Linguistic Research 36(1), 67-89
DOI: 10.17250/khisli.36.1.201903.003

Effects of gender on the use of voice onset time and fundamental frequency cues in perception and production of English stops*

Eunjin Oh
(Ewha Womans University)

Oh, Eunjin. 2019. Effects of gender on the use of voice onset time and fundamental frequency cues in perception and production of English stops. *Linguistic Research* 36(1), 67-89. This study investigated the effects of gender on the use of voice onset time (VOT) and fundamental frequency (F0) cues in perceiving and producing stop voicing in English. This line of inquiry stemmed from the consistent finding in previous studies that females produced longer mean VOT values than males for voiceless stops in English. The results of a forced-choice identification experiment showed that listener gender had no significant effect on the use of VOT and F0 cues in categorizing voiced and voiceless stops. The results of a production experiment found that females produced a smaller average VOT value for voiceless stops than males, contradicting the results of previous studies that females consistently showed longer mean VOT values. The statistical analyses did not identify any significant gender-based differences in VOT values and VOT/F0 distinctions between voiced and voiceless stops. The results of the perception and production experiments may indicate that the gender is not a factor in the use of VOT and F0 cues for stop voicing in English. **(Ewha Womans University)**

Keywords gender, VOT, F0, perception, production, stops, English

1. Introduction

Voice onset time (VOT) refers to the time between the burst of a stop and the onset of vocal cord vibration in the following segment. The VOT is the primary cue for voicing contrast in English in which voiced stops are realized as short-lag VOTs and voiceless stops as long-lag VOTs. Voiced stops can be

* I am grateful to anonymous reviewers for their helpful comments and suggestions, Meghan Sumner for the use of the phonetics laboratory (Experimental Linguistics Lab) at Stanford University, and the subjects for their participation. This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2016S1A5A2A01026875).

produced as fully voiced between voiced sounds, but are generally produced with lag VOT values in other contexts. Research has shown that fundamental frequency (F0) in the following vowel is the secondary cue for stop voicing in English (e.g., Whalen et al. 1993; Shultz et al. 2012). The F0 values are generally higher for voiceless stops than their voiced counterparts. Research has also indicated that F0 trajectory shapes differ depending on the voicing of the stops. For example, Whalen et al. (1993) and Shultz et al. (2012) based their perception studies on the assumption that the F0s exhibit flat or rising contours into the following vowel after the voiced stops while displaying falling contours after the voiceless stops.

VOT values reportedly vary depending on several phonetic or sociophonetic factors, such as vowel contexts, stop places of articulation, speaking rate, utterance types (e.g., spoken in isolated syllables or in sentences), speaker age, and speaker gender (Morris et al. 2008 and references cited therein). Previous studies regarding the effects of speaker gender on VOT in English voiceless stops have consistently found that females produced longer average VOT values than males (e.g., Sweeting and Baken 1982; Swartz 1992; Ryalls et al. 1997; Whiteside and Irving 1997, 1998; Morris et al. 2008). Researchers have interpreted this as resulting from physiological factors that make it relatively easier for male speakers with generally larger supraglottal cavities to form sub and supraglottal air pressure differences for vocal cord vibration, which shortens their VOTs (e.g., Smith 1978; Swartz 1992; Whiteside and Irving 1997, 1998; Koenig 2000; Whiteside et al. 2004).

On the other hand, research results regarding voiced stops in English have been inconsistent. While some studies have reported longer mean VOT values for females than males (Swartz 1992; Ryalls et al. 1997; Whiteside and Irving 1997; Morris et al. 2008), other studies have reported the opposite (Sweeting and Baken 1982; Whiteside and Irving 1998). The former results have been attributed to the physiological differences between genders, as in the case of voiceless stops. Meanwhile, researchers have interpreted the latter as stemming from differences in styles of speech between genders. In other words, they contend that females tend to speak more carefully than males, and therefore produce shorter VOTs for voiced stops and longer VOTs for voiceless stops in order to secure sufficient phonological contrasts between the two stops (e.g., Whiteside

and Irving 1997, 1998).

A number of cross-language studies have investigated the effects of speaker gender on the VOT values of stop consonants. The physiological factor described above evidently explains gender-related variations in the VOT values of voiced stops in Swedish (Helgason and Ringen 2008; n = 6), voiceless stops in Mandarin Chinese (Li 2013; n = 20), and voiceless aspirated stops in Madurese (Misnadin et al. 2015; n = 14), but not those of aspirated stops in Seoul Korean (Oh 2011; n = 38) or voiceless stops in Serbian (Sokolović-Perović 2012; n = 12). Meanwhile, the speech style factor according to which females tend to speak more carefully than males and produce more VOT distinctions between stop categories appears to explain gender-related variations in Mandarin Chinese (Li 2013), but not those in Swedish (Helgason and Ringen 2008) or Seoul Korean (Oh 2011). These results may indicate that the gender effects on VOT vary cross-linguistically.

Several studies have examined the effects of gender on VOT in the production of English stops as described above, but few have considered the effects of listener gender in perceiving the VOT cue of English stops. The main aim of this study was to investigate the effects of listener gender on the use of VOT and F0 cues in perceiving stop voicing in English. This study also included a production test that examined the effects of speaker gender in the use of VOT and F0 cues to determine whether the results of previous studies on production hold up and whether perception results are supported by production results. The research questions of this study were as follows: (1) Do male *vs.* female native listeners of English differ in the use of acoustic cues in their identification of stop voicing? (2) Do listeners show different perceptual patterns when listening to male *vs.* female stimuli? (3) Do male speakers show a smaller average VOT value than female speakers for voiceless stops as reported in previous studies? Is the gender difference statistically significant? (4) Do male *vs.* female speakers show significantly different F0 distinctions (i.e., F0 contrasts) between voiced and voiceless stops? Additionally, the study examined whether voiced and voiceless stops in English show rising and falling F0 contours, respectively, in following vowels, as assumed in previous studies. Sections 2 and 3 of this paper report the methods and the results of the production and perception experiments, respectively, and section 4 summarizes the main results of the study and discusses pertinent issues.

