Journal of the Korean Society of Speech Sciences Vol.6 No.4 (2014.12.31) www.speechsciences.or.kr
ISSN 2005-8063, pp. 47~52 http://dx.doi.org/10.13064/KSSS.2014.6.4.047

Effects of attention on the perception of L2 phonetic contrast

Lee, Hyunjung[1)]

ABSTRACT

This study investigated how the degree of attention modulates English learners' perception of Korean stop contrasts. The contributions of VOT and F0 in perceiving Korean stops were examined while availability of attentional resources was manipulated using a dual-task paradigm. Results demonstrated the attentional modulation in the use of VOT, but not in F0: under less attention, the contribution of VOT to the perception of aspirated stops decreased, whereas that of lenis stops increased, which suggests more native-like performance. This implies that the role of attention in perceiving non-native contrasts might differ depending on how equivalent the acoustic and perceptual cues are between L1 and target L2 contrasts.

Keywords: L2 speech perception, Korean stops, attention, working memory

1. Introduction

The coding of acoustic information into linguistic categories during speech perception varies depending on numerous factors. Using a dual-task paradigm (i.e., speech recognition while performing another task), some studies demonstrated that the amount of attention given to speech sounds is also a source of the variation (Gordon et al., 1993; Matty and Wiget, 2011). In the dual-task paradigm, the distracting situation to the primary speech recognition task is intentionally given by a secondary task (e.g., letter recall or arithmetic) that increases working memory demands. Working memory (WM) is defined as the information process system used in holding multiple information including visual images or verbal information. The importance of working memory for speech processing has been demonstrated in Francis and Nusbaum (2009) by showing that increasing working memory load slowed down spoken word recognition. Accordingly, the dual-task paradigm with a distracting situation for speech processing provides the condition increasing the WM load. This experimental paradigm is used to examine effect of attention on the relative contribution of multiple acoustic cues (Gordon et al., 1993) or lexical and sub-lexical information to speech recognition (Matty and Wiget, 2011). For example, Gordon et al. (1993) showed that in the perception of the voicing distinction and vowel contrasts in English, as the use of strong cues such as VOT and formant frequency decreases, weak cues such as F0 and vowel duration are given more weight; crucially, this trading relationship is stronger under distracting listening conditions, compared to non-distracting conditions. Through a simulated model treating the multiple acoustic cues and attention independently, Gordon et al. (1993) suggested that strong cues need careful attention, and argued that the distracting condition reduces competitions from primary cues (VOT, formant frequency), which allows listeners to rely more on secondary cues (F0, vowel duration). Mattys and Wiget (2011) tested how different attentional demands affect a Ganong effect. The Ganong effect is a lexical bias on phoneme identification; for example, an ambiguous sound between /k/ and /g/ is more likely to be perceived as /k/ when it is embedded in /_iss/, given that *kiss* is a word, but *giss* is not. Mattys and Wiget (2011) showed an increased Ganong effect under distracting conditions compared to non-distracting. That is, the distracting condition or the WM overload led listeners to rely more on lexical information compared to an optimal listening condition, and presumably the WM overload prevented listeners from accessing to acoustic

1) Hankyong National University, hyunjunglee123@gmail.com

Received: November 3, 2014
Revised: December 9, 2014
Accepted: December 14, 2014

details immediately. The effect of attention is also observed in the non-native speech perception. Asano (2014) noted that the incomplete knowledge in L2 relative to L1 may make L2 speech processing even more vulnerable under such distracting conditions. She found that German listeners' performance in discriminating Japanese consonantal length contrasts is affected by increasing task demands manipulated by increasing Inter-Stimulus-Intervals (ISI) and by adding another acoustic dimension (i.e., pitch) to sound stimuli in addition to the length component. Asano (2014) showed that the increased task demands by the additional pitch component reduced sensitivity to the L2 phonetic contrast. Asano (2014) concluded that L2 phonological contrasts can be established only under optimal listening conditions.

The purpose of this study is to examine if L2 speech perception is modulated by availability of attentional resources, particularly testing the effect of attention on the relative weights of multiple phonetic cues in L2 contrasts. This paper used a dual-task paradigm in which participants were asked to recognize non-native contrasts while memorizing a series of letters. Accordingly, the current experimental paradigm might closely mimic the ordinary listening situation in which people may devote less attention to listening to sounds than they typically do in a laboratory setting.

