Linguistic Research 42(2): 365-398
DOI: 10.17250/khisli.42.2.202506.005

English voicing contrast in Seoul Korean and Mandarin Chinese learners of English*

Yalin Wang** **· Harim Kwon*****
(Seoul National University)

Wang, Yalin and Harim Kwon. 2025. English voicing contrast in Seoul Korean and Mandarin Chinese learners of English. *Linguistic Research* 42(2): 365-398. The current study explores how L1 language background and L2 English proficiency influence Seoul Korean and Mandarin Chinese speakers' cue weighting patterns of VOT (voice onset time) and post-onset f0 (fundamental frequency) in distinguishing stop voicing contrast in their L2 English. English, Seoul Korean, and Mandarin Chinese differ in the phonetic implementation of their stop laryngeal contrasts. Stop VOT and post-onset f0 are related to the stop laryngeal contrasts in all three languages, but the role of these two phonetic cues differ (e.g., Whalen et al. 1993; Kang and Guion 2008; Guo and Kwon 2022). In speech production experiments including English voiceless and voiced stops, we find difference between Seoul Korean and Mandarin Chinese speakers, providing evidence for L1 influence on the learners' stop production. English proficiency also influences the phonetic implementation of the stop voicing contrast, but the proficiency effects are complicated, interacting with gender and phonological factors. The effects of L1 system on the L2 phonetic implementation, as well as the complexity of the proficiency effects, are explained by the phonetic distance between the target English sounds and the closest L1 sounds, in the Speech Learning Model (Flege 1995). **(Seoul National University)**

Keywords Stop voicing contrast, L1 background, L2 proficiency, voice onset time, fundamental frequency

---

* This work was supported by the Park Chung-Jip Scholarship Fund for the Next Generation in English literature and language at Seoul National University awarded to Yalin Wang in 2025. Part of the data appearing in this study were presented at 2024 Linguistic Society of Korea Winter Conference. The authors have no known conflict of interest to disclose.

** First Author

*** Corresponding author

1. Introduction

L2 (second language) learners' speech may show influence from their L1 (native language), and this influence has been investigated extensively. The influence of leaners' L1 on their L2 production may decrease as their leaners' L2 proficiency improves, but not all L2 sounds change in the same way in terms of their production accuracy. According to SLM (Speech Learning Model, Flege 1995), learners' speech production is influenced by the perceived similarity of the L2 sounds to L1 speech sounds. In general, an unfamiliar L2 phone that does not have a phonetically similar counterpart in the learners' L1 is more difficult to acquire than the familiar sound. However, this difficulty of the unfamiliar L2 phone seems to decrease as the learners' proficiency improves. This is because substituting the familiar L2 phone with a similar (but not identical) sound in the learners' L1 may result in inaccurate production patterns that may be even more difficult to overcome. Since this type of subtle phonetic differences can be the seed of foreign accented speech (e.g., Derwing and Munro 2005), it is necessary to understand how the L1 influence is manifested in familiar L2 sounds. This study aims to explore the L1 influence in English voicing contrast produced by Seoul Korean and Mandarin Chinese learners of English.

The stop laryngeal contrast of Seoul Korean and Mandarin Chinese differ from English stop voicing contrast. Phonologically, the two languages do not have voiced stops whereas English contrasts voiceless stops /p t k/ and voiced stops /b d ɡ/. Phonetically, all three languages implement their laryngeal contrasts with stop VOT (voice onset time) and post-onset f0 (fundamental frequency), but the role of these phonetic properties in maintaining the contrast differs (e.g., Whalen et al. 1993; Kang and Guion 2008; Guo and Kwon 2022). This makes it theoretically interesting to explore the phonetic implementation of L2 English stops produced by Seoul Korean and Mandarin Chinese learners of English.

In light of this, the current study focuses on how Seoul Korean and Mandarin Chinese learners of different English proficiency use and weigh the two phonetic properties relevant to the English stop voicing contrast, namely stop VOT and post-onset f0.

1.1 Laryngeal contrasts in English, Mandarin Chinese, and Seoul Korean

English maintains a phonological contrast between voiceless stops /p t k/ and voiced stops /b d ɡ/. Phonetically, English voiced stops can be produced with a voicing lead (with vocal cords vibrating during the stop closure) or a voicing lag (without vocal cords vibrating during closure) in word-initial positions. The "voiced" stops produced with a voicing lag are essentially devoiced (/b d ɡ/ → [b̥ d̥ ɡ̊]) and are typically unaspirated while the voiceless stops are aspirated with longer voicing lag (e.g., Lisker and Abramson 1964; Keating 1984; Hanson 2009). Table 1 exhibits the average VOT values of English stops in previous studies.

Table 1. Mean VOT (ms) of English stops (ranges in parenthesis, when available)

	Voiced		Voiceless
	Voicing lead	Short-lag (unaspirated)	Long-lag (aspirated)
Lisker and Abramson (1964)	-97 (-20~-155)	9 (0~35)	69 (20~135)
Docherty (1992)	N/A	21	56
Allen et al. (2003)	N/A	N/A	91
Dmitrieva et al. (2015)	-107	12	64
Chodroff and Wilson (2017)	N/A	21 (11~42)	95 (46~156)

Although stop VOT is the primary phonetic property for distinguishing English stop voicing contrast in word onsets, previous studies have also demonstrated that post-onset f0 plays a non-negligible role in signaling the stop voicing contrast (e.g., Hanson 2009; Shultz et al. 2012; Chodroff et al. 2019). In production, the vowels that are preceded by a voiceless stop have a higher f0 than those preceded by a voiced stop. Listeners' voicing perception is also influenced by the post-onset f0, in perception. When stop VOT is ambiguous between voiced and voiceless stops, listeners rely on post-onset f0 to determine whether the stop is voiced or not, and they are more likely to hear the stops followed by high f0 as voiceless than those followed by low f0 (e.g., Abramson and Lisker 1985; Whalen et al. 1990). Moreover, when hearing the voicing of long-lag VOT stops are followed by low f0 (a pattern that is atypical in natural English production), English listeners need more time to judge than when they hear typical patterns (i.e., long-lag VOT + high post-onset f0 or short-lag VOT + low post-onset f0) (Whalen et al. 1993).

Mandarin Chinese has an aspiration contrast, which differs from English voicing contrast, with the aspirated series /p^h t^h k^h/ having long-lag VOT and the unaspirated series /p t k/ having short-lag VOT. VOT-wise, Mandarin and English seem to have similar laryngeal contrast defined by a VOT difference (long-lag VOT vs. short-lag VOT). However, Mandarin unaspirated stops are rarely produced with voicing lead, even in intervocalic positions (e.g., Deterding and Nolan 2007), and therefore, it has been argued that Mandarin has a phonological contrast based on aspiration, rather than voicing (e.g., Ahn et al. 2024). In addition, previous findings suggest that Mandarin aspirated stops have longer VOT than English voiceless stops (see Tables 1 and 2). According to Cho and Ladefoged's (1999) classification, English voiceless stops fall in the "aspirated" range while Mandarin aspirated stops of Mandarin can be classified as "highly aspirated" range. VOT values of Mandarin unaspirated and aspirated stops from previous studies are shown in Table 2.

Table 2. Mean VOT (ms) of Mandarin Chinese stops (ranges in parenthesis, when available)

	Unaspirated	Aspirated
Rochet and Fei (1991)	-	103 (90~110)
Liao (2005)	22	82
Liu et al. (2008)	-	100
Li (2013)	19 (11~34)	85 (69~96)
Guo (2020)	17 (8~38)	111 (46~231)

The role of post-onset f0 as a secondary cue for Mandarin aspiration contrast is controversial. F0 distinguishes lexical meanings in Mandarin, serving as the primary phonetic property for lexical tone (e.g., Yang 2015). Despite this phonological function of f0 in Mandarin, previous studies have shown that post-onset f0 is influenced by the preceding stop's laryngeal category. For example, according to Xu and Xu (2003), f0 in unaspirated contexts is higher than that in aspirated contexts for all four tones. This effect is greater in tones beginning in a low f0 (Tone 2 and Tone 3, henceforth T2 and T3) than those beginning in a high f0 (Tone 1 and Tone 4, henceforth T1 and T4). Luo (2018), in contrast to Xu and Xu (2003), reports that post-aspirated f0 is higher than post-unaspirated f0. Guo and Kwon (2022), on the other hand, report difference between high-initial tones (T1 and T4) and low-initial tones (T2 and T3). Post-aspirated f0 is higher than post-unaspirated f0 in high-initial tones whereas, in low-initial tones, the effect of onset aspiration on following f0 is in the opposite direction.

