# INDIVIDUAL DIFFERENCES IN NON-NATIVE PHONOLOGICAL CONTRAST LEARNING: THE ROLE OF PERCEPTUAL SENSITIVITY TO SUB-PHONEMIC VARIATION IN NATIVE CATEGORIES

by

Jieun Lee

A Dissertation Submitted in
Partial Fulfillment of the
Requirements for the Degree of

Doctor of Philosophy
in Linguistics

at

The University of Wisconsin-Milwaukee
May 2022

# ABSTRACT

# INDIVIDUAL DIFFERENCES IN NON-NATIVE PHONOLOGICAL CONTRAST LEARNING: THE ROLE OF PERCEPTUAL SENSITIVITY TO SUB-PHONEMIC VARIATION IN NATIVE CATEGORIES

by

Jieun Lee

The University of Wisconsin-Milwaukee, 2022
Under the Supervision of Professor Hanyong Park

The current study explored individual differences in sensitivity to sub-phonemic variation of acoustic cues in the perception of a native language (L1) category in order to test the hypothesis that second language (L2) learners' different sensitivity to the L2-relevant acoustic dimension in L1 perception could explain individual variability in nonnative phonological contrast learning. In addition, this study investigated whether the modified High Variability Phonetic Training (HVPT) paradigm could aid in nonnative phonological contrast learning. The cue-attention switching training was added to the typical HVPT paradigm with multiple talkers, expecting to reallocate learners' attention away from the less relevant acoustic dimension to the more informative acoustic dimension in the perception of the target nonnative contrast(s). The present study targeted two groups of learners with different L1 backgrounds: naïve adult English learners of Korean and intermediate adult Korean learners of English. The multiple HVPT sessions trained English learners of Korean on a Korean three-way laryngeal contrast in stop (/p'/-/p/-/p$^h$/) and trained Korean learners of English on three English vowel contrasts, /i/-/ɪ/, /ɛ/-/æ/, and /ʊ/ -/u/.

The Visual Analogue Scaling (VAS) task measured English adult listeners' sensitivity to sub-phonemic acoustic details in the perception of English stop voicing contrast with a stimuli continuum of English voiced and voiceless stops (/b/-/p/) varying in VOT and f0 at vowel onset. For Korean adult listeners, the AXB oddity task quantified learners' sensitivity to within-category differences induced by spectral and duration cue changes, using a set of stimuli belonging to the Korean /i/ vowel but with different spectral and duration properties. The results of the HVPT training in experiments 1 and 3 revealed that in both groups, L2 learners with higher sensitivity to L2-relevant acoustic cues in L1 perception had an initial advantage in L2 contrast learning and showed more nativelike cue utilization during and after the HVPT. On the other hand, learners with less sensitivity to the "right" acoustic cues failed to systematically use those cues in perceiving the target L2 contrast(s).

Learners who received the modified HVPT with the cue-attention switching training with L1 stimuli in experiments 2 and 4 demonstrated more native-like use of acoustic cues in L2 perception than learners who received only the typical HVPT with multiple talkers. English learners of Korean with relatively less sensitivity to f0 cues in the perception of English voicing contrast performed similarly to those with relatively high sensitivity to f0 cues. For Korean learners of English, the benefit of the cue-attention switching training was observed in learning the English /i/-/ɪ/ contrast, but not in more challenging /ɛ/-/æ/ and /ʊ/ -/u/ contrasts. Korean learners of English with the cue-attention switching training showed more reliance on spectral than duration cues like English native listeners.

This study showed the relation between individual differences in sensitivity to sub-phonemic acoustic details in L1 and the nonnative novel phonological contrast learning and a possible type of training to overcome disadvantages due to the individual differences. The results suggest the transfer of L1 cue sensitivity to L2 cue utilization. That is, how successfully L2 learners progress to become more nativelike listeners can be predicted in terms of to what degree they have sensitivity to the L2 informative acoustic cue in L1 speech perception. This implies that individual differences in the L2-relevant cue sensitivity may determine the initial stage of learning and to what extent learners can benefit from L2 training. Moreover, this study emphasizes the importance of considering individual differences to predict L2 learners' learning outcomes and provide appropriate L2 training to learners whose perceptual abilities may place them at a disadvantage. The VAS and AXB oddity tasks showed possibilities as pretraining assessments to predict the acquisition of L2 phonological contrasts and L2 cue-weighting strategies.

*This page intentionally left blank*

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# LIST OF ABBREVIATIONS

| | |
|---|---|
| **CP** | Categorical Perception |
| **F1** | First formant |
| **F2** | Second formant |
| **HS** | High Sensitivity |
| **HVPT** | High Variability Phonetic Training |
| **ID** | Identification |
| **L1** | First (native) Language |
| **L2** | Second Language |
| **LS** | Low Sensitivity |
| **M** | Mean |
| **New Con** | New Consonant |
| **SD** | Standard Deviation |
| **VAS** | Visual Analogue Scaling |
| **VOT** | Voice Onset Time |

# ACKNOWLEDGEMENTS

Words cannot express my deepest gratitude to my advisor, Dr. Hanyong Park. I could not have undertaken this journey as a Ph.D. student without Dr. Park's invaluable patience, guidance, and feedback. With his mentoring and training, Dr. Park always inspired me to brainstorm and think about issues in Phonetics as a researcher and showed all his support to turn my research into something to contribute to the field of my study. In addition to his academic encouragement, he cared about my well-being as a Ph.D. student living in a foreign country. His positive comments and sincere care expressed in formal and informal ways improved me as a person and as a researcher.

I would like to extend my sincere gratitude to my defense committee, Dr. Anne Pycha, Dr. Jae Yung Song, Dr. Fred Eckman, and Dr. Jong-mi Kim, for their generously provided knowledge and expertise to make my research more insightful and meaningful. I am deeply indebted to my committee members' unconditional support and advice for my academic and career success. I am also grateful to the faculty members in the Department of Linguistics, Dr. Garry Davis, Dr. Nicholas Fleisher, Dr. Hamid Ouali, and Dr. Sandra Liliana Pucci. They provided a warm and productive environment for linguistic study, and their courses broadened my understanding and knowledge of Linguistics.

I had the pleasure of studying with my classmates and cohort members in the department. I enjoyed taking classes together, discussing and giving feedback on our research ideas, and sharing our concerns as graduate students with a bond of sympathy for each other. Special thanks go to my dear Korean friends, especially Okgi Kim, Dong-Jin Kim, Yunhui Kim, and Seulkee Park, for their academic help, feedback, and moral support. They never hesitated to participate in my pilot experiments and gave me constructive feedback on how to improve my

studies. Countless discussions and late-night hanging out brightened my academic years at UWM. I look forward to the exciting time ahead with them as colleagues and friends.

This endeavor would not have been possible without the generous support from Dr. Jong-mi Kim at Kangwon National University. She was the one who first introduced me to Linguistics, experienced me in linguistic research, and encouraged me to pursue my postgraduate degree in the field of Phonetics. She also generously allowed me to conduct my experiments for this dissertation in her Phonetics Lab. I deeply appreciate all her advice and sincere support.

My heartfelt appreciation goes to my family for their unconditional love, support, and prayers. I sincerely thank my parents, Hee Sang Lee and Hyo Sook Kim, my grandparents, Bu Geun Lee, Ok Soon Kim, Jae Sung Kim, and Chun Ja Jong, and my sisters, Jimin Lee and Yule Lee. Especially, I thank to my grandfather, Jae Sung Kim, for giving me strength to overcome difficult times. I miss him very much and wish him ultimate peace and comfort.

Special thanks to my friends in Korea. Ever since I decided to come to the US for my degree, they showed their biggest support for my future and wished me all the best. Whenever I visited Korea, they always welcomed me and devoted their time to making good memories with me. I also would like to acknowledge my friend, Rosamayra Zamora, for her editing help which made my writing better in many ways.

Lastly, it would be remiss of me not to mention my gratitude to my participants, as my dissertation would not have been possible without their serious participation. Although we were all coping with a difficult time due to the pandemic, my participants volunteered and continuously participated in my multiple-day training study.

Without the help and support from my loved one, I would not be where I am today. Thank you so much from the bottom of my heart.

# CHAPTER 1. INTRODUCTION

## 1.1. Categorical perception (CP) in speech sounds and within-category cue sensitivity

Categorical Perception (CP) in speech sounds refers to the phenomenon that the acoustic stimuli, which vary along a physical continuum of equal intervals, are perceived as discrete categories, and the differences between categories are more discriminable than within categories (e.g., Cheng & Chen, 2020; Harnad, 2003; Reetz & Jongman, 2020; Zhang et al., 2016). Liberman, Harris, Hoffman, and Griffith (1957) first established experimental evidence of CP, including identification and discrimination tasks to label phonemes. The generalized characteristics of CP are first, the identification task sharply defined the categorical boundary, and second, a marked accuracy peak in the discrimination task was close to the position of categorical boundary (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Liberman, et al., 1957). In early research on human speech perception, it was widely considered that listeners perceive speech sounds categorically, and continuous details of underlying acoustic forms are mainly lost in favor of discrete, categorical representation (e.g., Pisoni, 1973; Studdert-Kennedy, Liberman, Harris, & Cooper, 1970). The earlier research has helped us understand how listeners effectively and efficiently parse continuous and highly variable acoustic information in speech signals as discrete units.

In contrast to the view of categorical perception, a large body of evidence suggests gradient encoding of speech categories, in which listeners are sensitive to subtle acoustic differences in cues as within-category information (e.g., Massaro & Cohen, 1983; Miller, 1994; Pisoni & Lazarus, 1974; Toscano & Lansing, 2019). Previous research has found that listeners can discriminate speech stimuli within a category remarkably well, resulting in gradient categorization behavior which may allow for flexibility in how acoustic cues are mapped onto sound categories. In addition, training studies have shown that category boundaries can undergo changes, that categorical discrimination can be attenuated, and that previously indistinguishable stimuli can be separated into novel categories (e.g., Carney, Widin, & Viemeister, 1977; McClelland, Fiez, & McCandliss, 2002; Pisoni, Aslin, Perey, & Hennessy, 1982; Samuel, 1977). Researchers used tasks, such as a 4IAX task and a same-different task, to test listeners' sensitivity to low-level phonetic details (i.e., sub-phonemic acoustic details) in speech perception (Carney et al., 1977; Pisoni & Lazarus, 1974). Pisoni and Lazarus (1974) conducted a 4IAX discrimination test with English bilabial stop consonant stimuli which differed in voice onset time (VOT), varying from 0 to 60 ms VOT in 10 ms intervals. The 4IAX procedure asked listeners to compare which of two pairs of stimuli was the same, the first pair or the second pair. This procedure was conducted in order to force listeners to make a pairwise comparison, and thus respond to the magnitude of difference between pairs of stimuli. This measure of discrimination revealed that listeners are sensitive to within-category distinctions. The 4IAX shows a less categorical mode of discrimination in the sense that listeners can discriminate within-category differences much better than would be expected if they based their discrimination decision on only absolute identification (Pisoni & Lazarus, 1974, p. 5). In addition, it was found that the 4IAX test performance was higher than the performance of the traditional ABX discrimination task at some points along the stimulus continuum. This result indicated that the 4IAX test is more suitable for "within" category comparisons than for the between category comparisons. Taken together, the authors suggested that the results support the notion that listeners can discriminate the acoustic cues that underlie the voicing feature in a more nearly continuous mode of perception. Other research on native language (L1) speech perception also suggested listeners' usage of sub-phonemic acoustic details by showing that less canonical acoustic values of a category slowed down listeners' reaction times and lowered goodness rating scores of the stimuli (e.g., Gerrits & Schouten, 2004; Schouten, Gerrits, & Van Hessen, 2003).

In the domain of lexical processing, a number of previous studies have demonstrated the effects of within-category sub-phonetic variation on spoken word recognition (e.g., Andruski, Blumstein, & Burton, 1994; McMurray, Aslin, Tanenhaus, Spivey, & Subik, 2008; McMurray, Tanenhaus, & Aslin, 2002; Utman, Blumstein, & Burton, 2000). For instance, McMurray et al. (2008, 2002) examined within-category sensitivity using a measure of lexical activation based on the Visual World Paradigm and monitored eye movements throughout the experiment. Participants heard stimuli from one of six VOT continua (*beach/peach, bear/pear, bale/pail, bomb/palm, bump/pump, butter/putter*) instantiated as word-initial consonants ranging from 0 to 40 ms in 5 ms steps. After each token, participants selected the corresponding picture from four pictures, which were the target, its voicing competitor, and two unrelated items. The results showed that as VOT approached the category boundary, participants were more likely to look at the competitor picture, indicating gradient effects of the magnitude of the lexical competitor activation. Thus, this research evidenced the listeners' gradient lexical processing and sensitivity to within-category acoustic details.

In addition to previous research on gradient speech perception and listeners' sensitivity to sub-phonemic acoustic details, a number of previous studies have shown that certain speakers present more gradient perception and higher sensitivity to within-category acoustic details. For example, developmental studies have reported that children relative to adults, and children with language or reading disorders (e.g., dyslexic) relative to typically developing children, have exhibited a less steep switch between speech categories in the identification task with a two-answer forced-choice (2AFC) paradigm. Moreover, they showed better ability in discriminating within-category stimuli (e.g., Hazan & Barrett, 2000; Joanisse, Manis, Keating, & Seidenberg, 2000). The foregoing evidence led to a consensus that speech categories are fundamentally graded and that listeners actually attend to sub-phonetic acoustic details in speech perception. Importantly, gradient processing and sensitivity to fine-grained acoustic details can be advantageous in a certain context, such as online spoken word recognition as suggested in McMurray et al. (2008). McMurray et al. (2008) argued that sensitivity to fine-grain acoustic detail may allow listeners to anticipate upcoming phonetic information in the speech signal and enhance word recognition by dynamically compensating for acoustic variability.

## 1.2. Individual differences in within-category cue sensitivity

The aforementioned findings on categorical perception raise important questions: First, do individual listeners vary in their categorization gradiency? Second, how are these individual differences associated with sensitivity to continuous acoustic details? Recent studies have examined individual differences in categorization gradiency and found that listeners' sensitivity to within-category acoustic information is closely related to how gradient listeners' speech categorization is (e.g., Kapnoula, Winn, Kong, Edwards, & McMurray, 2017; Kim, Clayards, & Kong, 2020; Kong, 2019; Kong & Edwards, 2011, 2016). As a measure of gradiency of phoneme categorization, these studies utilized a visual analogue scaling (VAS) task, a continuous measure of phonetic categorization (Massaro & Cohen, 1983). According to Massaro and Cohen (1983), listeners' responses are less categorical on the VAS task compared to the traditional 2AFC task since the VAS task gives more options to respond. Rather than forcing listeners to choose two options for sound categorization (i.e., 2AFC task), the VAS task asks listeners to mark their choice anywhere along a continuous line between two options. Previous studies captured substantial individual differences in the VAS task and showed that listeners' VAS task response patterns are tightly linked to sensitivity to subtle acoustic differences in within-category information. Kong and Edward (2011, 2016), for example, examined English native listeners' perception of the stop voicing contrast (/da/-/ta/) with the VAS task and whether individual differences in responses are associated with listeners' sensitivity to multiple acoustic cues using an anticipatory eye movement paradigm (AEM) task. The multiple acoustic cues included VOT, the time-lapse between the release of the stop closure and the onset of voicing in the following vowel (Lisker & Abramson, 1964). The results of the VAS task showed apparent individual differences in gradiency of listeners' categorical perception; some exhibited a more categorical pattern in favor of endpoint responses, while others showed a more gradient pattern using an entire VAS scale. Furthermore, listeners who showed a more gradient response pattern were more sensitive to changes in the secondary cue of f0 during the AEM Task. The authors argued that listeners with more gradient response strategies are likely to attend to multiple cues (i.e., both VOT and f0). The authors suggested the possible benefit of gradient speech perception, especially under challenging listening context in which engaging a redundant cue (in this case, f0) is necessary. Real-life listening context cannot be always ideal enough to guarantee a perfectly audible primary cue. Some contexts may include the situation when a primary cue such as VOT is obscured due to background noise or interruptions to the speech signal; therefore, the use of f0 cue is necessary. According to Kong and Edward (2016), listeners who showed gradient perceptual patterns might have an advantage in real-life listening contexts due to higher sensitivity to the redundant acoustic cue. Kapnoula et al. (2017) also reported the relation between listeners' gradient VAS response patterns and their sensitivity to the secondary cue in the perception of English stop voicing contrast. The sensitivity was measured by the 2AFC task with the two continua with the minimum and maximum values of f0. In the graph of the 2AFC results, which show the proportion of /p/ responses, the differences in crossover point for the two continua were calculated. These differences served as a measure of secondary cue use that is independent of the VAS task (see Appendix A). In terms of vowel perception, Kim et al. (2020) examined whether previously observed patterns of individual differences in phoneme categorization gradiency are also observed for vowel perception and, if so, whether those patterns are related to secondary cue use. Native English listeners' gradiency were measured through the VAS task with English /ɛ/-/æ/ contrast stimuli varying in seven duration and seven spectral steps. Spectral and duration dimensions are used as primary and secondary cues, respectively, by native English listeners. Listeners also completed the 2AFC task with the same set of stimuli to measure differences in crossover point for the two continua (short and long vowel) as in Kapnoula et al. (2017). The crossover differences offered a measure of duration cue. The results found a significant association between categorization gradiency and secondary cue use. The authors suggested that listeners with a more gradient pattern are more sensitive to fine-grained acoustic details and thus better at utilizing subtle acoustic differences across multiple cues. In sum, previous studies in consonants and vowels have commonly suggested that individual listeners are different in their sensitivity to within-category acoustic details, and the differences are shown as a more gradient speech perception.

Kong (2019) and Kong and Kang (in press) extended the aforementioned studies by investigating a different group of listeners, L2 learners. In a more recent work, Kong and Kang (in press) examined how the degree of gradiency in L2 learners' perceptual judgments of L2 categories is related to learners' reliance on language-specific within-category acoustic details.

Korean learners of English completed L1 Korean stop perception tasks, which consist of a VAS task and a 3AFC task with three-way Korean stop stimuli (/t’a/-/ta/-tʰa/) varying in 5 steps of VOT and f0. The same listeners completed L2 English stop perception tasks, which consisted of a VAS task and a 2AFC task with English stop voicing stimuli (/da/-/ta/) varying in 6 steps of VOT and 5 steps of f0. In terms of native stop perception (i.e., Korean), the result showed that listeners with categorization gradiency (i.e., shown by less categorical decisions in the VAS task) tended to rely more on a secondary cue (VOT) in the Korean lax-aspirated stop perception. This result is consistent with the existing findings in English voiced-voiceless stop perception and English /ɛ/-/æ/ vowel contrast perception (e.g., Kim et al., 2020; Kong & Edwards, 2016). Regarding nonnative stop perception (i.e., English), Korean listeners who showed gradient categorization in L2 English stop perception tended to use more the secondary cue (i.e., f0) in the perception of English stop voicing contrast. The authors argued that this result should be understood differently from the result of the L1 Korean stop perception tasks. Kong and Kang (in press) further revealed the association between less gradient categorization in the L2 VAS task and learners’ higher L2 proficiency. This relationship implies that L2 learners’ gradient categorization might be due to their immature phonetic encoding where low L2 proficiency derives poor (i.e., nonnative-like) performance in utilizing a primary acoustic cue (VOT) for the target English contrast. More importantly, this study supports the possible cross-linguistic evidence for the relationship between secondary cue uses and gradient categorization. In sum, this study highlights differences between L1 and L2 perception regarding the benefits and hindrances of listeners’ sensitivity to acoustic details of a secondary cue. In L1 speech perception, sensitivity to acoustic details could be helpful to listeners to be resilient to poor properties in speech sounds. However, L2 learners’ sensitivity to secondary cues might be because they have not acquired the target language enough to utilize multiple acoustic cues in L2 speech perception properly.

As Kong and Kang (in press) demonstrated, within-category cue sensitivity can be beneficial or detrimental depending on the type of speech listeners perceive. However, it is still unknown how listeners' within-category sensitivity in L1 affects the perception of the different types of speech, L2. Suppose such sensitivity to sub-phonemic details in L1 perception can aid listeners to be resilient to poor or degraded acoustic properties in L1 speech sounds. Would it be possible for some listeners to take advantage of their sensitivity to sub-phonemic details to flexibly perceive exotic L2 speech signals? To examine this case, L2 phonological contrast learning could provide an ideal experimental context. By investigating how individual learners' within-category cue sensitivities in L1 influence their acquisition of L2 phonological contrast, we can reveal whether L2 learners can get benefits from their sensitivities to successfully acquire nativelike way of multiple cue utilization. In particular, for such sensitivity to be helpful, L2 phonological contrast should require learners to more actively engage a secondary cue in L1 as a primary cue in L2 or require learners to utilize an acoustic cue in L1 in a more continuous, gradient way in L2 speech perception.

### 1.3. Influence of L1 within-category cue sensitivity in nonnative contrast learning

Previous studies have found that when adult listeners distinguish L2 phonological contrasts that are not employed by their L1, listeners experience significant difficulties since the auditory space for their L1 heavily influences their L2 perception (e.g., Grenon, Kubota, & Sheppard, 2019; Holliday, 2015; Kim, Clayards, & Goad, 2018; Kondaurova & Francis, 2010; Schertz, Cho, Lotto, & Warner, 2015, 2016; Tremblay, Broersma, & Coughlin, 2018). Notably, when the relative informativeness of cues in L1 phonological contrasts does not align with the informativeness for L2 phonological contrasts, adult L2 learners often fail to recognize the target sounds in a nativelike manner. For instance, Japanese learners of English often perceive English vowels based on their duration, whereas English native listeners use spectral cues as a primary dimension. Grenon et al. (2019) suggested that Japanese learners rely on duration to categorize the English lax-tense vowels due to the use of duration cues for Japanese phonological length contrasts for both consonants and vowels. Since duration dimension is informative in Japanese, learners are sensitive to acoustic duration differences and exclusively utilize them to perceive English lax-tense vowel contrast. Tremblay et al. (2018) studied the effect of L1 cue-weighting strategies on the use of acoustic cues in L2 speech segmentation. Targeting English- and Dutch-speaking L2 learners of French, they investigated the influence of L1 on their use of f0 cues to signal word-final boundaries in French (target language). Even though both English and Dutch have lexical stress, the use of f0 cue for lexical identity is different. English stress is signaled by the contrast between full and reduced vowels (i.e., segmental information), while Dutch stress is signaled by prosodic information, such as f0 and duration (i.e., non-segmental information). The authors predicted that native Dutch learners would make greater use of f0 in L2 French due to the greater functional weight of f0 in Dutch. They revealed that native Dutch learners of French relied more on the f0 cue in French than native English learners of French. Thus, their results proved the transfer of L1 cue-weighting strategies to the perception of L2 contrasts.

I concur with the view that how L2 learners use acoustic cues in L2 is largely affected by how and which acoustic cues they use in L1. However, as we discussed in the previous section (see 1.2), there is a wide range of individual differences in acoustic cue sensitivity in L1 speech perception. Thus, I feel that there is a need to consider the effect of these individual differences in L2 learning, rather than assuming that L2 learners with the same L1 would show similar patterns in learning of L2 cue-weighting strategy.

In the present study, I sought to extend previous research on the L1 influence on L2 acquisition by considering two points: (1) L2 learners' individual differences in how they utilize within-category acoustic details in their native category perception and (2) its association with to what degree learners successfully perceive a nonnative sound contrast. Specifically, the present study is primarily concerned with the following question: Are individual differences in the sensitivity to multiple acoustic cues in L1 category perception related to nonnative phonological contrast learning? To answer this question, I examined two different L2 learning cases: adult L1 English learners of L2 Korean in the acquisition of Korean three-way stop contrast and adult L1 Korean learners of L2 English in the acquisition of English vowel contrasts.

The first learning case is when the relative cue weights in L2 are reversed compared to the weights in L1, indicating that the target L2 contrast requires the use of the same acoustic cues in learners' L1 but with different weights. In this learning condition, the relative informativeness of acoustic dimensions in L1 (VOT vs. f0) should be changed such that the most informative dimension in L1 (i.e., VOT) is no longer useful, but the role of the secondary cue in L1 (i.e., f0) is enhanced in perceiving L2 contrast. The second case is when L2 learners lack sensitivity to the L2-relevant acoustic cue due to the absence of such contrasts in L1. In this learning condition, among available acoustic cues (spectral vs. duration), learners' relatively low sensitivity to the L2-relevant acoustic dimension (spectral) should be increased to successfully distinguish target L2 speech contrasts.

### 1.3.1. *First learning case: Acquisition of Korean three-way stop contrast by naïve English learners of Korean*

The primary acoustic cue for the English voicing contrast in stops (e.g., /t/ versus /d/) is VOT. English native speakers use the VOT dimension in preference to the f0 dimension in classifying voicing in syllable-initial stop consonants. The VOT dimension is sufficient to differentiate English voiced stops from voiceless stops in production and perception: short-lag VOT values are associated with a voiced type, and long-lag VOT values are associated with a voiceless type (e.g., Francis, Kaganovich, & Driscoll-Huber, 2008; Harmon, Idemaru, & Kapatsinski, 2019). In previous studies (e.g., Kapnoula et al., 2017; Kong & Edwards, 2016), however, some native English listeners showed more gradient categorization in the VAS task with a set of English voicing contrast stimuli systematically varying in VOT and f0, and those gradient listeners were more sensitive to a redundant acoustic cue (i.e., f0). The f0 dimension in English secondarily characterizes the English stop voicing contrasts in that lower f0 values are associated with a voiced type, and higher f0 values are associated with a voiceless type. Although this may seem like a mere idiosyncrasy, it has potentially significant implications for learning an L2, especially one (such as Korean) which uses f0 as an equally important dimension as VOT, rather than as a secondary cue regarding lenis-aspirated stop contrast.

Unlike English initial stop contrasts, Korean has three-way stop contrasts: fortis, lenis, and aspirated. In the perception of Korean fortis stops, the shortest VOT is used as a primary cue along with mid-to-high f0. However, VOT is not a primary cue for Korean lenis and aspirated stops (e.g., Kang & Guion, 2008; Kong & Kang, 2017; Schertz et al., 2015). Notably, younger Korean speakers are merging VOT values for these two types of stops and using f0 as the primary acoustic cue (Kim, 2004). Moreover, Schertz et al. (2015) showed that Korean stop stimuli, which have the same VOT value but different f0 values, elicited different perceptual patterns in the Korean stop contrast classification task: stimuli with lower f0 values got more lenis stop responses, but stimuli with higher f0 values received more aspirated stop responses.

The difference in how VOT and f0 are used in the perception of English and Korean stop contrasts implies that for native English speakers, the active use of f0 along with VOT is essential in learning Korean stop contrast, especially for lenis-aspirated contrast. In the cross-language category mapping task (Schmidt, 2007), English listeners overall labeled the aspirated and lenis voiceless Korean stops as the corresponding aspirated voiceless English stops. The Korean fortis stops were often labeled as homorganic English voiced stops. For example, Korean lenis /p/ and aspirated /$p^h$/ stops with the Korean /a/ vowel were mapped on the English /p/ stop with 99% and 100%, respectively. Korean fortis /p'/ was labeled as the English /b/ stop with 84% in front of the /a/ vowel.

Based on the difference between Korean and English in terms of the relative importance of VOT and f0 cues and the results of the cross-language mapping task, the key to the successful learning of the Korean stop category is the reallocation of learners' attention from VOT to f0 by recognizing the informativeness of the f0 dimension. In this respect, native English learners of Korean who already have more sensitivity to the f0 dimension in perceiving their corresponding native stop contrast may have an advantage in learning Korean stop contrast by actively involving f0 cues. The first case aimed to examine whether some native English learners of Korean with more within-category sensitivity to f0 dimension in the perception of English stop voicing contrast may result in nativelike systematical use of f0 dimension earlier than learners with relatively low sensitivity to f0 dimension.

### 1.3.2. *Second learning case: Acquisition of English vowel contrasts by Korean learners of English*

American English has a larger vowel inventory with 10 monophthongs [i, ɪ, ɛ, æ, ɑ, ʌ, ə, ɔ, ʊ, u] (Ladefoged & Johnson, 2014; Nishi & Kewley-Port, 2008) than Korean vowel inventory with seven monophthongs [a, e, i, o, u, ɨ, ʌ] (Sohn, 2001). Comparing the English and Korean vowel systems, one significant difference lies in the distinctions between English tense vowels and lax counterparts. English has two high front vowels /i/ (tense) and /ɪ/ (lax), and two back vowels /u/ (tense) and /ʊ/ (lax). However, the Korean vowel system lacks such tense and lax vowel distinctions. English native listeners distinguish the tense and lax vowel contrasts with two major acoustic cues: spectral (frequencies of the first (F1) and the second (F2) formants) and duration cues. Previous studies showed that tense /i/ and /u/ vowels are spectrally distinct from and tend to have a longer duration than the corresponding /ɪ/ and /ʊ/ lax counterparts (Hillenbrand, Getty, Clark, & Wheeler, 1995; Hillenbrand, Clark, & Houde, 2000). The tense /i/ vowel has a lower F1 and higher F2 than its lax counterpart, and the tense /u/ vowel has a lower F1 and a little higher F2 than its lax counterpart. In terms of duration, American English vowel /i/ averages about 41% longer than /ɪ/, and /u/ is about 51% longer on average than /ʊ/ (Crystal & House, 1988). Speech perception studies have demonstrated that native American English listeners relied primarily on spectral cues and secondarily on duration cues when they categorized English vowel stimuli varying in spectral and duration cues (e.g., Flege, Bohn, & Jang, 1997; Kondaurova & Francis, 2008; Lee, 2008). Hillenbrand et al. (2000) showed that /i/-/ɪ/ and /u/-/ʊ/ contrasts, which were manipulated to vary in duration, were minimally affected by that duration manipulation in the vowel identification task. Hillenbrand et al. (2000) suggested that a relatively small influence of duration might be because these vowel contrasts are sufficiently well separated based on spectral cues. Lee (2008) also showed that native English listeners distinguished English /i/-/ɪ/ and /ʊ/-/u/ vowel contrasts solely by spectral cues, regardless of duration cues.