This paper examined English-speaking learners' perception of the three-way laryngeal contrast among the voiceless stops in Korean, testing how the degree of attention allocated to non-native contrasts affects the relative contributions of VOT and F0, which are also used for their L1 English voicing contrast, but with different cue weightings. The current investigation will help clarify the relationship between attention and cue weightings in L2 as well as L2 speech perception in everyday life.

It is well known that in English, VOT alone reliably distinguishes voiced stops from voiceless in the syllable-initial position, though VOT correlates with F0 at the onset of the following vowel. Voiceless stops in English have significantly longer VOT and slightly higher F0 than voiced ones (e.g., Lisker and Abramson, 1964; Hombert, 1978). Korean has the three-way distinction among voiceless stops (i.e., aspirated-/ph/, lenis-/p/ and fortis-/p'/) word-initially. Whereas English primarily uses VOT for the voicing distinction, in Korean F0 is as important as VOT for distinguishing the three voiceless stops (e.g., Cho et al., 2002; Francis and Nusbaum, 2002; Lee and Jongman, 2012). Korean fortis stops are produced with shorter VOT than lenis or aspirated stops, whereas VOT between lenis and aspirated stops substantially overlaps. Lenis and aspirated stops are instead distinguished mainly by F0: F0 is low at the onset of a vowel following a lenis stop, and high following an aspirated stop. In perception studies (Kim, 2004; Lee et al., 2013), fortis stops are perceived mainly relying on very short VOTs (e.g., 10–20ms), whereas lenis and aspirated stops are perceived showing a phonetic trading relation between VOT and F0. That is, on the relatively long VOT range (e.g., 30–100ms), F0 at the category boundary becomes lower as VOT becomes longer. Regarding cross-linguistic acoustic similarities between Korean and English, Kang and Guion (2006) reported that Korean fortis stops had an average VOT similar to that of English voiced stops (11ms and 0.98ms, respectively). On the other hand, Korean aspirated and lenis stops (68ms and 63ms, respectively) both were comparable to the English voiceless category (72ms). Given that L2 learners of Korean primarily rely on VOT in acquiring the Korean stop contrast (Chang et al., 2011), it seems reasonable to expect that English learners of Korean can identify fortis from lenis and aspirated stops better than they distinguish the contrast between lenis and aspirated stops. To attain native-like performance, therefore, English-speaking learners of Korean would need to become sensitive to F0, especially for the lenis-aspirated contrast. In fact, there is evidence that F0 may influence English-speaking learners' perception of Korean aspirated and lenis stops. Schmidt (2007) showed that Korean fortis stops were categorized as English voiced stops, as expected; notably, while both Korean lenis and aspirated stops were classified as English voiceless stops by English learners of Korean, the learners tended to perceive the aspirated stops, rather than lenis stops, as being closer to English voiceless stops. That is, the higher F0 in Korean aspirated and English voiceless stops than Korean lenis and English voiced stops might play a role in the identification despite the similar VOT range between the two categories. Overall, the previous cross-linguistic studies indicate that while VOT is crucial for English learners' perception of the Korean stops, F0 might also play a role when it comes to the choice between lenis and aspirated stops.

Based on the cross-linguistic acoustic and perceptual characteristics and the findings in Gordon et al. (1993), this study tested the following hypotheses. First, given the similar properties of VOT and F0 in English voiceless and Korean aspirated stops (long VOT and high F0), English-speaking learners of Korean would show a similar effect of attention as has been shown in English (i.e., under processing load conditions that allow less attention, as the salience of VOT as a cue

decreases then the contribution of F0 will increase). Second, for lenis stops, whose VOT and F0 properties are conflict insofar as analogy to English stops is concerned (i.e., similar to English voiceless stops in VOT, but to voiced stops in F0), two hypotheses would be possible; (1) if the learners rely more on F0 than VOT, the effect of attention would be opposite from the aspirated percept as they would treat VOT as a secondary cue, and (2) if the learners treat VOT as a primary cue as in L1, a similar attentional effect to aspirated or L1 voicing percept is expected. Finally, for fortis stops English learners of Korean would show little to no effect of attention, given its reliable categorization to English voiced stops and little effect of F0 in both L1 (Kim, 2004; Lee et al., 2013) and L2 (Schmidt, 2007). Overall, these predictions imply that English learners of Korean would perform more like native Korean listeners when they are less attentive to sounds; the predictions might be contradictory to Asano (2014) who argued that native-like good performance in an L2 phoneme identification can be achieved only under the optimal listening environment with no distracting condition to speech recognition.

