



Why is Korean lenis stop difficult to perceive for L2 Korean learners?

Boram Lee, Cécile Fougeron, Naomi Yamaguchi

Laboratoire de Phonétique et Phonologie (UMR7018 CNRS/Univ. Sorbonne-Nouvelle, France)

{lee.boram, cecile.fougeron, naomi.yamaguchi}@sorbonne-nouvelle.fr

Abstract

This study investigates how French learners, whose native language has only a binary laryngeal contrast, acquire the Korean three-way laryngeal contrast in stops by focusing on cue weighting. We tested how 21 French learners of Korean identify fortis/lenis/aspirated Korean stops over eight monthly sessions. Learners were the most successful at identifying aspirated stops. The identification of lenis stops was the most challenging, with no improvement over the 8 months, whereas the perception of aspirated and fortis stops improved. In order to explain the difficulty with lenis stops, we tested the relative contribution of VOT and f_0 to the perception of the contrast on synthesized stimuli. Learners showed different cue-weighting strategies: VOT was used to distinguish between aspirated and fortis/lenis, and f_0 was used to differentiate between fortis and lenis, implying a two-way rather than a three-way contrast. Furthermore, the larger weight given to VOT compared to f_0 by the French learners can explain the poor identification of lenis stops, whose main contrastive cue in Korean is a lowering of f_0 on the following vowel. Based on these findings, the acquisition of the three-way contrast necessitates a reorganization of the cues' relative weight.

Index Terms: Second language acquisition, acoustic cues, longitudinal study, L2 Korean, L1 French

1. Introduction

It has been well documented that in second language (L2) speech perception, L2 learners often struggle with L2 contrasts. For example, the English /ɹ/-/l/ contrast is problematic for Japanese learners (e.g. [16]) and Korean learners have difficulty identifying English tense and lax vowel contrast (e.g. [6]). One particular difficulty for L2 Korean learners is to perceive and produce the Korean stops contrast due to their typological rarity: Korean stops employ a three-way contrast (lenis, fortis, aspirated), whereas the majority of languages employ a two-way contrast. In a perceptual assimilation task conducted by [1] English listeners assimilated the Korean lenis and aspirated stops to the English voiceless stop, and the fortis to the voiced stop. In addition, [13] and [9] demonstrated that the Korean fortis stops were generally perceived as the unaspirated stops of Mandarin, and the Korean aspirated and lenis stops were perceived as Mandarin aspirated stops, while Spanish learners perceived all three Korean stops as the voiceless stop of Spanish. Taken together, these results show different perceptual patterns based on VOT type. According to [11], there exist three different types of VOT: 1) voicing lead (negative VOT value), 2) short voicing lag (voicing begins shortly after the stop release, resulting in voiceless, unaspirated stops), 3) long voicing lag (voicing starts late after the stop release, resulting in voiceless, aspirated stops). Languages with a short-lag vs. long-lag VOT contrast, such as English and Chinese, showed different patterns in perception compared to languages with a lead vs.

short VOT contrast, such as French or Spanish. Spanish learners assimilated all Korean stops to their L1 short-lag category, while English and Chinese learners assimilated the Korean lenis and aspirated stops to their L1 long-lag category and Korean fortis stop to their L1 short-lag category, although their short-lag phonemes are different (/p/ in Chinese, /b/ in English).

Meanwhile, VOT alone is not enough to understand why some categories in Korean stops are more difficult to acquire. [14] showed that Mandarin learners were good at discriminating Korean aspirated and fortis stops, whereas they less accurately identified the lenis stops. A similar finding was observed in Spanish learners who discriminated more accurately the fortis from the lenis and aspirated stops, than they discriminated the lenis from the aspirated stop ([13]). They explained these results by a lack of attention to f_0 cues on the following vowel. Indeed, previous studies on Korean stops have demonstrated that the three-way distinction can be captured both in production (e.g. [12]) and perception (e.g. [10]) using a combination of VOT and f_0 at the onset of the vowel following the stop consonant, suggesting that both play important roles in distinguishing the categories. More specifically, fortis stops have short-lag VOT and high f_0 , lenis have long-lag VOT and low f_0 , and aspirated have long-lag VOT and high f_0 . Thus, VOT contributes to the distinction of the fortis stop (shortest VOT) from both aspirated and lenis stops. In turn, f_0 contributes to the distinction of the lenis stop (lowest f_0) from the aspirated and fortis stops. Consequently, f_0 is a primary cue in Korean for the distinction of lenis stops. L2 Korean learners thus have to perceive and acquire not only VOT but also the f_0 cue for the three-way stops contrast, as well as their relative weight, referred to as 'cue weighting', for each category.