2. Perception experiment

2.1 Methods

Ten male (age mean 20.6; range 18-26) and ten female (age mean 18.7; range 18-21) subjects participated in the perception experiment as listeners. All of them also participated in the production experiment of this study. The perception experiment was conducted approximately one week after the production experiment (see section 3.1 for information regarding the participants of the production experiment). All participants were native speakers of American English born and raised in the U.S. and studying as undergraduate or graduate students at a university located in the state of California. Only two of the twenty participants had resided in a foreign country for more than six months (one male and one female had resided in Europe for two years and one year, respectively). No participants had foreign language fluency above an intermediate level.

One male and one female "tan" (/tæn/) tokens were selected among the production data pool to be used as base tokens for creating perception stimuli. These tokens were selected based on their relatively long VOTs and clear acoustic patterns. The VOT of the male base token was 78 ms, its onset F0 was 107 Hz, and its mid F0 was 96 Hz. The VOT of the female base token was 80 ms, its onset F0 was 235 Hz, and its mid F0 was 212 Hz.

Stimuli were manipulated using the PSOLA function of Praat (Boersma and Weenink 2016). Stimuli were generated in 7 VOT steps and 7 F0 steps. The VOTs for the male stimuli were generated in logarithmic steps of 10 ms (step 1), 14 ms, 20 ms, 28 ms, 40 ms, 56 ms, and 80 ms (step 7), and the mid F0s as 70 Hz (step 1), 80 Hz, 90 Hz, 100 Hz, 110 Hz, 120 Hz, and 130 Hz (step 7) featuring 10 Hz differences (Kong and Yoon 2013 for the logarithmic steps). The VOTs for the female stimuli were manipulated in logarithmic steps of 8 ms (step 1), 12 ms, 18 ms, 27 ms, 40 ms, 60 ms, and 90 ms (step 7), and the mid F0s as 170 Hz (step 1), 183 Hz, 196 Hz, 209 Hz, 222 Hz, 235 Hz, and 248 Hz (step 7) in 13 Hz differences. The total number of generated stimuli was 98 (7 VOTs * 7 F0s * 2 genders). The relative duration function was used to manipulate VOTs in which the original VOT was shortened or lengthened to the intended VOT

values. For the F0 manipulation, the F0 at the vowel midpoint was used as a reference point, and all F0 points within the vowel and the following nasal were lowered or raised together with designated ratios. The F0 contour shapes were maintained so the stimuli sounded more natural.

The perception test was conducted in a soundproof room in a phonetics laboratory at the university the participants attended, using the MFC function of Praat. Participants listened to the stimuli through an Audio-Technica ATH-M40X headphone. The response categories were "Tan" (/tæn/) and "Dan" (/dæn/) which were presented on a computer screen. Listeners were instructed to choose the word that sounded more similar. There was no time limit, but they were not allowed to change their answers. Each listener was presented with the stimuli five times in a randomized order, resulting in a total of 490 trials (98 unique stimuli * 5 repetitions). A practice session was conducted with 8 stimuli randomly chosen from the test pool. The task took each listener approximately 20 minutes.

2.2 Results

2.2.1 Effects of listener gender on the perception of English stops

This section discusses the findings of this study regarding differences in male and female listeners' use of the acoustic cue (VOT and F0) patterns in distinguishing between voiced and voiceless stops in English. Figure 1 displays the percentages of the voiceless responses (/tæn/) to the stimuli, which manipulated 7 steps of VOTs (left) and F0s (right) as a function of listener gender for male stimuli. The percentage values represent the ratio of the sum of voiceless responses at each VOT and F0 step. The figures indicate that both male and female listeners relied mainly on VOT in distinguishing between the voiced and voiceless stops. The identification curves are steep in relation to the change of VOT values and gentle in relation to the change of F0 values.

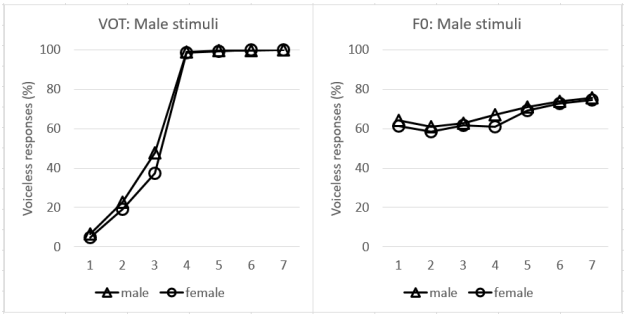

Figure 1. Percentages of voiceless responses (y-axis) by 7 steps of VOTs and F0s (x-axis) as a function of listener gender: Responses to male stimuli.

Table 1. Results of the mixed effects logistic regression for the listener gender model: Responses to male stimuli.

Male stimuli	Estimate	Std. error	z value	Pr(>|z|)
(Intercept)	2.62574	0.27885	9.416	<0.0001
VOT	2.39510	0.11166	21.449	<0.0001
F0	0.45345	0.04639	9.775	<0.0001
ListenerGender	0.20660	0.39557	0.522	0.6010
VOT*ListenerGender	−0.13947	0.15146	−0.921	0.3570
F0*ListenerGender	−0.06509	0.06377	−1.021	0.3070

Mixed effects logistic regression was used to model the binary response data. The program R (R Core Team 2016) and the package "lme4" (Bates et al. 2015) were used for the data analyses. The data were analyzed separately for male stimuli and female stimuli. The VOT and F0 steps were coded as levels (−3, −2, −1, 0, 1, 2, 3). Listener gender models were specified with VOT, F0, and Listener Gender as the fixed effects, and Subject as the random effect [Response~VOT+F0+ ListenerGender+VOT*ListenerGender+F0*ListenerGender+(1|Subject)]. The optimizer "bobyqa" was used. The reference for Listener Gender was male.

Table 1 presents the outputs of the listener gender model for the male stimuli. Both VOT (p < 0.0001) and F0 (p < 0.0001) were significant predictors, indicating that both cues are perceptually effective. The estimate coefficients were approximately 2.40 for VOT and 0.45 for F0, indicating heavier reliance on VOT than F0 in categorizing voiced and voiceless stops. The total number of voiceless responses were 1663 (67.9 %) for male listeners and 1608 (65.6 %) for female listeners, and the effect of Listener Gender in this case was not statistically

significant (p = 0.6010). Likewise, the interaction terms of VOT*Listener Gender (p = 0.3570) and F0*Listener Gender (p = 0.3070) were not statistically significant. As Figure 1 shows, for VOT, the differences between the highest and lowest percentages of the voiceless responses were 93.4 % (100 % at step 7 − 6.6 % at step 1) for male listeners and 95.1 % (100 % at steps 6 and 7 − 4.9 % at step 1) for female listeners. Regarding F0, the differences were 14.6 % (75.7 % at step 7 − 61.1 % at step 2) for male listeners and 16.0 % (74.6 % at step 7 − 58.6 % at step 2) for female listeners. These statistical results indicate that male and female listeners' use of VOT and F0 cues in distinguishing between voiced and voiceless stops did not differ when listening to the male stimuli.