To manipulate the attentional condition, participants were asked to perform a 3-alternative forced choice task (3AFC) while sometimes also performing a letter recall task; trials with and without the letter recall task provide the high and low working memory demands (High-WM vs. No-WM), respectively. The low task demands of the No-WM condition allows participants to focus only on speech sounds given no need to hold the letter information during the identification.

2. Methods

2.1 Participants

Fourteen native English-speaking beginning L2 learners of Korean (6 males, 8 females) participated in the experiment. Their age range from 18-23, with a mean of 19.5 (SD = 1.31). They were all native speakers of American English and had taken an eight week of Introductory Korean course at the University of Chicago at the time of the study participation. The participants had no experience with Korean through formal education or at home with family, though some of them had been exposed to the Korean language through the media. This study was approved by the IRB committee of the University of Chicago. All subjects gave informed consent and were paid for their participation.

2.2 Stimuli

The stimuli were synthetic CV syllables systematically varying both VOT and F0. A male native speaker of Seoul Korean produced one base token of /pa/ of which VOT and F0 at the onset of the following vowel was 57ms and 90Hz, respectively. First, a 5-step VOT continuum was created fixing the range from 10 to 98ms in 22ms steps (10, 32, 54, 76 and 98ms). VOT was manipulated by compressing and expanding the VOT portion of the base token using the Praat manipulation function. Then, at each of five VOT steps, the onset F0 was lowered to 85Hz and heightened to 160Hz in 15Hz steps, yielding a 6-step F0 continuum (85, 100, 115, 130, 145 and 160Hz). The endpoints of the VOT and F0 continua fall into one of the three stop categories in Korean according to a previous production studies (Kang and Guion, 2006; Lee and Jongman, 2012). A total of 30 different stimuli varying VOT and F0 dimensions were created for the 3AFC identification task.

2.3 Design and procedure

Each participant responded a total of 240 trials (30 stimuli x 2 WM conditions x 4 repetitions) over E-Prime in a sound-proof booth. The participants completed half of the trials with the letter recall task (High-WM) along with the identification task and the other half without the recall task (No-WM). The 30 sound stimuli were randomly presented within each of eight experimental blocks, divided into the two WM demands (four blocks for each WM condition). The presentation of the High-WM and No-WM blocks was alternated with each other, starting with the High-WM block for all participants. On each trial in the four High-WM blocks, the participants were first asked to memorize five random alphabet letters presented one at a time on a computer monitor for 1000ms. After 300ms delay, the 3AFC identification task followed; the participants gave their response to a sound stimulus by pressing one of the three response keys written as ‘pha’, ‘pa’, and ‘ppa’ within 10 seconds. After responding to the sound stimulus, a screen showed up asking the participants to type the memorized letters in the order using a keyboard; pressing the ‘enter’ key ended the trial, triggering the next trial, and therefore there was no time pressure for the letter recall part. The other four No-WM blocks were conducted without the letter recall distractor. The entire experiment including a prior short practice took approximately 30 min.

3. Results

Overall, the accuracy for the letter recall task itself (High-WM only) ranged from 58% to 94% across participants with an average of 78%. Given the fact that the inaccurate letter recall does not mean 'high attention' to sound stimuli, all the data were included in the analysis regardless of the correct or incorrect letter recall response.

Identification responses were modeled using a binomial mixed-effects logistic regression model with three predictors, VOT, F0, and Working Memory (High-WM, No-WM); the model for lenis /p/ also included VOT^2 as a term given that the lenis response proportion by VOT is not monotonic. The continuous variables, VOT and F0, were centered. Three models were obtained for each of the three dependent variables, aspirated, lenis and fortis responses (e.g., 1=aspirated, 0=non-aspirated; 1=lenis, 0=non-lenis, etc.). The model was fitted in *R*, using the *lmer()* function from the *lme4* package (Bates and Bolker, 2011). Each model maximally included the three predictors and all possible interactions with a by-subject random intercept, and the final model was obtained by forward selection. <Table 1> summarizes the parameter estimate for each of the fixed effects in the selected models for aspirated (/p^h/), lenis (/p/) and fortis (/p'/).