Unlike the two languages, Seoul Korean features a typologically unusual three-way laryngeal contrast for its stop consonants. All three types of stops are voiceless in word-initial positions. Traditionally, the contrast has been described that aspirated stops /p^h t^h k^h/ generally have strong aspiration, lenis (or lax) stops /p t k/ are produced as slightly aspirated, and fortis (or tense) stops /p* t* k*/ which are unaspirated. However, Seoul Korean has experienced a tonogenesis-like sound change (e.g., Silva 2006), with the shift from stop VOT to post-onset f0 as the primary phonetic property for distinguishing aspirated and lenis stops. Consequently, in contemporary Seoul Korean, VOT values become nondistinctive between lenis and aspirated stops (e.g., Kang and Guion 2006; Kang 2014; Kwon 2019), and speakers use the post-onset f0 as the primary phonetic property and the VOT as the secondary property to distinguish aspirated and lenis stops. For example, Kang and Guion (2006) report that younger speakers primarily use post-onset f0, while older speakers rely primarily on stop VOT, when they enhance the word-initial laryngeal contrast in clear speech. Nonetheless, the role of the secondary phonetic property (stop VOT) here is still contrastive (not redundant) because the full three-way laryngeal contrast cannot be maintained without VOT (e.g., Silva 2006; Kwon 2021). After the tonogenesis-like sound change, Seoul Korean phrase-initial stops are described as the aspirated stops having long-lag VOT and high post-onset f0, lenis stops having long-lag VOT and low post-onset f0, and fortis stops having short lag VOT and high post-onset f0. Table 3 presents the mean VOT of Korean initial stops from previous studies. Note the diachronic changes in VOT differences between the aspirated and lenis categories.

Table 3. Mean VOT (ms) of Seoul Korean stops (ranges in parenthesis, when available)

	Fortis	Lenis	Aspirated
Kagaya (1974)	15	60	160
Kim (1994)	9 (9~11)	51 (15~78)	78 (75~87)
Kang and Guion (2006)	16	61	79
Oh (2011)	15	52	70
Lee and Jongman (2012)	18	65	81
Yu (2018)	17	71	79

In sum, previous studies suggest the roles of stop VOT and post-onset f0 in defining the stop laryngeal contrasts differ in English, Mandarin Chinese, and Seoul

Korean. Although both English and Korean are non-tonal languages and they use stop VOT and post-onset f0 for their respective stop laryngeal contrast, the primacy and contrastiveness of the two phonetic properties differ in the two languages. Mandarin Chinese, similar to English, has a laryngeal contrast primarily maintained by the stop VOT; however, the role of post-onset f0 differs from English due to its significant role in lexical tones. This raises an interesting question regarding the stop productions by Seoul Korean and Mandarin Chinese learners of English. When English learners from Seoul Korean or Mandarin Chinese native language background produce English stops, how do their productions show influence from their native language?

1.2 Native language influence on L2 English stops

Producing a segment involves manipulating multiple phonetic properties related to the segment. For instance, when producing a /p/, speakers manipulate (both consciously and subconsciously) many phonetic properties including stop VOT and post-onset f0. It can be a challenge for learners of English to acquire the specific combination of the acoustic properties that characterize English voicing contrast. Previous studies have shown that L2 learners from various L1 backgrounds tend to rely heavily on familiar primary cues, such as VOT, when acquiring English stop contrasts (e.g., Alves and Luchini 2017; Hamzah et al. 2020). However, some studies have reported that under certain conditions, learners may also begin to utilize secondary cues like post-onset f0 (e.g., Van Alphen and Smits 2004). These findings call for a systematic investigation of learners of English from typologically different L1 backgrounds, which the present study aims to address by examining Seoul Korean and Mandarin Chinese learners, whose L1s differ markedly from English both in cue types and cue weighting.

Previous studies on Seoul Korean learners of English show that the learners rely heavily on post-onset f0 for English stop voicing contrast. For example, Schertz et al. (2015) examine how Seoul Korean speakers use stop VOT and post-onset f0 to distinguish stop laryngeal contrasts in L1 Seoul Korean and L2 English. They find that while Seoul Korean speakers use stop VOT as the primary cue for distinguishing English stop voicing, they also rely heavily on post-onset f0. This strong dependence on f0 likely stems from its role as the primary cue in distinguishing aspirated-lenis contrast in the learners' native language.

Native Seoul Korean speakers who learn English later in life associate English voiceless stops with Korean aspirated stops. In contrast, their English voiced stops appear to form a distinct category, differing from both Korean fortis and lenis stops, as English voiced stops are similar to fortis stops in VOT but align more closely with lenis stops in post-onset f0 (e.g., Kang and Guion 2006). Furthermore, Seoul Korean learners of English do not merger stop VOT between their L2 English voiced and voiceless stops, but rather the two categories are distinct in their VOT as well as their post-onset f0 (Kim 2012). Kong and Yoon (2013) investigate English stops produced by high school students of Seoul Korean learning English in English as Foreign Language (EFL) setting. Comparing learners of different proficiency based on the standardized high school English proficiency test, their findings indicate that lower-proficiency learners depend more heavily on post-onset f0 and less on stop VOT than higher-proficiency learners to distinguish English voiceless and voiced stops.

On the other hand, Mandarin learners of English face a different challenge. Although stop VOT is the phonetic property that is primarily contrastive in their native language, the typical VOT ranges for English voiceless stops and Mandarin aspirated stops are different. Also, Mandarin unaspirated stops are rarely produced with voicing lead. Finally, the association between post-onset f0 and the stop's laryngeal categories (voicing or aspiration) is less evident in Mandarin than in English. Due to these differences, Mandarin learners typically produce English voiceless stops with unusually longer VOT. For example, Kato and Baese-Berk (2021) examine English stops produced by Mandarin learners of English and find that the learners produce longer VOT for English voiceless stop /p/ than native English speakers. However, as the learners become more proficient in English, the VOT ranges become closer to those of native English speakers, indicating the reverse relation between L2 proficiency and L1 influence.

Previous research on Mandarin learners' English stop productions focuses on stop VOT while post-onset f0, a non-negligible secondary cue for English voicing contrast (e.g., Whalen et al. 1993), has been largely neglected. One exception is Guo (2020) who investigates stop VOT and post-onset f0 in Mandarin speakers' production of L1 Mandarin and L2 English stops. Based on Mandarin learners of English living in the US, she finds that, first, the VOT differences between their Mandarin aspirated and unaspirated stops are greater than those between their English voiceless and voiced stops, and second, the learners produce consistent post-onset f0 patterns for their English stops (high after voiceless, low after voiced). Notably, these learners' post-onset f0 patterns for their native aspirated and unaspirated stops differ depending on the tones (as reported in Guo and Kwon 2022), but their English stops show the f0 pattern in line with the high-initial tones in their L1, suggesting that their L1 tonal experience may shape their L2 post-onset f0 realization for stop voicing contrast. It is yet to be known how Mandarin learners of English in EFL setting produce the secondary cue, post-onset f0, in relation to English stop voicing contrast.

Thus, the current study aims to investigate how EFL learners of different native language (Seoul Korean and Mandarin Chinese) and varying English proficiency produce English voicing contrast, focusing on the phonetic implementation of the voicing contrast using two acoustic properties: the primarily contrastive stop VOT and a secondary property, post-onset f0. The research questions are as follows:

1. Does the implementation of the two phonetic properties related to English stop contrasts (stop VOT and post-onset f0) show L1 influence?
2. How does the L1 influence vary with the learners' English proficiency?

2. Methods

2.1 Participants

Twenty-four adult native speakers of Seoul Korean (SK) (12 female, 12 male; mean age = 22.8, range = 19~30) were recruited from Seoul, Korea. They were born and raised in Seoul or Gyeonggi Province in Korea, with no experience of living elsewhere for more than six months. Twenty-four Mandarin Chinese (MC) speakers (13 female, 11 male; mean age = 20.7, range = 18~25) were recruited from Dalian, China. All MC participants were born and raised in Northern China and had no residence history outside this region for more than six months. Both SK and MC participants learned English in an EFL setting and reported no knowledge of additional languages at the time of recording. Twelve native English (NE) speakers (7 female, 5 male; mean age = 26.8, range = 19~35) were recruited from Seoul as the control group. These NE speakers were born and raised in the United States or Canada. To minimize the exposure to Korean, we recruited NE speakers with less than 1 year experience of living in Korea. They reported not being fluent in other languages including Korean. All participants reported no history of speech or hearing disorders.

SK and MC participants were classified into different proficiency groups. For the SK learners, classification was based on their most recent English proficiency test scores from New TEPS (Test of English Proficiency by Seoul National University) or TOEIC (Test of English for International Communication). The MC participants were classified based on their most recent scores of the College English Test (CET4 or CET6). To classify all learners into comparable proficiency groups, New TEPS scores of SK speakers were converted to TOEIC scores using the official conversion table from the TEPS website[1], and all TOEIC scores were then mapped to the Common European Framework of Reference for Languages (CEFR) bands according to the British Council Korea guidelines[2]. The CET scores of the MC learners were also mapped to CEFR bands following the criteria in Jin et al. (2022). The participants' English learning backgrounds (i.e., the age at which they began formal instruction) and their proficiency are summarized in Tables 4-5.