The present study is part of a larger project investigating the acquisition (in production and perception) of this three-way contrast within a longitudinal perspective, through the observation of 21 French students during their first year of Korean L2 studies. Our purpose in this paper is to report on the evolution in the perceptual identification of the three stops and in the use of f_0 and VOT cues by the French learners, who are supposedly conditioned by the use of VOT as a primary and sole cue in their L1 to distinguish voiced (lead VOT) and unvoiced (short VOT) stop consonants.

2. Perception experiment

Longitudinal data were collected every month for a total of 8 sessions in the context of the first year of Korean studies. Each session consisted of a production task (not presented here) and two perception tasks described below ; 1) a three-alternative forced choice identification task with natural stimuli to examine the acquisition of the three Korean stops, and 2) a three-alternative forced choice identification task with synthesized stimuli to determine perceptual cue weights.

2.1. Participants and Stimuli

Twenty-one French female students (mean age: 18.2 years) in their first year of Korean studies, with no prior knowledge of the language, participated in this study in the context of a class ‘Korean Speaking 1&2’ conducted by the first author. Participation was on a voluntarily basis. None of the participants reported any speech or hearing problems or having spent more than one month in Korea. Twenty-five female native Korean speakers (mean age: 29.2 years) participated as a control group. They were all born and raised speak Standard Seoul Korean (from in Seoul or Gyeonggi Province).

The natural stimuli used in the first perception task were produced by a Seoul-born female native Korean speaker (the first author) and consisted of stop-initial syllables /CVpa/, with C: /t/, /tʰ/, /t*/ and V: /a/, /i/, /o/.

The synthesized stimuli used in the second perception task were generated on Praat [15] using the PSOLA algorithm. A token lenis stop /ta/ produced by first author was used as baseline. Following the procedure used by [11], the VOT duration of this token was compressed or expanded from 8 ms to 80 ms in 12 ms steps (8ms, 20ms, 32ms, 44ms, 50ms, 62ms, 80ms). For each VOT step, we manipulated f0 dimension by replacing the original f0 value during the vowel of /a/ with one of the five f0 values: 200Hz, 225Hz, 250Hz, 275Hz, 300Hz. In total, 35 different stimuli (5 VOT steps x 7 f0 steps) were created.

2.2. Task and Procedure

The perception experiments were conducted during the 2020-2021 academic year over eight sessions, separated by 4 weeks. Both perception tasks were accessible online (www.psytoolkit.com) [5]. Participants should perform the first experiment then the second one on the same day. In both experiments, participants listened to the stimuli and had to decide which consonant category was heard. They were asked to respond by typing on a keyboard ‘q’ for lenis, ‘m’ for fortis and ‘g’ for aspirated stops. Three consonants labels (known by the learners) were presented in Korean alphabet on the screen.

For the first task with natural stimuli, the set of 9 targets (3 consonants x 3 vowel contexts) and 18 filler words was repeated three times in a random order, yielding 81 trials (9 targets + 18 fillers x 3 times) per session. For the second task with synthesized stimuli, the set of 35 targets was repeated three times randomly, yielding 105 trials (35 targets x 3 times) per session. The inter-stimulus interval was of 5 seconds. Before each task, 12 trials with non-target stimuli were presented for familiarization. Each session lasted approximately 10 minutes.

3. Results

3.1. Natural stimuli

The accuracy in the identification of the three stops was analyzed with a mixed-effects logistic regression model [17] using the *glmer* function in the *lme4* library [4] for the R statistical package [18]. The model examined the relationship between the binary categorical responses (correct response: 1 vs. incorrect response: 0) and the predictors: stop category (fortis/lenis/aspirated), the session (sess1 to 8); by-participant random intercepts and by-contrast slopes were included in the model.

We found a main effect of stop category, with less hits for the Lenis stops ($\beta=-1.6428$, $SE=0.2126$, $z(4536)=-7.727$,

$p<0.001$) than for the Aspirated stops ($\beta=1.8396$, $SE=0.3134$, $z(4536)=5.870$, $p<0.001$), and the Fortis stops ($\beta=1.204$, $SE=0.2988$, $z(4536)=4.033$, $p<0.001$). We found also a main effect of the session ($\beta=0.5973$, $SE=0.3134$, $z(4536)=5.870$, $p<0.001$). Post-hoc comparisons across sessions, separately for the lenis, aspirated and fortis category, were done with Bonferroni *p*-value adjustments in *eameans* library.