Figure 2 displays the percentages of the voiceless responses by the 7 steps of VOTs (left) and F0s (right) as a function of listener gender for female stimuli. Again, the identification curves are steep in relation to the change of VOT values and gentle in relation to the change of F0 values, indicating that both male and female listeners relied more on VOT in categorizing voiced and voiceless stops.

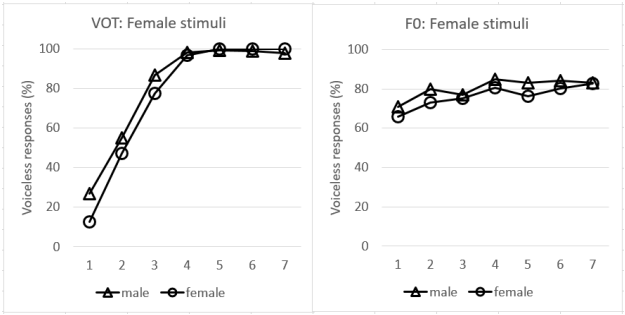

Figure 2. Percentages of voiceless responses (y-axis) by 7 steps of VOTs and F0s (x-axis) as a function of listener gender: Responses to female stimuli.

Table 2. Results of the mixed effects logistic regression for the listener gender model: Responses to female stimuli.

Female stimuli	Estimate	Std. error	z value	Pr(>|z|)
(Intercept)	4.18960	0.40721	10.289	<0.0001
VOT	2.25518	0.11266	20.017	<0.0001
F0	0.38615	0.04458	8.662	<0.0001
ListenerGender	−0.86568	0.55463	−1.561	0.1186
VOT*ListenerGender	−0.78728	0.13311	−5.914	<0.0001
F0*ListenerGender	−0.16370	0.05762	−2.841	0.0045

Table 2 presents the outputs of the listener gender model for the female stimuli. As with the male stimuli, both VOT ($p < 0.0001$) and F0 ($p < 0.0001$) were significant predictors. The estimate coefficients were approximately 2.26 for VOT and 0.39 for F0, indicating greater reliance on VOT than F0 in categorizing voiced and voiceless stops when listening to the female stimuli. The total number of the voiceless responses were 1969 (80.4 %) for male listeners and 1868 (76.2 %) for female listeners, and, consistent with the case of the male stimuli (Table 1), this Listener Gender effect was not statistically significant ($p = 0.1186$). The interaction between VOT and Listener Gender was statistically significant ($p < 0.0001$). As Figure 2 shows, for VOT, the differences between the highest and lowest percentages of the voiceless responses were 72.5 % (99.4 % at step 5 − 26.9 % at step 1) for male listeners and 87.4 % (100 % at steps 5, 6, and 7 − 12.6 % at step 1) for female listeners. The interaction between F0 and Listener Gender was also statistically significant ($p = 0.0045$). Regarding F0, the differences between the highest and lowest percentages of the voiceless responses were 14.0 % (84.9 % at step 4 − 70.9 % at step 1) for male listeners and 16.6 % (82.6 % at step 7 − 66.0 % at step 1) for female listeners. These statistical results may indicate that female listeners used the VOT and F0 cues more sensitively than male listeners when listening to the female stimuli.

2.2.2 Perceptual responses to the voicing cues by individual listeners

Figure 3 displays the percentages of the voiceless responses by 7 steps of VOT (left) or F0 (right) as a function of individual listeners. The upper graphs present the responses to the male stimuli and the lower graphs present the responses to the female stimuli. The results show no deviance in listeners' use of the voicing cues in that they all demonstrated sharper slopes for VOT than F0.

Binary logistic regression analyses for individual listeners were conducted to model the relations between the responses and the acoustic cues in categorizing voiced and voiceless stops at the individual level. Table 3 provides beta coefficient values for the VOT and F0 cues of individual listeners. Larger beta-coefficient values represent larger cue effects in the perception of stop voicing. The beta-coefficient values of VOT were larger than those of F0 for all

listeners, indicating that VOT served as a better predictor for distinguishing between the two stops across all individual listeners. For male listeners, the mean values of the beta coefficients for VOT and F0 were 2.635 *vs.* 0.495 for male stimuli, and 2.067 *vs.* 0.326 for female stimuli, respectively.[1] For female listeners, the mean values of the beta coefficients for VOT and F0 were 2.790 *vs.* 0.474 for male stimuli and 2.583 *vs.* 0.440 for female stimuli, respectively.

Independent *t*-tests indicated that the differences in the beta coefficient values between male and female listeners were not statistically significant for VOT (p = 0.769) and F0 (p = 0.845) with male stimuli and VOT (p = 0.179) and F0 (p = 0.307) with female stimuli. These results from the individual listener analyses indicate that listener gender had no significant effects in the use of the stop voicing cues.

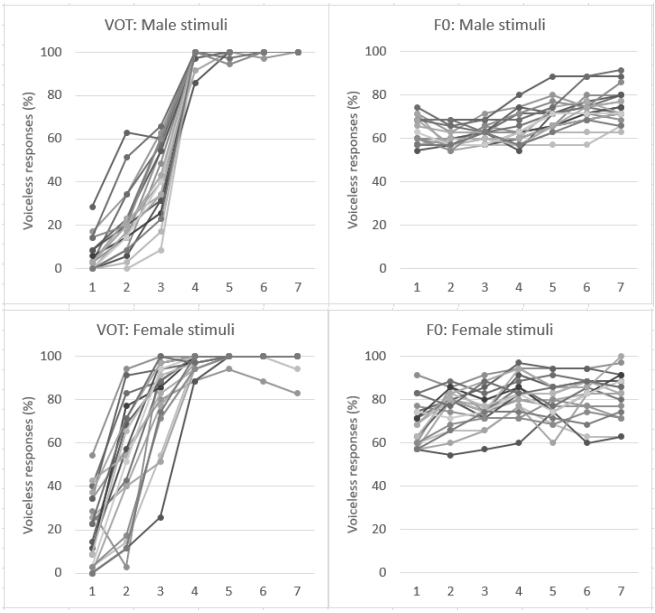

Figure 3. Percentages of voiceless responses (y-axis) by 7 VOT and 7 F0 steps (x-axis) by individual listeners: Male (upper) and female (lower) stimuli.