Table 4. English proficiency of SK speakers (mean values with range in parenthesis)

Proficiency group	N	Age (years)	English learning onset (years)	English proficiency test scores	CEFR
SK Group 1	11	20.82 (19~22)	6.91 (5~10)	NEW TEPS: 227.9 (210~250) TOEIC: 623.3 (600~660)	B1
SK Group 2	9	24.56 (19~30)	8.67 (5~14)	NEW TEPS:400.2 (365~417) TOEIC: 856.7 (785~935)	B2
SK Group 3	4	24.5 (20~27)	6.75 (5~10)	NEW TEPS: 499.3 (480~510) TOEIC:980 (980)	C1

Table 5. English proficiency of MC speakers (mean values and range in parenthesis)

Proficiency group	N	Age (years)	English learning onset (years)	English proficiency test scores	CEFR
MC Group 1	11	20.91 (18~25)	8.09 (5~11)	CET-4: 449.5 (426~519) CET-6: 428 (428)	B1
MC Group 2	13	20.54 (18~25)	7.07 (4~10)	CET-4: 569.5 (551~588) CET-6: 502.1 (450~571)	B2

1 Seoul National University Language Education Institute. (n.d.). TEPS *official website*. Retrieved 27 December 2024, from https://en.teps.or.kr/

2 British Council Korea. (n.d.). *British Council Korea official website*. Retrieved December 27, 2024, from https://www.britishcouncil.kr/

2.2 Procedure

The data presented in current study were collected as a part of a larger research project on simulated human-machine interactions consisting of two tasks, (1) reading words on a computer screen and (2) responding to programmed prompts. Here, we report only the data from the first task, namely, word reading. MC participants were recorded in Dalian, China, and SK and NE participants in Seoul, Korea. Participants were seated in a sound-treated booth, saw each stimulus word at the center of a laptop screen via E-Prime 3.0, and were asked to read the word aloud as naturally as possible. Their speech was acoustically recorded using a Zoom H6 recorder (44.1 kHz sampling rate).

2.3 Stimuli

Target stimuli were 54 monosyllabic English words beginning on a stop consonant. Half of them began on voiced stops /b d ɡ/ while the other half on voiceless stops /p t k/. Target words were balanced in terms of the place of articulation and consisted of minimal pairs, as shown in Table 6. The word selection was constrained by lexical availability, as minimal pairs differing only in voicing while holding place constant are limited in English. Due to these lexical availability constraints, it was not feasible to strictly match lexical frequency across the experimental items. Still, to minimize potential effects of word frequency on speech production, moderate to high-frequency words were selected when possible. The selected target words were presented with 216 monosyllabic fillers, including minimal pairs differing either in sonorant onsets (e.g., *lace-race*) or nucleus vowels (e.g., *match-much*). Fillers aimed to prevent the participants from figuring out the purpose of the experiment and hyper-articulating the stop voicing contrast. Each target word was repeated twice, yielding a 1:2 target-to-filler ratio. All stimuli were presented in a random order at the center of a laptop screen.

Table 6. List of target stimuli

/p/-/b/	*pace-base, pain-bane, paste-baste, peak-beak, peas-bees, peep-beep, Pete-beat, plight-blight, pond-bond*
/k/-/g/	*came-game, cap-gap, cape-gape, cave-gave, coal-goal, coat-goat, coop-goop, kale-gale, kill-gill*
/t/-/d/	*tale-dale, tame-dame, team-deem, teen-dean, tie-die, time-dime, toes-doze, tomb-doom, tune-dune*

2.4 Measurements

All acoustic measurements were taken in Praat (Boersma and Weenink 2020). Before taking the measurements, tokens with disfluency (e.g., extra-verbal interruptions such as coughing or clearing the throat, self-correction, or uttering a different word) were removed from the dataset. After removing disfluent tokens (4.2% of the total productions), the VOT of onset stop consonants and the post-onset f0 were measured for the remaining tokens. Stop VOT (ms) was measured from the onset of the stop burst to the onset of vocal fold vibration shown as glottal pulsing in the waveform and/or the voicing bar observed in the spectrogram. English voiced stops can sometimes be prevoiced with voicing lead, in which case, VOT values were negative, measured from the onset of vocal fold vibration, as indicated by the low-frequency periodicity during stop closure in the waveform, continuing directly into the stop burst. Post-onset f0 (Hz) was measured at the temporal midpoint of the vowel in each word using the pitch tracking function in Praat. Tracking errors, including f0 doubling or halving, were hand-corrected. For cross-gender comparisons, f0 values were converted into semitones using the formula: $\log_2(Hz/100) * 12$ (e.g., Whalen and Levitt 1995).

3. Results

3.1 L1 effects

To address the first research question about the effects of the learners' native language on the realization of phonetic cues related to English stop voicing contrast, we analyzed stop VOT and post-onset f0 in a series of linear mixed effects models, implemented in *lme4* (Bates et al. 2015) in R (R Core Team 2021). The data were analyzed by fitting separate models to each dependent variable: VOT (ms.) and f0 (semitone). Predictors included native LANGUAGE (**NE**, SK, MC), ONSET voicing (**voiceless**, voiced), and GENDER (**female**, male). All predictors were dummy-coded, with the reference levels bold-faced. The random-effects structure included by-word and by-speaker intercepts, determined by the forward best path algorithm (Barr et al. 2013). Initial models included the full interactions among the predictors and the best-fitting models were selected by the likelihood ratio test of the model with an interaction term in question compared to the one without. Any significant interactions were further analyzed by post-hoc Tukey's HSD tests conducted in the *emmeans* package (Lenth 2020). P-values for each coefficient were obtained using Satterthwaite approximations implemented in the *lmerTest* package (Kuznetsova et al. 2017). If a predictor turned out to be significant in multiple interactions (or a main effect and interactions), only the highest-level interaction was reported.

3.1.1 VOT

Before fitting the models to the VOT of word-initial stops, voiced stops produced with voicing lead were excluded as it could be misleading to include both positive and negative VOT in the same models. The proportion of prevoiced tokens differed across the three language groups (see Table 7). Compared to the results of NE speakers who produced 52.3% of the total voiced stops with pre-voicing, SK learners exhibited a moderate prevoicing tendency, with 240 prevoiced tokens (20.4%). In contrast, MC learners demonstrated little prevoicing, with only 12 tokens (0.9%) produced with the voicing lead.

Table 7. Number (proportion) of prevoiced voiced stops

SK	MC	NE
240 (20.4%)	12 (0.9%)	338 (52.3%)

Excluding these prevoiced tokens, the outcome of the best-fitting model on VOT is in Table 8. The significant interaction among Language * Onset * Gender [χ^2 = 251.7, $p < 0.001$], plotted in Figure 1, was further investigated with post-hoc Tukey comparisons. As shown in Table 9, the results show that, for voiceless stops, female

SK speakers did not show a significant difference in VOT compared to NE counterparts [$p = 0.89$], whereas male SK speakers produced significantly longer VOT [$\beta = -10.77$, $t = -6.36$, $p < 0.001$]. Similarly, male MC speakers also had significantly longer VOT than NE speakers [$\beta = -13.71$, $t = -7.13$, $p < 0.001$], while female MC speakers showed no significant difference [$p = 0.74$]. For voiced stops, female SK speakers produced significantly longer VOT than NE speakers [$\beta = -5.18$, $t = -4.05$, $p < 0.05$], whereas male SK speakers did not differ significantly [$p = 0.31$]. Among MC speakers, male participants again showed significantly longer VOT than NE speakers, but female MC speakers did not differ significantly from NE speakers.

Table 8. Summary of the linear mixed-effects models for VOT

Fixed Effects	Estimate	t-val.	p-val.
(Intercept)	93.69	38.53	< 0.001 ***
Gender	-6.79	-4.23	< 0.001 ***
Language (NE vs. MC)	3.45	2.91	< 0.01 **
Language (NE vs. SK)	1.23	0.96	0.34
Onset	-74.39	-25.85	< 0.001 ***
Gender * Language (NE vs. MC)	7.26	3.73	< 0.001 ***
Gender * Language (NE vs. SK)	12.54	6.06	< 0.001 ***
Gender * Onset	5.72	2.86	< 0.01 **
Language (NE vs. MC) * Onset	-2.03	-1.26	0.21
Language (NE vs. SK) * Onset	3.95	2.42	< 0.05 *
Language (NE vs. MC) * Onset * Gender	-11.88	-4.85	< 0.001 ***
Language (NE vs. SK) * Onset * Gender	2.81	1.73	< 0.05 *

Table 9. Post-hoc Tukey test comparing VOT by Language * Onset * Gender

Language	Onset	Gender	Estimate	t-val.	p-val.
NE vs. SK	Voiceless	Female	1.23	0.96	0.89
		Male	-10.77	-6.36	< 0.001 ***
	Voiced	Female	-5.18	-4.05	< 0.05 *
		Male	-4.84	-3.28	0.31
NE vs. MC	Voiceless	Female	-2.45	-2.91	0.74
		Male	-13.71	-7.13	< 0.001 ***
	Voiced	Female	-1.42	-1.20	0.83
		Male	-7.49	-6.02	< 0.01 **

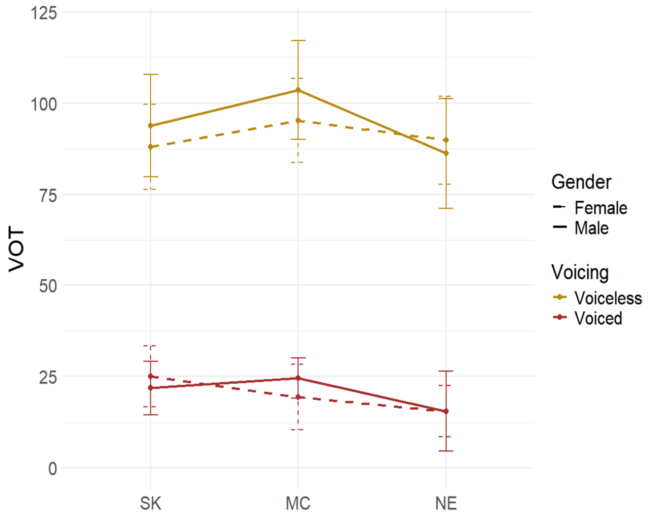

Figure 1. Stop VOT (ms.) of SK, MC, and NE speakers

3.1.2 Post-onset f0

The best-fitting model for post-onset f0 (semitone) included a significant Language * Onset * Gender interaction [χ^2 = 8785.2, $p < 0.001$]. The model outcome is in Table 10.