Another discrepancy between Korean and English vowel systems is that English has /ɛ/-/æ/ contrast while Korean has only one /ɛ/ ('에') category. Korean phonology has undergone the merger of /ɛ/ ('에') and /æ/ ('애') distinction, and this merger is prevalent in the Seoul dialect as well as in many others (e.g., Eychenne & Jang, 2015; Hwang & Moon, 2005; Moon, 2007; Shin, 2015; Shin, Kiaer, & Cha, 2012; Yoon & Kang, 2014). Similar to English /i/-/ɪ/ and /ʊ/-/u/ vowel contrasts, native English listeners based their identification of the front vowel contrast /ɛ/-/æ/ primarily on spectral quality and secondarily on vowel duration (Bohn & Flege, 1990). The mid-front vowel /ɛ/ has a lower F1 and higher F2 than the low-front vowel /æ/, and the vowel /æ/ vowel averages about 18% longer than the /ɛ/ vowel (Crystal & House, 1988). Hillenbrand et al. (2000) showed that unlike English /i/-/ɪ/ and /ʊ/-/u/ contrasts, the English /ɛ/-/æ/ contrast showed a robust duration effect in the vowel identification task. They argued that the English /ɛ/-/æ/ contrast shows a greater degree of overlap in their spectral properties and, as a consequence, duration plays a more critical role in recognizing these vowels.

These three English vowel pairs are among the most challenging contrasts for Korean learners of English to acquire (e.g., Flege et al., 1997; Grenon et al., 2019; Ingram & Park, 1997; Kim, Clayards, & Goad, 2017; Kim et al., 2018; Kondaurova & Francis, 2010; Tsukada et al., 2005). For example, Kim et al. (2018) conducted a longitudinal study to investigate the developmental changes in perceptual cue weighting of two English vowel contrasts (/i/-/ɪ/ and /ɛ/-/æ/) by adult and child Korean learners of English during their first year of immersion in Canada. Although adult learners used spectral cues more than child learners at earlier time points, native English listeners still made greater use of spectral cues than Korean learners. As for the pattern of duration cue weighting, both adult and child learners made more use of duration cues than native English listeners. After one year of massive exposure to English in Canada, both groups of learners used spectral and duration cues to distinguish /i/-/ɪ/ contrast while they used only duration dimension to distinguish /ɛ/-/æ/, suggesting the relative difficulty of English vowel contrasts in the acquisition. In other words, although learners did not reach the same level of using the spectral dimension to distinguish the English vowel contrasts, the Korean learners were able to use spectral differences only for the /i/-/ɪ/ contrast not for the /ɛ/-/æ/ contrast as they were exposed more to English. Lee and Cho (2018) reported that Korean learners of English less accurately identified the English /ʊ/-/u/ contrast than the /i/-/ɪ/ contrast. The learners performed worst on the mid-front vowel /ɛ/, and this vowel was mostly confused with the low front vowel /æ/. Korean learners of English who participated in this current study were expected to experience a great deal of difficulty in the acquisition of these three English vowel contrasts due to the discrepancy between the vowel inventories of Korean and English and the strong influence of learners' L1 to L2 learning (Tsukada et al., 2005). Late learners (i.e., adult learners) often perceive and produce at least some L2 vowels and consonants as instances of L1 sounds (e.g., Baker & Trofimovich, 2005; Best, 1995; Flege, 1995), especially in the beginning stages of L2 learning.

When it comes to perceptual cue-weighting of English vowel contrasts by L2 learners of English, previous studies have shown that L2 learners weigh the duration dimension more heavily than the spectral dimension to identify English vowel contrasts. This phenomenon has been observed by L2 learners with different L1 backgrounds, including languages that use vowel duration contrastively, such as Japanese (e.g., Grenon et al., 2019), and languages that do not use duration relevantly to distinguish L1 vowel contrasts, such as Spanish and Mandarin (e.g., Bohn, 1995; Escudero & Boersma, 2004; Kondaurova & Francis, 2010). Native Japanese listeners' reliance on duration can largely be accounted for by L1 transfer. However, for Spanish and Mandarin listeners, L1 transfer cannot solely explain why they relied primarily on duration dimension to identify English vowel contrasts (e.g., /i/-/ɪ/ and /ɛ/-/æ/). To resolve this issue, Bohn (1995) proposed *Desensitization Hypothesis*. This hypothesis argued that listeners of languages that do not use a certain area(s) of the vowel space become "*desensitized*" to variations in formants within that space and, therefore, use psychologically more salient temporal cues (i.e., duration cues) instead (see also Bohn & Flege, 1990). In case of Spanish and Mandarin listeners, it was stated that "native speakers of Spanish or Mandarin which have only one vowel category in the high-front area of the acoustic vowel space where English has two, may be said to be linguistically *desensitized* to spectral differences between vowels in that area" as in Figure 1 (Bohn, 1995, p. 295). It was additionally proposed that even listeners who lack experience with contrastive temporal distinctions in their L1 can still exploit duration dimension in the L2. Bohn (1995) stated that "whenever spectral differences are insufficient to differentiate vowel contrasts because previous linguistic experience did not sensitize listeners to these spectral differences, duration differences will be used to differentiate the nonnative vowel contrast" (p. 294-295).

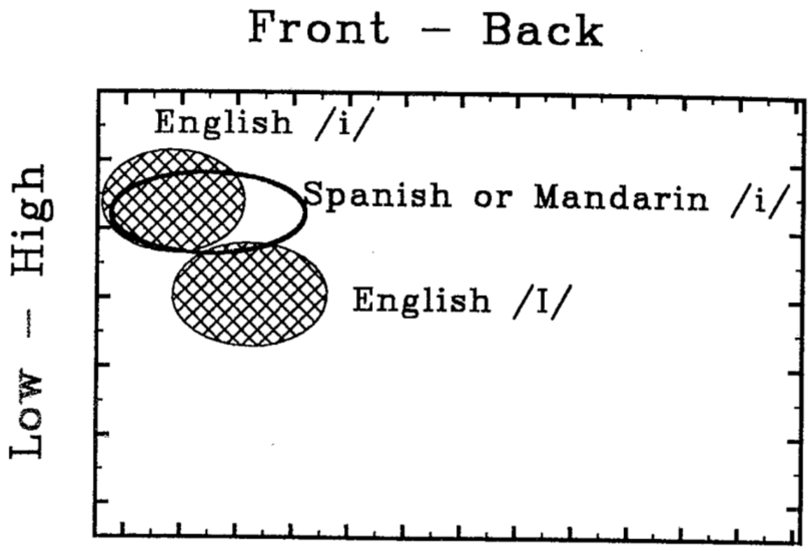


Figure 1. Illustration of the approximate arrangement of English /i/ and /ɪ/ and Spanish and Mandarin /i/ in the acoustic vowel space (Bohn, 1995, p. 286).

Figure 2 is comparable to Figure 1 in terms of English and Korean vowels. Figure 2 shows the illustration of the approximate arrangement of English /i/-/ɪ/, /ɛ/-/æ/, and /u/-/ʊ/ contrasts and Korean /i/, /ɛ/, and /u/ in the acoustic vowel space. Figure 2 plotted based on the vowel identification data in Ryu (2018) (see Appendix B). This vowel identification data is collected by 23 native female English speakers and eight native female Korean speakers who were born and educated in the Seoul/Kyeonggi region. Figure 2 demonstrates a similar pattern to Figure 1 in that Korean has only one vowel category in the acoustic vowel space where English has two. This observation suggests that Korean learners of English may not employ small-scale spectral differences in given portions of the vowel space since the organization of the L1 vowel space desensitizes the L2 learners to spectral differences. Therefore, following Bohn’s hypothesis, Korean learners of English are expected to show relatively lower sensitivity to spectral cues and difficulties in discerning differences between target English vowels in vowel quality. Consequently, they may rely on the duration dimension exclusively to identify English vowel contrasts, as shown in previous studies (Kim et al., 2017, 2018).

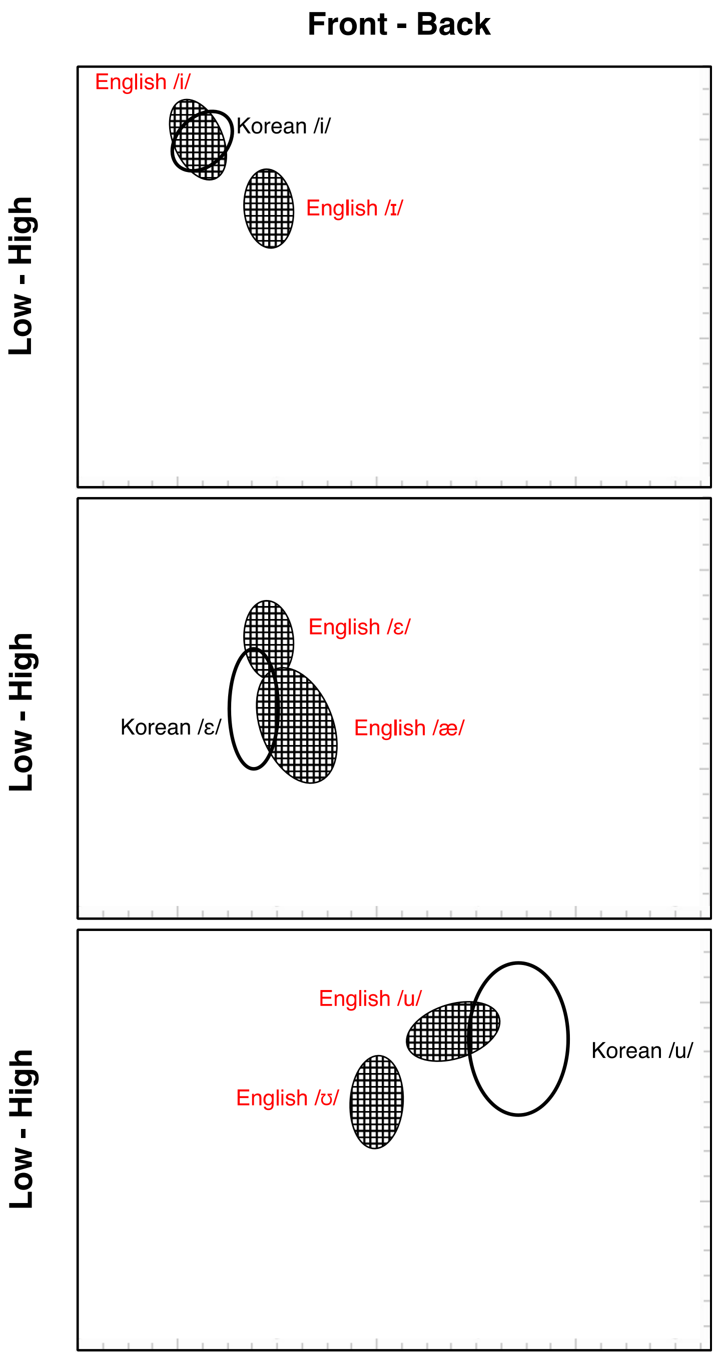


Figure 2. Illustration of the approximate arrangement of English /i/-/ɪ/, /ɛ/-/æ/, and /u/-/ʊ/contrasts and Korean /i/, /ɛ/, and /u/ in the acoustic vowel space.

In accordance with Bohn (1995), the relation of the L1 vowel space on the sensitivity to spectral cues is also evidenced in research examining the effect of L1 vowel inventory size on L2 vowel identification (e.g., Ryu, 2018; Souza, Carlet, Jułkowska, & Rato, 2017). For example, Souza et al. (2017) investigated how Danish, Portuguese, Catalan, and Russian learners of English identify the English /i/-/ɪ/ contrast. It was found that Danish learners of English (with a largest vowel inventory, 20+) demonstrated the most nativelike vowel perception with the primary reliance on spectral dimension while Russian learners of English (with a smallest vowel inventory, 5) over-relied on duration cues more than other learners. The authors argued that their results support Bohn's *Desensitization Hypothesis* in that the smaller vowel inventory of Russian *desensitized* the Russian participants to the small spectral differences present in the English/i/-/ɪ/, forcing them to rely on temporal cues instead.

In a similar sense, Figure 3 is presented below to compare English and Korean vowel spaces drawn based on Korean vowel production data collected from five native female English speakers and five native female Korean speakers (Ahn, 2004). As we can see, English vowels are more crowded in the vowel space than Korean vowels. Considering Bohn (1995) and Souza et al. (2017), Korean learners of English, once again, are expected to experience difficulties in utilizing spectral dimension as a primary cue and rather use duration dimension since the small size of Korean vowel inventory may *desensitize* L2 learners to spectral differences.

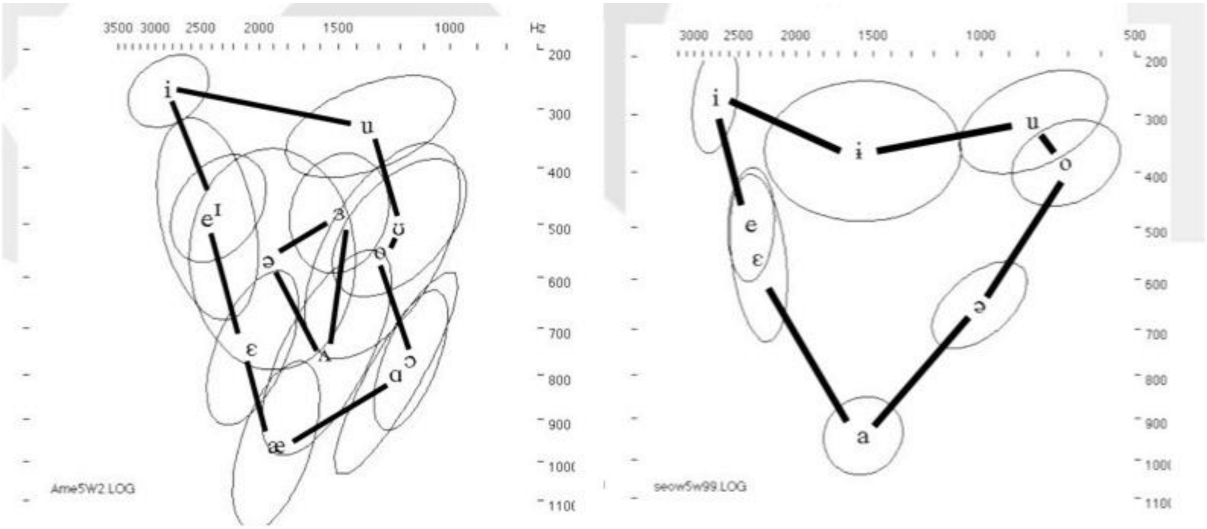


Figure 3. American English female speakers' F1 × F2 vowel space (left) and Korean female speakers' F1 × F2 vowel space (right) (Ahn, 2004, pp. 51 and 55).

Recall that this study examines whether L2 learners' sensitivity to within-category acoustic cues regarding the L2-relevant acoustic cues is beneficial in learning target L2 phonological contrasts. The foregoing comparison between Korean and English vowel systems expects Korean learners of English to have less sensitivity to spectral details since Korean does not have English-like spectral distinctions for the vowel pairs targeted in this study. Relating individual differences in sensitivity acoustic cues and Korean learners' difficulty in learning English vowel contrasts, I ask the following question: if some Korean learners of English are sensitive to within-category acoustic cues regarding spectral details, would they have an advantage of learning English vowel pairs?

The perceived relation between vowels in English and Korean additionally suggests two points: Korean English learners' difficulties in distinguishing English vowels in pairs and the possible advantage of within-category sensitivity to spectral dimension in learning. As mentioned in Flege et al. (1997), the only reliable way to determine the perceived relation between vowels in two language is empirically through a cross-language mapping study (Ingram & Park, 1997; Lee & Cho, 2018). The cross-language vowel mapping task between English and Korean (Lee & Cho, 2018) found that native Korean listeners unequivocally labeled English /i/-/ɪ/ and /u/-/ʊ/ contrasts as a single Korean vowel category, /i/ or /u/, with relatively high goodness ratings (5.5/7 for Korean /i/ and 5/7 for Korean /u/). This result indicates that the distinction between the two English vowels in each contrast can confuse Korean learners of English. Lee and Cho (2018) showed that English vowels /ɛ/ and /æ/ were mapped onto two Korean vowel categories /ɛ/ ('에') and /æ/ ('애'). It seems that Korean listeners differentiated two English vowels in terms of two different Korean vowel categories. However, the perceptual mapping patterns of English /ɛ/ and /æ/ to Korean /ɛ/ and /æ/ partially overlapped with similar goodness ratings. English /ɛ/ was mapped onto Korean /ɛ/ about 55% with mean goodness ratings of 4.9/7 and was mapped onto Korean /æ/ about 43% with mean goodness ratings of 5.1/7. English /æ/ was mapped onto Korean /ɛ/ about 30% with mean goodness ratings of 4.4/7 and was mapped onto Korean /æ/ about 68% with mean goodness ratings of 4.8/7. These results indicate that Korean listeners perceived English /ɛ/ and /æ/ as similarly good instances of either Korean /ɛ/ or /æ/. As suggested in Lee and Cho (2018), this partial overlapping may be due to the merger of the mid-front vowels /æ/ into /ɛ/ in the Korean vowel system as shown in Figure 4 (e.g., Eychenne & Jang, 2015; Hwang & Moon, 2005; Jang, Shin, & Nam, 2015; Shin et al., 2012; Sohn, 2001; Yoon & Kang, 2014). The formant charts shown in Figure 4 demonstrate the merger of Korean /ɛ/ and /æ/. Suppose that Korean listeners do not have a clear distinction between Korean /ɛ/ and /æ/, it was possible for them to randomly map English /ɛ/ and /æ/ vowels to one of Korean /ɛ/ and /æ/ vowels, resulting in similar goodness ratings for both Korean vowels. Thus, in the current study, I consider that English /ɛ/-/æ/ contrast is perceptually assimilated to a single Korean vowel /ɛ/.

Taking together all evidence regarding the differences between Korean and English vowel systems and the perceptual relationship between Korean and English vowels, I hypothesize that learners who show more sensitivity to spectral differences in the perception of a single Korean vowel category perform better in acquiring vowel contrasts. L2 learners with higher sensitivity to spectral differences may have better ability to utilize the small spectral differences present in target English vowel contrasts, and consequently, be able to separate an area in the perceptual vowel space for a single Korean vowel into two based on spectral cues and result in better discrimination of two English vowels in contrast.

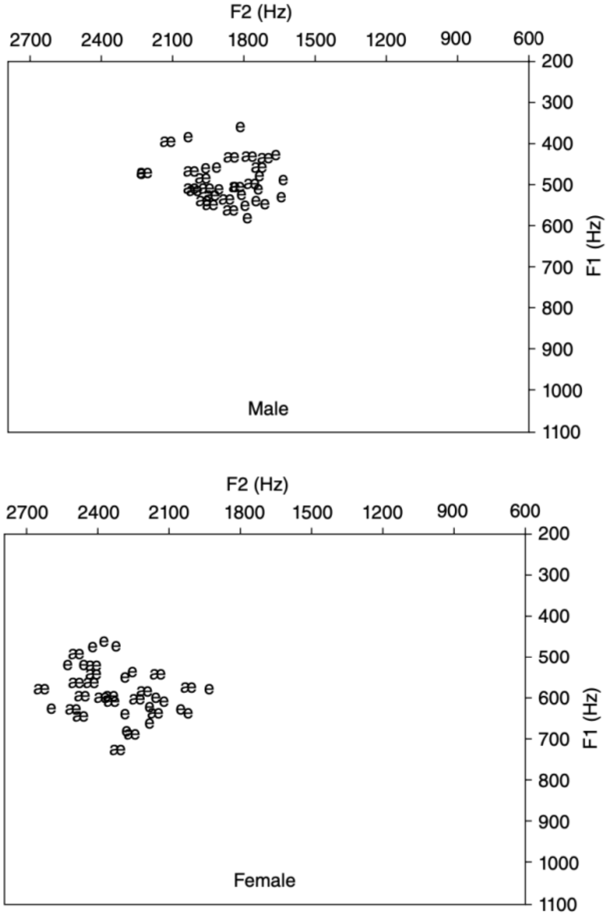


Figure 4. Formant plot of Korean 'ㅔ' (/e/) and 'ㅐ' (/æ/) based on Korean vowel production data collected from ten speakers of Standard Korean (Shin, 2015, p. 5). Korean /ɛ/ is often transcribed as /e/ although both refer to the mid front unrounded vowel.

## 1.4. Increase effectiveness of L2 phonetic training: stimuli distribution and feedback

The successful learning of nonnative contrasts in the current study will be shown as a nativelike use of acoustic cues after training. English learners of Korean should use the f0 and VOT dimensions as the primary and secondary cues, respectively, to identify the Korean three-way stop contrast; and Korean learners of English should predominantly rely on the spectral dimension to distinguish two vowels in each of English vowel contrast. In previous sections, it was hypothesized that participants who are more sensitive to the L2-relevant acoustic dimension (either f0 or spectral) in the perception of the corresponding L1 category (either English stop voicing contrast or Korean single vowel category) will perform better than participants with less sensitivity to sub-phonemic details of the L2-relevant dimension.

If this hypothesis is borne out, the next question is, "How can we aid learners who are predicted to perform relatively poorly in L2 contrast learning due to their less sensitivity to sub-phonemic details of the L2-relevant dimension?". The present study aims to test whether the modified version of High Variability Phonetic Training technique (HPVT) can reallocate attention to multiple acoustic cues in the desired way. Specifically, this study is concerned with examining whether the modified HVPT helps L2 learners redeploy their attention away from the L2 irrelevant acoustic dimension (e.g., VOT for Korean three-way stop contrast and duration for English vowel contrasts) to an L2-relevant acoustic dimension (e.g., f0 for Korean three-way stop contrast and spectral for English vowel contrasts).

Adult L2 learners must actively overcome interference from excessive attention directed toward specific acoustic cues that are detrimental in the target language (Kondaurova & Francis, 2010, p. 570). For example, previous research has demonstrated that Japanese learners of English pay attention primarily to the F2 to distinguish English /r/-/l/ contrast, which native English listeners do not employ for this contrast (e.g., Iverson et al., 2003; Yamada & Tohkura, 1992). Iverson et al. (2003) showed that Japanese listeners tended to ignore the variability along the F3 employed as a primary cue for the differentiation of /r/-/l/ contrast. They argued that although Japanese learners of English were sensitive to, and could detect, within-category differences along the F3 dimension, their attention was directed to the F2 frequency, interfering with their ability to recognize the English /r/-/l/ contrast in an English-like manner. Therefore, many previous studies have argued that L2 learners need to learn how to redirect (enhance) their attention to the L2-relevant acoustic dimension (e.g., F3 for English /r/-/l/ contrast) to achieve a nativelike perceptual pattern of the target contrast (Bradlow, Akahane-Yamada, Pisoni, & Tohkura, 1999; e.g., Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997; Escudero & Boersma, 2004; Kondaurova & Francis, 2010).

Previous studies have compared the effectiveness of short-term laboratory training techniques to investigate which training method can increase listeners' attention to the most informative acoustic cues over the less informative acoustic cues (e.g., Francis et al., 2008; Kabakoff, Go, & Levi, 2020; Kondaurova & Francis, 2010). For example, Kondaurova and Francis (2010) compared three laboratory training methods in acquiring English tense and lax vowels (i.e., /i/-/ɪ/) by native listeners of Spanish. The three training techniques include adaptive training for cue enhancement, inhibition training for cue inhibition, and natural correlation training. Adaptive training aimed to enhance listeners' attention to a target category-relevant dimension (i.e., spectral dimension for English /i/-/ɪ/ contrast) by presenting clearly distinguishable stimuli with exaggerated values compared to typical acoustic differences. Once native Spanish learners of English successfully identified the presented stimuli, stimuli with a reduced perceptual and acoustic difference were presented. The expected effect of adaptive training was to enhance the attention to the spectral dimension by stretching the perceptual distance between tokens so that learners acquire the distinctiveness between tokens through increased differentiability. In contrast, inhibition training introduced irrelevant variability along the less informative dimension (i.e., duration dimension for English /i/-/ɪ/ contrast), encouraging learners to ignore it in categorization. The expected effect of inhibition training was to withdraw learners' attention from the duration dimension by presenting a highly variable set of stimuli along that dimension, resulting in decreased perceptual distance between tokens and poorer differentiation ability. The natural correlation training presented listeners with stimuli clustered around prototypical values for both duration and spectrum cues. This training method was designed to be comparable to the HVPT with multiple talkers or phonetic environments. Kondaurova and Francis (2010) found that reliance on spectral dimension increased after training for all three training groups. However, the inhibition training group showed significantly higher scores in the post identification test of the target English vowel contrast in comparison to pretest scores, while the adaptive training group did not. These results suggest the effectiveness of cue-specific training, especially inhibition training.

Holt and Lotto (2006) showed that the distribution of training stimuli and feedback are examples of training properties that can be modified to restructure listeners' perceptual space to lead the reallocation of attention to multiple acoustic dimensions. The major finding of their research was that changes in weighting strategies can be affected by changes in input distribution parameters. Participants were trained to categorize sounds drawn from a two-dimensional acoustic space defined by the center frequency (CF) and modulation frequency (MF) of frequency-modulated sine waves. These two dimensions were psychophysically matched to be equally discriminable, but their distribution variances were different in training to induce a reversal of weighting (*MF* over *CF*). Training results demonstrated that when the variability of one perceptual dimension (*CF*) was increased (i.e., distributions spanned 15 stimulus steps), and the variability of the other dimension (*MF*) was decreased (i.e., distributions spanned five stimulus steps), the reversal of relative cue-weighting (*MF* over *CF*) was observed. The authors argued that increased stimulus variability of dimension *CF* induced the decreased informativeness of that dimension. Since the increased within-category variance across dimension *CF* creates a high degree of training stimuli distribution overlap, the perceptual distance between stimuli along this dimension can be reduced (i.e., stimuli across this dimension become similar to each other). As a result, this dimension is no longer informative in the categorization decision. On the other hand, by reducing the variance of dimension *MF*, there was not much overlap between stimuli, and the informativeness of this dimension was increased. Based on these results, they suggested that ineffective cue weighting strategies for L2 phonetic categories might be alleviated by manipulating the variance of uninformative dimensions in training stimuli. This implication regarding the acquisition of L2 speech categories aligns with the effectiveness of the inhibition training shown in Kondaurova and Francis (2010). Both studies showed that manipulating variance of acoustic dimensions or distribution of training stimuli can aid L2 learners to acquire the relative informativeness of acoustic dimensions in L2 speech categories.

The cue attention switching process can be effectively accomplished when feedback manipulation is accompanied. Feedback has been shown to be effective in shifting attention across perceptual cues to a phonological contrast in speech perception (Francis, Baldwin, & Nusbaum, 2000). Harmon, Idemary, and Kapatsinski (2019) examined how active learning with feedback could cause native listeners to shift which acoustic cues they used to identify an

English stop voicing contrast. Training stimuli varied in multiple VOT values but varied across only two levels of f0 (low vs. high f0). In order to make f0 perfectly predictive of voicing during training, for a stimulus, the voiced response was correct when f0 was low, and the voiceless response was correct when f0 was high, regardless of the VOT value. Feedback on participants' performances is given only based on the informative dimension (i.e., f0). It was shown that participants could change their perception to depend on the secondary cue (i.e., f0) and downweight the primary cue (i.e., VOT) when feedback reinforced the informativeness of the alternative f0 cue during the identification training. This study suggested that while ignoring the variability of the uninformative dimension, listeners may effectively downweight the uninformative dimension and upweight the informative dimension instead.

### 1.5. Additional training added to a typical high variability phonetic training (HVPT)

The current study designed an additional training to shift L2 learners' attention to the L2-relevant acoustic dimension. This additional training was added to the typical HVPT with multiple talkers to especially aid whose sensitivity to relevant L2 acoustic cues is relatively low in the perception of the native category. In the design of the additional training, I considered the importance of the modification of training stimuli distribution (Holt & Lotto, 2006; Kondaurova & Francis, 2010) and feedback (Harmon et al., 2019) to effectively redeploy learners' attention away from the less relevant acoustic cue in L2. The general format of the additional training resembles with *inhibition training* introduced in Kondaurova and Francis (2010). However, unlike Kondaurova and Francis (2010), the current study attempted to build the additional training with stimuli in learners' native language. Considering that L2 sounds are initially recognized as instances of L1 sounds, it was expected that the increased attention to the L2-

relevant acoustic dimension in the L1 perception helps learners identify target nonnative speech sounds by dominantly relying on that acoustic dimension. L1 stimuli for the additional training systematically varied in two acoustic dimensions, which act as cues in the target language. The two dimensions were VOT and f0 for English learners of Korean, and they were spectral and temporal qualities for Korean learners of English. The additional training presented stimuli which are highly variable along the L2 irrelevant dimension (i.e., either VOT or duration) but less variable along the L2-relevant dimension (i.e., either f0 or spectral). The feedback on learners' performances was determined only by the L2-relevant dimension (see 3.1.3.2 and 6.1.3.2 for more details). This training structure was expected to make the L2-relevant dimension predictive in identifying L1 stimuli while making learners ignore variations induced by the L2 irrelevant dimension.

I named this additional training *the cue-attention switching training*. It should be noted that the cue-attention switching training does not necessarily require L2 learners to lose their sensitivity to differences along the L2 irrelevant dimension (i.e., VOT and duration) since using VOT or duration as a primary cue would not result in a critical perceptual failure to identify target nonnative contrasts.

One may question the reliability of the additional training with L1 stimuli in directing listeners' attention to a certain acoustic dimension. The present study expects the cue-attention switching training to execute the intended changes in assigning attention to acoustic cues based on recent research on speech adaptation. The findings on speech adaptation research suggest the capability of changing listeners' reliance on multiple acoustic cues by manipulating the informativeness of cues (e.g., Idemaru & Holt, 2011, 2014; Kim et al., 2020; Schertz et al., 2015). For example, Kim, Clayards, and Kong (2020) examined native English listeners' adaptation to the unfamiliar speech of English vowels that deviate from English norms in the informativeness of the primary acoustic dimension. The canonical perception of English vowels is exclusively made based on the spectral dimension of speech, and duration information has only a minor influence as a secondary acoustic cue. During the speech adaptation task, a primary acoustic dimension (i.e., spectral quality) became uninformative while the secondary acoustic dimension (i.e., vowel duration) remained informative. Listeners mostly used spectral quality to signal vowel category at baseline consisting of stimuli with two distinctive spectral steps but could successfully adapt by utilizing durational information in the speech input when spectral quality was no longer diagnostic. For this reason, the present study devised the additional training by manipulating relative informativeness of acoustic dimension through stimuli distribution and feedback.