Figure 1 presents the % of correct identification of the three Korean stop categories over the 8 sessions. In session 1, learners identified relatively well the aspirated stops, with 57% correct identification. For the other categories, identification was close to chance level: 39% of the fortis and 34% of the lenis. The comparison between sessions presents a pattern of improvement which is also category-dependent. For the aspirated stops, learners start to improve significantly compared to session 1 from session 6 and onwards. For the fortis stops, improvement starts from the third session, and remains stable with a bit more than half of fortis stops correctly identified onwards. In session 8, 70% of the aspirated stops are correctly identified, while 53% of the fortis stops are correctly identified. For the lenis stops, identification remains around chance level for the 8 sessions.

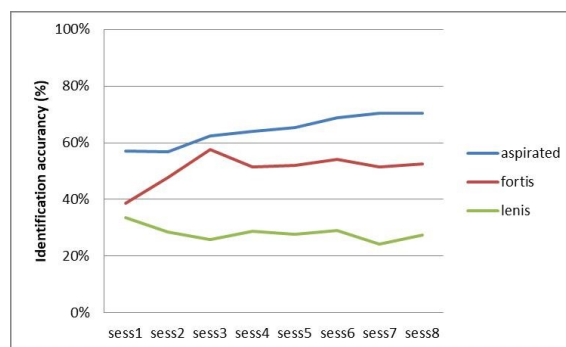


Figure 1 : Identification accuracy' rate of the Korean three-way category over time.

The response patterns are illustrated in Figure 2. Misidentified aspirated consonants are confused mostly with fortis consonants over the 8 sessions (24% of time on average). The improvement in the identification of the aspirated category is linked to a reduction of its confusion with the lenis category (20% in first session vs. 8% in last session). Fortis consonants are mainly confused with lenis stops (35% on average), and some are confused with aspirated stops (14% on average), and both confusions reduce across sessions. For the lenis stops, more than half of them are identified as aspirated, and 23% on average as fortis, and this confusion pattern remains stable over time.

In order to explain the poor identification of lenis stops and their confusion pattern, we examined the relative influence of the two acoustic dimensions VOT and f0 on the categorization of the three stop categories in the second experiment.

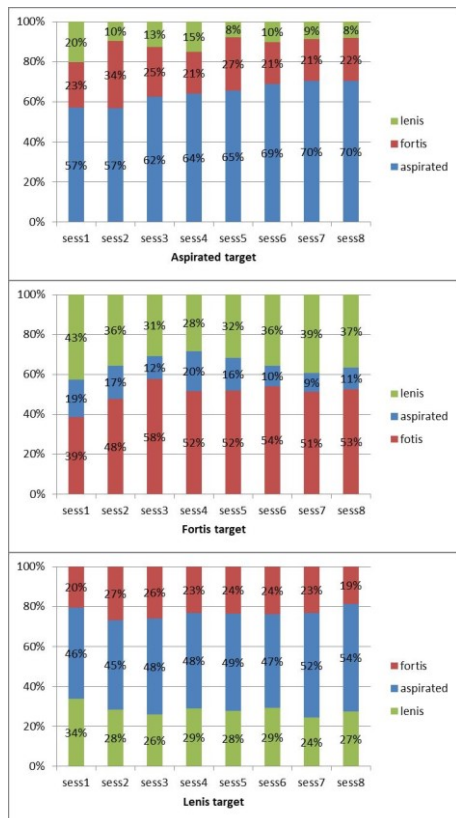


Figure 2 : Distribution of response (hits and confusions) for each target category (aspirated on upper panel, fortis on middle panel and lenis on lower panel).

3.2. Synthesized stimuli

Figure 3 presents the identification rate for the three categories as a function of VOT (upper panel) and f_0 (lower panel) for French learners in all sessions and for native Korean listeners. Korean listeners made use of VOT to differentiate fortis and aspirated stops: stimuli with short VOT (8 and 20 ms.) are identified as fortis while stimuli with VOT longer than 44ms are identified as aspirated. Manipulation of VOT did not affect the identification of the stimuli as belonging to the lenis category as expected. French learners also used VOT to identify aspirated stops, with around 60% of the stimuli with VOT longer than 44ms identified as aspirated. However, the stimuli with shorter VOT (8 and 20ms), are equally identified by French learners as either lenis or fortis.