Table 10. Summary of the linear mixed-effects models for post-onset f0

Fixed Effects	Estimate	t-val.	p-val.
(Intercept)	16.38	50.43	< 0.001 ***
Language (NE vs. SK)	-0.70	-4.55	< 0.001 ***
Language (NE vs. MC)	0.25	1.68	0.09 (*)
Onset	-1.59	-5.75	< 0.001 ***
Gender	-11.06	-55.95	< 0.001 ***
Language (NE vs. SK) * Onset	-0.53	-2.69	< 0.01 **
Language (NE vs. MC) * Onset	1.07	5.39	< 0.001 ***
Language (NE vs. SK) * Gender	1.91	7.78	< 0.001 ***
Language (NE vs. MC) * Gender	-1.81	-7.58	< 0.001 ***
Onset * Gender	0.13	0.52	0.60
Language (NE vs. SK) * Onset * Gender	-1.07	-3.54	< 0.001 ***
Language (NE vs. MC) * Onset * Gender	-0.55	-1.80	0.07 (*)

Table 11. Post-hoc Tukey test comparing post-onset f0 (voiceless - voiced stops)

	Gender	Estimate	t-val.	p-val.
SK	Female	2.12	8.32	< 0.001 ***
	Male	3.06	11.81	< 0.001 ***
MC	Female	0.82	3.04	< 0.05 *
	Male	0.94	3.55	< 0.05 *
NE	Female	1.59	5.75	< 0.001 ***
	Male	1.46	4.96	< 0.001 ***

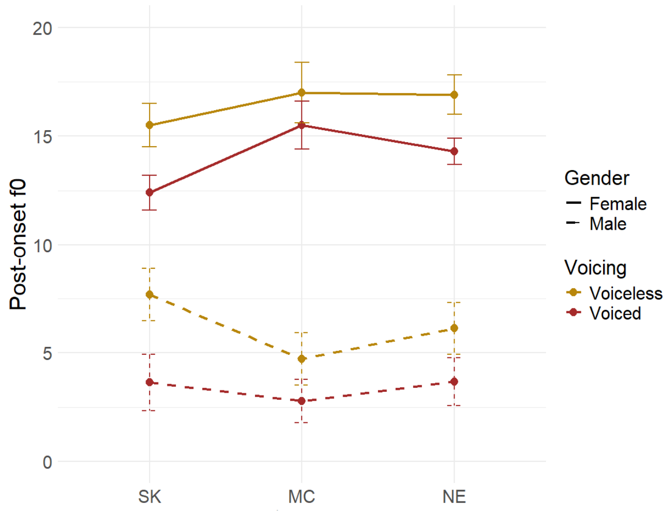

Figure 2. Post-onset f0 (semitone) of SK, MC, NE groups

The results of post-hoc Tukey comparisons on the significant interaction are in Table 11. As demonstrated in Figure 2, the results indicated that the post-voiceless f0 was higher than post-voiced f0 in all participant groups, but the f0 difference between voiceless stops and voiced stops was greater in the SK speakers than in the NE speakers, while the MC speakers showed smaller f0 difference between voiced and voiceless stops than the NE speakers.

3.1.3 Interim summary: Native language effect

We examined the effects of the learners' native language on the realization of two phonetic cues potentially related to English stop voicing contrast, stop VOT and post-onset f0. The findings suggest that speakers of different native language background produced English voicing contrast differently. Compared to native English control speakers, voiced stops produced by both SK and MC participants had less prevoicing, and male MC speakers and female SK speakers produced longer VOT for English voiced stops (when produced without voicing lead), exhibiting non-native-like phonetic patterns. These patterns appear to reflect transfer from the learners' L1 - both the unaspirated stops in Mandarin (Table 2) and word-initial lenis stops in Seoul Korean (Table 3) are typically produced with longer VOT than English voiced stop. In addition, both Seoul Korean and Mandarin Chinese do not have phonological voicing in their language, preventing the learners from producing English voiced stops with voicing lead. Furthermore, Seoul Korean has the inter-sonorant voicing of lenis stops (e.g., Jun 1998), but Mandarin does not (e.g., Deterding and Nolan 2007). Therefore, Mandarin learners may have even greater difficulty producing English voiced stops with prevoicing. The findings for voiceless stops' VOT were more complicated, showing gender effect. This will be further discussed in 4.1.

In terms of post-onset f0, SK speakers produced greater f0 differences for English voicing contrast while MC speakers produced smaller f0 differences. These are also in line with the learners' L1 system. In Seoul Korean, aspirated stops are distinguished by lenis stops primarily by post-onset f0, which could have been transferred to English voicing contrast. On the other hand, Mandarin is a tonal language in which f0 primarily signals lexical tones, and thus Mandarin learners are less likely to use f0 as a phonetic cue for stop voicing contrast in their L2 English.

3.2 English proficiency

To address the second research question about the effects of the learners' English proficiency on how stop voicing contrast was realized, we built additional series of linear mixed effects models. Stop VOT and post-onset f0 models were built separately for SK and MC speakers, with the fixed effects of L2 PROFICIENCY (SK models: **Group 1**, Group 2, Group 3; MC models: **Group 1**, Group 2), ONSET voicing (**voiceless**, voiced), and GENDER (**female**, male). Other detailed procedures of the statistical analyses were identical to those reported in the previous section. In this section, we focus only on the significant main effects or interactions involving

Proficiency, as they are directly relevant to the second research question.

3.2.1 Seoul Korean speakers

Figure 3 illustrates the VOT patterns for female and male SK speakers across different L2 proficiency levels.

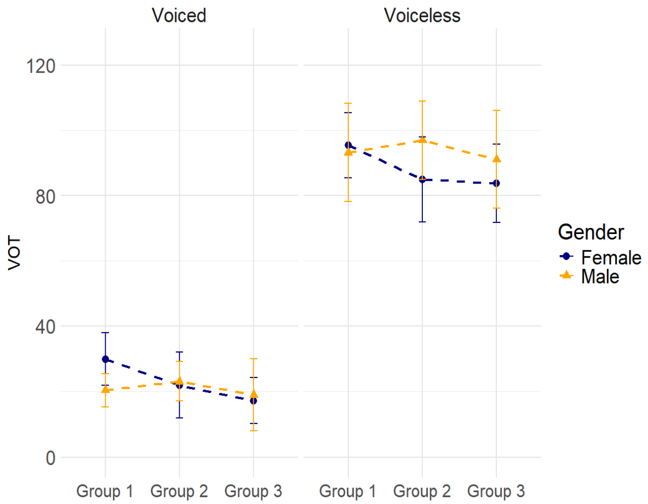

Figure 3. Stop VOT by SK speakers of different L2 English proficiency and gender

The outcome of the best-fitting model for SK speakers' VOT is in Table 12. The significant Onset * Proficiency interaction [χ^2 = 155.54, p <0.001], directly relevant to the research question, was further investigated by conducting post-hoc analyses. The post-hoc test results (Table 13) indicated that Group 1 SK speakers (corresponding to CEFR B1) produced significantly longer VOT (though borderline) than Group 3 speakers (corresponding to CEFR C1) both for voiceless stops [β = 7.23, t = 1.91, $p < 0.05$] and for voiced stops [β = 8.44, t = 2.17, $p < 0.05$]. Except for these differences between Group 1 and Group 3, the proficiency groups did not differ significantly. Specifically, Group 1 and Group 2 did not differ significantly in either voiceless [p = 0.56] or voiced stop production [p = 0.66]. Similarly, Group 2 and Group 3 showed no significant differences for either voiceless [p = 0.41] or voiced stops [p = 0.82].

These results suggest that English proficiency had an influence on VOT production among SK speakers, with lower-proficiency learners tending to produce longer VOT for both voiceless and voiced stops compared to higher-proficiency learners.

Table 12. Summary of the linear mixed-effects models for VOT (ms.)