For aspirated (/p^h/) responses, main effects of VOT, F0 and WM were significant; /p^h/ response increased with increasing VOT and F0, and it was greater in No-WM than High-WM. Crucially, the interaction between VOT and WM was significant, indicating that the effect of VOT in the /p^h/ response is modulated by the WM conditions; in <Figure 1> (top-left), the more gradual increase in /p^h/ responses in High-WM than No-WM suggests that the contribution of VOT in predicting /p^h/ decreases when there is less attention to sounds, consistent with the prediction. Contrary to predictions, however, the absence of interactions including F0 and WM indicates that the degree of attention does not significantly affect the use of F0 in predicting /p^h/.

For lenis (/p/) response, significant main effects were found for VOT, F0, and VOT^2; /p/ responses decreased with increasing VOT and F0, and the effect of VOT is not monotonic. The significant interaction between VOT and F0 indicates that the F0 effect is boosted as VOT increases and vice versa; in <Figure 1> (bottom-middle), for example, as F0 increases, English learners' perception of /p/ decreases more for long VOT stimuli than short VOT ones.

Importantly, the effect of VOT in /p/ response is also modulated by the WM conditions; in <Figure 1> (top- middle), even more decrease of /p/ response in No-WM than High-WM at low F0 (85Hz) suggests that the learners rely more on VOT change for /p/ response under more attentional condition to sounds. Notably, the relationship between VOT and WM in lenis responses is opposite from that in aspirated responses, which is in line with the first prediction for lenis. But, the absence of interactions including F0 and WM indicates that the degree of attention does not significantly affect the use of F0 in predicting /p/.

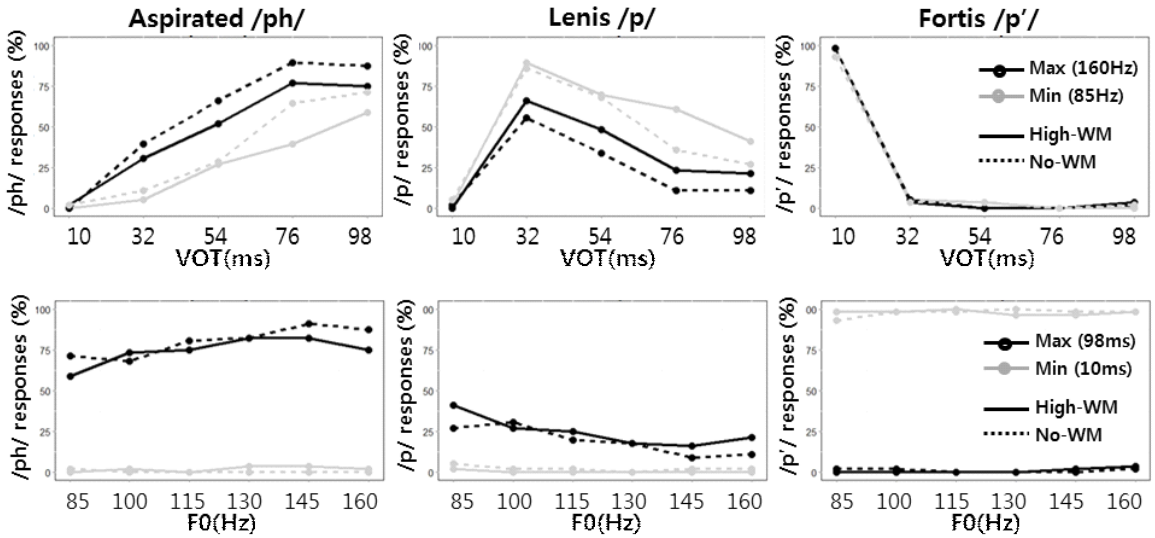

Figure 1. Mean percentage of each of /ph/, /p/ and /p'/ responses as a function of VOT (top) and F0 (bottom) in High-WM (solid line) and No-WM (dotted line) by the minimum (light color) and maximum (dark color) F0 and VOT values.