1 The mean values of the beta coefficients for male stimuli were calculated except those for M4. The binary logistic regression analysis for this listener produced abnormal results regarding the beta coefficients for male stimuli. The beta coefficient values were extremely large (97.123 for VOT and 13.683 for F0), while their *p*-values were 0.966 and 0.968, respectively.

Table 3. Beta coefficient values for VOT and F0 cues in the perception of English stops from binary logistic regression analyses for individual listeners.

Male listeners	Male stimuli		Female stimuli		Female listeners	Male stimuli		Female stimuli	
	VOT	F0	VOT	F0		VOT	F0	VOT	F0
M1	3.621	0.503	2.939	0.330	F1	2.596	0.576	2.366	0.300
M2	1.486	0.192	0.488	0.045	F2	1.814	0.446	2.528	0.289
M3	2.022	0.261	1.585	0.371	F3	2.632	0.242	2.274	0.673
M4	N.A.	N.A.	1.713	0.025	F4	2.498	0.573	1.195	0.277
M5	2.154	0.327	1.568	0.132	F5	5.368	0.436	2.652	0.586
M6	3.683	0.976	2.018	0.074	F6	3.128	0.533	3.726	0.601
M7	4.946	0.902	3.409	0.689	F7	1.946	0.640	2.387	0.258
M8	2.289	0.643	2.868	0.786	F8	2.392	0.360	2.376	0.667
M9	1.416	0.497	2.220	0.609	F9	1.840	0.499	2.217	0.172
M10	2.095	0.151	1.857	0.198	F10	3.683	0.439	4.111	0.577
Mean	2.635	0.495	2.067	0.326	Mean	2.790	0.474	2.583	0.440

2.2.3 Effects of stimulus gender on the perception of English stops

This section presents the findings of the study regarding differences in listeners' response patterns as a function of stimulus gender, i.e., when they listened to male *vs.* female stimuli. The response data for the male and female stimuli were merged in order to examine the effects of stimulus gender on the perception of English stops. Figure 4 displays the percentages of the voiceless responses by the 7 steps of VOT (left) or F0 (right) as a function of stimulus gender. Table 4 presents results of the mixed effects logistic regression for the stimulus gender model.

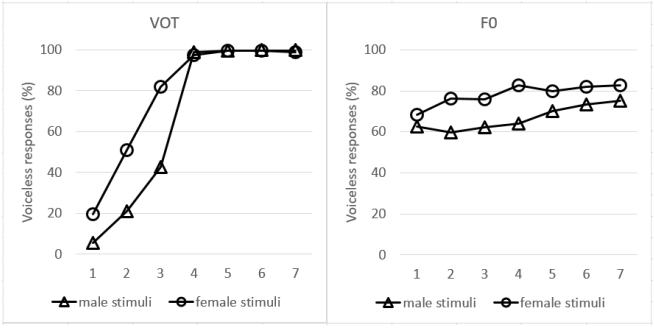

Figure 4. Percentages of voiceless responses by 7 steps of VOT (left) and F0 (right) as a function of stimulus gender.

Table 4. Results of the mixed effects logistic regression for the stimulus gender model: The VOT (upper) and F0 (lower) model.

VOT model	Estimate	Std. error	z value	Pr(>|z|)
(Intercept)	3.24677	0.19313	16.812	<0.0001
VOT	1.58354	0.05170	30.630	<0.0001
StimulusGender	−0.76548	0.13702	−5.587	<0.0001
VOT*StimulusGender	0.52539	0.07976	6.587	<0.0001

F0 model	Estimate	Std. error	z value	Pr(>|z|)
(Intercept)	1.321106	0.070021	18.867	<0.0001
F0	0.124644	0.017690	7.046	<0.0001
StimulusGender	−0.602713	0.046938	−12.841	<0.0001
F0*StimulusGender	−0.006216	0.023496	−0.265	0.7910

The mixed effects logistic regression models were run separately for VOT and F0. The VOT model was specified with VOT and Stimulus Gender as the fixed effects, and Subject as the random effect [Response~VOT+StimulusGender+ VOT*StimulusGender+(1|Subject)], and the F0 model with F0 and Stimulus Gender as the fixed effects, and Subject as the random effect [Response~F0+ StimulusGender+VOT*StimulusGender+(1|Subject)]. The optimizer "bobyqa" was used. The reference for Stimulus Gender was male.

The results of the VOT model (upper) and the F0 model (lower) in Table 4 indicate that VOT (p < 0.0001) and F0 (p < 0.0001) were significant predictors. The effect of Stimulus Gender was statistically significant in both the VOT (p < 0.0001) and F0 (p < 0.0001) models. The total numbers of voiceless responses were 3271 (66.8 %; male listeners 1663 + female listeners 1608) for male stimuli and 3837 (78.3 %; male listeners 1969 + female listeners 1868) for female stimuli, indicating that listeners chose significantly more voiceless stops when exposed to the female stimuli (see section 4 for discussion).

The interaction between VOT and Stimulus Gender was statistically significant (p < 0.0001), indicating that the main effect of VOT is modulated by Stimulus Gender. Listeners relied more on VOT differences in distinguishing voiced and voiceless stops when listening to male stimuli than female stimuli. As the left side of Figure 4 shows, the differences between the highest and lowest percentages of the voiceless responses by the 7 VOT steps were 94.3 % (100 % at step 7 − 5.7 % at step 1) for male stimuli and 80.0 % (99.7 % at step

5 – 19.7 % at step 1) for female stimuli, indicating that listeners categorized the voiced and voiceless stops by means of the VOT cue more sensitively when listening to male stimuli than when listening to female stimuli. The interaction between F0 and Stimulus Gender was not statistically significant ($p < 0.7910$). As the right side of Figure 4 shows, the differences between the highest and lowest percentages of the voiceless responses by the 7 F0 steps were 15.2 % (75.1 % at step 7 – 59.9 % at step 2) for the male stimuli and 14.3 % (82.7 % at steps 4 and 7 – 68.4 % at step 1) for the female stimuli, indicating that the listeners' use of the F0 cue in distinguishing between voiced and voiceless stops when listening to male *vs.* female stimuli did not differ.