Fixed Effects	Estimate	t-val.	p-val.
(Intercept)	97.73	24.00	< 0.001 ***
Gender	6.45	1.62	0.12
Onset	-70.65	-21.88	< 0.001 ***
Proficiency (Group 1 vs. Group 2)	-3.81	-0.88	0.40
Proficiency (Group 1 vs. Group 3)	-11.44	-2.17	< 0.05 *
Onset * Gender	-7.91	-5.25	< 0.001 ***
Onset * Proficiency (Group 1 vs. Group 2)	-0.73	-0.44	0.66
Onset * Proficiency (Group 1 vs. Group 3)	8.22	4.12	< 0.001 ***

Table 13. Post-hoc Tukey test comparing VOT by L2 proficiency

Proficiency	Onset	Estimate	t-val.	p-val.
Group 1 vs. Group 2	voiceless	4.53	1.04	0.56
	voiced	3.81	0.88	0.66
Group 1 vs. Group 3	voiceless	7.23	1.91	< 0.05 *
	voiced	8.44	2.17	< 0.05 *
Group 2 vs. Group 3	voiceless	4.64	1.30	0.41
	voiced	-1.31	-0.22	0.82

Figure 4 demonstrates the post-onset f0 by SK speakers of different English proficiency, and Table 14 reports the model outcome of the best-fitting model on post-onset f0. Again, the significant interaction between Onset * Proficiency [χ^2 = 218.7, $p < 0.001$] was further investigated in post-hoc tests. The post-hoc tests (see Table 15) showed the differences in f0 across all group comparisons were not statistically significant for voiceless and voiced stops. However, Group 1 produced marginally lower f0 compared to Group 3 for voiced stops [$\beta = -2.58$, $t = -2.28$, $p = 0.08$].

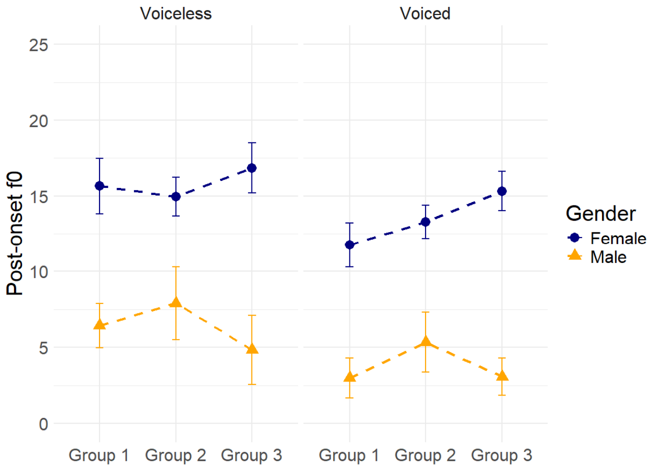

Figure 4. Post-onset f0 of SK speakers of different gender and English proficiency

Table 14. Summary of the linear mixed-effects models for post-onset f0

Fixed Effects	Estimate	t-val.	p-val.
(Intercept)	15.32	18.52	< 0.001 ***
Onset	-3.36	-11.36	< 0.001 ***
Proficiency (Group 1 vs. Group 2)	0.14	0.16	0.88
Proficiency (Group 1 vs. Group 3)	0.70	0.62	0.54
Gender	-8.75	-10.77	< 0.001 ***
Onset * Proficiency (Group 1 vs. Group 2)	1.53	10.29	< 0.001 ***
Onset * Proficiency (Group 1 vs. Group 3)	1.87	9.75	< 0.001 ***
Onset * Gender	-0.27	-1.99	< 0.05 *

Table 15. Post-hoc Tukey test comparing post-onset f0 by L2 proficiency

Proficiency	Onset	Estimate	t-val.	p-val.
Group 1 vs. Group 2	voiceless	-0.14	-0.16	0.98
	voiced	-1.67	-1.91	0.16
Group 1 vs. Group 3	voiceless	-0.70	-0.62	0.81
	voiced	-2.58	-2.28	0.08 (*)
Group 2 vs. Group 3	voiceless	-0.56	-0.53	0.86
	voiced	-0.91	-0.85	0.68

3.2.2 Mandarin Chinese speakers

Figure 5 shows the distribution of VOT among MC participants with varying English proficiency across different genders.

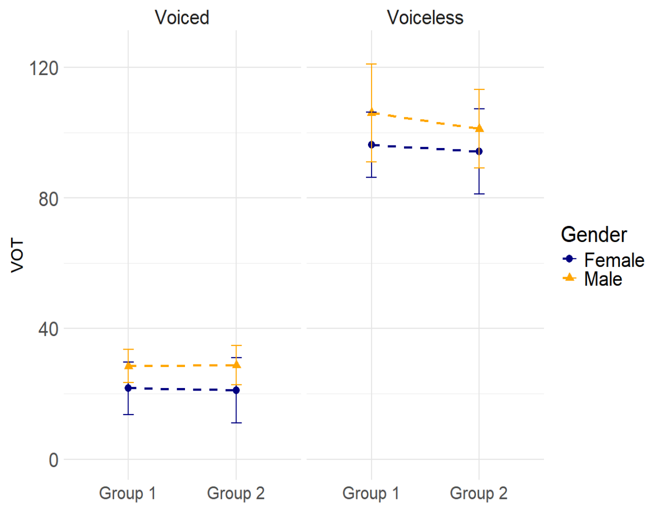

Figure 5. Stop VOT by MC speakers of different genders and L2 English proficiency

The best-fitting linear mixed-effects model of VOT for MC speakers included Onset * Proficiency * Gender interaction [χ^2 = 155.54, $p < 0.001$] (see Table 16). The significant three-way interaction was further investigated using Tukey's HSD tests. As summarized in Table 17, L2 proficiency did not significantly affect VOT production in Mandarin Chinese learners for either voiceless or voiced stops, across both male and female speakers. Regarding the gender effect (Table 18), male speakers produced significantly longer VOT than female speakers for voiceless stops in both Group 1 [β = -8.10, t = -5.30, $p < 0.01$] and Group 2 [β = -7.42, t = -2.41, $p < 0.05$]. For voiced stops, similar gender effect was found in Group 2 [β = -6.45, t = -2.87, $p < 0.05$], while no significant gender difference was found in Group 1 [p = 0.76].

Table 16. Summary of the linear mixed-effects models for VOT

Fixed Effects	Estimate	t-val.	p-val.
(Intercept)	0.19	24.75	< 0.001 ***
Proficiency	-0.02	-1.97	0.06 (*)
Gender	0.01	1.29	0.21
Onset	-0.14	-18.19	< 0.001 ***
Proficiency * Onset	0.01	2.62	0.15
Gender * Onset	0.02	2.24	< 0.05 *
Proficiency * Gender * Onset	-0.03	-5.66	< 0.001 ***

Table 17. Post-hoc Tukey test comparing VOT under the effect of proficiency

Onset	Gender	Estimate	t-val.	p-val.
voiceless	Female	2.75	0.45	0.82
	Male	3.27	1.28	0.24
voiced	Female	0.57	0.16	0.99
	Male	-0.67	-0.19	0.94

Table 18. Post-hoc Tukey test comparing VOT under the effect of gender

Onset	L2 proficiency	Estimate	t-val.	p-val.
voiceless	Group 1	-8.10	-5.30	< 0.01 **
	Group 2	-7.42	-2.41	< 0.05 *
voiced	Group 1	-3.10	-1.78	0.76
	Group 2	-6.45	-2.87	< 0.05 *

The best-fitting post-onset f0 model for MC speakers included significant Onset * Gender interaction [χ^2 = 405.5, $p < 0.001$]. The outcome of this model is shown in Table 19, and the corresponding data are plotted in Figure 6. This model revealed that the proficiency effects, and any interaction terms involving proficiency, did not significantly influence post-onset f0 of MC speakers.

Table 19. Summary of the linear mixed-effects models for post-onset f0

Fixed Effects	Estimate	t-val.	p-val.
(Intercept)	16.07	24.45	< 0.001 ***
Proficiency	-0.15	-0.19	0.85
Gender	-11.30	-14.13	< 0.001 ***
Onset	-0.51	-2.87	< 0.01 **
Gender * Onset	-0.44	-3.60	< 0.001 ***

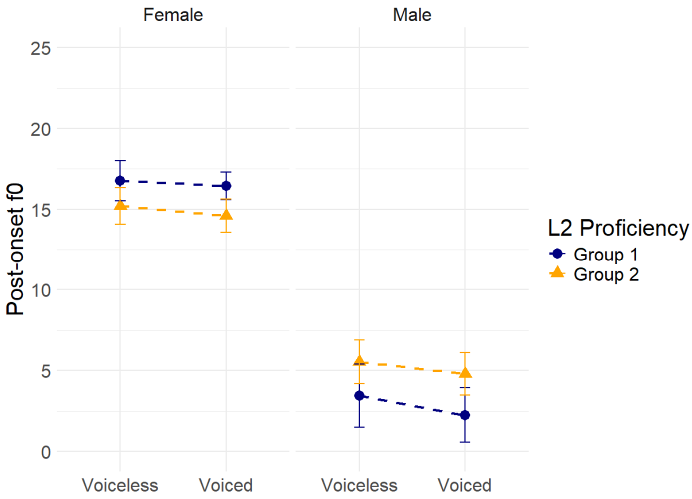

Figure 6. Post-onset f0 by MC speakers of different gender and English proficiency

3.2.3 Interim summary: Proficiency effects

To sum up, SK learners of English with lower English proficiency produced longer VOT for voiceless and voiced stops compared to those with higher proficiency. On the other hand, English proficiency does not have significant influence on f0 difference between voiced and voiceless stops except for the marginal difference between voiced stops produced by Group 1 and Group 3. In contrast, MC learners did not exhibit significant proficiency effects. Instead, gender effects were evident for stop VOT, with male speakers producing longer VOT for voiceless stops than female speakers in both groups, and longer VOT for voiced stops specifically in Group 2.