For learners who receive the modified version of HVPT with the cue-attention switching training, the predicted stages of successful learning are as follows. English learners of Korean should (1) switch their attention to the f0 dimension in the perception of the English stop voicing contrast before starting the L2 phonetic training, (2) utilize their increased attention to the f0 dimension in the acquisition of Korean three-way stop contrast, and (3) yield a native-like cue weighting strategy, which involves predominant reliance on the f0 dimension. The predicted stages for Korean learners of English are similar: (1) learners should switch their attention to spectral differences between stimuli falling in a Korean vowel category, (2) utilize their increased attention to the spectral dimension during the L2 phonetic training to rely on that dimension, and (3) identify target English vowel contrasts as native English listeners.

## 1.6. Research questions and predictions

This study is primarily concerned with whether and to what extent L2 learners' individual differences in sensitivity to within-category acoustic cues in L1 are related to their learning of nonnative phonological contrast, targeting English learners of Korean and Korean learners of English. More specifically, this study focuses on examining whether L2 learners who are sensitive to within-category acoustic cues regarding the L2-relevant acoustic dimension have an advantage of learning target nonnative phonological contrast. Two groups of L2 learners' sensitivity to the L2-relevant acoustic dimension were measured by the VAS task and the AXB oddity task (see 3.1.3.2 and 6.1.3.2 for more detail). This study further investigates the effectiveness of the cue-attention switching training in assisting learners to identify target nonnative speech sounds according to the informative acoustic dimension. Research questions and their specific predictions are provided below.

**RQ 1. Are individual differences in the sensitivity to within-category acoustic cues for L1 category perception related to novel phonological contrast learning?**

***Prediction 1.*** Native English learners of Korean displaying gradient response patterns in the VAS task, that is, who show greater sensitivity to the f0 dimension in the perception of their native stop voicing contrast, perform better in the acquisition of the Korean three-way stop contrast than learners displaying categorical response patterns.

***Prediction 2.*** Native Korean learners of English who exhibit more spectral cue sensitivity in the perception of a Korean vowel category outperform in the acquisition of the systematic use of spectral cues to discriminate the challenging

English vowel contrasts, /i/-/ɪ/, /ɛ/-/æ/, and /ʊ/-/u/, compared to learners with lower sensitivity to spectral differences.

**RQ 2. Can the cue-attention switching, which is added to HVPT and administered with L1 speech stimuli, decrease the possible learning gap due to individual differences?**

***Prediction 3.*** The cue-attention switching training would decrease the possible learning gap due to the learners' differences in their sensitivities to sub-phonemic acoustic details along the L2-relevant dimension (i.e., f0 or spectral) by intentionally locating learners' attention to the L2-relevant acoustic dimension.

# CHAPTER 2. EXPERIMENT 1

Experiment 1 focuses on the research question 1 with the prediction 1 targeting native English learners of Korean. The VAS task measures individual participants' within-category sensitivity to the f0 dimension in the perception of English stop voicing contrast (/d/-/t/). Participants complete a total of five-day experimental phase with a three-day HVTP to learn a Korean three-way laryngeal contrast in stops (/p'/-/p/-/$p^h$/).

As a reminder, the prediction 1 states that native English learners of Korean displaying gradient response patterns in the VAS task, that is, who show greater sensitivity to the f0 dimension in the perception of their native stop voicing contrast, perform better in the acquisition of the Korean three-way stop contrast than learners displaying categorical response patterns.

## 2.1. Methods

### *2.1.1. Participants*

Twenty-four adult native speakers of American English (fourteen female, ten male; mean age of 20.1 years, range of 18 – 24 years) participated in experiment 1 and completed the entire experiment phase including HVPT. The participants were undergraduate students at the University of Wisconsin-Milwaukee who were born and raised in the Midwest United States. Participants received extra course credit for participating and none of them reported having a history of hearing or speech impairment. Although some participants had some exposure to other languages as determined by a language background questionnaire, none of them considered themselves as bilingual. All participants did not have prior experience of Korean language learning either formally or informally. This indicates that the participants in this study were naïve

English learners of Korean at the time of participation. Average demographic and language background data on all participants in experiment 1 are presented in Table 1.

Table 1. Demographic and language background information of participants in experiment 1.

| Variables | Mean | *SD* |
|---|---|---|
| Sex | 14 F; 10 M | |
| Age (yrs. old) | 20.1 | 1.5 |
| Birthplace (state/country) | WI 17 | |
| | IL 7 | |
| Lived out of Midwest (yrs.) | 0.0 | 0.2 |
| Location (place, yrs.) | AZ 1 | |
| Father Birthplace (state) | WI 14 | |
| | IL 6 | |
| | NY 1 | |
| | MN 1 | |
| | PA 1 | |
| Father L1 (language, #) | English 24 | |
| Mother Birthplace (state) | WI 14 | |
| | IL 4 | |
| | MI 1 | |
| | MN 1 | |
| | NJ 1 | |
| | MA 1 | |
| | CA 1 | |
| | TX 1 | |
| Mother L1 (language, #) | English 24 | |
| L2 (language, #) | Spanish 12 | |
| | ASL 3 | |
| | Japanese 2 | |
| | French 1 | |
| | Italian 1 | |
| | German 1 | |
| (1 very poor to 7 very good) | | |
| L2 Speaking | 3.7 | 1.6 |
| L2 Understanding | 4.5 | 1.4 |
| L2 Reading | 4.1 | 1.7 |
| L2 Writing | 3.2 | 1.8 |

### *2.1.2. Stimuli*

#### 2.1.2.1. Stimuli for the Visual Analogue Scaling (VAS) task

The stimuli for the VAS task were pseudo-synthetic consonant-vowel (CV) syllables constructed to make a continuum from English /da/ to /ta/. They were created by recording natural productions of stop-initial CVC English words (*tot* and *dot*) in the context of carrier sentence ("I say ___ again."). One male native talker of English produced the English words, *tot* and *dot* multiple times. The talker was recorded in a sound-attenuated room in the Phonetics lab at the University of Wisconsin-Milwaukee with a Shure SM-10A microphone using a sampling rate of 44.1 kHz. For the stimulus manipulation in VOT and f0, the baseline token was first generated by selecting only initial voiceless stop /tɑ/ portion with a 250ms steady-state vowel /ɑ/ from one of *tot* tokens. Then, the baseline token was synthesized using Praat (Boersma, 2001) to create a series of stops covarying in VOT and f0, resulting in a set of stimuli spanning a two-dimensional acoustic space: seven steps of VOT by five steps of f0. The purpose of using the single baseline token /tɑ/ for the stimulus manipulation was to control unintended variables such as intensity or phonation type.

The VOT duration of the baseline token was manipulated to span the range from -40 ms to 40 ms in seven steps (-40ms, -26ms, -13ms, 0ms, 13ms, 26ms, and 40ms). The VOT steps fully encompassed the VOT range of the talker. Following (Lee et al., 2013), the shorter and longer VOTs were generated by compressing and expanding the VOT duration of the baseline token. The VOT manipulation procedure was as follows: first, we defined VOT as the interval between the release of the stop and the onset of voicing (i.e., the onset of periodicity in the acoustic waveform) and extracted the *Duration Tier* for this VOT portion using the "To

Manipulation…” function in Praat. Then, the original VOT value was changed into the desired outcome ratio value. For instance, to lengthen VOT from 13 ms to 40 ms, the ratio value 3.08 was input to the duration tier. At last, we generated a new token with modified VOT value by replacing the original duration tier with the new tier.

As for the f0 manipulation, the f0 range was fixed from 98 Hz to 130 Hz in five steps (98Hz, 106Hz, 114Hz, 122Hz, and 130Hz) (Kong & Edwards, 2016). The f0 manipulation procedure was as follows. First, we extracted the Pitch Tiers of seven VOT manipulate stimuli using the “To manipulation…” function in Praat. And then, we lowered or raised the f0 contours of the 250ms steady-state vowel /ɑ/ proportion by moving all pitch points in the pitch tiers, so that each manipulated pitch contour had one consistent pitch value. Lastly, the original f0 tier of the VOT manipulated stimuli was replaced with one of five manipulated f0 tiers. Thus, a total of 35 stimuli (7 steps of VOT X 5 steps of f0) were created as shown in Figure 6.

#### 2.1.2.2. Stimuli for the HVPT

We created the stimuli for HVPT which were pseudo-synthetic CV syllables spanning the Korean three-way bilabial stop contrasts with the vowel /a/, fortis (/p’a/), lenis (/pa/), and aspirated (/pʰa/). The procedure for the stimuli generation is described in the following paragraphs.

First, four adult female native Korean talkers speaking the Seoul dialect (mean age of 30.4 years, range of 24 – 34 years) recorded three CVCV pseudo words, /p’ata/, /pata/, and /pʰata/, at their normal speech rate. All the talkers were recorded in a sound-attenuated room with a Shure SM-10A microphone using a sampling rate of 44.1 kHz. Each recording of words was analyzed using Praat, and the VOT of the initial bilabial stop and the f0 at the start of the vowel periodic portion were measured. In line with the previous research (e.g., Kim, 2004), all four talkers produced the longest VOT for the aspirated stop and the shortest VOT for the fortis stop. In terms of f0, all talkers presented the highest values in /pʰata/ tokens, intermediate in /p'ata/ tokens, and lowest in /pata/ tokens. These acoustic measurements were conducted to confirm that all talkers showed a similar pattern in the production of Korean three-way bilabial stop contrast. Table 2 summarizes the acoustic analysis for each talker's production of Korean three-way bilabial stops.

Second, each talker's lenis stop token, /pata/, was chosen as a baseline token (Schertz et al., 2015). The initial /pa/ portion with a 250 ms steady-state vowel /a/ was excerpted and modified in Praat. Similar to the manipulation procedure for VAS stimuli, "To Manipulation…" function was used to generate a set of training stimuli covarying in seven steps in VOT and five steps in f0. The same VOT and f0 manipulation procedures for the VAS stimuli were utilized. It should be noted that the ranges of VOT and f0 values for the stimuli manipulation were determined based on each talkers' natural production pattern of Korean three-way bilabial stop contrast. Considering that none of the Korean talkers had the identical shortest/longest VOT and lowest/highest f0 values, we chose to set different f0 and VOT ranges for each talker instead of applying the same ranges for all four talkers. This decision was to ensure that the manipulated stimuli represent the production patterns of talkers without exaggerating the natural differences between fortis, lenis, and aspirated stops. The VOT range was set from 0 ms to each talker's longest VOT value (i.e., VOT of aspirated stop). The f0 range was set from each talker's lowest f0 value (i.e., f0 of lenis stop) to the highest f0 value of the same talker (i.e., f0 of aspirated stop). Once the ranges of VOT and f0 for each talker were calculated, the baseline /pa/ tokens were manipulated as follows. The VOT value of each talker's baseline token /pa/ was lengthened up to the longest VOT value of that speaker and shortened to 0 ms by the same intervals to create seven stimuli with different VOT values. Therefore, a total of seven VOT manipulated stimuli was generated for each talker. As for the next step, the f0 value of each of seven VOT manipulated stimuli was increased up to the highest f0 value of that speaker by the same interval to generate five f0 manipulated stimuli. Since the VOT manipulated stimuli were generated with the baseline token (lenis) with the lowest f0 value, we did not need to lower the f0 of the stimuli. We would like to highlight once again that the ranges of the VOT and f0 differed for each talker as well as the intervals between the manipulated stimuli.

Through the manipulation process, a set of 35 stimuli for each Korean talker was created, encompassing the full production range of VOT and f0 of that talker. As a result, there were four sets of stimuli consisting of 35 stimuli for each talker. Table 3 demonstrates each talker's VOT and f0 intervals, and Figure 6 shows the layout of stimuli with assigned numbers from 1 to 35.

Upon completion of stimuli manipulation, we conducted a forced-choice perception task to examine how native Korean listeners perceived the sets of manipulated stimuli. This task had two purposes: one was to examine Korean listeners' perceptual patterns in their perception of Korean stops as a function of VOT and f0, and the other was to determine the answers for each set of stimuli to give learners trial-by-trial feedback during the HVPT sessions. A total of 24 native listeners of Seoul Korean were presented with a forced-choice task in which they heard manipulated sets of stimuli encompassing the full range of the two-dimensional acoustic space (i.e., VOT and f0) as described above. All listeners completed the task at the Phonetics Laboratory at Kangwon National University in South Korea. For each stimulus, the listeners were asked to choose which of three sounds (i.e., fortis, lenis, and aspirated) best represent what they heard by clicking one of the buttons on the computer screen. The perception task consisted of four blocks and each block randomly presented a set of stimuli from only one talker with three repetitions of the 35 stimuli. The perception task was administered on Praat and took about 30 minutes (35 stimuli × 3 repetitions × 4 blocks = 420 trials). We analyzed listeners' responses for each stimulus and calculated the proportions of responses for each category as follows: the number of responses for one of choices (i.e., fortis, lenis, or aspirated) was divided by the total number of responses and multiplied by 100 to transform the numbers to percentage. The 75 percentage of agreement in responses was used as a cutoff to determine the answer for each manipulated stimulus.

The force-choice perception task results showed native Korean listeners' overall response patterns and their use of VOT and f0 dimensions in the identification of synthesized stimuli. The mixed-effect logistic regression analysis was conducted with listeners' response data to each talker's set of stimuli. The dependent variables were the binary responses of "바" /pa/ and "파" /pʰa/ (lax-aspirated) excluding the responses of "빠" /p'a/. The fixed effect variables were VOT and f0 steps of each stimulus as continuous variables, whose units were scaled to make beta-coefficients comparable between two acoustic dimensions. The results showed that listeners used f0 dimension as a primary cue and VOT dimension as a secondary cue, as shown by a larger f0 coefficient for all talkers (Talker 1: $\beta_{VOT}$: 1.75, $\beta_{f0}$: 4.31; Talker 2: $\beta_{VOT}$: 1.07, $\beta_{f0}$: 3.11; Talker 3: $\beta_{VOT}$: 2.46, $\beta_{f0}$: 3.72; Talker 4: $\beta_{VOT}$: 1.25, $\beta_{f0}$: 2.33). Figure 5 shows answers for each talker's set of stimuli determined by the 75% agreement criteria.

The stimuli from three talkers (Talker 1, 2, & 3) were designated for the HVPT, and those from the one remaining talker were used for the new talker generalization test. This test was taken by learners after completing the entire HVPT sessions.

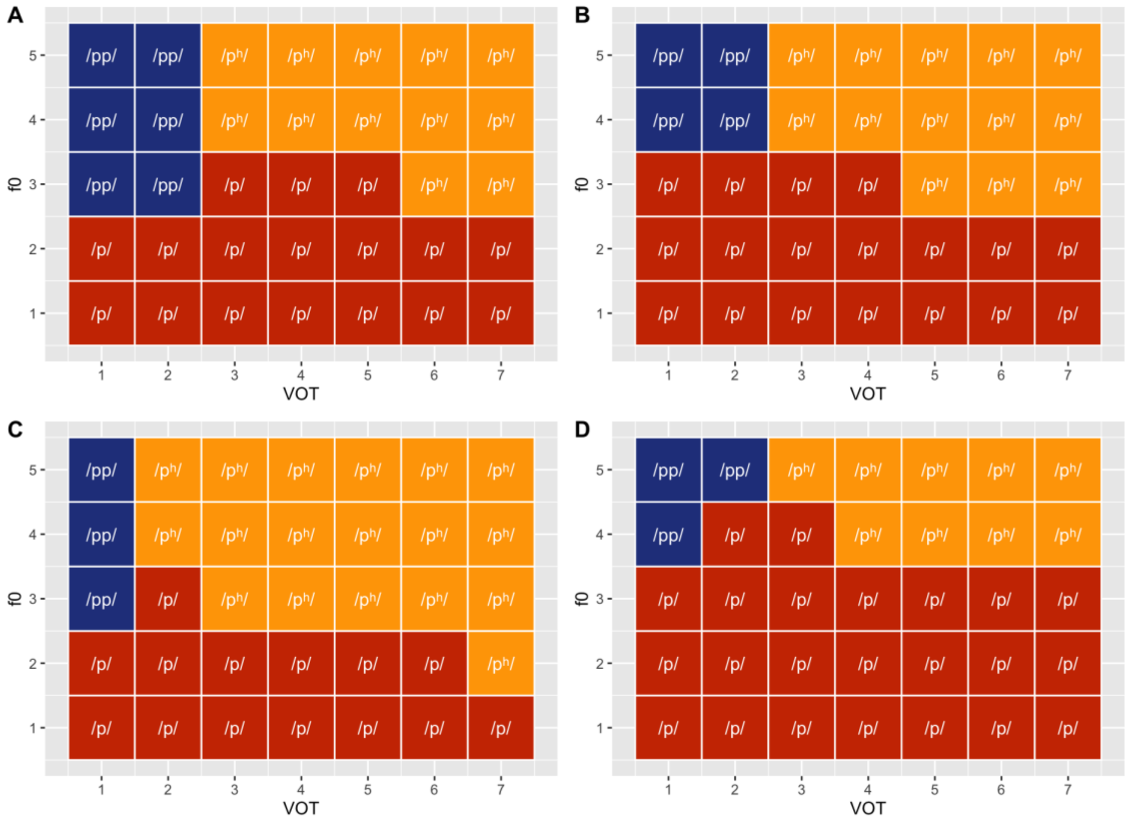


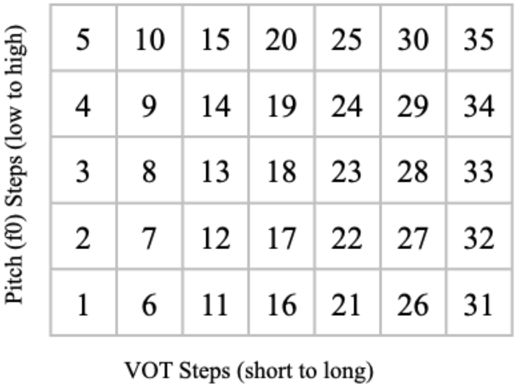
Figure 5. Answers for each set of stimuli for Talker 1 (A), Talker 2 (B), Talker 3 (C), and Talker (4). Blue cells indicate fortis (/p'/), red cells indicate lenis (/p/), and yellow cells indicate aspirated (/pʰ/) stops.

Table 2. Summary of the acoustic analysis of each native Korean talker.

| Talkers | VOT (ms) | | | f0 at vowel onset (Hz) | | |
|---|---|---|---|---|---|---|
| | /p'a/ | /pa/ | /pʰa/ | /p'a/ | /pa/ | /pʰa/ |
| Talker 1 | 7 | 65 | 80 | 211 | 189 | 239 |
| Talker 2 | 17 | 72 | 86 | 203 | 177 | 221 |
| Talker 3 | 19 | 84 | 91 | 234 | 199 | 260 |
| Talker 4 | 20 | 71 | 85 | 254 | 211 | 261 |

Table 3. VOT and f0 intervals of each talker used in the process of stimuli manipulation.

| Talkers | VOT interval (ms) | f0 interval (Hz) |
|---|---|---|
| Talker 1 | 13.28 | 12.35 |
| Talker 2 | 14.31 | 11.20 |
| Talker 3 | 15.09 | 15.32 |
| Talker 4 | 14.11 | 12.49 |



| Pitch (f0) Steps (low to high) | | | | | | | |
|---|---|---|---|---|---|---|---|
| | 5 | 10 | 15 | 20 | 25 | 30 | 35 |
| | 4 | 9 | 14 | 19 | 24 | 29 | 34 |
| | 3 | 8 | 13 | 18 | 23 | 28 | 33 |
| | 2 | 7 | 12 | 17 | 22 | 27 | 32 |
| | 1 | 6 | 11 | 16 | 21 | 26 | 31 |

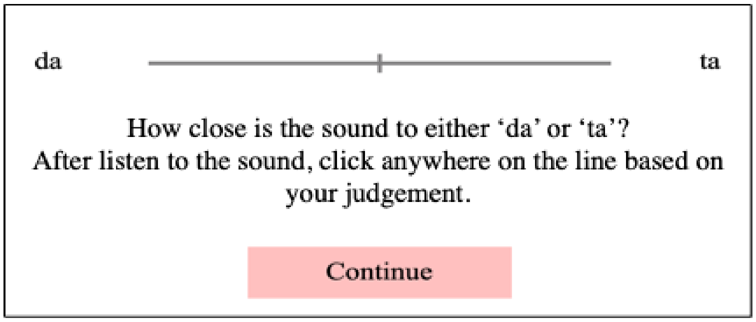
VOT Steps (short to long)

Figure 6. Stimuli from 1 to 35 manipulated in VOT and f0.

### *2.1.3. Procedure*

#### 2.1.3.1. The VAS task

The VAS task asked participants (i.e., naïve native English "learners" of Korean) to make a judgment after hearing each manipulated stimulus on how close the stimulus sounds to either English /da/ or /ta/. Participants heard 105 trials of the 35 manipulate stimuli (see section 2.1.3.2.1) with three repetitions in random order, using Praat. During the task, a line was displayed on the computer screen as in Figure 7. One end of the line was labeled as 'da', and the other end was labeled as 'ta'. Participants were instructed to click anywhere on the line that corresponded with the perception of proximity to ‘da’ or ‘ta’. The VAS task was completed in approximately 8 minutes.



How close is the sound to either ‘da’ or ‘ta’?
After listen to the sound, click anywhere on the line based on your judgement.

Figure 7. Illustration of visual analogue scaling task.

### 2.1.3.2. HVPT

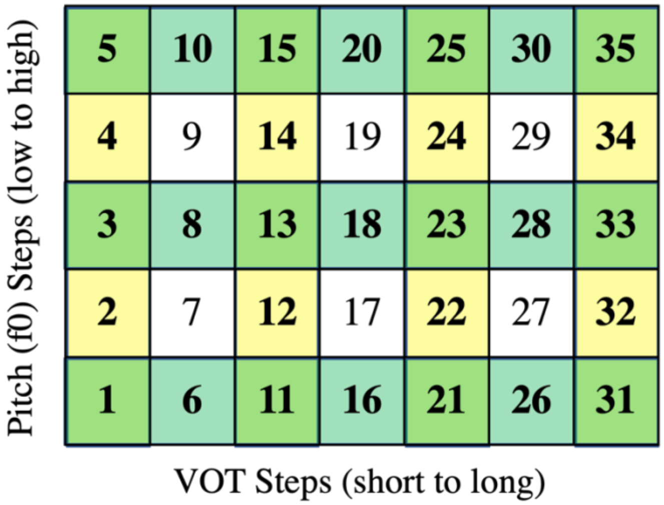
Experiment 1 was managed in the span of five days with three-day HVPT sessions. Each day took about 30 minutes to complete. Once participants started their first training session, their consistent participation until the last day of experiment 1 (i.e., Day 5) was required, indicating that all participants completed the sessions from *Day 2* to *Day 5* without a break. This regulation was to precisely evaluate participants’ performances during the HVPT by minimizing any discrepancy between participants.

Three Korean CV pseudowords containing either fortis, lenis, or aspirated Korean bilabial stop (i.e., /p’a/, /pa/, and /p$^h$a/) were trained. The training sessions were administered through Paradigm software (Paradigm Stimulus Presentation, 2007). All testing and training were carried out at the Phonetics Laboratory at University of Wisconsin-Milwaukee. Participants sat at a desk equipped with a desktop PC or laptop computer, and stimuli were presented over headphones (SONY, MDR-V6) connected to a PC at a comfortable listening level. No more than two participants were allowed to participate in the experiment simultaneously to reduce any interruption or distraction. The following paragraphs explain how each day of experiment 1 was managed and which tasks and tests were given to participants.

The HVPT design employed a pretest-post-test procedure closely modeled after the methods used by previous studies (e.g., Logan, Lively, & Pisoni, 1991; Strange & Dittmann, 1984). In this design, the effects of HVPT were assessed by comparing performance on a pretest and a post-test administered before and after a three-day training period. In addition, we assessed participants' improvements after every training session through everyday identification (ID) tests. Generalization to a new voice and a new Korean three-way stop contrast with a different place of articulation was examined after training.

### *2.1.3.2.1. Day 1: the VAS task and AXB pretest*

Participants started *Day 1* with the VAS task. Upon the completion of the VAS task, the AXB pre-test was given to participants with stimuli from one of the training talkers (Talker 1). On each trial of the pretest, participants heard three stimuli in a row and had to choose if the middle token (*X*) was the same token as the first token (*A*) or the last token (*B*). The possible combinations of *A* and *B* tokens were *AAB*, *BAA*, *BBA*, and *ABB*. *A* and *B* tokens differed from each other in terms of either VOT or f0 value. For example, if the *A* token was stimulus number 1, the *B* token could be either stimulus 11, which had the same f0 step but different VOT steps, or stimulus 2, which had the same VOT step but different f0 step. If the B token was 11, the four stimulus combinations were, *A* (1) *A* (1) *B* (11), *B* (11) *A* (1) *A* (1), *B* (11) *B* (11) *A* (1), and *A* (1) *B* (11) *B* (11). The difference in VOT steps between stimuli was two steps while f0 step difference was one step (see Figure 8). For stimulus pairs with *B* tokens having different VOT steps from *A* tokens, stimuli from first, third, or fifth f0 steps were used (i.e., stimuli in green panels in Figure 8). For stimulus pairs with *B* tokens having different f0 steps from *A* tokens, stimuli from first, third, fifth, or seventh VOT steps were used (i.e., stimuli in yellow panels in Figure 8). This decision was to prevent the AXB test from being too lengthy. A total of 124 trials ((5 VOT pairs × 3 f0 steps × 4 combinations) + (4 f0 pairs × 4 VOT steps × 4 combinations) = 124)) were presented in random order with a short break and the procedure required approximately 12 minutes to complete.



|  |  |  |  |  |  |  |
|---|---|---|---|---|---|---|
| **5** | **10** | **15** | **20** | **25** | **30** | **35** |
| **4** | 9 | **14** | 19 | **24** | 29 | **34** |
| **3** | **8** | **13** | **18** | **23** | **28** | **33** |
| **2** | 7 | **12** | 17 | **22** | 27 | **32** |
| **1** | **6** | **11** | **16** | **21** | **26** | **31** |

Figure 8. Stimuli pairs for the AXB pre and post-tests. Stimuli in bold were used in the AXB tests.

### *2.1.3.2.2. Day 2 to Day 4: 3-Day of HVPT and the new talker generalization ID test*

Participants received three-day computer-based audio training with trial-by-trial feedback from *Day 2* to *Day 4*. Every training session started with a daily familiarization phase, which auditorily presented three target training words with corresponding photographs. Since participants were naïve learners of Korean without any knowledge of Korean orthography, pictures were associated with training words. Six trials were presented during the familiarization phase (3 target words × Talker 1 × 2 repetitions = 6 trials). Each training session lasted approximately 30 minutes per day.

After the familiarization phase, the training phase in the form of an identification (ID) task was initiated. On each trial, an auditory stimulus was played first, and three photographs were presented on a computer screen as response options. After clicking one option among the three, trial-by-trial feedback was followed. If the participant responded correctly, a chime sounded with a photograph of a smiley face and the next trial was automatically presented. If the participant chose a wrong picture, the stimulus with the correct picture was repeated once more. The training session was separated into three blocks to give participants short breaks. Furthermore, stimuli from only a single talker were presented in each block rather than having all three of talkers' stimuli mixed together to enhance the effectiveness of HVPT by reducing the degree of variability from one stimulus to the next (Perrachione, Lee, Ha, & Wong, 2011). Each talker had a different number of stimuli falling into three stop categories based on the results of the force-choice perception task by Korean listeners (see Figure 5). To expose participants to an even number of stimuli for each category, some stimuli were repeated. For example, talker 1 had 6 fortis, 17 lenis, and 12 aspirated stop stimuli. Thus, randomly selected fortis and aspirated stimuli were repeated to have same number as lenis stimuli which has the greatest number of stimuli, resulting in 51 training stimuli for Talker 1 (17 stimuli × 3 categories = 51 stimuli). A total of 159 trials were presented during the training sessions (51 for Talker 1 + 54 for Talker 2 + 54 for Talker 3 = 159 trials). Since each talker had a different number of stimuli falling into three stop categories, some stimuli were repeated to present an even number of stimuli for each category. Figure 9 shows the overall demonstration of the training phase.

Each training session ended with an everyday ID test. The procedure was identical to the training phase except no feedback was given. The set of stimuli from Talker 1 was used for this test, yielding a total of 51 trials. The everyday ID test was conducted a total of three times during experiment 1.

After the last training session on *Day 4*, participants completed one test, the new talker generalization test. A total of 66 stimuli produced by a novel Talker 4 (i.e., a talker not used in either the pretest and the post-test phase or the training phase) was presented with two repetitions. The overall process was the same as the everyday ID test.

#### *2.1.3.2.3. Day 5: AXB post-test and the new consonant generalization ABX test*

After completing the entire HVPT, participants were tested again to assess the degree of improvements (i.e., the post-test) and the degree of generalization of learning to novel stimuli (i.e., generalization tests). The post-test, which was identical to the pre-test, was conducted. The first test of generalization test was a new talker generalization test, which was conducted on Day 4. A second generalization test, the new Korean three-way stop contrast generalization test, was conducted on Day 5. The test stimuli consisted of new words that the participants had not heard before. A Korean three-way stop contrast, fortis (/k’/), lenis (/k/), and aspirated (/kʰ/), was used to test whether participants generalized what they learned to the stop contrast with a different place of articulation (velar), which showed similar acoustic characteristics in terms of VOT and f0. Talker 1 produced /k’ata/, /kata/, and /kʰata/ and the talker displayed the longest VOT and the highest f0 for the aspirated (/kʰ/), the shortest VOT for the fortis (/k’/), and the lowest f0 for the lenis (/k/). The baseline token /ka/ was manipulated to generate 35 stimuli in the same way for producing stimuli for training (see section 2.1.2.2). This test had a format of AXB discrimination test, which was the same as the pre-test and the post-test. A total of 124 trials were presented with two repetitions. Table 4 demonstrates the timeline of 5-day training sessions.