Table 1: Estimates for predictors in mixed-effects models in predicting listeners' aspirated (/p^h/), lenis(/p/) and fortis (/p'/) responses. (*p < 0.05; **p < 0.01; ***p < 0.001).

	Aspirated /ph/		
	Coef. β	**SE(β)**	***z***
Intercept	-0.14	0.27	-0.51
VOT	2.55	0.27	9.57***
F0	0.45	0.15	3.05**
WM$_{high}$	-0.68	0.33	-0.07*
VOT:WM$_{high}$	-0.67	0.18	-3.62***
VOT:F0			
VOT^2			
F0:WM$_{high}$			
VOT:F0:WM$_{high}$			
	Lenis /p/		
	Coef. β	**SE(β)**	***z***
Intercept	0.46	0.25	1.82
VOT	-0.68	0.28	-2.39*
F0	-0.57	0.21	-2.79**
WM$_{high}$	0.41	0.32	1.27
VOT:WM$_{high}$	0.62	0.23	2.69**
VOT:F0	-0.22	0.11	-2.04*
VOT^2	-1.79	0.11	-16.92***
F0:WM$_{high}$			
VOT:F0:WM$_{high}$			
	Fortis /p'/		
	Coef. β	**SE(β)**	***z***
Intercept	-9.462	1.52	-6.24***
VOT	-8.90	1.33	-6.70***
F0			
WM$_{high}$			
VOT:WM$_{high}$			
VOT:F0			
VOT^2			
F0:WM$_{high}$			
VOT:F0:WM$_{high}$			

For fortis (/p'/) responses, main effects of VOT and WM were significant; /p'/ are less likely to be heard with increasing VOT, and in No-WM than High-WM. The main effect of F0 and any interactions were not significant. As we predicted, the effect of VOT on /p'/ response is not affected by F0 and WM conditions, and the short VOT alone is a strong predictor for /p'/ response for English learners of Korean.

4. Discussion and Conclusion

The current investigation revealed that the English learners' use of VOT in identifying Korean lenis and aspirated stops is influenced by the degree of attention. Confirming the predictions, the effect of attention on VOT was opposite between aspirated and lenis percept. The parallel effect of attention between L1 and L2 on aspirated percept implies that English learners of Korean establish a merged L1-L2 category of Korean aspirated and English voiceless stops, similar to late Korean learners of English (Kang and Guion, 2006). The fact that the effect of attention on the lenis percept was opposite from that observed for L1 English stops and for L2 Korean aspirated stops suggests that the learners may not rely primarily on VOT in perceiving lenis stops as they do in L1 voicing distinction, and they might become aware of the cue weightings in Korean lenis stops distinct from Korean aspirated or English voiced stops.

The findings in the aspirated and lenis perception suggest that L2 learners might establish knowledge about cue weightings in novel sounds in the eight week period of exposure to the L2.

Finally, regarding the fortis percept, short VOT seems robust enough to trigger fortis responses despite the increasing task demand by the letter recall task.

Regarding F0, however, the statistics did not support the prediction that the degree of attention modulates the contribution of F0 in aspirated and lenis percept. A methodological difference

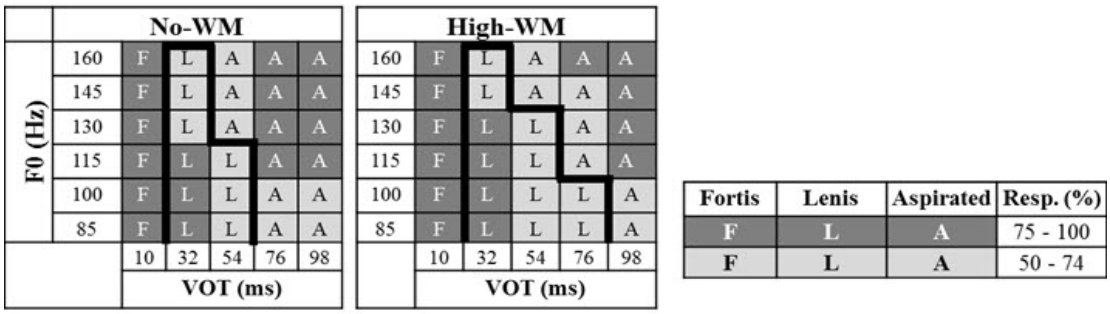

Figure 2. Identification rate (%) between No- and HighWM conditions for aspirated /ph/, lenis /p/ and fortis /p'/ stops as a function of VOT and F0.