3.3 Weighting of VOT and f0

To determine the relative importance, or cue weighting, of the two phonetic properties for distinguishing voiced and voiceless stops in SK and MC learners of English, Linear Discriminant Analysis (LDA) was performed using the *lda* function from the *MASS* package (Venables and Ripley 2002). To understand how the cue weighing patterns differed by speakers' native language and English proficiency, the coefficients for the

VOT and f0 differences were calculated separately for different language and proficiency groups. The distribution of the LDA coefficients is plotted in Figure 7. The outcome indicates that stop VOT is more important than post-onset f0 in differentiating English voiced and voiceless stops, across all groups of speakers. That is, all participants, including EFL learners and native speakers, used stop VOT primarily to make the English stop voicing contrast. However, the LDA coefficients for VOT in SK groups (especially the low proficiency group SK-1) are notably smaller than NE or MC groups.

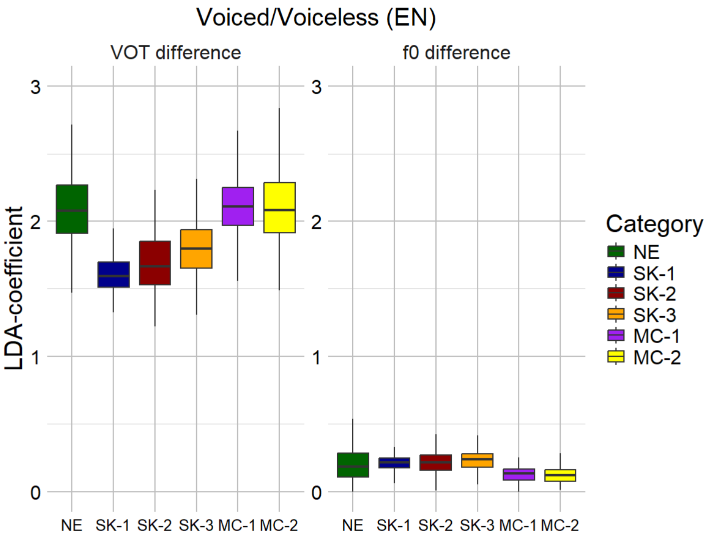

Figure 7. LDA coefficient differences in voiced/voiceless contrasts

Figure 8 visually presents the distribution of voiced and voiceless stops in terms of the z-standardized VOT and f0 values for each language and proficiency group. The visualization further confirms the LDA results: in all groups of speakers, regardless of their native language or English proficiency, English voicing contrast seems to be distinguished by stop VOT more reliably than post-onset f0 (indicated by greater separation between voiced and voiceless stops in the horizontal than vertical dimensions in all plots).

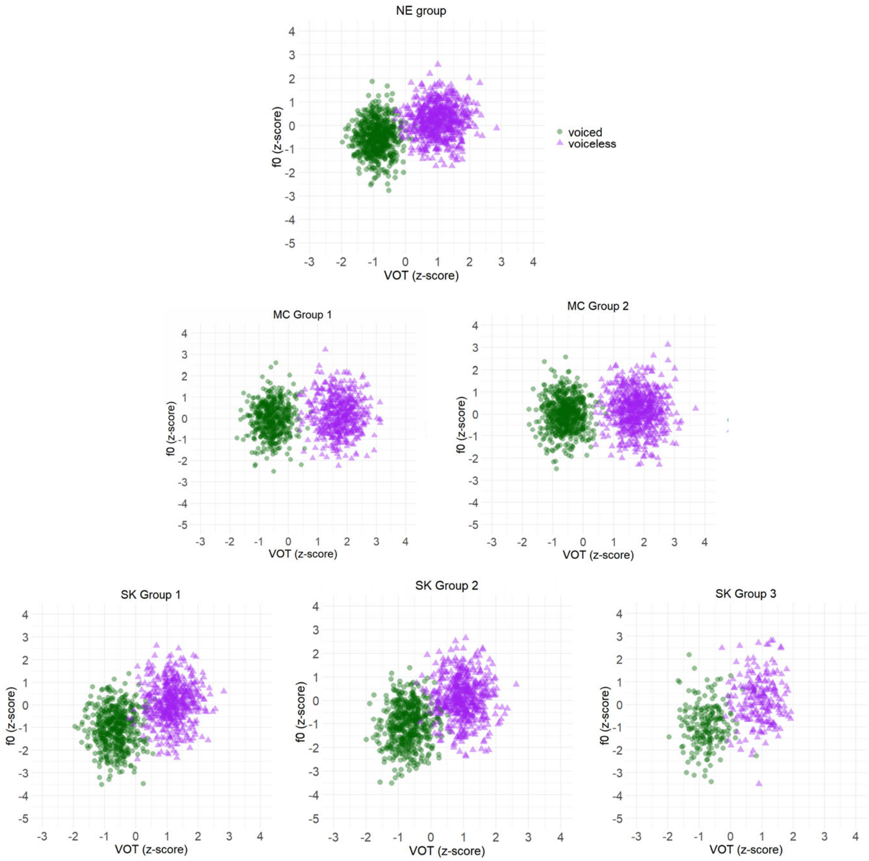

Figure 8. The distribution of VOT and f0 (z-scored) of different speaker groups

Closer inspection of different groups reveals that, first, for NE speakers, the voiced and voiceless stops are well separated by VOT while they overlap in f0. Still, NE speakers' voiceless stops have slightly higher f0 than voiced stops. Second, MC speakers show similar patterns to NE speakers, though their voiceless and voiced stops are more separated in terms of VOT and more overlapped in f0 than NE speakers. This pattern is consistent in both proficiency groups. Third, SK groups show relatively greater f0 separation than NE speakers. In terms of VOT, low proficiency groups (Groups 1 and 2) seem to have greater degrees of VOT overlap than other groups.

4. Discussion

The current study explored how language background and English proficiency impact EFL learners' production of English stop voicing contrast, focusing on realizations of two phonetic properties, stop VOT and post-onset f0. English words beginning on stop consonants produced by SK and MC learners of English were investigated, and the results indicated that learners of different native language background produced English stops differently. English proficiency also influenced stop production of SK speakers, but the proficiency effects were less evident in MC speakers.

4.1 L1 Influence on English stop productions

The current findings demonstrate that EFL learners' stop productions are influenced by their L1. The phonetic implementation of English stop voicing contrast differs in SK and MC learners of English. However, the L1 influence was not simple and interacted with the learners' gender.

First, for voiceless stops, male SK speakers produced longer VOT than NE speakers, and for voiced stops, female SK speakers produced longer VOT than NE speakers (see Figure 1 and Table 9). In Seoul Korean, female speakers produce shorter VOT for aspirated stops than male speakers, which can be attributed to the tonogenesis-like sound change in Seoul Korean (e.g., Oh 2011). As the VOT merger between Korean aspirated and lenis categories are more advanced in female than male speakers, female SK speakers often exhibit shorter VOT for aspirated stops than their male peers (e.g., Kang 2014). The longer VOT observed in English voiced stops produced by SK speakers suggests that they associate English voiced stops with Korean lenis, rather than fortis, stops, although Korean fortis stops have VOT values more comparable to English voiced stops (see Tables 1 and 3).

As for the MC speakers, male speakers produced longer VOT than the NE control speakers for both voiceless and voiced stops. The long VOT of voiced stops, again, provides evidence for L1 effects, as Mandarin unaspirated stops (Table 2) are typically produced with longer VOT than English voiced stops (Table 1). However, female MC speakers did not have significant difference from the NE counterparts. This is consistent with Li (2013) who also reports that male Mandarin speakers have longer VOT than female speakers when producing L2 English stops. Although the reason for this gender difference is unclear, the gender effects seem to be more evident in relatively more proficient learners (see Table 18 and Figure 5). This arguably suggests that male and female speakers may follow different learning trajectories or adopt different strategies in learning L2 speech sounds.

In addition, both SK and MC learners had less prevoicing for voiced stops than NE control speakers (Table 7), which also provides evidence for the L1 influence. Phonological voicing contrast is absent in both Seoul Korean and Mandarin Chinese, which could prevent the learners from producing English voiced stops with voicing lead. Furthermore, MC learners showed minimal prevoicing while SK speakers exhibited a moderate prevoicing tendency. Although both languages do not feature voiced stops in word-onset positions, Seoul Korean lenis stops are allophonically voiced in inter-sonorant positions (e.g., Jun 1998) while Mandarin unaspirated stops do not undergo such voicing (e.g., Deterding and Nolan 2007; Ahn et al. 2024). Presumably due to this difference in their L1, Mandarin learners may show little prevoicing when producing English voiced stops. It should be noted, however, that phrase-medial voicing of Korean lenis stops is not entirely consistent, as more recent phonetic studies (e.g., Shin 2021) report considerable variation depending on contexts. Nevertheless, the experience of producing voiced segments in at least some phonological environments, even if allophonic, may have led SK learners to produce prevoicing more frequently than MC learners, who lack such experience altogether.