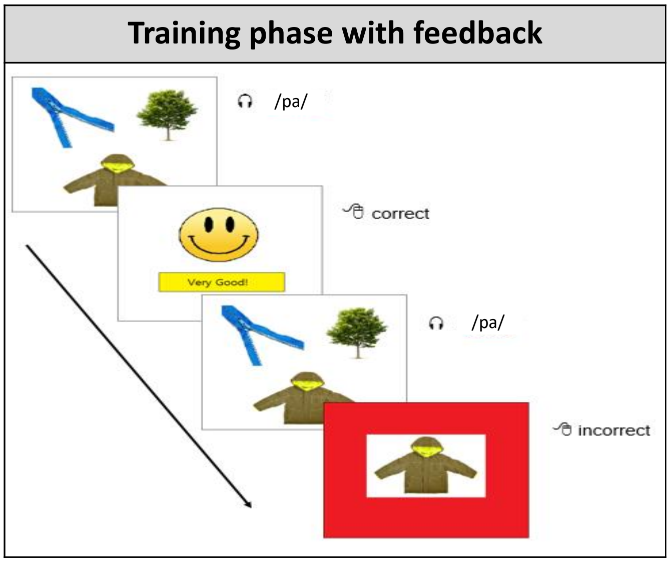


Figure 9. Description of the overall training phase.

Table 4. Timeline of the five-day experiment 1 with three-day HVPT sessions.

| | Participation Day | | | | |
|---|---|---|---|---|---|
| **Day** | Day 1 | Day 2 | Day 3 | Day 4 | Day 5 |
| **Tasks** | • VAS Task<br>• Pre AXB Test (n = 124)<br>• Language Questionnaire | • Daily Familiarization Phase<br>• HVPT Session (n= 159)<br>• Everyday Identification (ID) Test (n = 51) | | | • Post AXB Test<br>• New Consonant Generalization Test (n = 124) |
| | | | | + New Talker Generalization Test (n = 66) | |
| **Time** | About 30 minutes | | | | |
| **Location** | Laboratory | | | | |

### *2.1.4. Analysis*

#### 2.1.4.1. The VAS task

The analysis of the VAS task closely followed previous work (Kong, 2019; Kong & Edwards, 2011, 2016). The click location for each stimulus was converted to a VAS rating scale (0% – 100%). The click locations, closer to 0%, indicated more /da/-like perceptual responses, while the click locations, closer to 100%, indicated more /ta/-like perceptual responses. To quantify each participant's degree of gradiency of responses, we examined distributions of click location along the VAS rating scale (i.e., histograms) and quantified them using polynomial regression models. Polynomial regression models were made for each participant, of which the coefficients of the quadratic regression curves overlaid on the histograms of the VAS responses were used as a numerical index of gradiency for each participant. In other words, the coefficients served to exhibit spread or steepness of the increases of click locations toward the edges of the VAS rating scale. As in Kong and Edwards (2011, 2016), we expected that some listeners would result in bipolar peaks at the two ends of the line (categorical responses), while response patterns for other listeners would result in click distributions that were more evenly spread across the entire line (gradient responses). Therefore, larger coefficients from the curves were interpreted as a more categorical response pattern, while smaller coefficients were interpreted as a more gradient response pattern.

#### 2.1.4.2. HVPT

The percentages of correct answers were collected for all types of tests for each participant and used for the analysis. The tests included the AXB pretest and post-test, everyday

ID tests (Day 1 test, Day 2 test, and Day 3 test, henceforth), the new talker generalization test, and the new consonant generalization test.

### *2.1.5. Results*

#### 2.1.5.1. The VAS task

Figure 10 shows VAS responses averaged across all participants. Overall, participants used the entire line when making responses, although they responded more using the two endpoints of the line. Figure 11 and Figure 12 show the distributions of click locations of representative participants who made more categorical responses and more gradient responses. The curve fits (i.e., red lines) from the polynomial regression models were overlaid on top of the click distributions. The dotted lines in graphs represent the midpoints of the click locations. The 12 plots in Figure 11 and Figure 12 show that there were apparent individual differences across participants in how gradient their responses were. Some participants' VAS responses (e.g., Participants C1_1, C1_2, C1_4, C1_5, C1_7, C1_10) were heavily clustered around the two endpoints of the VAS scale. In contrast, some participants (e.g., Participants G1_2, G1_3, G1_4, G1_6, G1_9, G1_10) judged the stimuli more gradiently, indicating that VAS responses were more distributed across the entire VAS rating scale and were relatively less clustered around the two endpoints of the scale.

For the group analysis, we divided participants into two groups based on their response patterns of the VAS task and named those groups the Categorical and Gradient groups. More specifically, the criterion of group division was the proportion of the entire responses falling into the first and the last 20% of the VAS rating scale. Participants whose proportion of responses falling into these criteria exceeded 70% (i.e., preferably clicking the edges of the scale) were assigned to the Categorical group. Therefore, Participants C1_1, C1_2, C1_4, C1_5, C1_7, C1_10 in Figure 11 belonged to the Categorical group. In contrast, participants who did not mainly click the endpoints of the scale more than 70% were assigned to the Gradient group, as shown by Participants G1_2, G1_3, G1_4, G1_6, G1_9, and G1_10 in Figure 12. A total of 12 participants were assigned to each group. Figure 13 shows histograms of the averaged VAS responses for each group.

As mentioned earlier, the gradiency of VAS responses for each individual was quantified as the coefficients of the quadratic regression curves overlaid on the histograms of VAS responses. As illustrated in Figure 11 and Figure 12, the slopes were relatively steep concave curves for participants in the categorical group, while participants in the gradient group presented relatively shallow concave curves. The independent *t*-test showed that the quantified gradiency by the categorical group ($M = 0.00204912$, $SD = 0.0005326$) was significantly higher than the gradient group ($M = 0.000354273$, $SD = 0.000256881$) ($p < .001$).

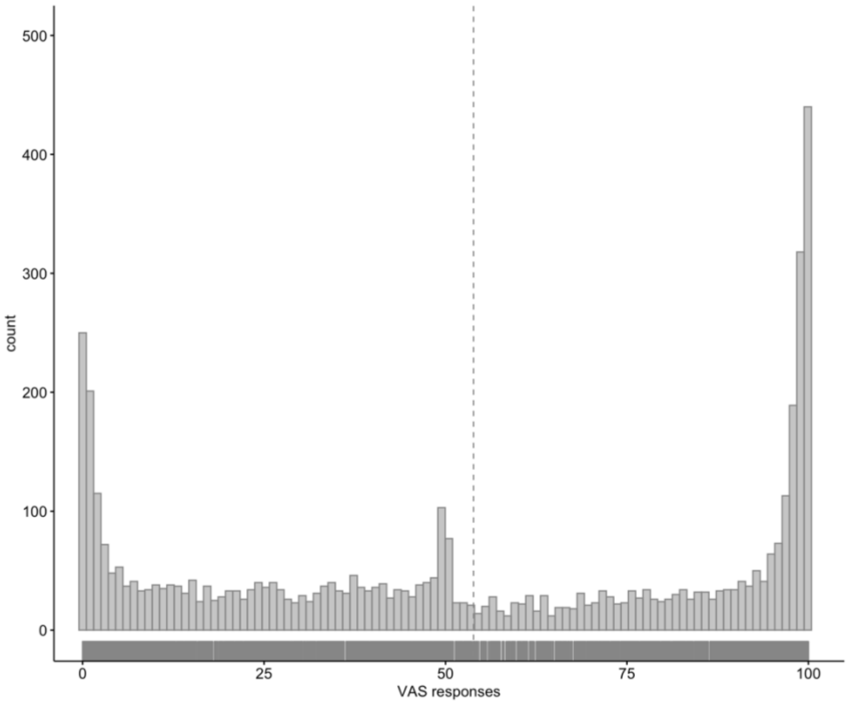


Figure 10. Distributions of the VAS task click locations averaged across all participants. The dotted line represents the midpoints of the click locations for all participants.

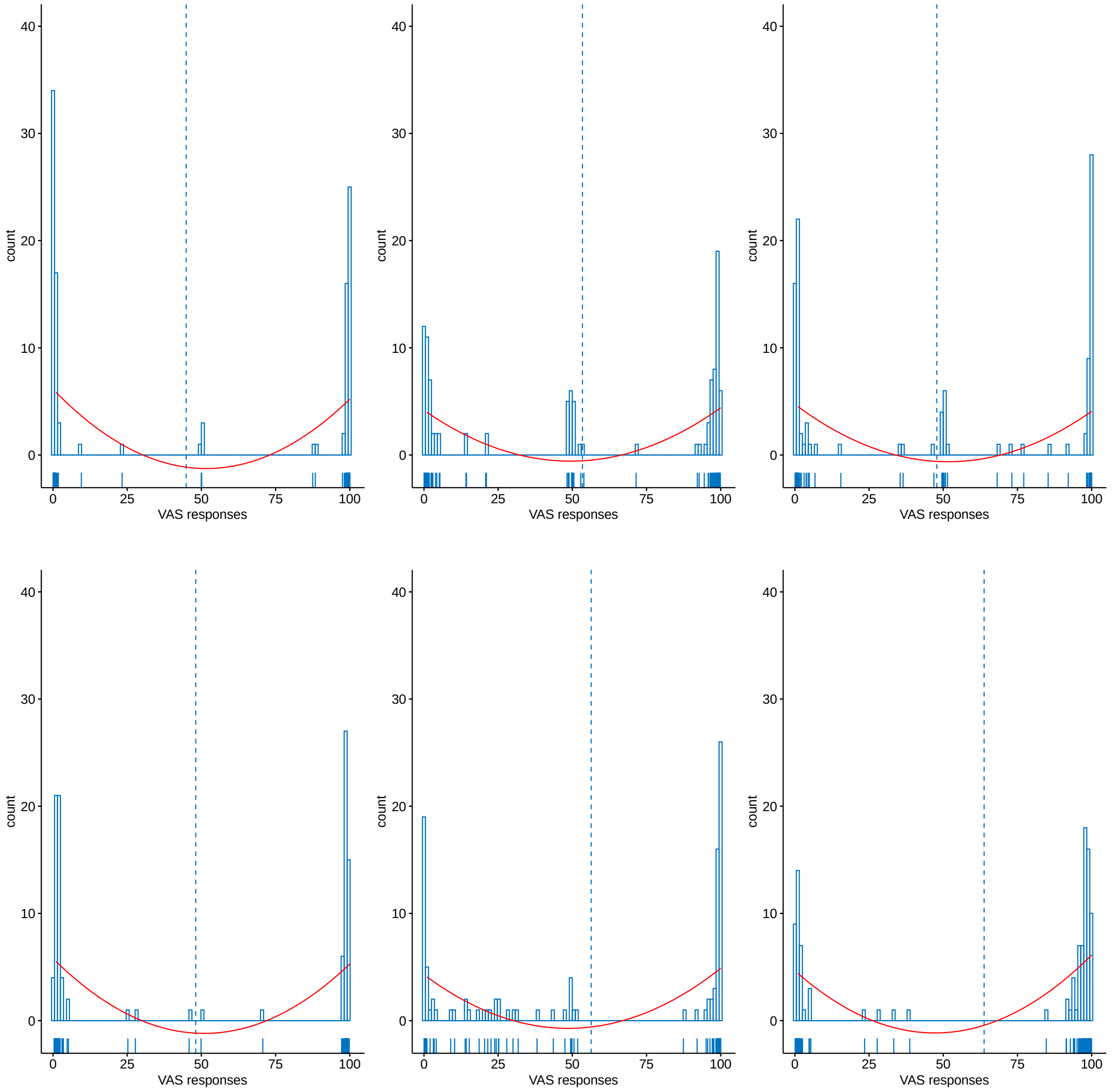


Figure 11. Distributions of the VAS task click locations of represented participants (C1_1, C1_2, C1_4, C1_5, C1_7, C1_10) who showed categorical responses. The curve fits (i.e., red lines) from the polynomial regression models were overlaid on the top of the click distributions. The dotted lines represent the midpoints of the click locations.

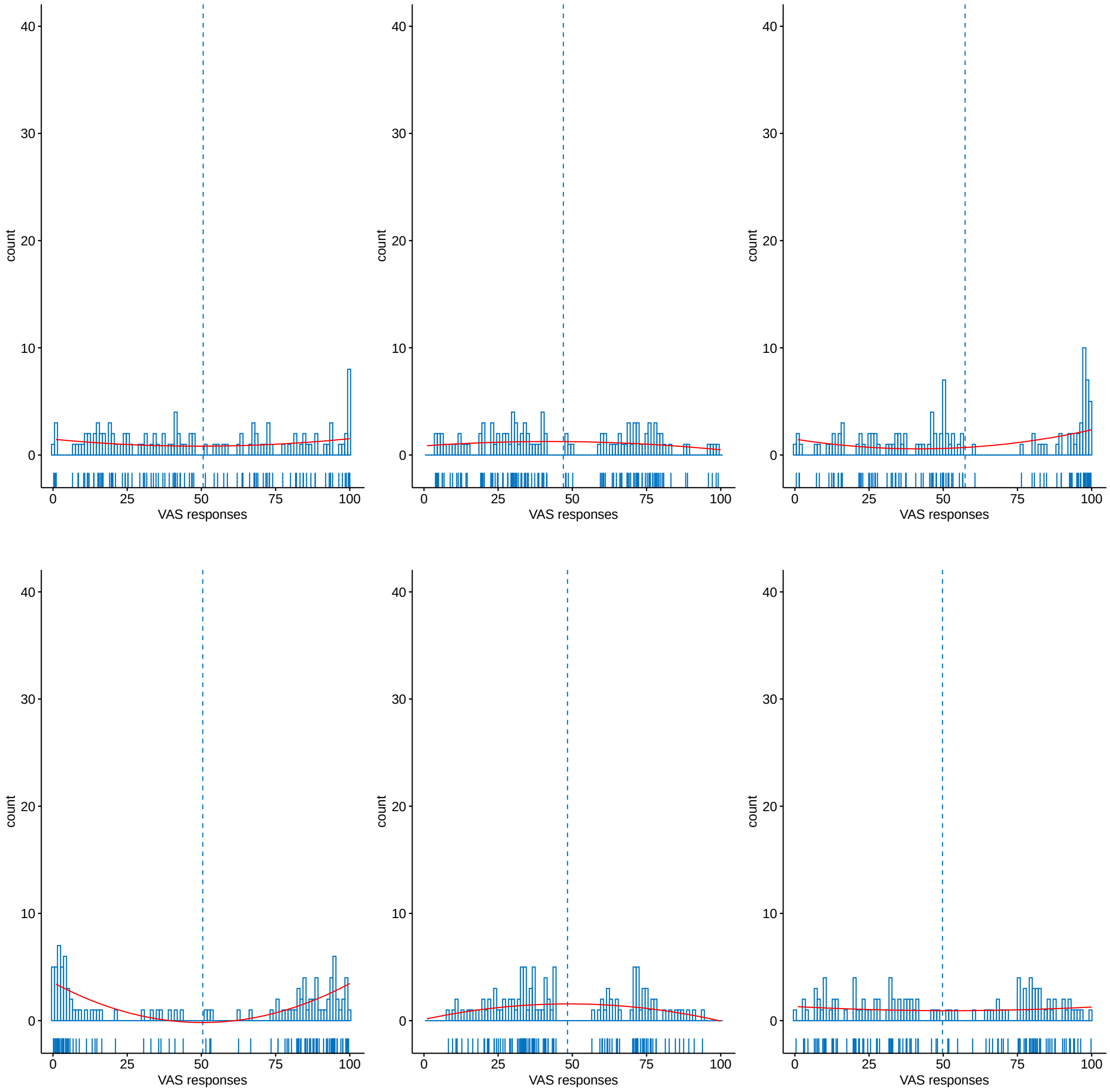


Figure 12. Distributions of the VAS task click locations of represented participants (G1_2, G1_3, G1_4, G1_6, G1_9, G1_10) who showed gradient responses. The curve fits (i.e., red lines) from the polynomial regression models were overlaid on the top of the click distributions. The dotted lines represent the midpoints of the click locations.

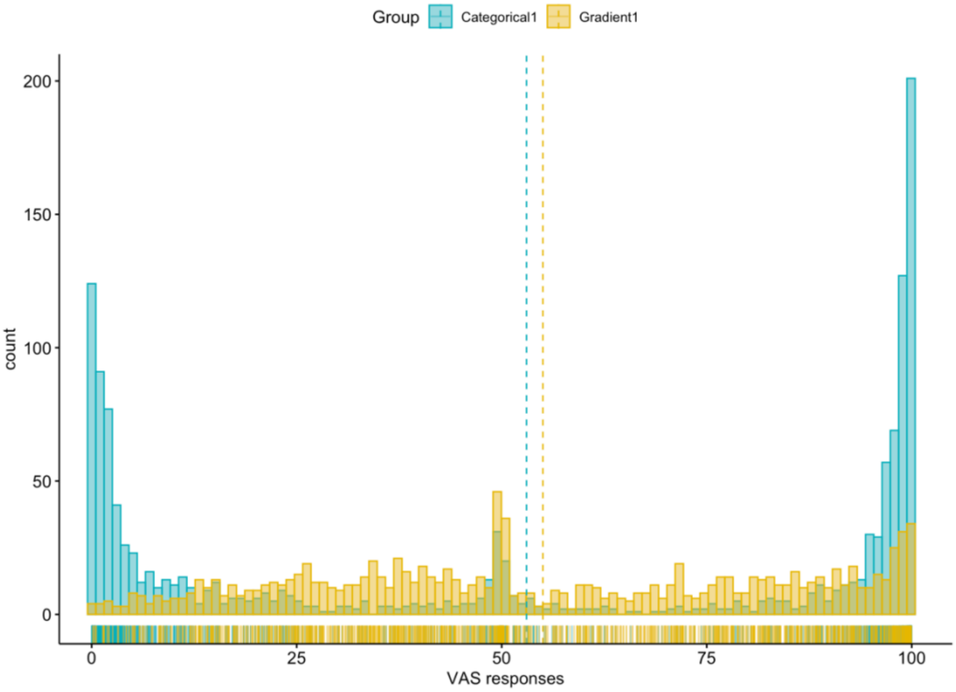


Figure 13. Distributions of the VAS task click responses averaged by groups (Categorical 1 vs. Gradient 1). The dotted lines represent the midpoints of the click locations for each group.

#### 2.1.5.2. Results of HVPT by groups (Categorical vs. Gradient)

To observe how the Categorical and Gradient groups were different in their performances, participants' performances on tests (i.e., test scores in the percentage of correct) were analyzed using linear mixed-effects regression models using the *lmer* function from the *lme4* package (version 1.1-27) (Bates, Mächler, Bolker, & Walker, 2015) in *R* (R Core Team, 2020). Two mixed-effects regression models were built, one for the pretest, post-test, and new consonant generalization tests results and the other for everyday ID tests and the new talker generalization test results. Building two separate models was motivated by the fact that the tests for the first model were in the discrimination test type while the tests for the second model were in the identification test type. Test scores were submitted to the models. The models included participant, item-level predictors (i.e., fixed effects), and their 2-way interactions. The participant-level predictor was Group (Categorical vs. Gradient), which was centered (-0.5 and 0.5) in order that main effects were evaluated as the average effects over all levels of Group (rather than at a specified reference level). The item level predictor included Test (i.e., test scores). Test was dummy coded using *lizContrasts* and *lizContrasts4* functions (Dong, Clayards, Brown, & Wonnacott, 2019; Wonnacott, Brown, & Nation, 2017) in R, which is a way to compare each level of a variable to the reference level. In the first model, Test had two levels, coded to Pre VERSUS Post and Pre VERSUS New Con with pretest scores (i.e., Pre) as the reference level. In the second model, Test had three levels, coded to Day1 VERSUS Day2, Day1 VERSUS Day3, and Day1 VERSUS New Talker with Day 1 ID test scores (i.e., Day1) as the reference level. The motivation of using *Dummy* contrasts is to examine learners' improvement at each time point compared to the first day of training. It should be noted that Kim et al. (2018) used *Helmert* contrast, which compares participants' performances at each time point with the mean of performances at previous time points. However, compared to Kim et al. (2018) and other previous training studies (e.g., Bradlow et al., 1997), the current study provided relatively short training (3-day training). Therefore, comparing learners' subsequent everyday ID test scores with their Day 1 ID test scores was expected to show learners' improvements more clearly. Both models included random intercepts for participants, along with random slopes for participants for Test variable. The *lme4* package provides *p*-values automatically for logistic mixed-effects models but not for linear mixed-effects models. For two models with a continuous outcome variable (i.e., test scores), *p*-values were calculated using the *lmerTest* package using Satterthwaite's degrees of freedom method. Following Giannakopoulou et al. (2017), all of the models reported converged with bound optimization by quadratic approximation (BOBYQA optimization; Powell, 2009).

### *2.1.5.2.1. Pre, post, and new consonant generalization tests*

As shown in Figure 14, the pretest, post-test, and new consonant generalization test performance differed between the two participant groups. Overall, the gradient group performed better in all types of tests. The gradient group achieved a higher percentage of correct answers in the post-test (*M* = 87.0, *SD* = 5.2) and the new consonant generalization test (*M* = 83.9, *SD* = 3.4) than the comparable pre-test result (*M* = 80.4, *SD* = 5.8). On the other hand, the results of the categorical group did not show such an improvement in the post-test (*M* = 70.9, *SD* = 10.8) and the new consonant generalization test (*M* = 68.3, *SD* = 11.1), compared to their performance in the pretest (*M* = 69.8, *SD* = 11.5).

Table 5 summarizes the linear mixed-effects model, including Group and Test (Pre VERSUS Post and Pre VERSUS New Con), and the interactions between them. There were main effects of Group and Pre VERSUS Post, reflecting the overall high performance of the gradient group and improved participants' performances after the training. Inspecting Figure 14, the main effect of Pre VERSUS Post seems to reflect that the difference between pretest and post-test scores emerges only by the gradient group. Post hoc pairwise comparisons revealed a significant improvement of post-test scores compared to pretest scores from the gradient group, *t*(26.2) = -4.18, *p* < .001, but not from the categorical group, *t*(26.2) = -0.75, *p* = .457. There were reliable interactions between Group and Pre VERSUS Post and Group and Pre VERSUS New Con. This result shows that the gradient group showed a greater improvement in the post-test and the new consonant generalization test than the categorical group.

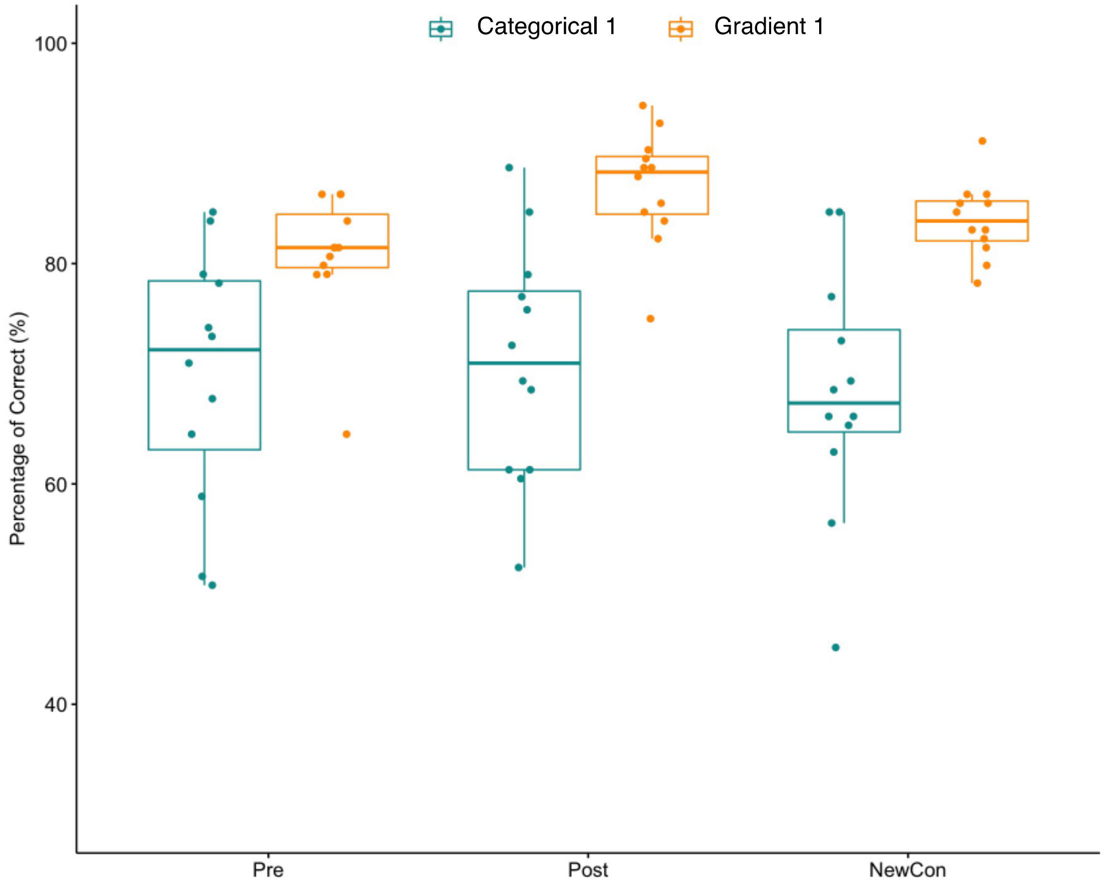


Figure 14. Results of the AXB pre-test, posttest, and new consonant generalization tests in percentage correct (%). *Boxplots: shaded region indicates interquartile range; whiskers extend to extreme values; solid bar indicates median; points indicate outliers.*

Table 5. Summary of fixed effects for the mixed-effects linear regression model of the pretest, post-test, and new consonant generalization tests scores.

| Predictor | Estimate ($\beta$) | Std. Error | *t-value* | *p-value* | |
|---|---|---|---|---|---|
| (Intercept) | 76.80 | 1.57 | 49.04 | <0.001 | *** |
| Group | 14.24 | 3.13 | 4.55 | <0.001 | *** |
| Pre_VERSUS_Post | 3.61 | 0.99 | 3.64 | 0.001 | ** |
| Pre_VERSUS_NewCon | 0.77 | 1.07 | 0.72 | 0.478 | |
| Group:Pre_VERSUS_Post | 5.01 | 1.98 | 2.53 | 0.018 | * |
| Group:Pre_VERSUS_NewCon | 4.64 | 2.15 | 2.16 | 0.041 | * |

### *2.1.5.2.2. Everyday ID tests and the new talker generalization test*

Figure 15 shows a similar trend as in Figure 14. The gradient group outperformed the categorical group in all three everyday ID tests. Compared to the performance in the Day 1 ID test (*M* = 69.6, *SD* = 8.3), the gradient group showed increase in the Day 2 (*M* = 75.9, *SD* = 8.7) and the Day 3 ID tests (*M* = 80.4, *SD* = 3.9). By contrast, the categorical group did not improve in the course of training; Day 1 (*M* = 58.01, *SD* = 12.64), Day 2 (*M* =57.5, *SD* = 13), and Day 3 (*M* = 59, *SD* = 12.3). It should be noted that the categorical group showed greater standard deviations on every ID test result, indicating a great deal of individual differences between participants in the categorical group compared to the gradient group. Even though both groups did not show much improvement in the new talker generalization test, the gradient group showed better performance (*M* = 70.5, *SD* = 8.6) than the categorical group (*M* = 56.1, *SD* = 13.3).

Table 6 summarizes the linear mixed-effects model, including Group and Test (Day1 VERSUS Day2, Day1 VERSUS Day3, and Day1 VERSUS New Talker), and their interactions. The results showed significant main effects of Group and Day1 VERSUS Day3, reflecting that the overall test scores were higher in the gradient group than the categorical group and the score of everyday ID test conducted after the last day of training (i.e., Day 3) showed was higher than the score of the test conducted after the first day of training (i.e., Day 1). As a follow-up, we conducted a post hoc pairwise comparison to examine whether the main effect of Day1 VERSUS Day3 came from only the gradient group or from both groups. The result showed that only the gradient group showed improvement on the Day 3 ID test, *t*(30.5) = -3.43, *p* = .002. There was a reliable interaction between Group and Day1 VERSUS Day3, reflecting that compared to the categorical group, the gradient group showed better performance which increased with training. Although the main effect of Group showed the overall higher performance of the gradient group, we did not find a reliable interaction between `Group` and `Day1 VERSUS New Talker`. Thus, there was no reliable evidence of better generalization of learning in the gradient group than the categorical group.