In summary, the perception experiment showed that both VOT and F0 were significant predictors in categorizing voiced and voiceless stops, but listeners relied more heavily on VOT than F0. The statistical analyses indicated that listener gender had no significant effect on the binary responses. The binary logistic regression analyses of individual listeners showed that male and female listeners did not differ in their use of VOT and F0 cues. The experiment also found that stimulus gender had a significant effect that listeners chose more voiceless stops when listening to female stimuli than when listening to male stimuli.

3. Production experiment

3.1 Methods

Eleven male (age mean 20.6; range 18-26) and eleven female (age mean 18.9; range 18-21) speakers participated in the production experiment. Among these, ten male and ten female speakers also participated in the perception experiment described in section 2.1. The participants in the production experiment had the same characteristics as the participants in the perception experiment.

Regarding test materials, the experiment used twelve monosyllabic words containing /CVn/ ("bin," "din," "ghin," "ban," "dan," "gan," "pin," "tin," "kin," "pan," "tan," and "can") in which C was either a voiced or a voiceless stop in three places of articulation (bilabial, alveolar, or velar) and V was either a high

or a low vowel (/ɪ/ or /æ/). The words were presented in a carrier sentence ('Say "/CVn/" to me') in a randomized order. Since VOT can vary depending on speaking rate, speakers read one sentence every two seconds to control to some extent for the effect of speaking rate (e.g., Miller et al. 1986; Pind 1995; Kessinger and Blumstein 1997, 1998).

Recordings were made using a Marantz PMD661 recorder and an Audio-Technica ATM75 Cardioid Condenser Headworn microphone in the same soundproof room in which the perception experiment was conducted (section 2.1). The microphone was placed approximately 10 cm from the speakers' mouth. A computer program automatically changed the screen to display one sentence every two seconds. Each participant initially read the sentences to practice, and subsequently, read them three times for the recording. Participants' first two utterances were analyzed, and their third utterances were analyzed when errors in VOT or F0, such as production errors or pitch tracking errors, were found. The recordings were digitized at a sampling rate of 44,100 Hz and stored as WAV files. Data was then analyzed using the speech analysis program Praat (Boersma and Weenink 2016).

VOTs were measured manually on the waveforms and wideband spectrograms, and confirmed aurally, from the release burst of the stop to the beginning of the first periodic cycle. All voiced stops were produced with lag VOTs except in one case where a male speaker produced /g/ in "gan" with prevoicing. In this case, the third utterance was analyzed. F0s were measured at two points. Onset F0s were measured 10 ms after the voice onset, and mid F0s at the midpoint of the vowel duration following the VOT. As an index of F0 contour shapes of either rising or falling, differences between the onset F0s and the mid F0s were identified. Positive values indicate falling F0 shapes, and negative values indicate rising F0 shapes. Larger absolute differences mean that the degrees of F0 changes are larger (section 3.2.2). The utterance time was limited to two seconds per sentence to control for the potential effect of speaking rate, but this would not completely exclude gender differences in speaking rate, and therefore, word durations were compared between genders. These were measured from the stop release of the target words to the beginning of the release of /t/ of the following word in the carrier sentence.

3.2 Results

3.2.1 Effects of speaker gender on VOT in English stops

Figure 5 exhibits the mean VOT values and standard deviations of voiced and voiceless stops as a function of speaker gender. The mean VOT values of the voiced stops were 24.7 ms (SD 10.0 ms) for male speakers and 22.7 ms (SD 13.0 ms) for female speakers. The mean VOT values of the voiceless stops were 86.6 ms (SD 17.9 ms) for males and 83.0 ms (SD 17.2 ms) for females. For both voiced and voiceless stops, female speakers showed smaller average VOT values than male speakers. Notably, the finding of a shorter mean VOT for the voiceless stops of female speakers contradicts the results of previous studies in which female speakers consistently showed longer mean VOT values than male speakers (section 1). The VOT distinctions between voiced and voiceless stops quantified by taking the differences between the mean VOTs of the two stops were 61.9 ms for males and 60.2 ms for females.

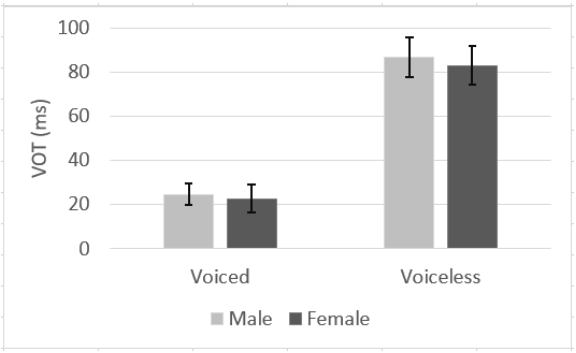

Figure 5. Means and standard deviations of VOT in ms by stop voicing and speaker gender.

Repeated-measures ANOVAs for VOT, VOT distinctions (i.e., VOT differences between voiced and voiceless stops), and word duration were conducted to investigate the effects of speaker gender on these features. Speaker Gender served as the between-subjects factor, and Stop Place and Vowel Context served as the within-subjects factors. An alpha level of 0.05 was specified to determine

statistical significance. Speaker gender differences in VOT values were not statistically significant for either voiced ($F(1, 42) = 1.515$, $p = 0.225$) or voiceless ($F(1, 42) = 0.978$, $p = 0.328$) stops. The gender difference in VOT distinctions between the two stop categories was not statistically significant, either ($F(1, 42) = 0.245$, $p = 0.623$).

The mean word durations in the context of the voiced stops were 318.5 ms (SD 67.4 ms) for males and 322.0 ms (SD 76.7 ms) for females. The mean word durations in the context of the voiceless stops were 358.5 ms (SD 70.0 ms) for males and 359.8 ms (SD 83.7 ms) for females. Gender differences in the word durations were not statistically significant for voiced ($F(1, 42) = 0.031$, $p = 0.862$) or voiceless ($F(1, 42) = 0.003$, $p = 0.955$) stops, indicating that the above results regarding the VOT values as a function of speaker gender were not related to speaking rate.