between the current study and Gordon et al. (1993) may explain the non-significant effect of attention on F0; Gordon et al. (1993) used a seven-step VOT continuum only at high (180Hz) and low F0 (100Hz), whereas this study used a six-step of F0 continuum. In fact, the greater use of F0 in High-WM was observed numerically; as illustrated in <Figure 2>, low F0s triggering lenis responses play a role across wider VOT ranges under High-WM than No-WM, and the perceptual pattern of High-WM is more similar to that of native Seoul Korean listeners reported in Lee et al. (2013).

Overall, the results demonstrate that L2 speech perception is also modulated by attentional demands. The attentional modulation depending on different cue weightings in L2 phonemic categories suggests that careful attention to non-native phonemic contrasts does not always guarantee native-like performance in L2 perception, and that the role of attention in perceiving non-native contrasts should be discussed considering the mutual equivalence of acoustic and perceptual characteristics between L1 and L2.

Acknowledgments

I would like to thank Alan Yu for insightful comments at all stages of this work. I am also grateful to James Kirby for helping me set-up stimuli and Hi-Sun Kim for recruiting participants.

References

Asano, Y. (2014). "Stability in perceiving non-native segmental length contrasts," in *Proceedings of the 7th International Conference on Speech Prosody*. Dublin, Ireland, 321–325.

Bates, M. M., and Bolker, B. (2011). LME4, R package version 0.999999-2.

Chang, S. E., Burge, M., and Choi, Y. (2011). "A cross-linguistic study of Korean laryngeal stops by the native speakers of Chinese, English, Korean, and Spanish," in *Proceedings of the 17th International Congress of Phonetic Sciences.* Berkeley, USA, 432–435.

Cho, T., Jun, S., and Ladefoged, P. (2002). "Acoustic and aerodynamic correlates of Korean stops and fricatives". *Journal of Phonetics* 30, 193–228.

Francis, A., and Nusbaum, H. C. (2002). Selective attention and the acquisition of new phonetic categories. *Journal Experimental Psychology: Human Perception and Performance* 28, 349–366.

Francis, A., and Nusbaum, H. C. (2009). Effects of intelligibility on working memory demand for speech perception. *Attention, Perception, & Psychophysics*, 71, 1360-1374.

Gordon, P. C., Eberhardt, J. L., and Rueckl, J. G. (1993). "Attentional modulation of the phonetic significance of acoustic cues", *Cognitive Psychology* 25, 1–42

Hombert, J. M. (1978). "Consonant types, vowel quality, and tone," in *Tone: A linguistic survey,* edited by V. A. Fromkin (New York: Academic Press), pp. 77–111.

Kang, K.-H., and Guion, S. (2006). "Phonological systems in bilinguals: Age of learning effects on the stop consonant systems of Korean-English bilinguals," *Journal of the Acoustical Society of America* 119, 1672–1683.

Kim, M. (2004). "Correlation between VOT and F0 in the perception of Korean stops and affricates," in *Proceedings of the 8th international conference on spoken language processing,* Jeju Island, Korea, 49–52.

Lee, H., and Jongman, A. (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, 145–169.

Lee, H., Politzer-Ahles, S., and Jongman, A. (2013). "Speakers of tonal and non-tonal Korean dialects use different cue weightings in the perception of the three-way laryngeal stop contrast," *Journal of Phonetics* 41, 117–132.

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

Mattys, S. L., and Wiget, L. (2011). "Effect of cognitive load on speech recognition," *Journal of Memory and Language* 65, 145–160.

Schmidt, A. M. (2007). "Cross-language consonant identification: English and Korean," in *Language Experience in Second Language Speech Learning: In Honor of Kames Emil Flege,* edited by O. S. Bohn and M. J. Munro (John Benjamins Publishing), pp. 185–200.

• **Lee, Hyunjung**
Department of English, Hankyong National University
327 Jung-ang-ro, Ansung-si, Kyenggi-do,
Tel: 031-670-5310 Fax: 031-670-5319
Email: hyunjunglee123@gmail.com
Areas of interest: phonetics, phonology, psycholinguistics
2014~present: lecturer