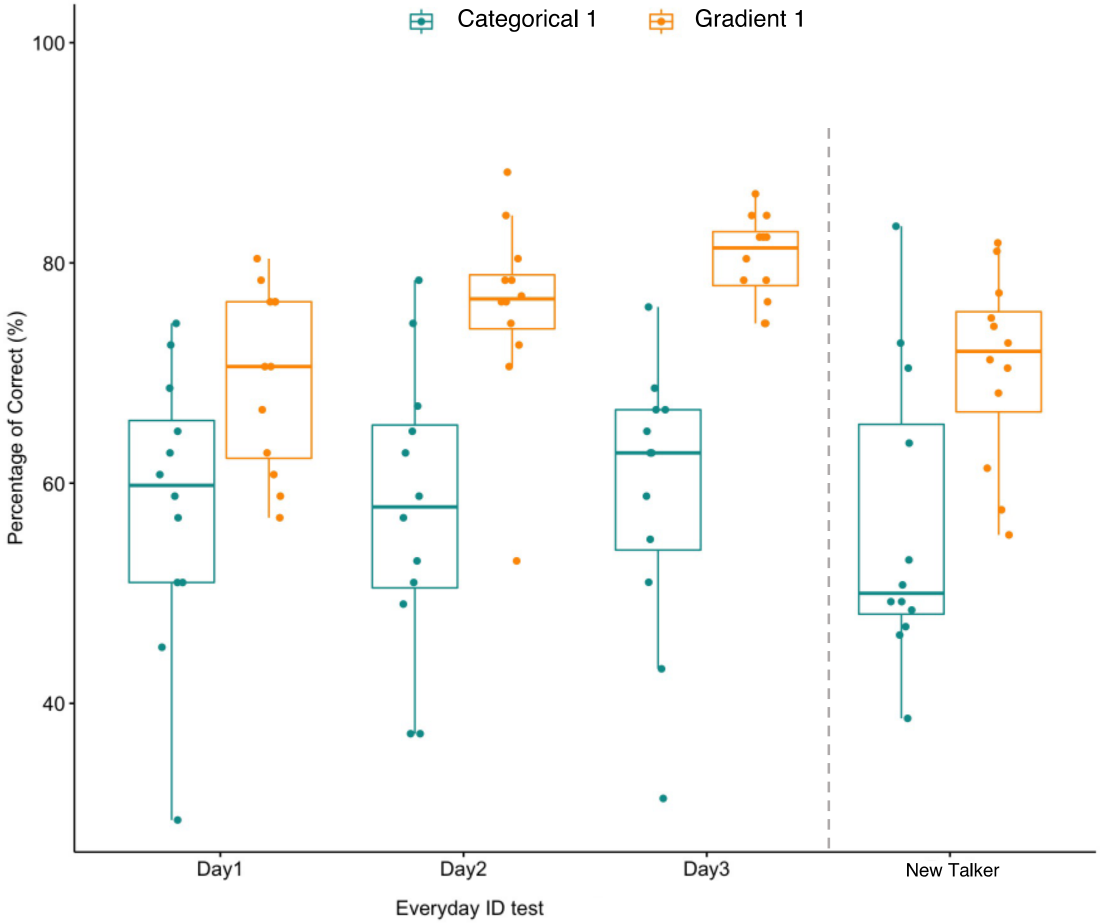


Figure 15. Results of everyday ID tests and the new talker generalization test in percentage correct (%). *Boxplots: shaded region indicates interquartile range; whiskers extend to extreme values; solid bar indicates median; points indicate outliers.*

Table 6. Summary of fixed effects for the mixed-effects linear regression model of everyday ID tests scores.

| Predictor | Estimate ($\beta$) | Std. Error | *t-value* | *p-value* | |
|---|---|---|---|---|---|
| (Intercept) | 65.87 | 1.55 | 42.54 | <0.001 | *** |
| Group | 16.45 | 3.10 | 5.31 | <0.001 | *** |
| Day1_VERSUS_Day2 | 2.90 | 2.13 | 1.36 | 0.180 | |
| Day1_VERSUS_Day3 | 5.86 | 2.13 | 2.75 | 0.008 | ** |
| Day1_VERSUS_New Talker | -0.52 | 2.31 | -0.23 | 0.823 | |
| Group:Day1_VERSUS_Day2 | 6.72 | 4.26 | 1.58 | 0.121 | |
| Group:Day1_VERSUS_Day3 | 9.84 | 4.26 | 2.31 | 0.025 | * |
| Group:Day1_VERSUS_ New Talker | 2.86 | 4.61 | 0.62 | 0.539 | |

### 2.1.5.3. Results of HVPT by individuals in experiment 1

The results presented in the previous sections have provided evidence for the relationship between participants' VAS response patterns and their performances in the HVPT. Group trends can roughly illuminate that participants with gradient VAS response patterns are better in their learning of Korean three-way stop contrast. Nevertheless, the results averaged across all participants are not sufficient to conclude the relationship between the gradiency (i.e., slope) of VAS responses and L2 learning performances since a wide range of individual differences in gradiency were not considered. Thus, the following section performed a close inspection of the VAS data from individual participants to explore whether individual differences in the gradiency of VAS responses are related to their performances in training, and, if so, to examine whether these individual differences are also related to participants' use of acoustic cues (i.e., VOT and f0) in identifying three target Korean stop consonants. More specifically, I investigated whether participants with more gradient VAS task response patterns benefited from their higher secondary cue sensitivity in the native stop perception and consequently yielded better perceptual learning of Korean lenis-aspirated stop distinction, which are primarily distinguished by the f0 dimension. Following earlier studies Kong and Edwards (2011, 2016) and our results on the greater sensitivity to f0 cues by listeners who showed gradient response patterns in the VAS task, a partial correlation test was conducted to investigate the relationship between individual differences in the gradiency of VAS responses and acoustic cue utilization. The purpose of this analysis was to unfold whether participants' sensitivities to f0 cues shown by the VAS task are responsible for their utilization of VOT and f0 cues in L2 Korean stop perception.

#### *2.1.5.3.1. Test scores*

The overall data analysis procedure with the test scores is identical to the group results analysis except that the individual differences in the VAS response patterns were considered. Two separate linear mixed-effects regression models were built; one with the pretest, post-test, and new consonant generalization tests results and the other with everyday ID test results. As in the previous models for group analysis, the models included the fixed effect variables, Test and Gradiency, and their 2-way interactions. Test was coded using *Dummy* contrasts, resulting in Pre VERSUS Post and Pre VERSUS New Con contrasts with pretest scores (i.e., Pre) as the reference level for the first model. The same variable was dummy coded to Day1 VERSUS Day2, Day1 VERSUS Day3, and Day1 VERSUS New Talker contrasts with Day 1 ID test scores (i.e., Day1) as the reference level for the second model. Gradiency was the continuous variable consisting of the quantified coefficients of the quadratic regression curves overlaid on participants' histograms of VAS responses (see 2.1.4.1). Gradiency was standardized by centering and dividing by 2 standard deviations, using the *rescale()* function from the *arm* package (Gelman et al., 2021) in *R* to reduce collinearity and make the intercept interpretable as mean gradiency of the VAS responses.

The model included random intercepts for participants and random slopes for participants for Test.

The results of the overall linear mixed-effects regression model with the pretest, post-test, and new consonant generalization tests scores are summarized in Table 7. There was a significant effect of Gradiency, indicating that participants with more gradient VAS response patterns (i.e., lower coefficients) were overall better in their performances in the tests. There was also a significant main effect of Pre VERSUS Post, suggesting that participants improved in the post-test compared to their performances in the pretest. A significant negative interaction effect between Gradiency and Pre VERSUS New Con was found; after receiving the HVPT, participants who made more gradient responses in the VAS task were related to greater improvements in their discrimination of the target Korean contrast in the new consonant generalization test.

The results of the linear mixed-effect model with the scores of everyday ID tests are presented in Table 8. As in the first model, there was a significant effect of Gradiency, showing that the gradiency was negatively related to the overall mean Everyday ID test scores. The significant main effect of Day1 VERSUS Day3 indicates that the scores of everyday ID tests conducted after the last day of training (i.e., Day 3) were overall higher. Although all participants improved after training, the amounts of improvement differed depending on their VAS response patterns. A significant two-way interaction between Gradiency and Day1 VERSUS Day3 and a trend towards significance between Gradiency and Day1 VERSUS Day2 show that participants with lower coefficients reached higher levels of achievements in identifying three target Korean stop consonants.

Table 7. Summary of fixed effects for the mixed-effects linear regression model of the pretest, post-test, and new consonant generalization tests scores.

| Predictor | Estimate ($\beta$) | Std. Error | *t-value* | *p-value* | |
|---|---|---|---|---|---|
| (Intercept) | 76.80 | 1.59 | 48.17 | <0.001 | *** |
| Gradiency | -14.01 | 3.21 | -4.37 | <0.001 | *** |
| Pre_VERSUS_Post | 3.61 | 1.07 | 3.39 | 0.002 | ** |
| Pre_VERSUS_NewCon | 0.77 | 1.06 | 0.73 | 0.474 | |
| Gradiency:Pre_VERSUS_Post | -3.26 | 2.14 | -1.52 | 0.141 | |
| Gradiency:Pre_VERSUS_NewCon | -4.85 | 2.14 | -2.27 | 0.032 | * |

Table 8. Summary of fixed effects for the mixed-effects linear regression model of everyday ID tests scores.

| Predictor | Estimate ($\beta$) | Std. Error | *t-value* | *p-value* | |
|---|---|---|---|---|---|
| (Intercept) | 65.87 | 1.48 | 44.48 | <0.001 | *** |
| Gradiency | -17.13 | 2.98 | -5.75 | <0.001 | *** |
| Day1_VERSUS_Day2 | 2.90 | 2.11 | 1.37 | 0.176 | |
| Day1_VERSUS_Day3 | 5.86 | 2.11 | 2.78 | 0.008 | ** |
| Day1_VERSUS_ New Talker | -0.52 | 2.31 | -0.22 | 0.824 | |
| Gradiency:Day1_VERSUS_Day2 | -7.65 | 4.24 | -1.80 | 0.078 | . |
| Gradiency:Day1_VERSUS_Day3 | -10.56 | 4.24 | -2.49 | 0.016 | * |
| Gradiency:Day1_VERSUS_ New Talker | -6.78 | 4.65 | -1.46 | 0.153 | |

#### *2.1.5.3.2. Relation between the quantified VAS gradiency and changes in acoustic cue reliance*

To examine the extent to which each acoustic dimension (i.e., VOT and f0) contributes to participants' identification of target Korean consonants and how they changed their reliance on VOT and f0 in the course of training, participants' responses in everyday ID tests conducted on the first and the last day of training sessions were analyzed with a mixed-effects logistic regression model, using the *glmer*() function from the *lme4* package (version 1.1-27) in *R*. The everyday ID tests were in the form of a three-alternative force choice task with three photographs representing either fortis, lenis, or aspirated Korean stop. In the model, the dependent variables were the binary responses of lenis /pa/ and aspirated /pʰa/ excluding fortis /p’a/ responses. The lenis /pa/ and aspirated /pʰa/ responses were coded as 0/1, respectively. The decision of this data processing was to clearly see whether participants learn how to systemically use f0 cues appropriately to distinguish training stimuli of Korean lenis and aspirated stops after the training, which are primarily differed in terms of f0 cues. Mixed-effects logistic regression models allow binary data to be analyzed. The model included fixed effects for Gradiency, Day, VOT, and F0, as well as interactions between Gradiency and F0, Gradiency and VOT, Gradiency and Day, Day and F0, Day and VOT, Gradiency, Day, and F0, and Gradiency, Day, and VOT. The approach for the analysis was only to inspect the model for effects and interactions between the experimental variables (i.e., fixed effects) where there are precise predictions. For example, the model included 2-way and 3-way interactions between Gradiency, Day, and each of the cues (VOT and F0) since they can indicate whether the use of the cues changes over time and did so differently depending on the participants’ gradiency. However, the interaction effect between Gradiency, VOT, and F0, for instance, was not included since this effect including both acoustic dimensions is not relevant for examining how individual differences in Gradiency is associated with a unique contribution of each acoustic dimension to target consonant identification responses after controlling for each other. VOT and F0 variables were standardized by centering and dividing by 2 standard deviations. Random intercepts for participants were included to account for participant-specific variability in binary responses. Random slopes for participants for Day, VOT, and F0 were also included to account for by-participant variability in the effect of each variable on their responses.

Figure 16 plots the estimated response curves drawn from the regression model investigating the effect of individual differences in gradiency of the VAS task responses on the proportion of Korean aspirated stop (i.e., /pʰa/) responses for the everyday ID test stimuli across days as a function of VOT and f0 steps. The left-side graph represents the estimated proportion of /pʰa/ responses for standardized five f0 steps, while the right-side graph shows the estimated proportion of /pʰa/ responses for the standardized five VOT steps. The results from the different everyday ID tests (i.e., Day 1 vs. Day 3) are represented in different colors. The overall mixed-effects logistic regression model results are summarized in Table 9.

The overall pattern of identification responses in Figure 16 shows that as the VOT and f0 steps of test stimuli become higher, participants identified them as Korean aspirated stops. In terms of participants' gradiency, Figure 16 demonstrates that participants' individual differences in gradiency are associated with the degrees of VOT and f0 cue utilizations in the Korean stop contrast identification. As participants' gradiency coefficients became lower (i.e., more gradient VAS response patterns), they utilized both VOT and f0 cues more natively by making more /pʰa/ responses as VOT and f0 steps increased. This identification pattern was observed in both Day 1 and Day 3 ID tests, which suggests that the negative relation between participants' gradiency coefficients and their cue utilization was shown even after the first day of training and remained intact until the end. The regression analysis found the positive main effects of `F0` and `VOT`, showing that both VOT and f0 dimensions significantly contribute to participants' identification responses in the everyday ID tests. The model also found significant two-way interactions between each of the cues (`F0` and `VOT`) and `Gradiency` and between `F0` and `Day`, indicating that the use of f0 cues changed over time and the effects of `VOT` and f0 cues in the identification of the target Korean stop contrast changed differently in relation to participants' gradiency of the VAS task responses. A significant interaction between `F0` and `Day` shows that participants identified Korean lenis and aspirated stop by using f0 cues more after the last session of HVPT (i.e., Day 3) than Day 1. More importantly, the model found significant two-way interactions between

Gradiency and each acoustic dimension (VOT and F0), indicating that greater uses of acoustic cues in everyday ID tests are associated with lower gradiency values (i.e., more gradient VAS task response patterns). The three-way interaction between F0, Day, and Gradiency was marginally significant, indicating that greater use of f0 cues in the Day3 ID test was associated with lower gradiency values.

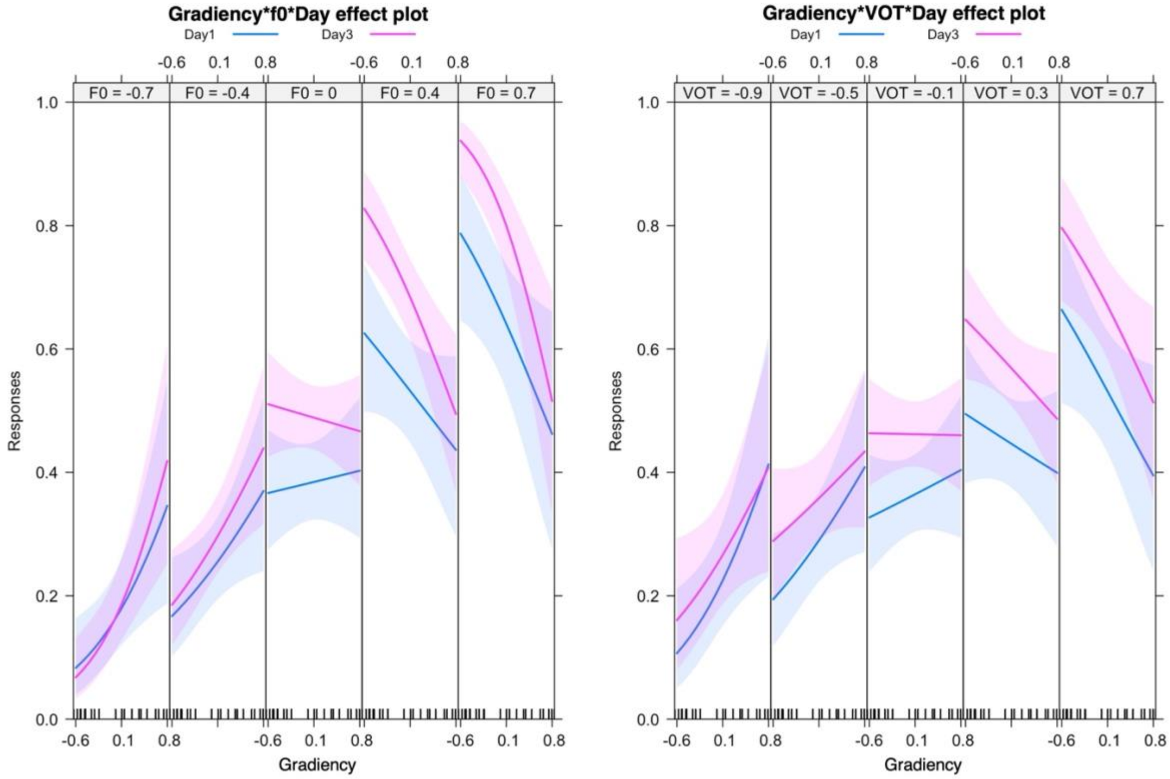


Figure 16. Predicted logit curves for perception of the target Korean stop contrast by native English participants. Predictions are drawn from a binary logistic regression model with Gradiency, Day, VOT, and F0 as predictors. Each panel represents standardized either f0 or VOT steps of test stimuli. The x-axis represents standardized individual participants' quantified gradiency. The y-axis of each panel represents the predicted probability of "aspirated" responses. Results from Day1 and Day 3 ID tests are shown in different colors.

Table 9. Summary of fixed effects for the mixed-effects logistic regression model predicting changes in f0 and VOT cue weights in everyday ID tests.

| Predictor | Estimate ($\beta$) | Std. Error | *z-value* | *p-value* | |
|---|---|---|---|---|---|
| (Intercept) | -0.25 | 0.09 | -2.72 | 0.007 | ** |
| Gradiency | -0.01 | 0.19 | -0.05 | 0.962 | |
| F0 | 1.99 | 0.25 | 7.98 | <0.001 | *** |
| VOT | 1.09 | 0.19 | 5.71 | <0.001 | *** |
| Day | 0.45 | 0.16 | 2.83 | 0.005 | ** |
| Gradiency:F0 | -2.09 | 0.49 | -4.31 | <0.001 | *** |
| Gradiency:VOT | -1.22 | 0.37 | -3.27 | 0.001 | ** |
| Gradiency:Day | -0.22 | 0.31 | -0.72 | 0.472 | |
| F0:Day | 0.68 | 0.26 | 2.59 | 0.010 | ** |
| VOT:Day | 0.24 | 0.25 | 0.94 | 0.350 | |
| Gradiency:F0:Day | -0.93 | 0.50 | -1.87 | 0.061 | . |
| Gradiency:VOT:Day | 0.10 | 0.49 | 0.21 | 0.838 | |

#### *2.1.5.3.3. Correlation between the quantified VAS gradiency and VOT and f0 cue uses*

Partial correlation tests were conducted to show whether there is a meaningful correlation between gradient responses in the VAS task, which utilized L1 speech stimuli, and VOT and f0 cue utilization in L2 phonetic training. For the partial correlation analysis between participants' use of VOT and f0 cues and their quantified gradiency of the VAS task, logistic regression coefficients were performed first to quantify how participants use VOT and f0 cues. A series of logistic regression models, using the *glmer()* function from the *lme4* package (version 1.1-27) in *R,* were fitted to each participants' responses on the Day 3 everyday ID test, which were converted into binary data (i.e., 0 = /pa/, 1= /p$^h$a/). The models included fixed effects for standardized VOT and f0 steps. As in the previous analysis, the fortis stop category was excluded. Individual participants' perceptual weights for VOT and f0 cues were calculated based on VOT and f0 coefficients ($\beta$) fitted to each participants' response data. The coefficients from the individual models were used as measures of the perceptual weight of the respective cues.

This method was adapted from previous studies, which examined participants' cue-weighting to multiple acoustic cues in the perception of non-native speech contrasts (e.g., Kim et al., 2017; Kong, 2019; Schertz et al., 2015). One thing to note here is that only data from the Day 3 ID test was used to calculate the coefficients since the primary interest of this analysis was to explore whether their success in learning how to weigh VOT and f0 cues is mediated by their different sensitivities to VOT and f0 cues in the perception of native stop category. Therefore, to take into consideration learners' gains from the training, participants' responses for the Day 3 ID tests were submitted to logistic regression models. The VOT-tuned logistic regression coefficients ($\beta_{\mathrm{VOT}}$) and f0-tuned coefficients ($\beta_{\mathrm{f0}}$) served as measures of the perceptual weights in the identification of Korean lenis and aspirated stops. Higher coefficients indicate more use of the corresponding acoustic cue, and negative values indicate that participants use the cue in the opposite direction from native listeners of Korean. The full results of participants' coefficients can be found in Appendix I.

Before the correlation analysis, a visual analysis of graphs showing participants' response patterns to everyday ID test stimuli was conducted in the levels of groups and individuals. Figure 17 shows the overall responses of everyday ID tests to each test stimulus with different VOT and f0 steps averaged across participants in each group (Gradient vs. Categorical) in order to reflect the group trend in the use of each acoustic dimension. The overall pattern of responses of the everyday ID test conducted after the training illustrates that the gradient and the categorical groups differ in their cue weighting strategies. Overall, the cue weighting strategies show that the participants in each group used VOT dimension in a relatively similar fashion in identifying Korean fortis stops, indicating that both gradient and categorical groups identified stimuli with the shortest VOT step as fortis stops. However, Figure 17 illustrates that participants in the gradient group made greater use of f0 cues than participants in the categorical group in identifying Korean lenis and aspirated stops. The overall group patterns displayed in Figure 17 do not reflect substantial individual differences in participants' use of VOT and f0 dimensions. Therefore, the following paragraph focuses on how individual participants are different in their response patterns to everyday ID tests.

Figure 18 and Figure 19 show the identification responses of the six selected participants. The panels display considerable differences in individual participants' cue weighting strategies, as well as how they changed their strategies after the completion of the training. The patterns of identification responses in Figure 19 demonstrate that participants (i.e., G1_1, G1_2, and G1_6) involved the active use of f0 cues with the VOT cues in their cue weighting strategies after the last training session (i.e., Day 3). More specifically, participants in Figure 19 predominantly relied on the f0 cues on the Day 3 ID test to map test stimuli into the Korean lenis or aspirated stop categories, which are mainly distinguished by the f0 dimension. Furthermore, they labeled stimuli with short VOT and mid-to-high f0 values as Korean fortis stops while categorizing stimuli with the same VOT step but the lower f0 step as Korean lenis stops on the Day 3 ID test. As for the participants in Figure 18 (C1_1, C1_2, and C1_3), the evidence of using the f0 cues for perceiving the target Korean stop consonants was not shown even after the entire HVPT. These learners seem to demonstrate considerable difficulty in identifying stimuli as Korean lenis or aspirated stops, as shown in Figure 18 with the mixed distributions of responses for those two stop categories. For example, in the Day 3 ID test, the C1_1 participant successfully used the VOT cues to distinguish Korean fortis stops from lenis and aspirated stops. However, he failed to utilize the f0 cues for accurately distinguishing stimuli with longer VOT values as either Korean lenis or aspirated stops by relying on the f0 dimension.

That is, some participants, such as in Figure 19, weighted both VOT and f0 dimensions equally importantly to distinguish Korean aspirated stops from the other two stop categories like native speakers of Korean. However, some participants, such as in Figure 18, systemically only used the VOT cues to differentiate Korean fortis stops from other two stop categories, lenis and aspirated stops, suggesting that they poorly realized the importance of f0 cues and failed to utilize both f0 and VOT dimensions in Korean stop perception. In sum, a close inspection of the graphs reveals that participants make different use of acoustic cues in classifying the target Korean stop contrast: some participants make use of VOT cues to distinguish Korean fortis stops from two other stop categories, whereas other participants utilize both VOT and f0 cues and are more attuned to f0 cues to distinguish Korean lenis and aspirated stops.

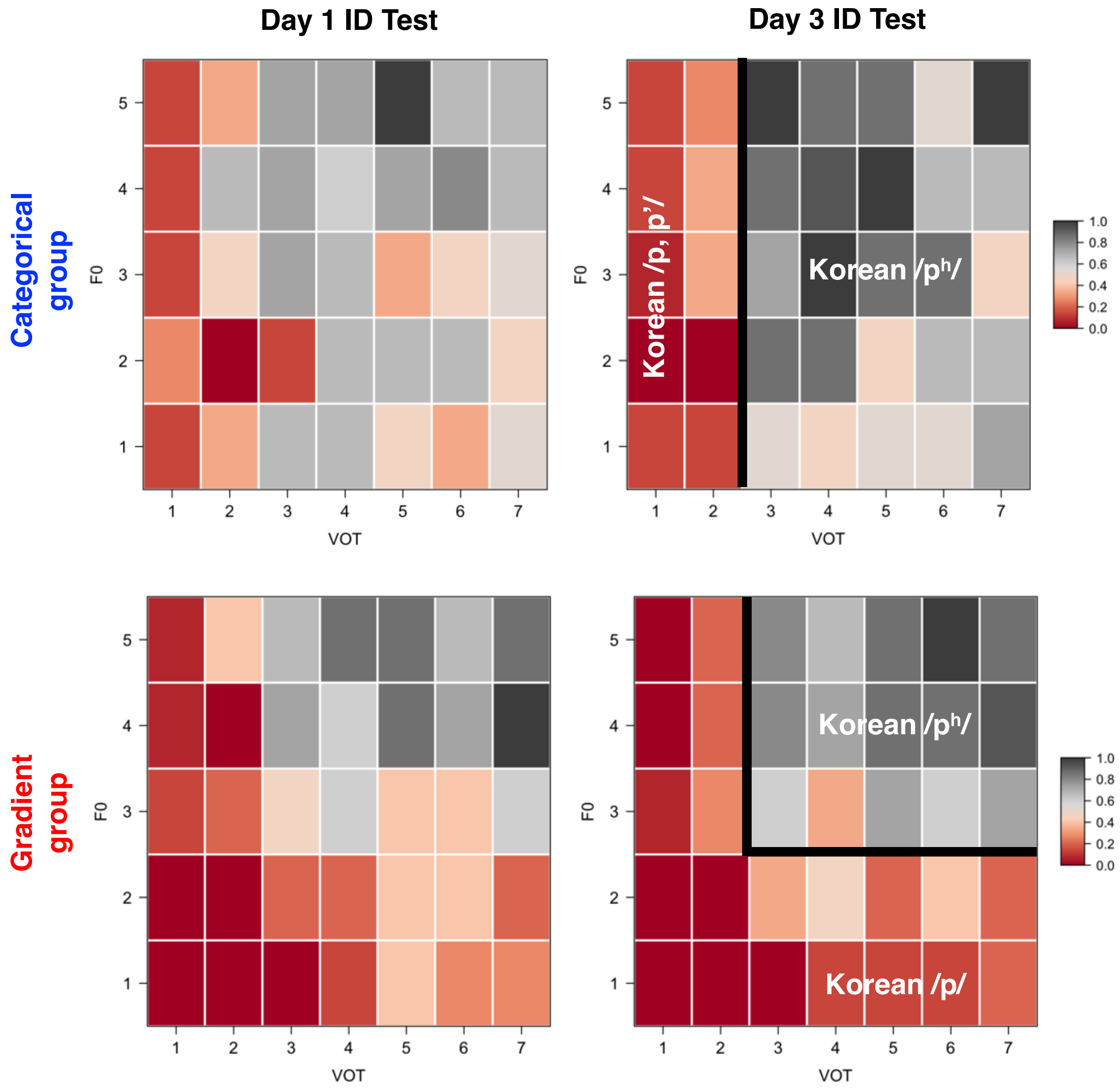
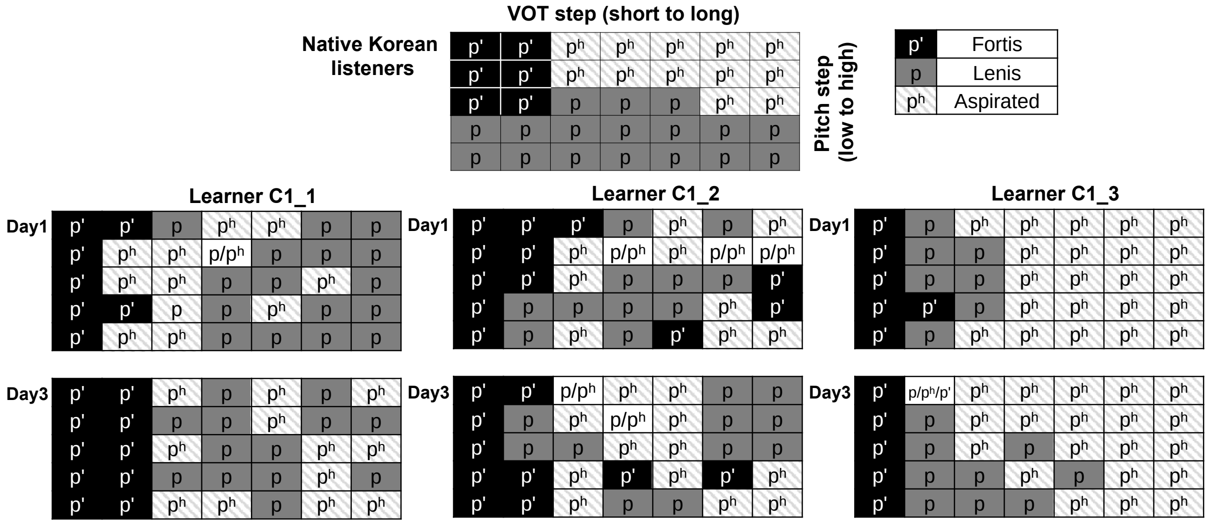


Figure 17. Heat plots of participants' overall responses of the Day 1 ID test and the Day 3 ID test to each combination of VOT and f0 (pitch) dimensions. The darkness of the cell represents the percentage of responses in a forced-choice task; the darkest grey cells elicited 100% Korean aspirated stop (/p$^h$/) responses, while the darkest red cells elicited 0% Korean aspirated stop (/p$^h$/) responses.