The post-onset f0 serves as a secondary cue for English voicing contrast (e.g., Whalen et al. 1993), and EFL learners show L1 influence on how they utilize this secondary cue in L2. SK speakers showed greater post-onset f0 differences between English voiced and voiceless stops than the NE control speakers while MC speakers exhibited smaller differences (see Table 11 and Figure 2). The greater f0 differences in SK speakers seem to be due to the fact that their native language utilizes post-onset f0 as the primary cue for its lenis-aspirated contrast. On the other hand, in Mandarin, f0 is the primary cue for lexical tones and the association between onset aspiration and post-onset f0 is more complicated (e.g., Xu and Xu 2003; Luo 2018; Guo and Kwon 2022). This may present greater challenges for MC learners as they acquire the association between English voicing contrast and f0. The heavier reliance on f0 of SK learners as the cue for onset laryngeal contrast (Figure 8) is consistent with Schertz's (2015) findings. Although SK learners use stop VOT as the primary cue for English voicing contrast, they still show relatively higher post-onset f0 discriminant weights compared to native English speakers, providing evidence for the transfer of their native language phonetic strategies to L2 English. The cue weighting patterns of the MC learners reflect MC speakers' primary reliance on aspiration for distinguishing stop consonants, contrasting with the voicing distinctions in English (Yang 2015; Ahn et al. 2024).

4.2 English proficiency effects

The current findings provide limited evidence for English proficiency effects on the EFL learners' production of English stop voicing. The proficiency effects were most prominent in the SK learners' implementation of voiced stop VOT, whereas MC speakers show little proficiency effects. These outcomes can be interpreted in SLM (Flege 1995), which predicts L2 learners' speech production based on the perceived similarity of the target L2 phones to L1 speech sounds.

For SK speakers, the phonetic similarity between English voiceless stops and Korean aspirated stops is greater than that between English voiced stops and Korean lenis (or fortis) stops. This means, according to SLM, English voiceless stops are easier for SK speakers to acquire than voiced stops. However, once English voiceless stops are substituted with Korean aspirated stops, the subtle phonetic differences between the L1 and L2 phones may be even more difficult to overcome. On the other hand, English voiced stops, with greater phonetic distances to the closest L1 phones, may pose greater challenges for beginners and show slower phonetic learning.

Our findings are not entirely consistent with these predictions. First, voiceless stops exhibited significant VOT reduction from intermediate (Group 1, corresponding to CEFR B1) to advanced (Group 3, corresponding to CEFR C1) learners (Table 13). However, no significant changes were found in post-onset f0 for voiceless stops, indicating that SK learners may still fail to fully acquire the subtle phonetic differences between English voiceless stops and Korean aspirated stops (e.g., English voiceless stops typically exhibit lower post-onset f0 than Korean aspirated stops).

On the other hand, Group 1 SK learners produced voiced stops with significantly longer VOT than Group 3 learners (Table 13), as well as marginally lower post-voiced f0 values (Table 15). This suggests that as SK learners become more proficient, their voiced stop VOT decreases and post-onset f0 increases. The separation between English voiced stops and Korean lenis stops seems to occur at later stages of L2 learning. Previous studies in SK learners of English also show similar patterns that low-proficiency speakers exhibited longer VOT but lower f0 values for voiced stops compared to high-proficiency speakers (e.g., Kim 2012; Kong and Yoon 2013). Whether even more advanced learners eventually develop a distinct category for English voiceless stops and produce them with lower post-onset f0 than Korean aspirated stops remains an open question for future research. It should be noted, however, that the number of participants in SK Group 3 was small (n = 4), and the observed proficiency effects should therefore be interpreted with caution.

For MC speakers, we do not find clear proficiency effects. Previously, however, it has been shown that low-proficiency MC learners produce longer VOT than high-proficiency learners (e.g., Kato and Baese-Berk 2021). We suspect that, for MC speakers, the phonetic distance between English voiceless stops and Mandarin aspirated stops, as well as the distance between English voiced stops and Mandarin unaspirated stops, is likely smaller than the distance between English voiced stops and Korean lenis or fortis stops. It may rather be comparable to—or slightly greater than—the distance between English voiceless stops and Korean aspirated stops. If so, MC learners associate English stops to their native category in earlier stages of L2 learning, and their production does not likely change with their proficiency between intermediate to upper-intermediate levels. The proficiency groups in the current study are B1 and B2. Investigating more advanced learners may reveal clearer proficiency effects in terms of English stop voicing contrast, capturing the later separations of L2 categories from the similar L2 sounds. However, in EFL setting without immersive L2 learning experience, we were not able to find more advanced learners.

It should also be noted that the proficiency levels of the MC learners in this study were determined based on the College English Test (CET). The CET is a comprehensive English proficiency test that assesses university students' listening, reading, and writing skills. However, the CET-SET (College English Test - Spoken English Test), which evaluates oral communication skills, remains optional (e.g., Zheng and Cheng 2008; Adamson and Xia 2011). As a result, CET-based proficiency classifications primarily reflect general English abilities and may not accurately represent students' speaking skills or pronunciation (Bachman and Palmer 1996). The optional nature of the CET-SET, combined with the general focus of the CET, may also lead MC learners to overlook subtle phonetic details during their learning process.

The lack of observable proficiency effects in MC learners in the current study may, in part, be attributed to this assessment limitation.

5. Conclusion

The current study provides insight into how Seoul Korean and Mandarin Chinese learners of English, across different proficiency levels, implement phonetic properties related to onset laryngeal contrast when producing L2 English stops. The results demonstrated cross-linguistic differences in the realization of stop VOT and post-onset f0, reflecting the effect of the learners' native language. The study also reports complicated findings on the effects of English proficiency on the learners' English stop productions, but the overall patterns are in line with the predictions of the Speech Learning Model (SLM; Flege 1995).

Though the current study primarily aimed to examine the production patterns of EFL learners, the findings have potential applications in second language instruction and the Automatic Speech Recognition (ASR) development. Understanding the phonetic patterns of EFL learners may inform more effective pronunciation teaching strategies tailored to specific L1 backgrounds. For example, instructors may focus on providing more targeted feedback or practice activities to address segment-specific challenges, such as the production of voiced stops by lower-proficiency SK learners, whose longer VOT may lead to perceptual confusion. In addition, phonemes that share similar characteristics are frequently confused with each other in ASR systems and stop consonants remain particularly error-prone (e.g., Choe et al. 2022). Moreover, the variability in non-native speech further challenges ASR performance (Vu et al. 2014). Integrating L2 speech characteristics, not just the primary cues like VOT but also the secondary cues such as post-onset f0, may help enhance the recognition accuracy. The current findings may also inform the refinement of Computer-Assisted Pronunciation Training (CAPT) tools that support self-paced learning and help learners identify and improve their pronunciation weakness more effectively.

References

Abramson, Arthur S. and Leigh Lisker. 1985. Relative power of cues: F0 shift versus voice timing. In Victoria Fromkin (ed.), *Phonetic linguistics: Essays in honor of Peter Ladefoged*, 25-33. Cambridge, MA: Academic Press.

Adamson, Bob and Beibei Xia. 2011. A case study of the College English Test and ethnic minority university students in China: Negotiating the final hurdle. *Multilingual Education* 1(1): 1-11.

Ahn, Suzy, Harim Kwon, and Matthew Faytak. 2024. Tongue position in Mandarin Chinese voiceless stops. *JASA Express Letters* 4(2): 1-9.

Allen, Sean J., Joanne L. Miller, and David DeSteno. 2003. Individual talker differences in voice-onset-time. *The Journal of the Acoustical Society of America* 113(1): 544-552.

Alves, Ubirata Kickhofel and Pedro Luis Luchini. 2017. Effects of perceptual training on the identification and production of word-initial voiceless stops by Argentinean learners of English. *Ilha do Desterro* 70(3): 15-32.

Bachman, Lyle F. and Adrian S. Palmer. 1996. *Language testing in practice: Designing and developing useful language tests (vol. 1)*. Oxford: Oxford University Press.

Barr, Dale J., Roger Levy, Christoph Scheepers, and Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68(3): 255-278.

Bates, Douglas, Martin Machler, Ben Bolker, and Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1): 1-48.

Boersma, Paul and David Weenink. 2020. Praat: Doing phonetics by computer (version 6.1.16) [Computer software]. Retrieved from https://www.praat.org.

Cho, Taehong and Peter Ladefoged. 1999. Variation and universals in VOT: Evidence from 18 languages. *Journal of Phonetics* 27(2): 207-229.

Chodroff, Eleanor and Colin Wilson. 2017. Structure in talker-specific phonetic realization: Covariation of stop consonant VOT in American English. *Journal of Phonetics* 61: 30-47.

Chodroff, Eleanor, Alessandra Golden, and Colin Wilson. 2019. Covariation of stop voice onset time across languages: Evidence for a universal constraint on phonetic realization. *The Journal of the Acoustical Society of America* 145(1): EL109-EL115.

Choe, June, Yiran Chen, May Pik Yu Chan, Aini Li, Xin Gao, and Nicole Holliday. 2022. Language-specific effects on automatic speech recognition errors for world Englishes. In Nicoletta Calzolari and Chu-Ren Huang (eds.), *Proceedings of the 29th International Conference on Computational Linguistics*, 7177-7186.