Concerning f_0 , the native Korean listeners used f_0 cue in order to distinguish the three stop categories: stimuli with a low f_0 (200-225Hz) are identified as lenis, and stimuli with high f_0 (>250Hz) are identified as aspirated stops. With f_0 higher than 250Hz, the responses toward fortis stops also increase but remains below 50%. In contrast, French learners appear to be insensitive to f_0 manipulation for their ‘aspirated’ responses which are globally in the majority for all stimuli with f_0 at 225Hz or higher. Stimuli with the lowest f_0 however, are either identified as belonging to the lenis or aspirated category, and not as fortis.

Chi-squared test for each category of VOT and f_0 was used to examine which cues French learners use with the same weight as those of Korean for identifying the synthesized stimuli. It revealed that the use of VOT for the categorization

of aspirated stops was similar for French learners and native Korean listeners ($\chi^2(6) = 58.45, p = 9.26$), but not for lenis and fortis stops (lenis: $\chi^2(6) = 86.82, p < .001$, fortis: $\chi^2(4) = 419.72, p < .001$). Concerning f_0 , French learners did not use f_0 cue as did Korean listeners for neither of the stop categories (aspirated: $\chi^2(4) = 636.56, p < .001$, lenis: $\chi^2(4) = 638.11, p < .001$, fortis: $\chi^2(4) = 95.339, p < .001$).

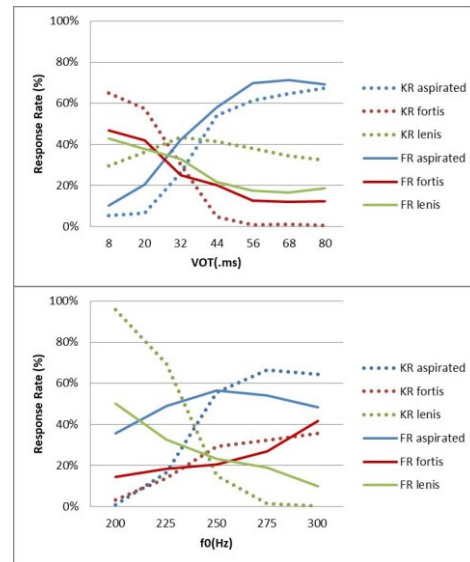


Figure 3 : Identification rate as ‘aspirated’, ‘fortis’ and ‘lenis’ according to stimuli manipulation on VOT (top) and f_0 (bottom) for French learners (FR, solid lines) and native Korean (KR, dotted lines).

In order to examine how the sensitivity of French learners to VOT and f_0 cues evolves over time, we analyzed the distribution of the responses with a Chi-squared test for each category of VOT and f_0 across sessions, and there was a significant difference only between the first three sessions and the last three sessions. That is why we merged data from the first to third session and from the sixth to eighth session. The first two panels of Figure 4 illustrate the identification rate as a function of VOT for the fortis (upper panel) and aspirated (lower panel) categories, with the Korean identification rate as a solid line for comparison.

Between the first three sessions and the last three sessions, the distribution of the fortis responses according to the increase of VOT in the synthesized stimuli evolve significantly ($\chi^2(20) = 36.6, p < .01$). Stimuli with a shorter VOT are better identified as fortis stops in the last sessions than in the first, and stimuli with longer VOT (>32ms) are less identified as fortis. Accordingly, the distribution of the aspirated responses also changed over time ($\chi^2(20) = 42.34, p < .01$): stimuli with longer VOT (>32ms) are better identified as aspirated in the last sessions. For the lenis stops responses according to VOT manipulation, we observe no change in the distribution over time ($\chi^2(20) = 20.639, p = 0.41$).

The third and fourth panels of figure 4 illustrate the identification rate for the fortis and lenis categories according to f_0 manipulation. Fortis responses evolve from the first sessions to the last ($\chi^2(14) = 41.49, p < .001$): in the last three sessions, high f_0 is used more to identify fortis, and low f_0 less. It is notable that for stimuli with f_0 higher than 250Hz, learners mirror the identification rates of Korean at the end of the year.

No change is found for the distribution of the lenis responses over time ($\chi^2(14) = 22.403, p = 0.07$): identification remains poor for the lenis responses expected to be higher for the stimuli with a low f_0 .