VOT step (short to long)

Pitch step (low to high)

| | | |
|---|---|---|
| p' | Fortis | ■ |
| p | Lenis | ▩ |
| p$^h$ | Aspirated | ▨ |

Native Korean listeners

| | | | | | | |
|---|---|---|---|---|---|---|
| p' | p' | p$^h$ | p$^h$ | p$^h$ | p$^h$ | p$^h$ |
| p' | p' | p$^h$ | p$^h$ | p$^h$ | p$^h$ | p$^h$ |
| p' | p' | p | p | p | p$^h$ | p$^h$ |
| p | p | p | p | p | p | p |
| p | p | p | p | p | p | p |

Learner C1_1

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Day1 | p' | p' | p | p$^h$ | p$^h$ | p | p |
| | p' | p$^h$ | p$^h$ | p/p$^h$ | p | p | p |
| | p' | p$^h$ | p$^h$ | p | p | p$^h$ | p |
| | p' | p' | p | p | p$^h$ | p | p |
| | p' | p$^h$ | p$^h$ | p | p | p | p |
| Day3 | p' | p' | p$^h$ | p | p$^h$ | p | p$^h$ |
| | p' | p' | p | p | p$^h$ | p | p |
| | p' | p' | p$^h$ | p | p | p$^h$ | p$^h$ |
| | p' | p' | p | p | p | p$^h$ | p |
| | p' | p' | p$^h$ | p$^h$ | p | p$^h$ | p$^h$ |

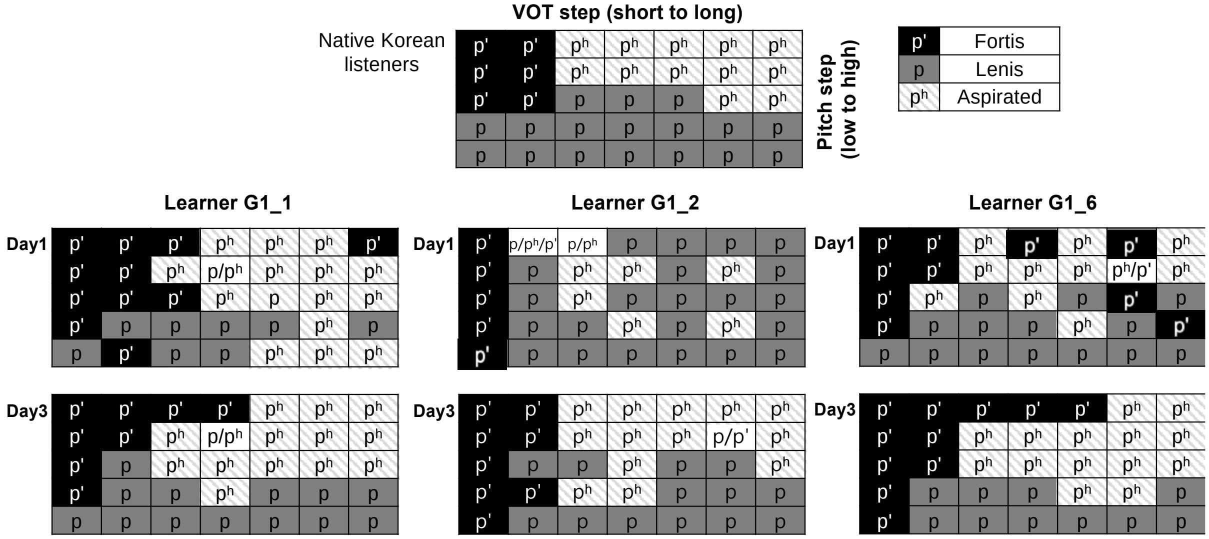
Learner C1_2

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Day1 | p' | p' | p' | p | p$^h$ | p | p$^h$ |
| | p' | p' | p$^h$ | p/p$^h$ | p$^h$ | p/p$^h$ | p/p$^h$ |
| | p' | p' | p$^h$ | p | p | p | p' |
| | p' | p | p | p | p | p$^h$ | p' |
| | p' | p | p$^h$ | p | p' | p$^h$ | p$^h$ |
| Day3 | p' | p' | p/p$^h$ | p$^h$ | p$^h$ | p | p |
| | p' | p | p$^h$ | p/p$^h$ | p$^h$ | p | p |
| | p' | p | p | p$^h$ | p$^h$ | p | p |
| | p' | p' | p$^h$ | p' | p$^h$ | p' | p$^h$ |
| | p' | p' | p$^h$ | p | p | p$^h$ | p$^h$ |

Learner C1_3

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Day1 | p' | p | p$^h$ | p$^h$ | p$^h$ | p$^h$ | p$^h$ |
| | p' | p | p | p$^h$ | p$^h$ | p$^h$ | p$^h$ |
| | p' | p | p | p$^h$ | p$^h$ | p$^h$ | p$^h$ |
| | p' | p' | p | p$^h$ | p$^h$ | p$^h$ | p$^h$ |
| | p' | p | p$^h$ | p$^h$ | p$^h$ | p$^h$ | p$^h$ |
| Day3 | p' | p/p$^h$/p' | p$^h$ | p$^h$ | p$^h$ | p$^h$ | p$^h$ |
| | p' | p | p$^h$ | p$^h$ | p$^h$ | p$^h$ | p$^h$ |
| | p' | p | p$^h$ | p | p$^h$ | p$^h$ | p$^h$ |
| | p' | p | p | p$^h$ | p | p$^h$ | p$^h$ |
| | p' | p | p | p | p$^h$ | p$^h$ | p$^h$ |

Figure 18. Native Korean listeners' patterns in the everyday identification test (top) and mapping plots of responses of everyday ID tests (Day 1 & Day 3) by three participants from the categorical group (C1_1, C1_2, and C1_3).



VOT step (short to long)

Pitch step (low to high)

| | | |
|---|---|---|
| p' | Fortis | ■ |
| p | Lenis | ▩ |
| p$^h$ | Aspirated | ▨ |

Native Korean listeners

| | | | | | | |
|---|---|---|---|---|---|---|
| p' | p' | p$^h$ | p$^h$ | p$^h$ | p$^h$ | p$^h$ |
| p' | p' | p$^h$ | p$^h$ | p$^h$ | p$^h$ | p$^h$ |
| p' | p' | p | p | p | p$^h$ | p$^h$ |
| p | p | p | p | p | p | p |
| p | p | p | p | p | p | p |

Learner G1_1

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Day1 | p' | p' | p' | p$^h$ | p$^h$ | p$^h$ | p' |
| | p' | p' | p$^h$ | p/p$^h$ | p$^h$ | p$^h$ | p$^h$ |
| | p' | p' | p' | p$^h$ | p | p$^h$ | p$^h$ |
| | p' | p | p | p | p | p$^h$ | p |
| | p | p' | p | p | p$^h$ | p$^h$ | p$^h$ |
| Day3 | p' | p' | p' | p' | p$^h$ | p$^h$ | p$^h$ |
| | p' | p' | p$^h$ | p/p$^h$ | p$^h$ | p$^h$ | p$^h$ |
| | p' | p | p$^h$ | p$^h$ | p$^h$ | p$^h$ | p$^h$ |
| | p' | p | p | p$^h$ | p | p | p |
| | p | p | p | p | p | p | p |

Learner G1_2

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Day1 | p' | p/p$^h$/p' | p/p$^h$ | p | p | p | p |
| | p' | p | p$^h$ | p$^h$ | p | p$^h$ | p |
| | p' | p | p$^h$ | p | p | p | p |
| | p' | p | p | p$^h$ | p | p$^h$ | p |
| | p' | p | p | p | p | p | p |
| Day3 | p' | p' | p$^h$ | p$^h$ | p$^h$ | p$^h$ | p$^h$ |
| | p' | p' | p$^h$ | p$^h$ | p$^h$ | p/p' | p$^h$ |
| | p' | p | p | p$^h$ | p | p | p$^h$ |
| | p' | p' | p$^h$ | p$^h$ | p | p | p |
| | p' | p | p | p | p | p | p |

Learner G1_6

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Day1 | p' | p' | p$^h$ | p' | p$^h$ | p' | p$^h$ |
| | p' | p' | p$^h$ | p$^h$ | p$^h$ | p$^h$/p' | p$^h$ |
| | p' | p$^h$ | p | p$^h$ | p | p' | p |
| | p' | p | p | p | p$^h$ | p | p' |
| | p | p | p | p | p | p | p |
| Day3 | p' | p' | p' | p' | p' | p$^h$ | p$^h$ |
| | p' | p' | p$^h$ | p$^h$ | p$^h$ | p$^h$ | p$^h$ |
| | p' | p' | p$^h$ | p$^h$ | p$^h$ | p$^h$ | p$^h$ |
| | p' | p | p | p | p$^h$ | p$^h$ | p |
| | p' | p | p | p | p | p | p |

Figure 19. Native Korean listeners' patterns in the everyday identification test (top) and mapping plots of responses of everyday ID tests (Day 1 & Day 3) by three participants from the gradient group (G1_1, G1_2, and G1_6).

Individual participants' $\beta_{\text{VOT}}$ and $\beta_{\text{f0}}$ coefficients obtained from a series of logistic regression analyses are plotted in Figure 20. Data points are different in color and shape to represent the group information of participants. The overall cue-weighting patterns by group show that the gradient group used both VOT and f0 dimensions relatively differently in identifying Korean lenis and aspirated stops compared to the categorical group. The most noticeable difference is that majority of participants in the gradient group weighted more on f0 cues while VOT cues had a much weaker effect on their identification of Korean lenis and aspirated stops (*M* (*SD*) of $\beta_{\text{VOT}}$ and $\beta_{\text{f0}} = 1.85$ (1.20), 3.91 (1.99)). On the other hand, participants in the categorical group seemed to present relatively equal use of VOT and f0 cues or more reliance on VOT cues than f0 cues. (*M* (*SD*) of $\beta_{\text{VOT}}$ and $\beta_{\text{f0}} = 1.43$ (1.98), 1.02 (2.01)). Figure 21 also confirms that the f0 dimension had a much stronger effect on the gradient group's identification than the categorical group. Figure 21 shows the proportion of responses of Korean aspirated stops across f0 steps of test stimuli. Thin lines are logistic curves that fit each participant's response data from Day 3 ID test, and thick lines are that that fit to averaged group data from the same test. As in Figure 20, it was observed that participants in the gradient group showed a stronger effect of f0 dimension in their identification responses compared to the participants in the categorical group. After controlling for the effect of the VOT dimension, Figure 21 shows that the gradient group participants demonstrated more native-like use of f0 cues in the identification of Korean lenis and aspirated stops. It should be noted that this plot displays considerable differences in individual learners' use of f0 cues.

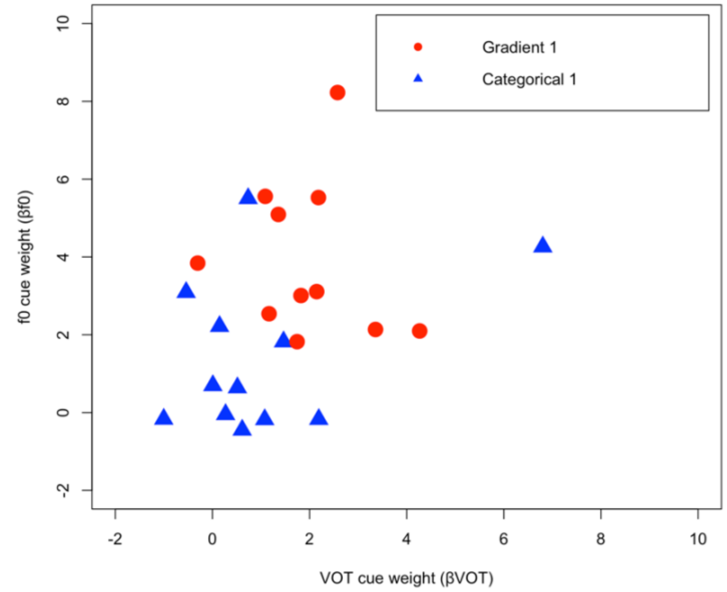


Figure 20. Individual participants' βVOT and βf0 coefficients obtained from a series of logistic regression analysis.

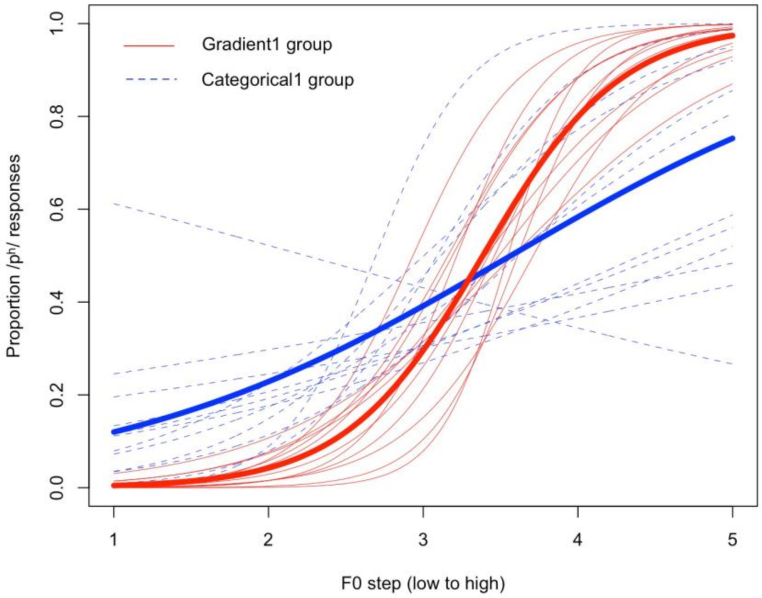


Figure 21. Proportion of /pʰ/ responses along f0 steps of test stimuli. Thin lines are logistic curves fit to each individual participant Day3 ID test data and thick lines are logistic curves fit to group-averaged data. Blue lines indicate categorical group, and red lines indicate gradient group.

Partial correlation tests were conducted to examine whether individual participants' quantified gradiency of the VAS task responses can explain the substantial individual differences in VOT and f0 cue weighting in identifying Korean lenis and aspirated stops. As shown in Figure 22, the analysis revealed no significant correlation between individual participants' gradiency and VOT cue weights ($\beta_{\mathrm{VOT}}$: $r = -.35$, $p = .09$). However, there was a significant negative correlation between gradiency and f0 cue weight ($\beta_{\mathrm{f0}}$: $r = -.59$, $p = .003$), indicating that participants with lower gradiency values attended to f0 dimension more than participants with higher gradiency values. These results suggest that participants who showed more sensitivity to f0 cues in the perception of their native stop voicing contrast (English /p/-/b/) could use f0 cues to identify Korean lenis and aspirated stops as native Korean listeners do. In accordance with earlier studies (e.g., Kapnoula et al., 2017; Kong & Edwards, 2011, 2016), the negative correlation between gradiency and f0 cue weight demonstrates that English native listeners' gradient responses were responsible for their f0 cue utilizations. However, this current research is different from previous studies in that participants' f0 cue utilization was measured in the perception of the nonnative category, not in the perception of the native stop category. This issue will be addressed in the discussion section (see 4.1) regarding transfer of cue sensitivities from native category perception to nonnative category perception. In sum, the quantified gradiency measured with L1 speech stimuli (i.e., coefficients of the quadratic regression curves overlaid on the histograms of the VAS responses) could account for individual variability in how listeners utilize acoustic cues in nonnative category perception.

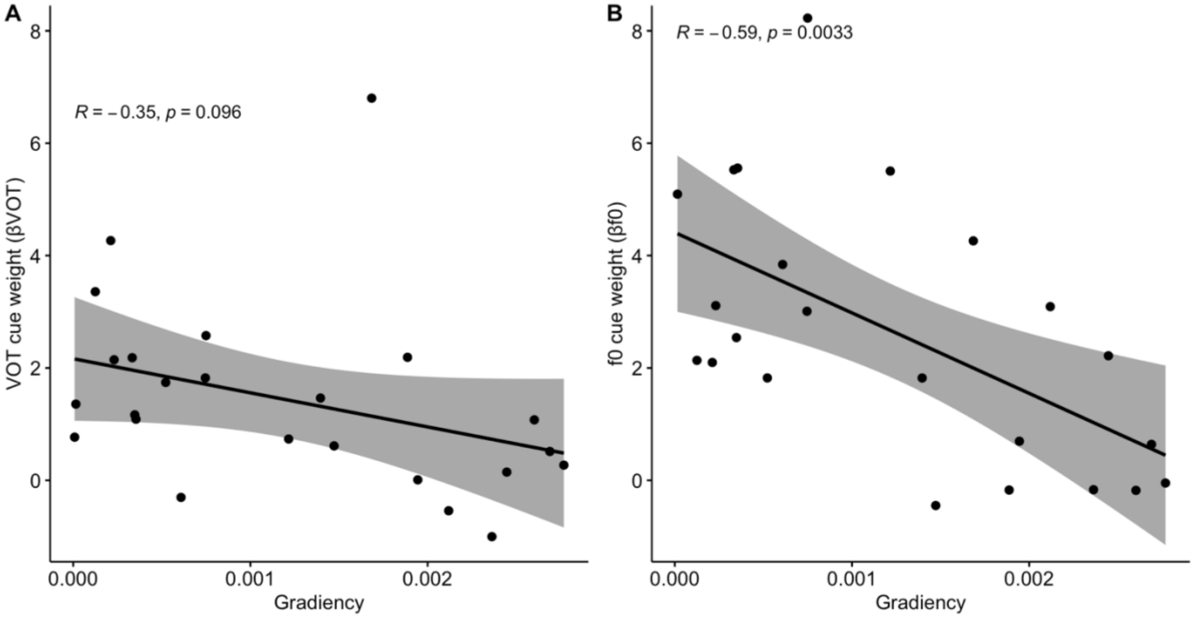


Figure 22. Correlation between individual participants' quantified gradiency of the VAS task and individual participants' beta-coefficients of either VOT (left) or f0 (right).

## 2.2. Summary of experiment 1

In experiment 1, participants were divided into two groups depending on their response patterns in the VAS task: the categorical 1 and the gradient 1 groups. The gradient 1 group demonstrated a larger benefit of HVPT than the categorical 1 group. As the gradient group 1 received more training, they demonstrated substantial gains in tests, everyday ID tests, the post-test, and the new consonant generalization test; however, the categorical 1 group did not improve in training over time. Considering that the primary goal of HVPT is to generate robust and highly generalized improvements in L2 speech perception, it is worth pointing out again that, in experiment 1, only the gradient 1 group responded to HVPT and presented their ability to generalize perceptual learning to the novel and untrained contrast stimuli.

The statistical analysis considering individual variability in the gradiency of the VAS responses showed that participants with more gradient response patterns in the VAS task resulted in higher scores in tests. The quantified gradiency could successfully predict individual differences in the success of nonnative contrast learning in response to HVPT and the ability to generalize their perceptual learning to a novel, untrained set of stimuli with evidence of greater improvement as gradiency values become lower. The logistic regression model results showed that VOT and f0 dimensions had a significant effect on the choice of lenis or aspirated category, but the effect of f0 dimension was larger for participants who had a more gradient response pattern on the VAS task (i.e., lower quantified gradiency). The f0 beta-coefficients from individual regression models for Korean lenis and aspirated contrast showed a meaningful correlation with the quantified gradiency. This correlation suggests that the more gradient participants' response pattern on the VAS task was, the more they relied on the f0 dimension to identify the challenging Korean lenis and aspirated contrast in a more nativelike way.

# CHAPTER 3.   EXPERIMENT 2

In experiment 1, learners who showed the categorical VAS response pattern did not successfully use the f0 cue in the perception of Korean lenis and aspirated stops. This problem persisted even after the last training session reflected as poorly mapped test stimuli into the three Korean stop categories (see Figure 18 & Figure 19). In addition, the categorical 1 group did not show their ability generalize their learning while the gradient 1 group did generalize their leaning in the identification of the novel set of stimuli produced by a new talker.

Then, how can the HVPT training paradigm be modified to assist the impaired learners? In experiment 2, the effect of additional training, named as cue-attention switching training, was tested. Especially, experiment 2 deals with the research question 2. Regarding this question, I predict that learners with relatively less sensitivity to f0 cues (i.e., the categorical group) would perform similarly to learners with relatively higher sensitivity to f0 cues (i.e., the gradient group). The logic behind this prediction is that if learners with relatively less sensitivity to f0 cues (i.e., learners with categorical VAS response pattern) can make a temporal change in the perception of their native contrast by up-weighting the relevant cue (f0) in the nonnative contrast and down-weighting the prominent native cue (VOT), learners can start their learning at a similar starting point as those who are already sensitive to the f0 dimension in their native language perception. I expected that such a short and straightforward additional training would decrease the possible learning gap due to the listeners' differences in their sensitivities to acoustic cues in the perception of native speech contrast.

## 3.1. Methods

### *3.1.1. Participants*

A new group of nineteen native speakers of American English was recruited for experiment 2 (eighteen female, one male: mean age of 20.0 years, range of 18 – 23 years). The participants were undergraduate students at the University of Wisconsin-Milwaukee. Except for one participant born in New Jersey, eighteen participants were born in the Midwest United States, either in Wisconsin or Illinois. All participants were raised in the Midwest United States and did not have experience residing outside the Midwest for more than one year at the time of the experiment. Participants reported no history of a speech or hearing impairment and received extra course credits upon completing the experiment. Like the participants in experiment 1, some of the participants in experiment 2 had some experience in foreign languages, but none of them considered themselves bilingual. None of the participants have prior experience in Korean language learning both formally and informally, which indicates that participants in experiment 2 were also naïve learners of Korean at the time of participation. The summary of demographic and language background data is shown in Table 10.

Table 10. Demographic information of participants in experiment 2.

| Variable | Mean | *SD* |
|---|---|---|
| Sex | 18 F; 1 M | |
| Age (yrs. old) | 19.9 | 1.3 |
| Birthplace (state, #) | WI 16 | |
| | IL 1 | |
| | MN 1 | |
| | NJ 1 | |
| Lived out of Midwest (yrs.) | 0.0 | 0.0 |
| Father Birthplace (state/country, #) | WI 15 | |
| | IL 1 | |
| | OH 1 | |
| | MN 1 | |
| | England 1 | |
| Father L1 (language, #) | English 19 | |
| Mother Birthplace (state/country, #) | WI 16 | |
| | IL 1 | |
| | PA 1 | |
| | MN 1 | |
| Mother L1 (language, #) | English 19 | |
| L2 (language, #) | Spanish 5 | |
| | German 2 | |
| | ASL 2 | |
| | French 1 | |
| | Chinese 1 | |
| (1 very poor to 7 very good) | | |
| L2 Speaking | 3.3 | 0.9 |
| L2 Understanding | 3.5 | 1.6 |
| L2 Reading | 4.1 | 1.7 |
| L2 Writing | 3.3 | 1.7 |

### *3.1.2. Stimuli*

Stimuli for the VAS task and HVTP training were identical to the stimuli used in experiment 1.

### 3.1.2.1. Stimuli for the cue-attention switching training

The only difference between experiments 1 and 2 is implementing the cue-attention switching training. The stimuli for this training were English pseudo-synthetic consonant vowel (CV) syllables that were constructed to make a continuum from /ba/ to /pa/. The English bilabial voicing stop contrast was chosen for the additional training since this contrast is equivalent to the Korean three-way laryngeal stop contrast (Schmidt, 2007). As mentioned earlier, native English listeners predominantly map Korean fortis, lenis, and aspirated stops to either English voiced or voiceless category. Therefore, it was predicted that using stimuli falling into the perceptually corresponding category to the L2 category would be appropriate to enhance the effectiveness of the additional training and directly examine whether modified acoustic cue sensitivity in L1 by the additional training would benefit nonnative contrast learning. Another reason for selecting English /ba/ to /pa/ for the additional training was to equally expose participants in experiments 1 and 2 to the stimuli for the VAS task, which were a continuum from English /da/ to /ta/.

A set of 35 English /ba/ to /pa/ tokens was created, out of which a subset was extracted for the cue-attention switching training. All tokens consisted of a synthetic consonant ranging from /b/ to /p/ with different VOT and f0 values and a 250 ms steady-state vowel /a/. The 35 original stimuli were created according to the following procedure: a 28-year-old female native talker of a Midwestern dialect of American English produced several examples of English pseudo words ‘ba’ and ‘pa’ in the context of the carrier sentence (“I say _____ again”). All tokens were recorded in a double-walled sound booth in the Phonetics Lab at the University of Wisconsin-Milwaukee using a Marantz PMD 661 digital audio recorder and a Shure SM-10A head-worn microphone at a sampling rate of 44.1 kHz with 16-bit resolution. The microphone was positioned on a boom approximately 2 to 3 cm away from the corner of the speaker’s mouth.

The talker was instructed to produce the tokens with clear pronunciation at a normal speech rate. From the recorded set, one token of ‘pa’ was selected using the criterion that there be enough aspiration portion of /p/, no abrupt changes in format movement throughout the vowel /a/ phonation of the token, no abrupt changes in fundamental frequency, no clicks and minimal extraneous noise throughout the recording. The /p/ portion with a 250 ms steady-state vowel /a/ portion was manually extracted. In this baseline token, the /p/ was defined as starting at the beginning of the burst release/aspiration and ending at the end of the aspiration before starting the periodic vowel portion. The start of the periodicity was defined as a point at the first zero crossing where the first period of the waveform began. The baseline token was then manipulated to consist of seven VOT steps and five f0 steps, yielding 35 different CV stimuli. The detailed stimuli manipulation process is identical to generating the set of HVPT stimuli (see 2.1.2.2). The baseline token /pa/ was manipulated in VOT by using “To Manipulation…” in Praat. The range of VOT for the stimuli manipulation was set from the shortest 0 ms to the longest 72 ms, which was the VOT value of naturally produced English /pa/ token from the speaker. The VOT step interval was set to 12 ms. And then, in each VOT step, the f0 dimension was manipulated by replacing the original f0 value of the baseline token with one of five f0 values (181 Hz, 194 Hz, 207 Hz, 220 Hz, and 233 Hz) using the “To manipulation…” function in Praat. Similar to the VOT range, the lowest and highest f0 values were based on the speakers’ natural productions of English /ba/ and /pa/, which are 81 Hz and 233 Hz, respectively. The f0 interval between steps was 13 Hz. This procedure yielded 35 CV stimuli (seven VOT steps × five f0 steps).

### *3.1.3. Procedure*

The apparatus and procedures for the VAS task and the HVPT were identical to those used in experiment 1.

#### 3.1.3.1. HVPT

The overall procedure of HVPT was identical to experiment 1, except participants completed the cue-attention switching training before the beginning of each Korean training session. Experiment 2 consisted of five days (i.e., from Day 1 to Day 5) and involved three stages: pretest, training, and post-test. The pretest contained three tasks as experiment 1 (see Table 4). Day 1 began with the VAS task, followed by the pre AXB task. Participants received training from Day 2 to Day 4 that ended with an everyday ID test. As in experiment 1, the HVPT included training stimuli of three Korean CV pseudowords, fortis, lenis, or aspirated Korean bilabial stop (i.e., /p'a/, /pa/, and pʰa/), produced by three female talkers and everyday ID tests were performed with a set of training stimuli from Talker 1. After the last training session (Day 4), each participant completed the new talker generalization test with stimuli produced by a novel talker (i.e., not a training talker). The post-test stage on Day 5 consists of two tests: the post AXB test and the new consonant AXB generalization test. The pretest, post-test, and new consonant generalization tests were in the AXB discrimination task format. The stimuli for the new consonant generalization were a set of stimuli of Korean three-way stop contrast in velar stops (i.e., fortis (/k'/), lenis (/k/), and aspirated (/kʰ/)) produced by Talker 1. The overall timeline of five days of the experimental phase is summarized in Table 11.

Table 11. Timeline of five-day experiment 2 with three-day of HVPT.

| | Participation Day | | | | |
|---|---|---|---|---|---|
| **Day** | Day 1 | Day 2 | Day 3 | Day 4 | Day 5 |
| **Tasks** | • VAS Task<br>• Pre AXB Test (n = 124)<br>• Language Questionnaire | • Cue-Attention Switching Training (n = 112)<br>• Daily Familiarization Phase<br>• HVPT Session (n= 159)<br>• Everyday Identification (ID) Test (n = 51) | | | • Post AXB Test<br>• New Consonant Generalization Test (n = 124) |
| | | | | + New Talker Generalization Test (n = 66) | |
| **Time** | About 30 - 40 minutes | | | | |
| **Location** | Laboratory | | | | |

### 3.1.3.2. Cue-attention switching training

In addition to each Korean training session from Day 2 to Day 4, participants in experiment 2 received the additional cue-attention switching training. This additional training had the two-alternative forced-choice (2AFC) paradigm with two response choices, either English ‘ba’ or ‘pa.’ In this task, participants were instructed that they will hear speech sound files spoken by an English native speaker and identify whether the sound file they just heard is either English ‘ba’ or ‘pa.’ After each stimulus was played, the computer screen presented two words ‘ba’ and ‘pa’, and participants were required to click one of the words based on their judgments. Before the next stimulus played, a burst of white noise was presented for 1,250 ms to mask the echoic memory of the previously presented token. The trial-by-trial feedback was provided on every trial. If participants identified the stimulus correctly, a smiley face with the chime sound was presented at the center of the computer screen. When participants wrongly identified the stimulus, the correct answer (i.e., either ‘ba’ or ‘pa’ word) was shown on the computer screen with a repetition of the stimulus.

This additional training phase aimed to switch participants’ attention from VOT to f0 cues in the perception of English bilabial stop voicing contrast, which the Korean three-way bilabial stop contrast is mapped onto. To increase irrelevant variability along the relatively less important dimension (i.e., VOT) to encourage participants to ignore it in the identification of English /ba/ and /pa/, the cue-attention switching training only used stimuli varying along with all seven VOT steps but only with the lowest and highest f0 steps were presented. This choice of stimuli resulted in a subset of fourteen stimuli which consisted of seven with the lowest f0 step and seven with the highest f0 step. The trial-by-trial feedback was determined only based on the f0 step of the stimuli. The correct responses for the seven stimuli with the lowest f0 step were always set to English 'ba', and the correct responses for the seven stimuli with the highest f0 step were always set to English 'pa'. In other words, regardless of the VOT steps of the stimuli, the English 'ba' responses were always correct as long as the stimuli had the lowest f0 step while the English 'pa' responses were always correct to the stimuli with the highest f0 step. This feedback was designed to make the f0 as an informative and predictive dimension for identifying English /ba/ and /pa/ during the training while conveying the message to participants that the VOT dimension is no longer informative for their decisions due to the random relationship between the VOT steps and English voicing contrast. It should be noted that sometimes, the feedback was not matched with participants' canonical identification patterns. For example, participants were highly likely to identify the stimuli with the lowest f0 step with relatively long VOT values as English /pa/. However, the feedback was always 'ba' for that stimulus due to the lowest f0 step. It was considered that this type of disagreement might intervene in some participants and make them not learn the informativeness of the f0 dimension if participants had high doubt on the validation of stimulus and feedback pairs. Therefore, the instruction informed participants that the speaker who produced the stimuli might speak a different dialect of English (e.g., British, Canadian, Irish, Scottish English, regional dialect, etc.). The cue-attention switching training was run using Paradigm software on laptop or desktop computers in the Phonetics Lab at the University of Wisconsin-Milwaukee. Each cue-attention switching training session took approximately 8 to 9 minutes. The fourteen stimuli were presented with eight times repetition (i.e., 112 trials). Together with the HVPT sessions, participation from Day 2 to Day 4 took about 25 to 30 minutes to complete.

### *3.1.4. Analysis*

The analytical methods for the VAS task and the HVPT were identical to those used in experiment 1.

