kmrg@mail.kcu.ac
Areas of interest: Phonetics, Phonology, Storytelling