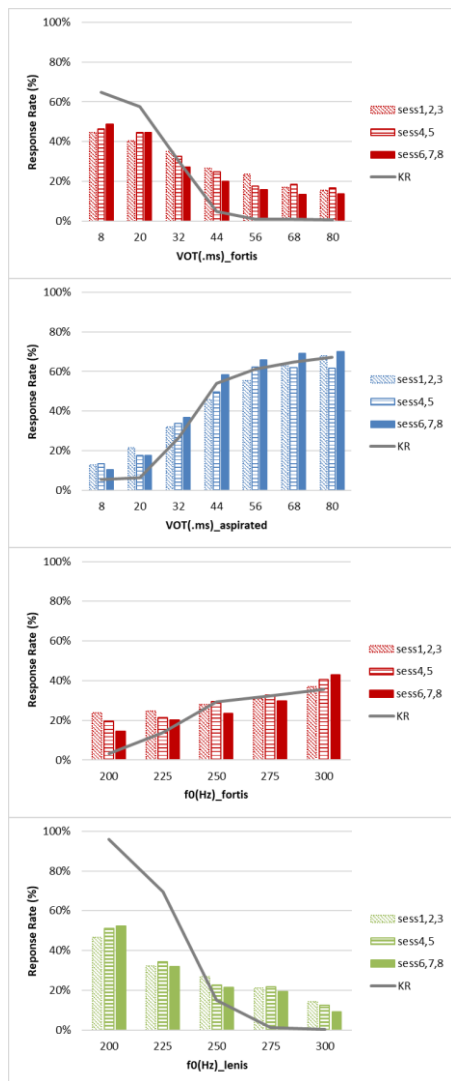


Figure 4 : Illustration of the identification rate as a function of VOT and f_0 averaged over the first three sessions compared to the last three sessions for the fortis responses (panel 1 and 3), the aspirated responses (panel 2) and fortis responses (panel 4). The results of the 4th & 5th sessions are presented as well. Native Korean listeners' responses (KR) are illustrated in solid line for comparison.

4. Discussion

This study is, to the best of our knowledge, the first longitudinal report on the acquisition of L2 Korean stops by French learners. Previous studies on L2 Korean perception has been interpreted using the Perceptual Assimilation Model (PAM, [3]) or the Speech Learning Model (SLM [7] & SLM-r [8]). Both models posit that the perceived phonetic distance between L1 and L2 phonemes influences the perception of non-native segments, and that this distance predicts which phoneme will be easy or difficult to learn for L2 learners at the beginning of learning.

The SLM-r proposes that the dissimilarity of the L1-L2 sound predicts whether a new category will be formed for the L2 sound. [2] showed that French learners were more successful at perceiving Korean aspirated because Korean aspirated stop has a long-lag VOT, whereas French has a lead vs. short VOT contrast. According to the PAM, for French learners, the perception of Korean aspirated and fortis stops would be easy because they perceive the aspirated as French voiceless stops and the fortis as French voiced stops. However, the perception of Korean lenis stops would be hard since Korean lenis would be perceived as both French voiced and voiceless stops. This pattern was confirmed in our first experiment: we found that aspirated stop was the easiest category for learners to identify, followed by fortis, whereas lenis stops were poorly identified. Moreover, the lenis category was mostly confused with the aspirated category (more than 50% of the time), because the lenis and the aspirated stops have similar VOT values (long VOT). The fortis was misidentified as the lenis, which was an unexpected result. Because the fortis has a short VOT and the lenis has a long VOT, explaining this confusion with the VOT type is not possible. As mentioned in introduction, the f_0 cue is more important for distinguishing between lenis and the other categories in Korean. As a result, results need to be interpreted with the two cues (VOT and f_0) in mind. The second experiment allows us to examine cue weighting, which might be related to the difficulty of identifying lenis stops. [11] demonstrated that Korean relied on f_0 to perceive lenis stops, and primarily on VOT for fortis stops, and f_0 and VOT for aspirated stops. In our second experiment, we observed the same pattern for Korean L1 speakers. Meanwhile, for French learners, the VOT cue alone is used to identify the aspirated vs. fortis/lenis contrast. This pattern explains why learners misidentified the fortis as the lenis. They accorded nearly equal weight to VOT for the fortis and lenis categories. Regarding f_0 , this cue alone is used to distinguish the lenis vs. fortis contrast, unlike in Korean, where the f_0 is used more to identify the lenis vs. aspirated contrast. As a result, for learners, short VOT and low f_0 induce the perception of the lenis stop, while for Korean speakers, regardless of VOT value, low f_0 is the most relevant cue to distinguish the lenis stop from the fortis and aspirated stops. The larger weight given to VOT compared to f_0 by the French learners can explain the poor identification of lenis stops.