#### 3.1.4.1. Cue-attention switching training

The percentage of correctness was calculated for each participant. Figure 23 shows the results of the cue-attention switching training in the percentage of correct. Participants' accuracy was high in the first session of the cue-attention switching training; however, there was a wide range of individual differences in the first session (*M* = 86.2, *SD* =12.7), indicating that some participants had a relatively hard time identifying the stimuli by relying only on the f0 dimension. However, participants hit the ceiling by the final training session (*M* = 97.1, *SD* =5.05), suggesting that the cue-attention switching training successfully trained participants to identify the training stimuli with the f0 cues.

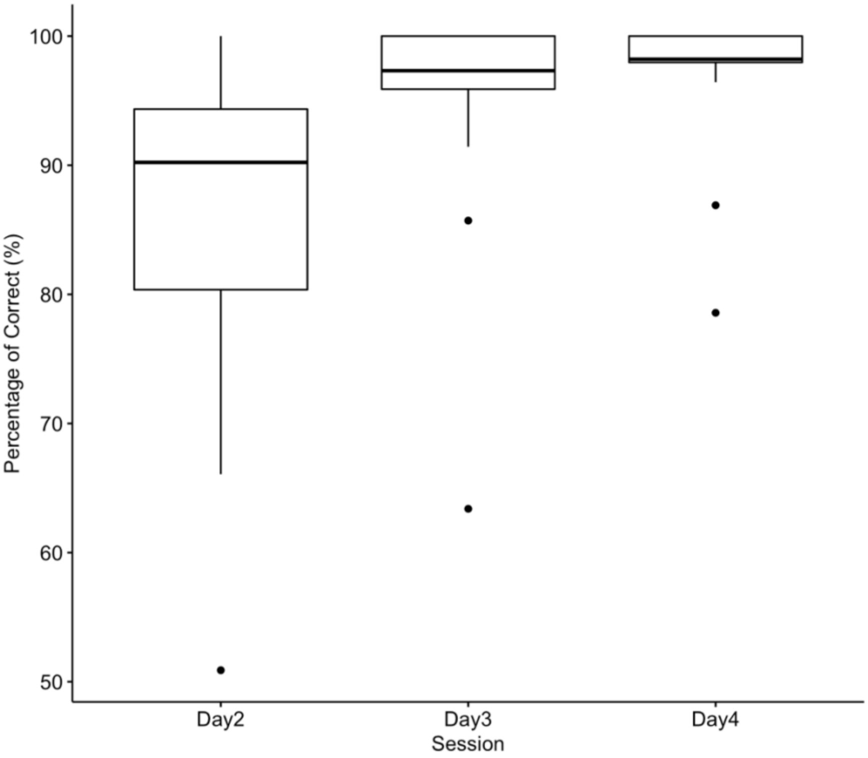


Figure 23. Results of the cue-attention switching training in the percentage of correct (%). *Boxplots: shaded region indicates interquartile range; whiskers extend to extreme values; solid bar indicates median; points indicate outliers.*

### *3.1.5. Results*

#### 3.1.5.1. The VAS task

As shown in Figure 24 and Figure 25, participants in experiment 2 showed individual variability in their VAS response patterns like participants in experiment 1. The curve fits (i.e., red lines) from the polynomial regression models were overlaid on top of the click distributions. The dotted lines in graphs represent the midpoints of the click locations. Some participants (i.e., C2_1, C2_2, C2_5, C2_6, C2_7, C2_8 in Figure 24) showed more categorical response patterns, most of the responses centered around the edges of the VAS scale, than others whose responses were distributed along the entire scale (i.e., G2_1, G2_2, G2_4, G2_5, G2_6, G2_8 in Figure

25). The quantified gradiency of the VAS responses was calculated for each participant in the same way in experiment 1 (i.e., coefficients of the quadratic regression curves overlaid on the histograms of the VAS responses. Participants were divided into two groups for the group analysis based on whether more than 70% of total responses were in the first and last 20% of the VAS scale. A total of ten participants were assigned to the categorical group, and nine participants were assigned to the gradient group. The groups were named Categorical 2 and Gradient 2 respectively. The quantified gradiency of participants in the gradient 2 group ($M = 0.000434499$, $SD = 0.0002800998$) was smaller than the ones of the categorical 2 participants ($M = 0.0020179016$, $SD = 0.0007130153$), which indicates that the categorical 2 group's VAS response patterns were more concave than the gradient 2 group. The independent $t$-test reached its significance ($p < .001$).

Two independent $t$-tests were performed to check whether the quantified gradiency between gradient groups and categorical groups in experiments 1 and 2 are not significantly different. The result showed that the quantified gradiency between gradient 1 and gradient 2 groups was not significantly different ($p = .50$), and the same result was found for the categorical groups as well ($p = .91$).

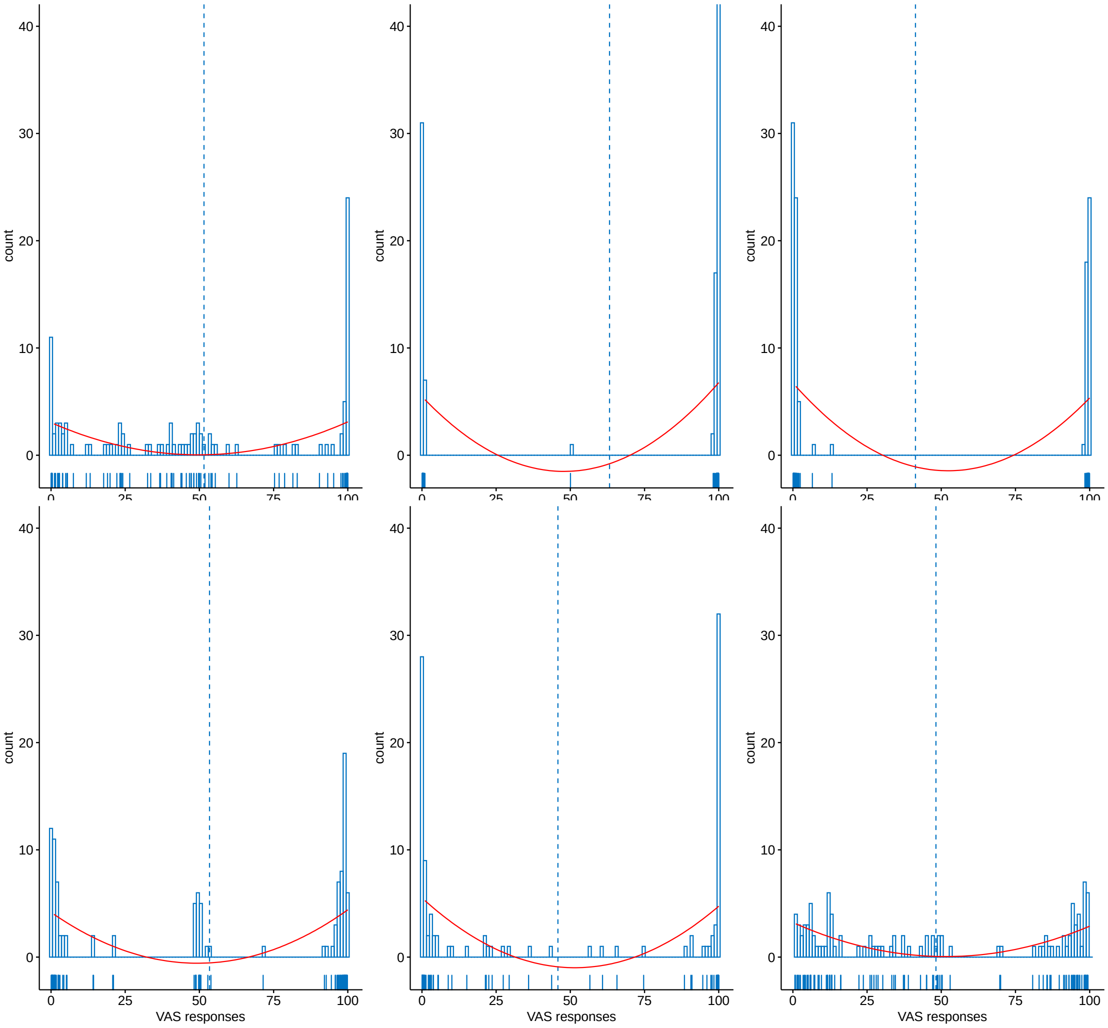


Figure 24. Distributions of the VAS task click locations of represented participants (C2_1, C2_2, C2_5, C2_6, C2_7, C2_8) who showed categorical responses. The curve fits (i.e., red lines) from the polynomial regression models were overlaid on the top of the click distributions. The dotted lines represent the midpoints of the click locations.

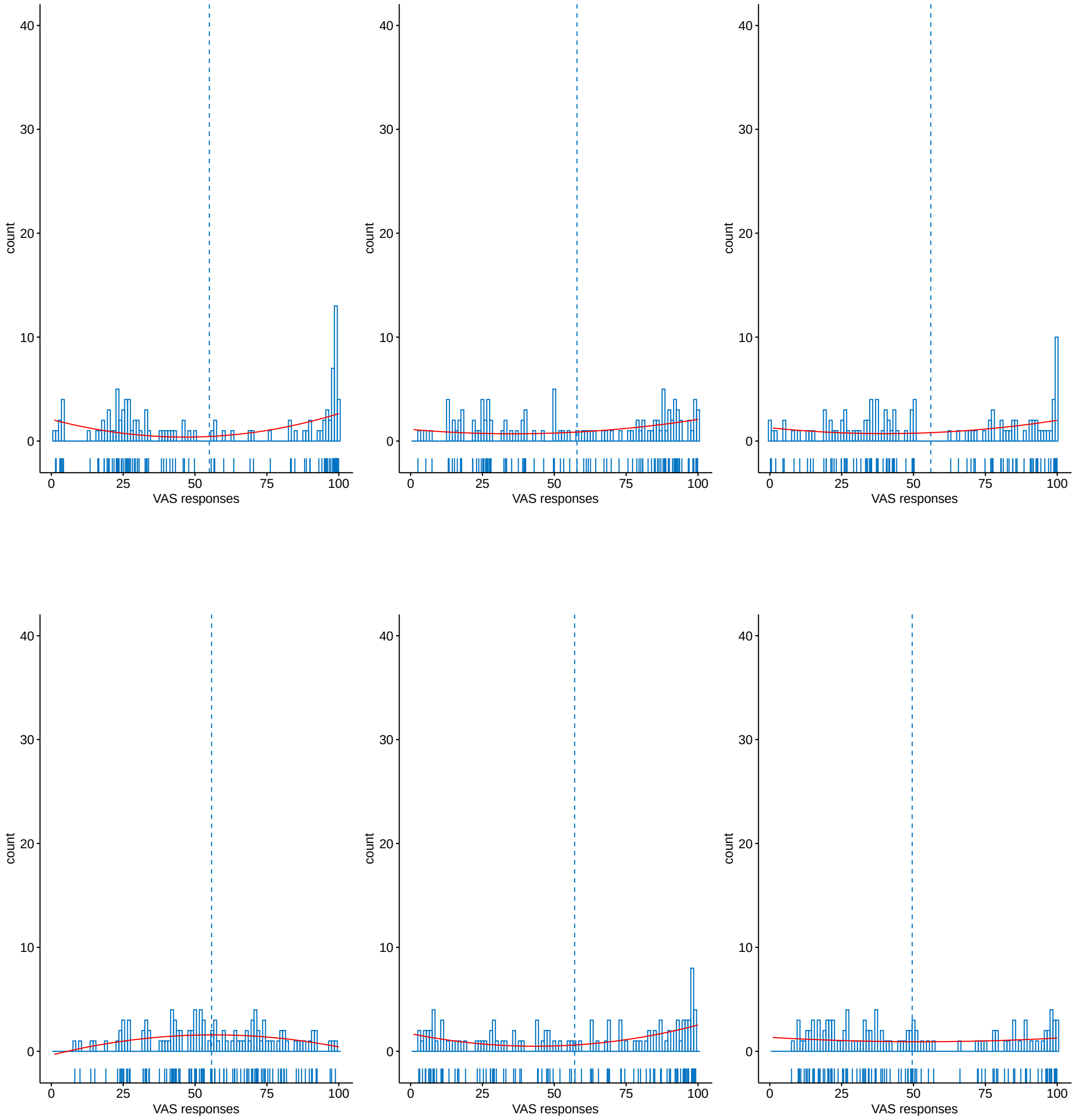


Figure 25. Distributions of the VAS task click locations of represented participants (G2_1, G2_2, G2_4, G2_5, G2_6, G2_8) who showed gradient responses. The curve fits (i.e., red lines) from the polynomial regression models were overlaid on the top of the click distributions. The dotted lines represent the midpoints of the click locations.

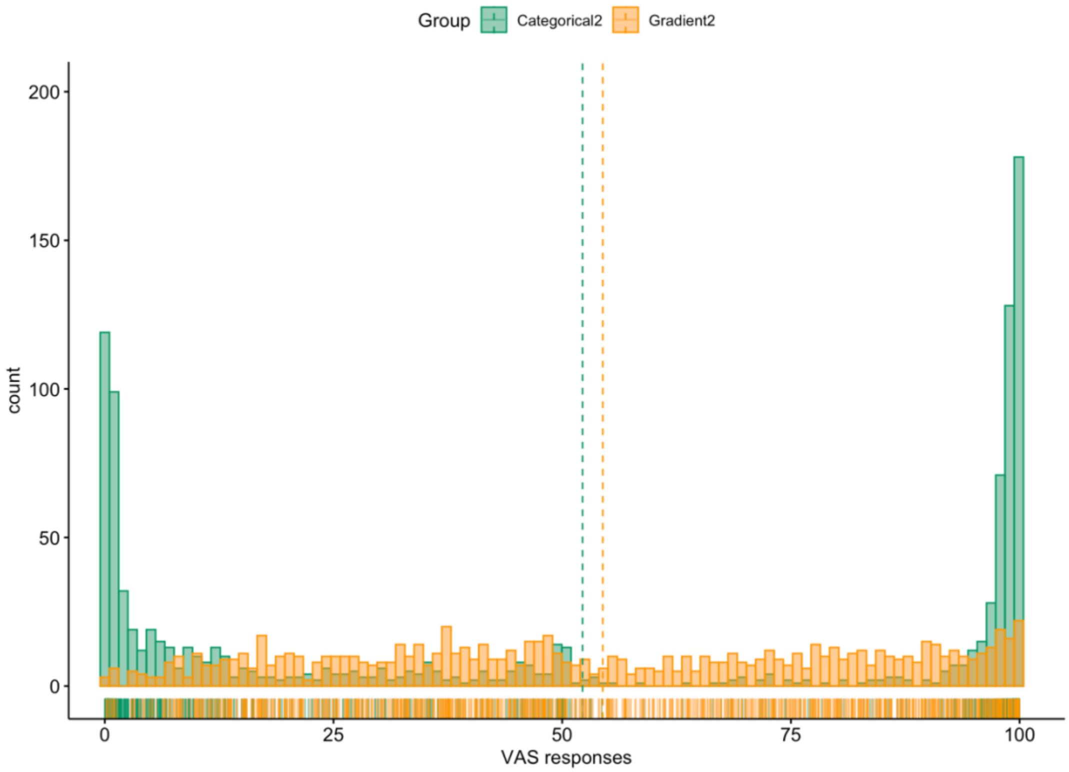


Figure 26. Distributions of the VAS task click locations averaged by groups (Categorical 2 vs. Gradient 2). The dotted lines represent the midpoints of the click locations for each group.

### 3.1.5.2. Results of HVPT by group (Categorical 2 vs. Gradient 2)

We used linear mixed-effect regression models with the *lmer* function from the *lme4* package (version 1.1-27) in *R*. Identical to experiment 1, two separate mixed-effect regression models were built. The first model was run for the pretest, post-test, and new consonant generalization test results, which had the AXB discrimination task type. Participants' accuracy in selecting the correct answer, either A or B, was submitted to the model with `Group` (Categorical 2 vs. Gradient 2), `Test`, and their 2-way interactions as fixed effects and random intercepts for participant and random slopes for participants for `Test` as random effects. The `Group` variable was centered to -0.5 and 0.5, and the `Test` variable was dummy coded to `Pre`

VERSUS Post and Pre VERSUS New Con contrasts so that post-test and new consonant generalization test scores were compared to pretest scores. The second model was constructed with the results of everyday ID test scores, which were participants' accuracy in selecting the correct answer from three choices. With everything identical to the first model, the Test variable was dummy coded to Day1 VERSUS Day2, Day1 VERSUS Day3, and Day1 VERSUS New Talker contrasts with Day 1 ID test scores (i.e., Day1) as the reference level.

Recall that experiment 2 was conducted to examine whether participants who show categorical VAS response patterns could overcome their difficulties learning the target Korean stop contrast through the cue-attention switching training. Therefore, the test results in the percentage of correct responses from the categorical 1 and the categorical 2 groups were compared via separate linear mixed-effect models. All fixed and random effects entered were identical, except that the Group variable was changed to Categorical 1 vs. Categorical 2. All linear mixed-effects models reported converging with bound optimization by quadratic approximation (BOBYQA optimization; Powell, 2009).

A mixed-effect logistic regression model was employed to examine whether the cue-attention switching training resulted in the changes of weighting of VOT and f0 dimensions from the beginning to the end of the training in the categorical 2 group compared to the categorical 1 group. It was expected that if the cue-attention switching training successfully shifted the categorical 2 group's attention to the f0 dimension, the use of f0 cues in identifying Korean lenis and aspirated stops would increase compared to the use of f0 cue by the categorical 1 group. Thus, the categorical 1 and the categorical 2 groups' Day 1 and Day 3 ID test responses were submitted to the mixed-effect logistic regression model. Only Korean lenis and aspirated responses were included and coded to 0 and 1, respectively. The fixed effects included standardized F0, VOT, and centered Day variables. The resulting beta-coefficients for a given variable show the change in log odds of the relevant response given a single standard-deviation increase in that variable.

#### *3.1.5.2.1. Pre, post, new consonant generalization tests*

Before discussing Figure 27, an independent *t*-test was conducted to check the differences in pretest scores between categorical 2 and gradient 2 groups. Result showed that the pretest scores between these two groups were not significantly different from each other ($p = .45$, $p = .08$).

Figure 27 shows the percentage of correctness of the pretest, post-test, and new consonant generalization test by the categorical 2 and the gradient 2 groups. Overall, both groups performed similarly, suggesting that the gradient 2 group's performances did not exceed the categorical group's performances by a wide margin. The average scores of these three tests by the categorical 2 group were 73.3%, 79.5%, and 77.3%, respectively, which did not show a large difference to the ones by the gradient 2 group, 75.6%, 81.5%, and 77.8%.

Table 12 summarizes linear mixed-effects models, including Group, Test (Pre VERSUS Post and Pre VERSUS New Con), and their interactions. The Group variable had no reliable main effect, suggesting no difference in test scores between categorical 2 and gradient 2 groups. There were reliable effects of Pre VERSUS Post and Pre VERSUS New Con, indicating an effect of training through improvements from pretest to post-test and from pretest to the new consonant generalization test; however, critically, there was no reliable interaction between Group and Test contrasts. Thus, unlike experiment 1, there was no reliable evidence of better learning in the gradient 2 group than in the categorical group. For significant main effects, post hoc pairwise comparisons showed a significant improvement of post-test scores compared to pretest scores from both categorical 2 and gradient 2 groups (*t*(21.2) = -3.37, *p* = .01 and *t*(21.2) = -3.0, *p* = .03, respectively). *P*-value adjustment was done by the Tukey method for comparing a family of 4 estimates.

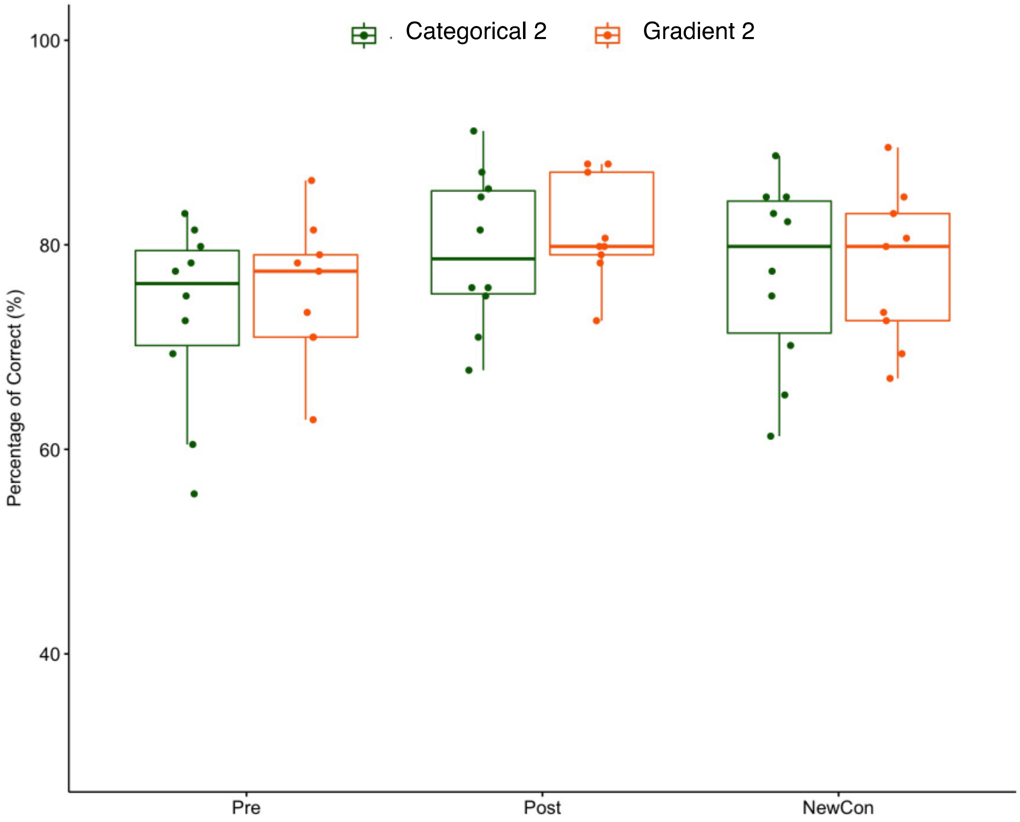


Figure 27. Results of the AXB pre-test, posttest, and new consonant generalization tests in percentage correct (%). *Boxplots: shaded region indicates interquartile range; whiskers extend to extreme values; solid bar indicates median; points indicate outliers.*

Table 12. Summary of fixed effects for the mixed-effects linear regression model of the pretest, post-test, and new consonant generalization tests scores.

| Predictor | Estimate ($\beta$) | Std. Error | *t-value* | *p-value* | |
|---|---|---|---|---|---|
| (Intercept) | 77.45 | 1.52 | 51.08 | <0.001 | *** |
| Group | 1.59 | 3.04 | 0.52 | 0.606 | |
| Pre_VERSUS_Post | 6.03 | 1.27 | 4.77 | <0.001 | *** |
| Pre_VERSUS_NewCon | 3.10 | 1.27 | 2.45 | 0.019 | * |
| Group:Pre_VERSUS_Post | -0.39 | 2.53 | -0.15 | 0.880 | |
| Group:Pre_VERSUS_NewCon | -1.80 | 2.53 | -0.71 | 0.482 | |

#### *3.1.5.2.2. Everyday ID tests & new talker generalization test*

Figure 28 shows a similar trend as in Figure 27. Inconsistent with the results from experiment 1, categorical 2 and gradient 2 groups were not much different in everyday ID tests. Notably, both groups behaved quite similarly on all three everyday ID tests. The average ID test scores of the categorical 2 group were 71.2%, 75.5%, and 78.2%, and the average scores of the gradient 2 group were 68.2%, 75%, and 82.6%. It should be noted that Figure 28 seems to suggest a wider range of individual variability in learning in the categorical 2 group than in the gradient group. If we compare the interquartile ranges of box plots on Day 2 and Day 3 ID tests, the ranges of the categorical 2 group were larger than the ones of the gradient 2 group. This visual inspection suggests individual differences in the effeteness of the cue-attention switching training. Some categorical learners might benefit from the cue-attention switching training and result in a higher learning outcome, while others might not and still show their struggles during training. This issue is addressed in more detail in the discussion section (see 4.2).

Although both groups showed similar average scores on the new talker generalization test (67.3% for categorical 2 and 70.2% for gradient 2), in terms of the new talker generalization test, it seems that the generalization test scores by both groups were very similar to their average scores of the Day 1 ID test (67.3% for categorical 2 and 70.2% for gradient 2). This result suggests that both groups failed to show their ability to generalize their learning to novel stimuli.

Table 13 summarizes linear mixed-effects models, including Group, Test (Day1 VERSUS Day2, Day1 VERSUS Day3, and Day1 VERSUS New Talker), and their interactions. There was no reliable main effect of Group, reflecting that the categorical 2 and the gradient 2 groups did not show different performances across everyday ID tests and the new talker generalization test. The model also found significant main effects of Day1 VERSUS Day2 and Day1 VERSUS Day3, showing that participants improved overall in training over time (Day 1: 69.8%, Day 2: 75.3%, and Day 3: 80.3%). However, participants did not show greater improvement on the new talker generalization test than their performances on the Day1 ID test. More importantly, Group variable was not involved in significant interactions with any combination of Day1 VERSUS Day2, Day1 VERSUS Day3, and Day1 VERSUS New Talker. Thus, unlike experiment 1, there was no reliable evidence of better learning in the gradient group than in the categorical group.

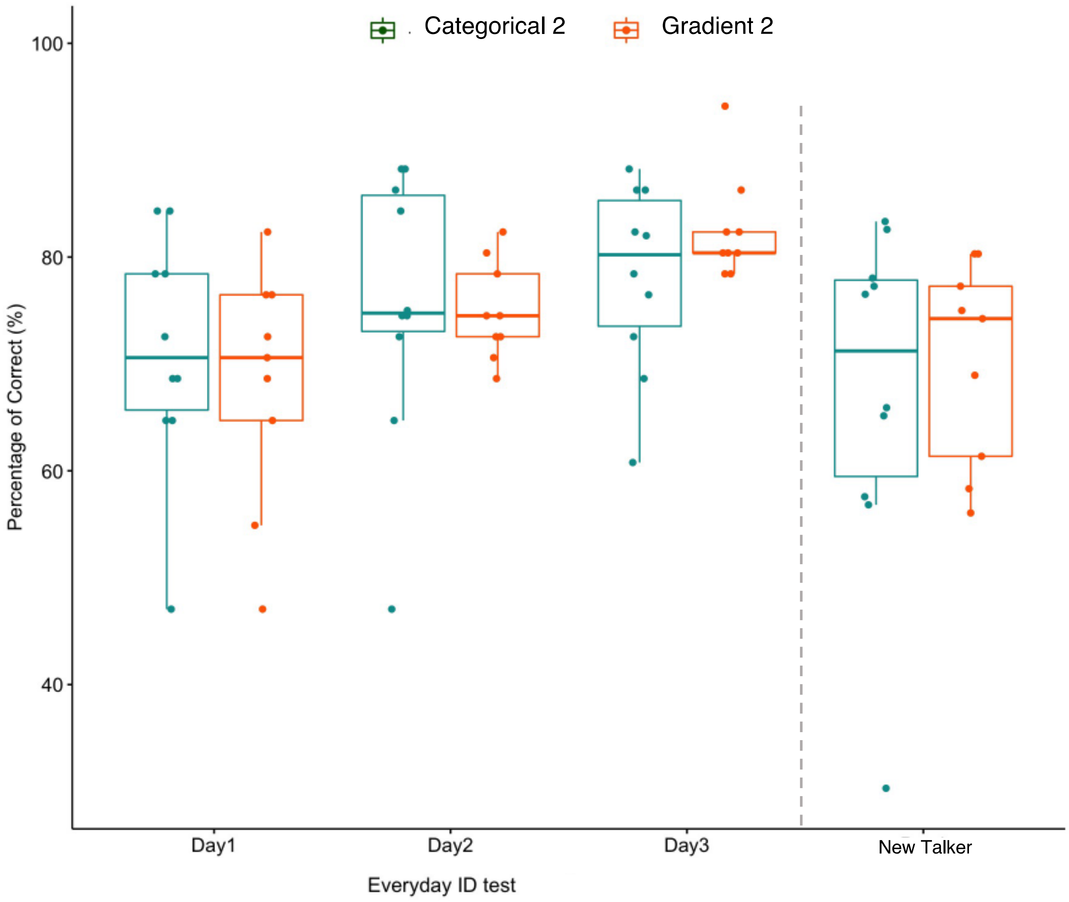


Figure 28. Results of everyday ID tests and the new talker generalization test in percentage correct (%). *Boxplots: shaded region indicates interquartile range; whiskers extend to extreme values; solid bar indicates median; points indicate outliers.*

Table 13. Summary of fixed effects for the mixed-effects linear regression model of everyday ID test scores.

| Predictor | Estimate ($\beta$) | Std. Error | *t-value* | *p-value* | |
|---|---|---|---|---|---|
| (Intercept) | 73.50 | 1.74 | 42.22 | <0.001 | *** |
| Group | -0.14 | 2.98 | -0.05 | 0.963 | |
| Day1_VERSUS_Day2 | 5.50 | 1.93 | 2.85 | 0.007 | ** |
| Day1_VERSUS_Day3 | 10.51 | 1.93 | 5.45 | <0.001 | *** |
| Day1_VERSUS_New Talker | -1.06 | 2.99 | -0.36 | 0.725 | |
| Group:Day1_VERSUS_Day2 | 3.32 | 4.01 | 0.83 | 0.413 | |
| Group:Day1_VERSUS_Day3 | -1.49 | 4.01 | -0.37 | 0.713 | |
| Group:Day1_VERSUS_New Talker | 2.46 | 5.96 | 0.41 | 0.682 | |

#### *3.1.5.2.3. Comparison of the categorical 1 and the categorical 2 groups*

Since I was specifically interested in group differences between the categorical 1 group in experiment 1 and the categorical 2 group in experiment 2, I included additional analysis comparing two categorical groups for the discrimination tasks (pretest, post-test, new consonant generalization test) and the training (everyday ID tests). Linear mixed-effect regression model was run over the categorical 1 and 2 groups data, predicting their percentage of correctness with fixed factors of Group (Categorical 1 vs. Categorical 2) and Test (Pre VERSUS Post and Pre VERSUS New Con) and their interactions. Table 14 summarizes the results. The model revealed a reliable main effect of Group and significant interactions between Group and Pre VERSUS Post and Pre VERSUS New Con. Another linear mixed-effect regression model was run with Test variable coded to Day1 VERSUS Day2, Day1 VERSUS Day3, and Day1 VERSUS New Talker, and the results are summarized in Table 15. The model found a significant main effect of Group and a significant interaction between Group and Day1 VERSUS Day3. These results indicate that the categorical 2 group which received the additional cue-attention switching training with HVPT, learned more than the categorical 1 group which was trained only with HVPT sessions.