3.2.2 Effects of speaker gender on F0 contrasts between English stops

This section reports results regarding F0 values and F0 distinctions (i.e., F0 contrasts) of voiced and voiceless stops. Table 5 reports the F0 values in Hz and in semitone (St) for a normalization across speaker genders (Whalen and Levitt 1995). The conversion of Hz into St was made using the formula "12[ln(Hz/100)/ln2]" with the reference of 100 Hz = 0 St. As Table 5 and Figure 6 show, both males and females demonstrated higher F0 values for voiceless stops than voiced stops at both vowel onset and midpoint. Paired *t*-tests indicated that these differences were significant. At the vowel onset, the mean F0 values of male speakers were 112.3 Hz (1.82 St) for the voiceless stops and 102.8 Hz (0.32 St) for the voiced stops ($t(131) = -11.601$, $p < 0.001$), and those of female speakers were 216.3 Hz (13.27 St) for the voiceless stops and 200.2 Hz (11.91 St) for the voiced stops ($t(131) = -12.390$, $p < 0.001$). At the vowel midpoint, the mean F0 values of male speakers were 104.2 Hz (0.54 St) for the voiceless stops and 100.1 Hz (−0.16 St) for the voiced stops ($t(131) = -6.948$, $p < 0.001$), and those of female speakers were 197.8 Hz (11.73 St) for the voiceless stops and 191.5 Hz (11.14 St) for the voiced stops ($t(131) = -5.256$, $p < 0.001$). These results indicate that both male and female speakers produced significantly different F0 values for voiced and voiceless stops.

The F0 contrasts between voiced and voiceless stops calculated by taking the differences between the mean F0 values of the two stops were larger for males than females in St (1.50 St *vs.* 1.36 St at the onset; 0.70 St *vs.* 0.59 St at the midpoint). Repeated-measures ANOVAs for the F0 contrasts were conducted to investigate the effects of speaker gender, using the F0 differences between voiced and voiceless stops as input data. Speaker Gender served as the between-subjects factor, and Stop Place and Vowel Context served as the within-subjects factors. The gender difference in the F0 contrasts in St was not statistically significant for the vowel onset ($F(1, 42) = 0.357$, $p = 0.553$) or the vowel midpoint ($F(1, 42) = 0.346$, $p = 0.560$).[2]

Table 5. Means, standard deviations, and ranges of onset and mid F0s in Hz (upper) and St (lower) by stop voicing and speaker gender.

		Male			Female		
		Voiced	Voiceless	Diff.	Voiced	Voiceless	Diff.
Onset Hz	Mean	102.8	112.3	9.5	200.2	216.3	16.1
	SD	14.3	17.1		22.4	21.1	
	Range	67~142	68~170		157~252	173~261	
Mid Hz	Mean	100.1	104.2	4.1	191.5	197.8	6.3
	SD	14.1	15.4		20.7	18.5	
	Range	65~141	73~147		153~230	160~232	
Onset St	Mean	0.32	1.82	1.50	11.91	13.27	1.36
	SD	2.39	2.53		1.96	1.72	
	Range	−6.93~ 6.07	−6.68~ 9.19		7.81~ 16.00	9.49~ 16.61	
Mid St	Mean	−0.16	0.54	0.70	11.14	11.73	0.59
	SD	2.43	2.46		1.89	1.63	
	Range	−7.46~ 5.95	−5.45~ 6.67		7.36~ 14.42	8.14~ 14.57	

2 Excluding the two speakers (one male and one female) who did not participate in the perception experiment (see sections 2.1 and 3.1), all statistical results on production remained the same. These statistical results need to be confirmed when production and perception results are compared. These results include gender differences in the VOT values for the voiced ($F(1, 38) = 2.581$, $p = 0.116$) and voiceless ($F(1, 38) = 2.932$, $p = 0.095$) stops; in the VOT distinctions ($F(1, 38) = 1.055$, $p = 0.311$); in the word duration for voiced ($F(1, 38) = 0.001$, $p = 0.979$) and voiceless ($F(1, 38) = 0.034$, $p = 0.856$) stops; and in the F0 contrasts in St at the onset ($F(1, 38) = 0.278$, $p = 0.601$) and at the midpoint ($F(1, 38) = 0.585$, $p = 0.449$).

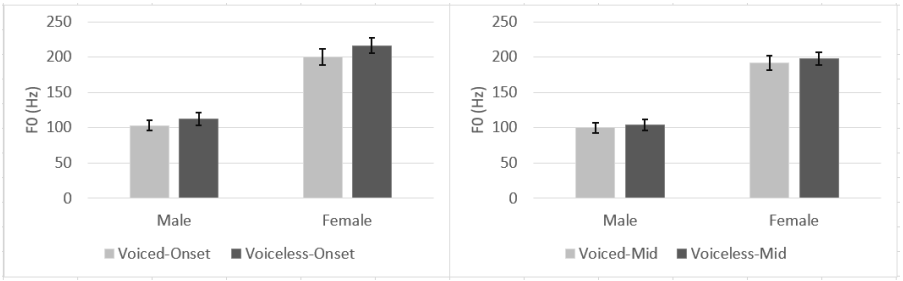

Figure 6. Means and standard deviations of onset (left) and mid (right) F0s in Hz by stop voicing and speaker gender.

As pointed out in section 1, researchers have previously assumed that the F0s exhibit flat or rising contours into the following vowel after voiced stops and falling contours after voiceless stops (Whalen et al. 1993; Shultz et al. 2012). However, the analyses of the production data in this study indicated that both voiced and voiceless stops tended to display falling contours. First, for the voiced stops, the mean F0s were 102.8 Hz (0.32 St) at the onset and 100.1 Hz (−0.16 St) at the midpoint for male speakers, and 200.2 Hz (11.91 St) at the onset and 191.5 Hz (11.14 St) at the midpoint for female speakers, indicating that the F0s tended to fall. As for the voiceless stops, the mean F0s were 112.3 Hz (1.82 St) at the onset and 104.2 Hz (0.54 St) at the midpoint for male speakers, and 216.3 Hz (13.27 St) at the onset and 197.8 Hz (11.73 St) at the midpoint for female speakers, likewise indicating that the F0s tended to fall. The degrees of the F0 descent calculated by taking the differences between the mean F0 values at the onset and at the midpoint were larger for the voiceless stops than the voiced stops for both male (8.1 Hz *vs.* 2.7 Hz) and female (18.5 Hz *vs.* 8.7 Hz) speakers, indicating that the F0 fell more steeply for the voiceless stops than for the voiced stops.