Derwing, Tracey M. and Murray J. Munro. 2005. Second language accent and pronunciation teaching: A research-based approach. *TESOL Quarterly* 39(3): 379-397.

Deterding, David and Francis Nolan. 2007. Aspiration and voicing of Chinese and English plosives. In Jürgen Trouvain and William John Barry (eds.), *Proceedings of the 16th*

International Congress of Phonetic Sciences, 385-388.

Dmitrieva, Olga, Fernando Llanos, Amanda A. Shultz, and Alexander L. Francis. 2015. Phonological status, not voice onset time, determines the acoustic realization of onset f0 as a secondary voicing cue in Spanish and English. *Journal of Phonetics* 49: 77-95.

Docherty, Gerard J. 1992. *The timing of voicing in British English obstruents*. Berlin, New York: De Gruyter Mouton.

Flege, James E. 1995. Second language speech learning: Theory, findings, and problems. In Winifred Strange (ed.), *Speech perception and linguistic experience: Issues in cross-language research,* 233-277. London: York Press.

Guo, Yuting. 2020. *Production and perception of laryngeal contrasts in Mandarin and English by Mandarin speakers*. PhD Dissertation. George Mason University.

Guo, Yuting and Harim Kwon. 2022. Production and perception of Mandarin laryngeal contrast: The role of post-plosive f0. *Frontiers in Communication* 7: 1-15.

Hamzah, Mohd Hilmi, Ahmed Elsayed Samir Madbouly, Hasliza Abdul Halim, and Abdul Halim Abdullah. 2020. The production of the English stop voicing contrast by Arab L2 speakers of English. *Indonesian Journal of Applied Linguistics* 10(2): 434-444.

Hanson, Helen M. 2009. Effects of obstruent consonants on fundamental frequency at vowel onset in English. *The Journal of the Acoustical Society of America* 125(1): 425-441.

Jin, Yan, Wei Jie, and Wei Wang. 2022. A study on the alignment between the College English Test Band 4 and 6 (CET-4 and CET-6) and language proficiency standards [In Chinese]. *Foreign Language World* 209(2): 24-32.

Jun, Sun-Ah. 1998. The accentual phrase in the Korean prosodic hierarchy. *Phonology* 15(2): 189-226.

Kagaya, Ryohei. 1974. A fiberscopic and acoustic study of the Korean stops, affricates and fricatives. *Journal of Phonetics* 2(2): 161-180.

Kang, Kyoung-Ho and Susan G. Guion. 2006. Phonological systems in bilinguals: Age of learning effects on the stop consonant systems of Korean-English bilinguals. *The Journal of the Acoustical Society of America* 119(3): 1672-1683.

Kang, Kyoung-Ho and Susan G. Guion. 2008. Clear speech production of Korean stops: Changing phonetic targets and enhancement strategies. *The Journal of the Acoustical Society of America* 124(6): 3909-3917.

Kang, Yoonjung. 2014. Voice onset time merger and development of tonal contrast in Seoul Korean stops: A corpus study. *Journal of Phonetics* 45: 76-90.

Kato, Misaki and Melissa M. Baese-Berk. 2021. Contextually-relevant enhancement of non-native phonetic contrasts. *Journal of Phonetics* 88: 1-12.

Keating, Patricia A. 1984. Phonetic and phonological representation of stop consonant voicing. *Language* 60(2): 286-319.

Kim, Mi-Ran Cho. 1994. *Acoustic characteristics of Korean stops and perception of English stop consonants*. PhD Dissertation. The University of Wisconsin-Madison.

Kim, Mi-Ryoung. 2012. L1-L2 transfer in VOT and f0 production by Korean English learners: L1 sound change and L2 stop production. *Phonetics and Speech Sciences* 4(3): 31-41.

Kong, Eun Jong and In Hee Yoon. 2013. L2 proficiency effect on 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.

Kuznetsova, Alexandra, Per B. Brockhoff, and Rune HB Christensen. 2017. lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software* 82(13): 1-26.

Kwon, Harim. 2019. The role of native phonology in spontaneous imitation: Evidence from Seoul Korean. *Laboratory Phonology* 10(1): 1-24.

Kwon, Harim. 2021. A non-contrastive cue in spontaneous imitation: Comparing mono- and bilingual imitators. *Journal of Phonetics* 88: 1-19.

Lee, Hyunjung and Allard Jongman. 2012. Effects of tone on the three-way laryngeal distinction in Korean: An acoustic and aerodynamic comparison of the Seoul and South Kyungsang dialects. *Journal of the International Phonetic Association* 42(2): 145-169.

Lenth, Russell V. 2020. Emmeans: Estimated marginal means, aka least-squares means. R Package Version 1.5.3. CRAN.

Li, Fangfang. 2013. The effect of speakers' sex on voice onset time in Mandarin stops. *The Journal of the Acoustical Society of America* 133(2): EL142-EL147.

Liao, Shi-Jer. 2005. *Interlanguage production of English stop consonants: A VOT analysis.* Unpublished Master's Thesis. National Kaohsiung Normal University.

Lisker, Leigh and Arthur S. Abramson. 1964. A cross-language study of voicing in initial stops: Acoustical measurements. *Word* 20(3): 384-422.

Liu, Hanjun, Manwa L. Ng, Mingxi Wan, Supin Wang, and Yi Zhang. 2008. The effect of tonal changes on voice onset time in Mandarin esophageal speech. *Journal of Voice* 22(2): 210-218.

Luo, Qian. 2018. *Consonantal effects on F0 in tonal languages*. PhD Dissertation. Michigan State University.

Oh, Eunjin. 2011. Effects of speaker gender on voice onset time in Korean stops. *Journal of Phonetics* 39(1): 59-67.

R Core Team. 2021. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rochet, Bernard L. and Yanmei Fei. 1991. Effect of consonant and vowel context on Mandarin Chinese VOT: Production and perception. *Canadian Acoustics* 19(4): 105-106.

Schertz, Jessamyn, Taehong Cho, Andrew Lotto, and Natasha Warner. 2015. Individual differences in phonetic cue use in production and perception of a non-native sound contrast. *Journal of Phonetics* 52: 183-204.

Shin, Ji-young. 2001. The phonological environments and patterns of Korean voiceless vowelization [In Korean]. *Hangugeo Hak (Korean Linguistics)* 14: 235-259.

Shultz, Amanda A., Alexander L. Francis, and Fernando Llanos. 2012. Differential cue weighting in perception and production of consonant voicing. *The Journal of the Acoustical Society of America* 132(2): EL95-EL101.

Silva, David J. 2006. Acoustic evidence for the emergence of tonal contrast in contemporary Korean. *Phonology* 23(2): 287-308.

Sun, Sylvia Henel. 1997. *The development of a lexical tone phonology in American adult learners of standard Mandarin Chinese*. PhD Dissertation. University of Hawai'i at Manoa.

Van Alphen, Petra M. and Roel Smits. 2004. Acoustical and perceptual analysis of the voicing distinction in Dutch initial plosives: The role of prevoicing. *Journal of Phonetics* 32(4): 455-491.

Venables, William N. and Brian D. Ripley. 2002. *Modern Applied Statistics with S*. New York, NY: Springer.

Vu, Ngoc Thang, Yuanfan Wang, Marten Klose, Zlatka Mihaylova, and Tanja Schultz. 2014. Improving ASR performance on non-native speech using multilingual and crosslingual information. In Haizhou Li, Helen M. Meng, Bin Ma, Engsiong Chng, Lei Xie (eds.), *Fifteenth Annual Conference of the International Speech Communication Association (INTERSPEECH), 11-15.*

Whalen, Douglas H. and Andrea G. Levitt. 1995. The universality of intrinsic F0 of vowels. *Journal of Phonetics* 23(3): 349-366.

Whalen, Douglas H., Arthur S. Abramson, Leigh Lisker, and Maria Mody. 1990. Gradient effects of fundamental frequency on stop consonant voicing judgments. *Phonetica* 47(1-2): 36-49.

Whalen, Douglas H., Arthur S. Abramson, Leigh Lisker, and Maria Mody. 1993. F0 gives voicing information even with unambiguous voice onset times. *The Journal of the Acoustical Society of America* 93(4): 2152-2159.

Xu, Changxin and Yi Xu. 2003. Effects of consonant aspiration on Mandarin tones. *Journal of the International Phonetic Association* 33(2): 165-181.

Yang, Bei. 2015. *Perception and production of Mandarin tones by native speakers and L2 learners*. Berlin: Springer.

Yu, Hye Jeong. 2018. Tonal development and voice quality in the stops of Seoul Korean. *Phonetics and Speech Sciences* 10(4): 91-99.

Zheng, Ying and Liying Cheng. 2008. Test review: College English Test (CET) in China. *Language Testing* 25(3): 408-417.

Yalin Wang
PhD Candidate
Department of English Language and Literature
Seoul National University
1 Gwanak-ro, Gwanak-gu,
Seoul, 08826, Korea

E-mail: yalin1108@snu.ac.kr

Harim Kwon
Associate Professor
Department of English Language and Literature
Seoul National University
1 Gwanak-ro, Gwanak-gu,
Seoul, 08826, Korea
E-mail: harimkwon@snu.ac.kr

Received: 2025. 02. 14.
Revised: 2025. 05. 12.
Accepted: 2025. 05. 13.