Furthermore, the French learners' use of VOT to identify aspirated stops was similar to that of native Korean. It explains why, from the beginning of learning, learners perceived quite well the aspirated stop. By examining the cue weighting over time, we observed that learners assigned more VOT cue to identify the fortis and aspirated stops and more f_0 cue for the fortis stops. It explains why the identification of lenis did not improve in the first experiment.

On the basis of these findings, we suggest that French learners should be taught to pay more attention to f_0 , which is the most relevant cue to distinguish the lenis from the aspirated and the fortis stops for native Korean. In light of our findings, one question remains: how do French learners employ these cues in production to distinguish the three-way stops contrast? We expect that production would reflect perception. In the future, we plan to investigate production data from the same participants to better understand the influence of cue weighting in perception and production in the L2 acquisition of Korean.

5. Acknowledgements

This work is partially supported by ANR-10-LABX-0083.

6. References

- [1] A. M. Schmidt, “Cross-language identification of consonants. Part 1. Korean perception of English,” *The Journal of the Acoustical Society of America*, vol. 99, no. 5, pp. 3201–3211, May, 1996.
- [2] B. Lee, N. Yamaguchi, and C. Fougeron, “Perception des occlusives du coréen L2 : réorganisation du cue weighting au cours du temps,” in *Proceedings of JEP 2022 – 34èmes Journées d’Etudes sur la Parole*, Île de Noirmoutier, France, June, 2022.
- [3] C. T. Best, G. W. McRoberts, and E. Goodell, “Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener’s native phonological system,” *The Journal of the Acoustical Society of America*, vol. 109, no. 2, pp. 775–794, Feb. 2001.
- [4] D. Bates, M. Mächler, B. Bolker, and S. Walker, “Fitting Linear Mixed-Effects Models Using lme4,” *J. Stat. Soft.*, vol. 67, no. 1, pp. 1–48, Oct. 2015.
- [5] G. Stoet, “PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments,” *Teaching of Psychology*, vol. 44, no. 1, pp. 24–31, Nov. 2017.
- [6] J. E. Flege, O.-S. Bohn, and S. Jang, “Effects of experience on non-native speakers’ production and perception of English vowels,” *Journal of Phonetics*, vol. 25, no. 4, pp. 437–470, Oct. 1997.
- [7] J. E. Flege. “Assessing constraints on second-language segmental production and perception,” in *Phonetics and phonology in language comprehension and production: Differences and similarities*, New York: De Gruyter Mouton, pp. 319–355, 2003.
- [8] J. E. Flege and O.-S. Bohn, “The revised speech learning model (SLM-r),” in *Second Language Speech Learning*, Cambridge University Press, pp. 3–83, 2021.
- [9] J. J. Holliday, “The perceptual assimilation of Korean obstruents by native Mandarin listeners,” *The Journal of the Acoustical Society of America*, vol. 135, no. 3, pp. 1585–1595, Mar. 2014.
- [10] J. Schertz, T. Cho, A. Lotto, and N. Warner, “Individual differences in perceptual adaptability of foreign sound categories,” *Attention, Perception, & Psychophysics*, vol. 78, no. 1, pp. 355–367, Sep. 2015.
- [11] L. Lisker and A. S. Abramson, “A Cross-Language Study of Voicing in Initial Stops: Acoustical Measurements,” *WORD*, vol. 20, no. 3, pp. 384–422, Jan. 1964.
- [12] M. Kim, “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, Oct. 2004.
- [13] M.T.M. García and J. J. Holliday, “The perception of Korean stops by native speakers of Spanish,” in *Proceedings of the 19th International Congress of Phonetic Sciences*, Melbourne, Australia, pp. 2585–2589, Aug. 2019.
- [14] N. Ryu, “Effects of L2 proficiency on Mandarin speakers’ perception of Korean stops and affricates,” *Journal of Korean Language Education*, vol. 27, no. 4, pp. 35–56, Dec. 2016.
- [15] P. Boersma and D. Weenink, “Praat: doing phonetics by computer, version 6.0.37,” 2018. [Online]. Available: <http://www.praat.org/>
- [16] P. Iverson et al., “A perceptual interference account of acquisition difficulties for non-native phonemes,” *Cognition*, vol. 87, no. 1, pp. B47–B57, Feb. 2003.
- [17] R. H. Baayen, D. J. Davidson, and D. M. Bates, “Mixed-effects modeling with crossed random effects for subjects and items,” *Journal of Memory and Language*, vol. 59, no. 4, pp. 390–412, 2008.
- [18] Team, R. Core. “R: A language and environment for statistical computing,” 2013.