Figure 7 displays scatterplots of all data points for the onset F0s (x-axis) and mid F0s (y-axis). The figures on the left show the voiced stops, and the figures on the right show the voiceless stops. The upper figures present the data for male speakers and the lower figures present the data for female speakers. Based on the dotted lines representing "y = x," data points on the upper left side indicate rising F0 contour shapes in which an onset F0 corresponds to a mid F0 with a higher value. Meanwhile, data points on the lower right side indicate falling F0 contour shapes in which an onset F0 corresponds to a mid F0 with a

lower value. Figure 7 depicts the linear regression fits of the mid F0s in relation to the onset F0s and the linear equations of the fitted lines.

In Figure 7, the slope values of the fitted lines for both voiced and voiceless stops are less than 1, and smaller slope values indicate larger degrees of F0 descent. The slope values of the male speakers were 0.9358 for voiced stops and 0.8407 for voiceless stops, and those of the female speakers were 0.8557 for the voiced stops and 0.7162 for the voiceless stops. These results indicate that F0 contours generally fall for both stop categories. Slope values were smaller for voiceless stops than voiced stops for both genders, indicating that the F0 contours fall more sharply for voiceless stops than voiced stops. These facts can also be illustrated by the higher number of data points on the upper left side based on the "y = x" lines for voiced stops compared to voiceless stops. Regarding the function of speaker gender, female speakers showed sharper degrees of F0 descent than male speakers for both voiced (slopes 0.8557 for females *vs.* 0.9358 for males) and voiceless (slopes 0.7162 for females *vs.* 0.8407 for males) stops.

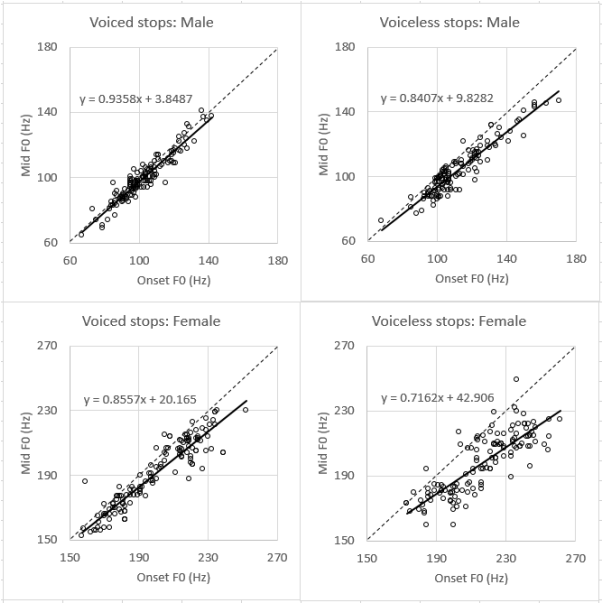

Figure 7. Scatterplots of onset (x-axis) and mid (y-axis) F0s in Hz by stop voicing and speaker gender. Dotted lines indicate "y = x," and solid lines indicate fitted lines.

4. Summary and discussion

This section summarizes the experimental results of this study. The results of the perception experiment showed that both VOT and F0 were significant predictors in the categorization of voiced and voiceless stops, but listeners relied more heavily on VOT than F0. The binary logistic regression analyses for individual listeners indicated that male and female listeners did not differ in their use of VOT and F0 cues in distinguishing between voiced and voiceless stops. Stimulus gender had a significant effect that listeners chose more voiceless stops when listening to female stimuli than when listening to male stimuli. Presumably, the relatively higher F0s in female stimuli generated more voiceless responses to female stimuli than to male stimuli. The production experiment showed that female speakers had a smaller average VOT value for voiceless stops than male speakers, contradicting the findings of previous studies that have consistently shown larger average VOT values for female speakers than male speakers. VOT values and VOT/F0 distinctions between voiced and voiceless stops were not significantly different between the two genders.

As discussed in section 1, several studies have reported that the F0 cue is used in the identification of stop voicing in English only when the VOT is ambiguous (e.g., Whalen et al. 1990). However, Whalen et al. (1993) reported that inappropriate F0s delayed response times even when the VOT was unambiguous, and "[t]his was true both of F0 contours and level F0 differences (p. 2152)." The researchers interpreted this result as evidence that all perceptual cues provided to listeners are fully used even when the cues are not primary. The results of both the perception and production experiments in the present study support this interpretation. In the perception study, the F0 cue was a significant predictor in the identification of stop voicing, and this result supports the finding of Whalen et al. (1993) in that the F0 cue contributes to the perception of stop voicing in English. In the production study, the F0 values differed significantly between voiced and voiceless stops for both male and female speakers.

As for the F0 contour shapes as a function of stop voicing, previous studies stated that voiced and voiceless stops show different F0 contour shapes as rising and falling, respectively, and reflected this difference in the F0 contours in their manipulation of perception stimuli (e.g., Whalen et al. 1993; Shultz et al. 2012;

see section 1). However, the results of this study indicate that F0 contour shapes do not show contrastive patterns for stop voicing in English. Like voiceless stops, voiced stops tended to show falling F0 contour shapes; that the F0 contour shapes are the perceptual cue for stop voicing, as previous studies have assumed, thus appears less likely. The slopes of F0 falling and the consistency of the falling shapes were higher for the voiceless stops than the voiced stops. That is, the F0 contours tended to fall more sharply following voiceless stops than voiced stops. This result may indicate that, in contrast to the assumptions of previous studies, native English listeners do not use falling or rising F0 patterns, but use a slope of F0 falling, in identifying stop voicing.

The perception stimuli manipulation methods used in the present study may have impacted the results via voiceless bias, since potential perceptual cues for voiceless stops other than VOT and F0 such as the closure duration, the amplitude of aspiration noise, the burst amplitude, the onset frequency of F1, and the vowel duration remain cues for voiceless stops in the manipulated stimuli (e.g., Repp 1979). However, manipulating other voicing cues together with VOT and F0 would have made it difficult to determine which factors affected listener responses other than VOT and F0 and produce potential confusion in interpretation resulting from mismatches among different cues for voiced and voiceless stops. Moreover, both genders were in the same situation regarding the potential effects of voiceless bias in the present study. It will be interesting to see in future studies if the perception results remain constant when the base tokens are selected from a voiced stop.

Regarding the effects of speaker gender on VOT in English voiceless stops, previous studies have consistently reported that female speakers produced longer mean VOT values than male speakers, as explained in section 1. Regarding statistical significance, the results of previous studies varied. Some studies reported that the difference between genders was statistically significant (e.g., Swartz 1992; Ryalls et al. 1997), while other studies reported no significant difference between genders in the VOT of voiceless stops (e.g., Sweeting and Baken 1982; Morris et al. 2008). However, no studies have reported the reverse pattern in which males display longer mean VOT values than females for voiceless stops in English. As pointed out in section 1, researchers have interpreted these differences as resulting from physiological differences between

genders (e.g., oral cavity size differences producing differences in the formation of transglottal pressure). However, the results of this study did not support previous findings in this regard; in fact, female speakers showed a shorter average VOT value for voiceless stops than male speakers. This indicates that gender-based physiological factors may not be relevant to the production of VOT in English stops.

Morris et al. (2008) investigated the effects of speaker gender on VOT with 40 male and 40 female native speakers of American English. The main result of the study was that while females produced a longer mean VOT for voiceless stops than males, the gender effect was not statistically significant. The gender effect on VOT of voiced stops was not significant, either. They stated that the gender difference in VOT concerning English stops was not a factor and, therefore, does not need to be controlled in the context of isolated syllables. The present study showed no significant gender difference in the VOT of English stops produced within a carrier sentence. These findings may indicate that there is also no need to control the gender factor when the stops are produced within a carrier sentence.

As described in section 1, several cross-language studies have investigated the effects of speaker gender on VOT values, and their various findings have indicated that gender-based effects on VOT may vary from language to language. Although the present study supports the notion that the gender-based effects on VOT do not exist in the case of English stops, this does not exclude the possibility that significant gender differences in VOT exist in other languages. Considering the fact that the results of previous cross-linguistic studies varied and the numbers of participants were relatively small, researchers need to undertake cross-linguistic investigations into the effects of gender on VOT and attempt to explain the phenomena with larger numbers of participants in subsequent studies.

References

Bates, Douglas, Martin Maechler, Ben Bolker, and Steven Walker. 2015. *lme4: Linear mixed-effects models using Eigen and S4*. R package version 1.1-9.

Boersma, Paul and David Weenink. 2016. Praat: Doing phonetics by computer. Software retrieved from <http://www.fon.hum.uva.nl/praat/>.

Helgason, Pétur and Catherine Ringen. 2008. Voicing and aspiration in Swedish stops. *Journal of Phonetics* 36: 607-628.

Kessinger, Rachel H. and Sheila E. Blumstein. 1997. Effects of speaking rate on voice-onset time in Thai, French, and English. *Journal of Phonetics* 25: 143-168.

Kessinger, Rachel H. and Sheila E. Blumstein. 1998. Effects of speaking rate on voice-onset time and vowel production: Some implications for perception studies. *Journal of Phonetics* 26: 117-128.

Koenig, Laura L. 2000. Laryngeal factors in voiceless consonant production in men, women, and 5-year-olds. *Journal of Speech, Language, and Hearing Research* 43: 1211-1228.

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. *Journal of the Korean Society of Speech Sciences* 5: 81-90.

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

Miller, Joanne L., Kerry P. Green, and Adam Reeves. 1986. Speaking rate and segments: A look at the relation between speech production and speech perception for the voicing contrast. *Phonetica* 43: 106-115.

Misnadin, Misnadin, James P. Kirby, and Bert Remijsen. 2015. Temporal and spectral properties of Madurese stops. In *Proceedings of the 18th International Congress of Phonetic Sciences*, Glasgow, United Kingdom.

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

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

Pind, Jörgen. 1995. Speaking rate, voice-onset time, and quantity: The search for higher-order invariants for two Icelandic speech cues. *Perception and Psychophysics* 57: 291-304.

R Development Core Team. 2016. *The R project for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Repp, Bruno H. 1979. Relative amplitude of aspiration noise as a voicing cue for syllable-initial stop consonants. *Language and Speech* 22: 173-189.

Ryalls, John, Allison Zipprer, and Penelope Baldauff. 1997. A preliminary investigation of the effects of gender and race on voice onset time. *Journal of Speech, Language, and Hearing Research* 40: 642-645.

Shultz, Amanda A., Alexander L. Francis, and Fernando Llanos. 2012. Differential cue weighting in perception and production of consonant voicing. *Journal of the Acoustical*

Society of America 132: EL95-EL101.
Smith, B. 1978. Effects of place of articulation and vowel environment on voice stop consonant production. *Glossa* 12: 163-175.
Swartz, Bradford L. 1992. Gender difference in voice onset time. *Perceptual and Motor Skills* 75: 983-992.
Sweeting, Patricia M. and Ronald J. Baken. 1982. Voice onset time in a normal-aged population. *Journal of Speech and Hearing Research* 25: 129-134.
Sokolović-Perović, Mirjana. 2012. The voicing contrast in Serbian stops. PhD Dissertation, Newcastle University.
Whalen, Douglas H., Arthur S. Abramson, Leigh Lisker, and Maria Mody. 1990. Gradient effects of fundamental frequency on stop consonant voicing judgments. *Phonetica* 47: 36-49.
Whalen, Douglas H., Arthur S. Abramson, Leigh Lisker, and Maria Mody. 1993. F0 gives voicing information even with unambiguous voice onset times. *Journal of the Acoustical Society of America* 47: 36-49.
Whalen, Douglas H. and Andrea G. Levitt. 1995. The universality of intrinsic F_0 of vowels. *Journal of Phonetics* 23: 349-366.
Whiteside, Sandra P. and Caroline J. Irving. 1997. Speakers' sex differences in voice onset time: Some preliminary findings. *Perceptual and Motor Skills* 85: 459-463.
Whiteside, Sandra P. and Caroline J. Irving. 1998. Speakers' sex differences in voice onset time: A study of isolated word production. *Perceptual and Motor Skills* 86: 651-654.
Whiteside, Sandra P., Luisa Henry, and Rachel Dobbin. 2004. Sex differences in voice onset time: A developmental study of phonetic context effects in British English. *Journal of the Acoustical Society of America* 116: 1179-1183.

Eunjin Oh
Professor
Dept. of English Language and Literature
Ewha Womans University
52 Ewhayeodae-gil, Seodaemun-gu
Seoul 03760, Republic of Korea
E-mail: ejoh@ewha.ac.kr

Received: 2018. 07. 26.
Revised: 2019. 02. 08.
Accepted: 2019. 02. 22.