Table 14. Summary of fixed effects for the mixed-effects linear regression model for the pretest, post-test, and new consonant generalization tests scores (Categorical 1 vs. Categorical 2).

| Predictor | Estimate ($\beta$) | Std. Error | *t*-value | *p*-value | |
|---|---|---|---|---|---|
| (Intercept) | 73.43 | 1.92 | 38.16 | <0.001 | *** |
| Group | 5.98 | 3.85 | 1.55 | 0.135 | |
| Pre_VERSUS_Post | 2.96 | 1.28 | 2.32 | 0.027 | * |
| Pre_VERSUS_NewCon | 0.28 | 1.18 | 0.23 | 0.817 | |
| Group:Pre_VERSUS_Post | 5.96 | 2.56 | 2.33 | 0.026 | * |
| Group:Pre_VERSUS_NewCon | 6.74 | 2.37 | 2.85 | 0.010 | ** |

Table 15. Summary of fixed effects for the mixed-effects linear regression model for everyday ID test scores (Categorical 1 vs. Categorical 2).

| Predictor | Estimate ($\beta$) | Std. Error | *t-value* | *p-value* | |
|---|---|---|---|---|---|
| (Intercept) | 65.13 | 2.12 | 30.78 | <0.001 | *** |
| Group | 14.55 | 4.23 | 3.44 | 0.002 | ** |
| Day1_VERSUS_Day2 | 0.74 | 1.95 | 0.38 | 0.708 | |
| Day1_VERSUS_Day3 | 2.10 | 1.95 | 1.08 | 0.288 | |
| Day1_VERSUS_New Talker | -3.96 | 2.87 | -1.38 | 0.178 | |
| Group:Day1_VERSUS_Day2 | 6.65 | 3.90 | 1.71 | 0.096 | . |
| Group:Day1_VERSUS_Day3 | 9.02 | 3.90 | 2.31 | 0.026 | * |
| Group:Day1_VERSUS_New Talker | 0.24 | 5.73 | 0.04 | 0.967 | |

#### *3.1.5.2.4. Weighting of VOT and f0 cues (Categorical 1 vs. Categorical 2)*

A mixed-effects logistic regression model was built with the results of Day 1 and Day 3 ID tests. Participants' responses for the lenis /pa/ and aspirated /pʰa/ stops were coded as 0 and 1, respectively, excluding fortis /p'a/ responses. The model included fixed effects for `Group` (Categorical 1 vs. Categorical 2), `Day` (Day1 vs. Day3), `VOT`, and `F0`, as well as their interactions.

Figure 29 shows participants' predicted response curves from the mixed-effect logistic regression model across different levels of VOT (left graph) and f0 (right graph). The y-axis represents the estimated proportion of Korean /pʰa/ responses controlling the effect of f0 or VOT dimension. Each panel represents either Day 1 or Day 3 ID test (coded to -0.5 and 0.5), and the results from the different groups; Categorical 1 vs. Categorical 2 are represented in different colors. Figure 29 compares how each group was different in the weight of VOT and f0 dimensions when participants identified Korean lenis and aspirated stops. In particular, the graphs on the right-side show that the categorical 2 group's pattern for stimuli with low f0 steps are mainly predicted to be identified as Korean lenis stops, while stimuli with high f0 steps were classified as Korean aspirated stops. Although the categorical 2 groups showed a similar pattern of using VOT dimension in identifying Korean lenis and aspirated stops, they primarily relied on f0 dimensions as evidenced by steeper lines of the categorical group 2 in the right panel than the ones in the left graph. In terms of group comparison, it is clearly shown that the categorical 2 group more strongly utilized the f0 dimension in the identification of Korean stops than the categorical 1 group in both ID tests. Note that the categorical 1 group did not even seem to increase the perceptual weighting of the f0 dimension after the last day of HVPT (i.e., Day 3).

The results of the regression analysis are shown in Table 16. The model found significant two-way and three-way interactions, suggesting that categorical 1 and categorical 2 groups were different in the use of f0 and VOT dimensions, and the use of dimensions changed differently over the training in each group. A significant two-way interaction between `Group` and `F0` shows that the categorical 2 group relied more on the f0 dimension than the categorical group 1 to identify Korean lenis and aspirated stops. The two-way interaction between `F0` and `Day` suggests an increase in the perceptual weighting of the f0 dimension after training. More importantly, the model found a significant three-way interaction between `Group`, `F0`, and `Day`, which indicates that the increase in the weighing of the f0 dimension was larger for the categorical 2 group than the categorical 1 group.

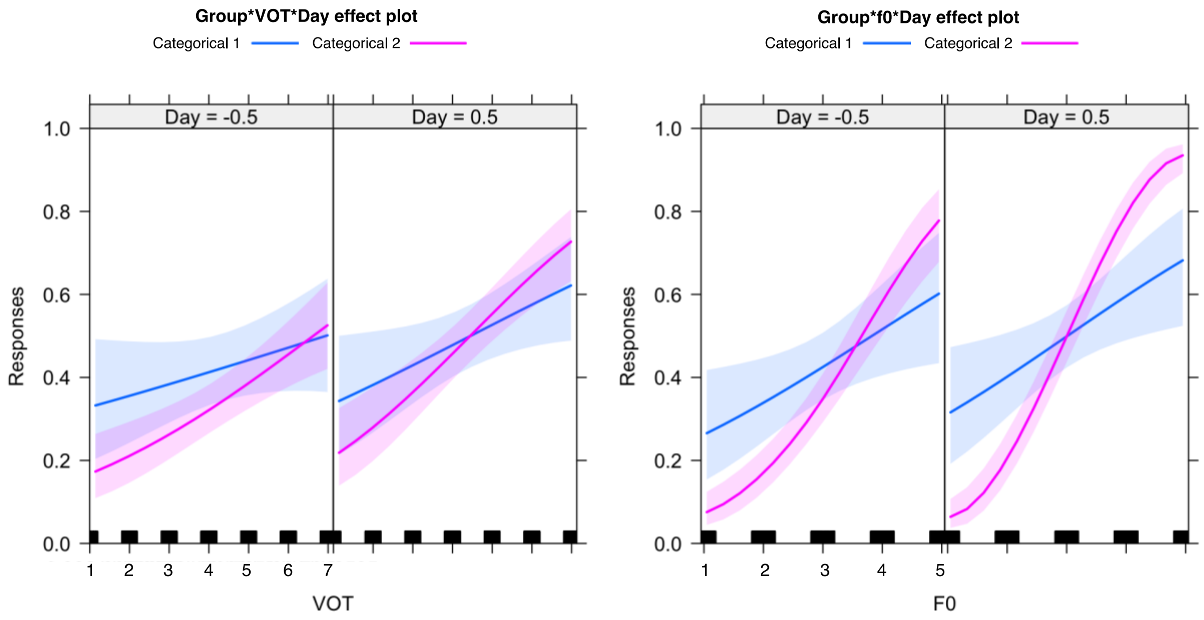


Figure 29. Predicted logit curves for perception of the target Korean stop contrast by native English participants. Predictions are drawn from a binary logistic regression model with `Group`, `Day`, `VOT`, and `F0` as predictors. Each panel represents the results from Day 1 (-0.5) and Day 3 (0.5) ID tests. The y-axis of each panel represents the predicted probability of “aspirated” responses. Results from Categorical 1 and Categorical 2 groups are shown in different colors.

Table 16. Summary of fixed effects for the mixed-effects logistic regression model predicting changes in f0 and VOT cue weights in everyday ID tests (Categorical 1 vs. Categorical 2).

| Predictor | Estimate ($\beta$) | Std. Error | *z-value* | *p-value* | |
|---|---|---|---|---|---|
| (Intercept) | -0.26 | 0.08 | -3.25 | 0.001 | ** |
| Group | -0.16 | 0.16 | -0.99 | 0.321 | |
| F0 | 2.50 | 0.25 | 10.00 | <0.001 | *** |
| VOT | 1.00 | 0.18 | 5.66 | <0.001 | *** |
| Day | 0.51 | 0.13 | 4.07 | <0.001 | *** |
| Group:F0 | 2.20 | 0.51 | 4.32 | <0.001 | *** |
| Group:VOT | 0.65 | 0.36 | 1.79 | 0.074 | . |
| Group:Day | 0.32 | 0.25 | 1.26 | 0.208 | |
| F0:Day | 0.79 | 0.23 | 3.43 | 0.001 | *** |
| VOT:Day | 0.34 | 0.21 | 1.61 | 0.108 | |
| Group:F0:Day | 1.07 | 0.44 | 2.43 | 0.015 | * |
| Group:VOT:Day | 0.09 | 0.42 | 0.22 | 0.824 | |

### 3.1.5.3. Results of HVPT by individuals in experiment 2

In the following sections, I will discuss the results of analyses, including an individual-level factor, which is the gradiency of the VAS responses. How participants' quantified gradiency is related to their performance in learning the target Korean three-way stop contrast is examined. Unlike experiment 1, participants in experiment 2 received the additional cue-attention switching training. Thus, it was expected that participants' quantified gradiency would not affect training as much as it had in experiment 1. The results of three analyses are presented below: linear mixed-effect regression models investigating the effect of individual differences in gradiency of the VAS task responses on the test scores, a mixed-effects logistic regression model examining the relationship between participants' quantified gradiency and their changes of weighting of VOT and f0 dimension in the identification of Korean lenis and aspirated stops, and the correlation between individual differences in gradiency and VOT- and f0-tuned coefficients ($\beta_{\mathrm{VOT}}$ and $\beta_{\mathrm{f0}}$). The statistical methods for these analyses were identical to those adapted in experiment 1.

#### *3.1.5.3.1. Test scores*

Two separate linear mixed-effects regression models were built with fixed factors of Gradiency (quantified gradiency of the VAS responses), Test, and their 2-way interactions. The Test variable was coded to Pre VERSUS Post and Pre VERSUS New Con contrasts for the first model and Day1 VERSUS Day2, Day1 VERSUS Day3, and Day1 VERSUS New Talker contrasts for the second model. Gradiency was standardized prior to analysis.

Results of the regression analysis with participants' pretest, post-test, and the new consonant generalization test scores are summarized in Table 17. There was no reliable main effect of Gradiency, indicating that individual differences in the VAS task responses did not affect participants' scores of three discrimination tests. The model found significant main effects of Pre VERSUS Post and Pre VERSUS New Con, suggesting that all participants in experiment 2 improved from the pretest to the post-test and generalized their learning to discriminate novel Korean three-way stop contrast. Most importantly, the model did not find any significant two-way interactions between Gradiency and Test variables. This result indicates that participants resulted in successful learning of the target contrast and a successful transfer of training to the perception of the new Korean stop contrast, and individual differences in Gradiency did not seem to affect the degree of the improvement in the post-test or generalization of learning in the new consonant generalization test.

The results of the second regression model with everyday ID test scores are shown in Table 18. In line with the results of the first model, the second model did not find a reliable main effect of Gradiency, suggesting that individual differences in Gradiency did not influence participants' improvement in training. There were significant main effects of Day1 VERSUS Day2 and Day1 VERSUS Day3, indicating that participants improved their identification of Korean three-way stop contrast. However, Day1 VERSUS New Talker did not reach its significance level, showing that participants failed to transfer training to identify stimuli produced by a novel talker. Critically, the model did not find significant two-way interactions between Gradiency and Test. This result demonstrates that participants with a relatively steeper response pattern of the VAS task did not perform differently from participants with a relatively gradual response pattern.

Table 17. Summary of fixed effects for the mixed-effects linear regression model of the pretest, post-test, and new consonant generalization tests scores.

| Predictor | Estimate ($\beta$) | Std. Error | *t*-value | *p*-value | |
|---|---|---|---|---|---|
| (Intercept) | 77.45 | 1.52 | 50.93 | <0.001 | *** |
| Gradiency | 1.23 | 3.06 | 0.40 | 0.692 | |
| Pre_VERSUS_Post | 6.03 | 1.26 | 4.79 | <0.001 | *** |
| Pre_VERSUS_NewCon | 3.10 | 1.26 | 2.46 | 0.019 | * |
| Gradiency:Pre_VERSUS_Post | 0.56 | 2.53 | 0.22 | 0.826 | |
| Gradiency:Pre_VERSUS_NewCon | 2.35 | 2.53 | 0.93 | 0.359 | |

Table 18. Summary of fixed effects for the mixed-effects linear regression model of everyday ID tests.

| Predictor | Estimate ($\beta$) | Std. Error | *t*-value | *p*-value | |
|---|---|---|---|---|---|
| (Intercept) | 73.50 | 1.74 | 42.22 | <0.001 | *** |
| Gradiency | -0.14 | 2.98 | -0.05 | 0.963 | |
| Day1_VERSUS_Day2 | 5.50 | 1.93 | 2.85 | 0.007 | ** |
| Day1_VERSUS_Day3 | 10.51 | 1.93 | 5.45 | <0.001 | *** |
| Day1_VERSUS_New Talker | -1.06 | 2.99 | -0.36 | 0.725 | |
| Gradiency:Day1_VERSUS_Day2 | 3.32 | 4.01 | 0.83 | 0.413 | |
| Gradiency:Day1_VERSUS_Day3 | -1.49 | 4.01 | -0.37 | 0.713 | |
| Gradiency:Day1_VERSUS_New Talker | 2.46 | 5.96 | 0.41 | 0.682 | |

#### *3.1.5.3.2. Relation between the quantified VAS gradiency and changes in acoustic cue reliance*

A mixed-effects logistic regression analysis was implemented with the *glmer*() function to analyze participants' perceptual patterns in their perception of Korean stops as a function of VOT and f0 dimensions. Specifically, the primary interest of this analysis was to examine how individual differences in gradiency influence changes in participants' weights in VOT and f0 dimensions during the training to identify confusing Korean lenis and aspirated stops. It was expected that if the additional cue-attention switching training effectively locates participants' attention to the f0 dimension, all participants would show an increase in the use of f0 cues in the perception of Korean stops regardless of the differences in their response patterns in the VAS task. Participants' responses in Day 1 and Day3 ID tests were submitted to the analysis, excluding responses for the Korean fortis stop category, with fixed factors of Gradiency, Day, VOT, and F0 with their two- and three-way interactions of our interests. All variables (Gradiency, Day, VOT, and F0) were standardized prior to analysis such that the resulting beta-coefficients for a given variable show the change in log odds of the aspirated /pʰa/ response given a single standard-deviation increase in that variable. The overall mixed-effects logistic regression model results are summarized in Table 19.

Figure 30 shows the estimated response curves from the regression model investigating the effect of individual differences in gradiency of the VAS task responses on the proportion of Korean aspirated stop (i.e., /pʰa/) responses for the everyday ID test stimuli across days as a function of VOT and f0 steps. The left-side graphs represent the estimated proportion of /pʰa/ responses for standardized five f0 steps, while the right-side graphs show the estimated proportion of /pʰa/ responses for the standardized five VOT steps. The results from the different everyday ID tests (i.e., Day 1 vs. Day 3) are represented in different colors. The visual inspection of Figure 30 shows that test stimuli with low f0 value (i.e., f0 = -0.7) are mainly classified as Korean lenis stops; on the other hand, test stimuli with a high f0 value (i.e., f0 = 0.7) are identified as Korean aspirated stops. The reliance on the f0 dimension seems to increase on the Day 3 ID test compared to the Day 1 ID test, which was shown by a larger difference of the estimated /pʰa/ responses between the lowest and highest f0 steps in Day 3 ID test. In addition, Figure 30 critically shows is that individual differences in the gradiency do not seem to be associated with participants' reliance on the f0 dimension in identifying Korean lenis and aspirated stops. In comparison to Figure 16, which shows the influence of individual participants' gradiency on the use of f0 dimension in the perception of lenis and aspirated contrast, Figure 30 indicates that all participants used the f0 dimension as the major cue to distinguish the lenis vs. aspirated contrast.

The model in Table 19 shows that VOT and F0 significantly influenced participants' responses to the test stimuli. However, F0 was a more important dimension than VOT in separating aspirated responses from lenis since the beta-coefficient for F0 was larger than the one for VOT ($\beta_{f0} = 3.53$, $\beta_{VOT} = 0.93$). The model shows no reliable two-way interactions between Gradiency and each of the acoustic dimensions, suggesting that regardless of the level of gradiency, participants in experiment 2 demonstrated greater use of f0 than VOT in classifying lenis and aspirated stops. There was a reliable two-way interaction between Day and F0, indicating that participants got to rely more on the f0 dimension in the Day 3 ID test than the Day 1 ID test, as shown by the positive $\beta_{f0}$ value (1.40).

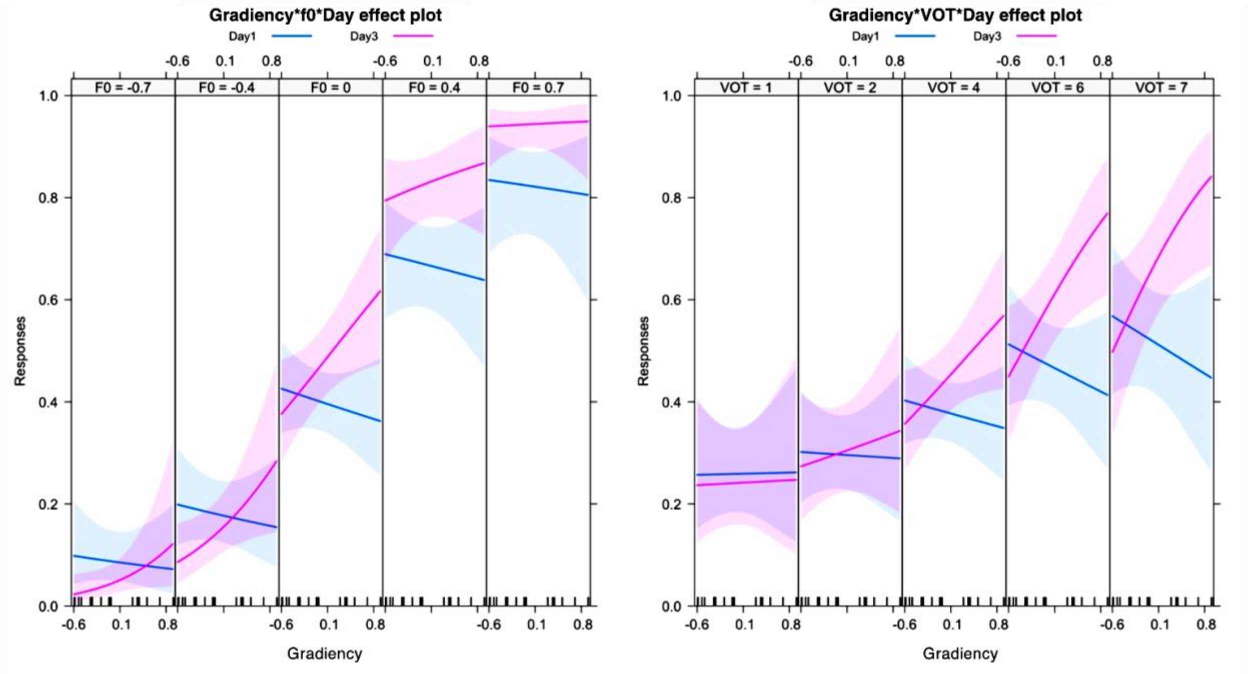


Figure 30. Predicted logit curves for perception of the target Korean stop contrast by native English participants. Predictions are drawn from a binary logistic regression model with VOT and F0 as predictors. The left-side graphs represent the estimated proportion of aspirated stop (i.e., /pʰa/) responses for standardized five f0 steps, while the right-side graphs show the estimated proportion of /pʰa/ responses for the standardized five VOT steps. The results from the different everyday ID tests (i.e., Day 1 vs. Day 3) are represented in different colors.

Table 19. Summary of fixed effects for the mixed-effects logistic regression model predicting changes in f0 and VOT cue weights in everyday ID tests.

| Predictor | Estimate ($\beta$) | Std. Error | *z*-value | *p*-value | |
|---|---|---|---|---|---|
| (Intercept) | -0.27 | 0.10 | -2.73 | 0.006 | ** |
| Gradiency | 0.22 | 0.20 | 1.13 | 0.257 | |
| F0 | 3.53 | 0.36 | 9.76 | <0.001 | *** |
| VOT | 0.93 | 0.23 | 4.12 | <0.001 | *** |
| Day | 0.30 | 0.16 | 1.88 | 0.060 | . |
| Gradiency:F0 | -0.35 | 0.63 | -0.55 | 0.580 | |
| Gradiency:VOT | 0.28 | 0.43 | 0.64 | 0.524 | |
| Gradiency:Day | 0.88 | 0.34 | 2.60 | 0.009 | ** |
| F0:Day | 1.40 | 0.36 | 3.94 | <0.001 | *** |
| VOT:Day | 0.41 | 0.31 | 1.33 | 0.184 | |
| Gradiency:F0:Day | -0.85 | 0.80 | -1.06 | 0.289 | |
| Gradiency:VOT:Day | 0.81 | 0.66 | 1.23 | 0.220 | |

### *3.1.5.3.3. Correlation between the quantified VAS gradiency and VOT and f0 cue uses*

As in experiment 1, individual participants' cue uses were computed via separate logistic regression analysis for each participant's Day 3 ID test responses with VOT and F0 as predictors of perceived Korean lenis and aspirated stops. VOT and F0 were standardized before the analysis. The beta-coefficients ($\beta_{f0}$ and $\beta_{VOT}$) from each logistic model were taken as an approximation of a given participant's reliance on a given cue for the relevant comparison, and since they are based on standardized values for each dimension, they can be compared to one another in order to determine the relative weighting of each dimension in predicting response patterns of Day 3 ID test. And then, the partial correlation analysis was performed to investigate the relationships between the beta-coefficients of each acoustic dimension and the quantified gradiency of the VAS responses. The full results of participants' coefficients can be found in Appendix I.

Before the correlation analysis, I conducted a visual inspection of plots showing participants' categorization patterns across each combination of VOT and f0 dimensions of everyday ID test stimuli. Figure 31 shows categorical 2 and the gradient 2 groups' responses to the test stimuli, and each cell represents each test stimulus in the combinations of VOT and f0 steps. Overall, both groups showed similar patterns of identifying Korean stops. Both groups classified stimuli with mid-to-long VOT values based on their f0 values; stimuli with low f0 were consistently classified as lenis stops, while stimuli with high f0 were consistently classified as aspirated stops. The results suggest a benefit of the cue-attention switching training in that it reduced the group differences in the use of the f0 dimension to identify Korean lenis and aspirated stops. More specifically, for both categorical 2 and gradient 2 groups, f0 dimension had a significant effect on the choice of stop type, especially lenis and aspirated category, with higher f0 eliciting more aspirated responses.

For visual inspection at an individual level, Figure 32 and Figure 33 show the identification responses of the six selected participants. They showed a similar identification pattern to some extent with the very reliable use of the f0 dimension to distinguish lenis and aspirated stops. The six participants in experiment 2 seemed to actively utilize the f0 dimension as a primary cue to identify lenis and aspirated contrast, which some of the participants in experiment 1 struggled to distinguish in terms of f0 differences between the test stimuli (see Figure 18). If we recall the perceptual patterns of selected participants in experiment 1, there was considerable variability between participants in the weights of f0 cues to identify lenis and aspirated contrast. However, the individual variability seemed to be reduced among participants in experiment 2, especially after completing the whole HVPT period (i.e., Day 3 ID test). Together with the group results reported in the previous paragraph, individual results also suggest a benefit of the cue-attention switching training to rely on the f0 cues as native listeners of Korean.

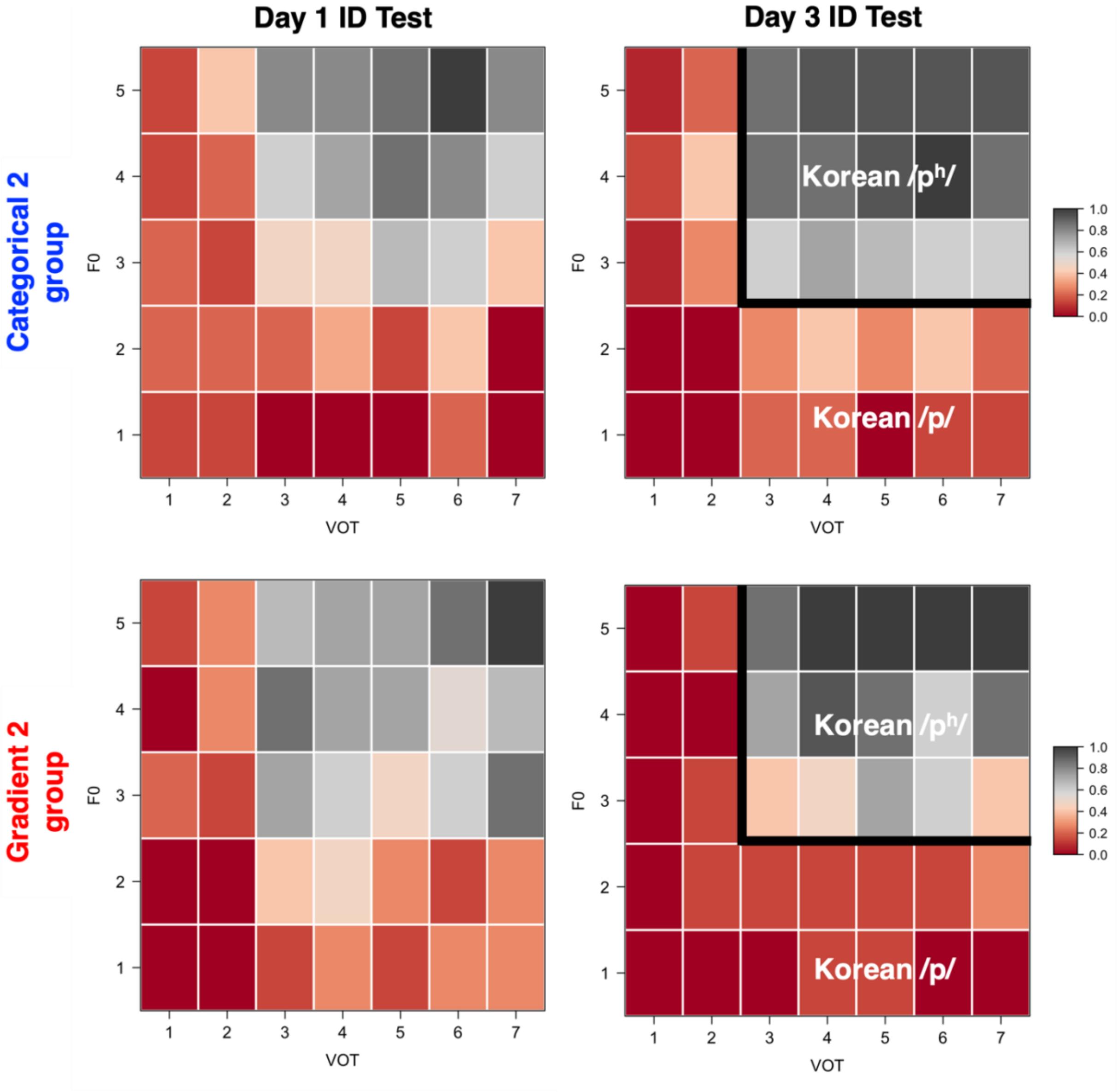
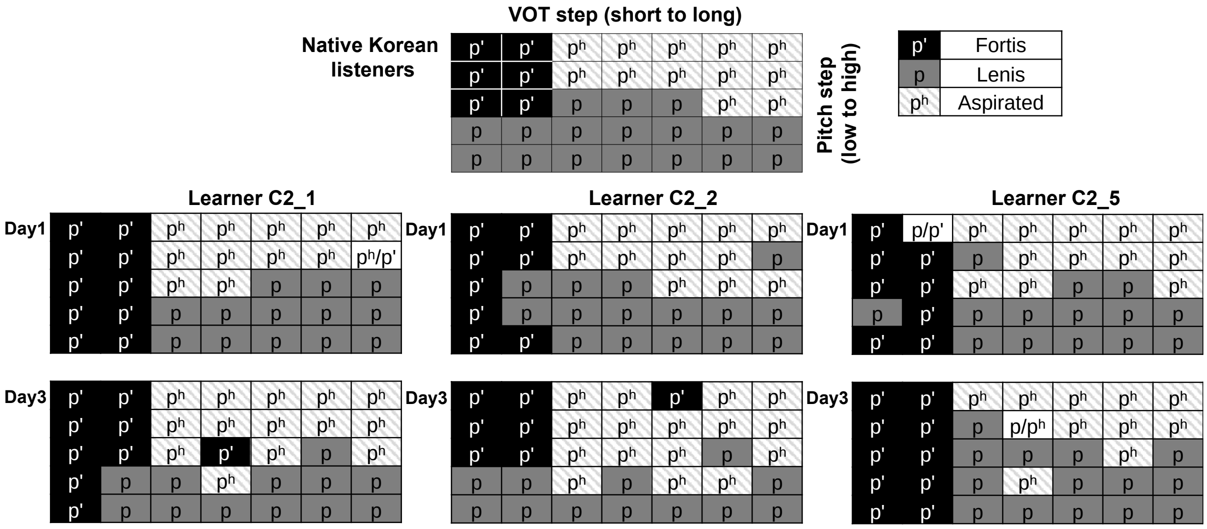


Figure 31. Heat plots of participants' overall responses of the Day 1 ID test and the Day 3 ID test to each combination of VOT and f0 (pitch) dimensions. The darkness of the cell represents the percentage of responses in a forced-choice task; the darkest grey cells elicited 100% Korean aspirated stop (/p$^h$/) responses, while the darkest red cells elicited 0% Korean aspirated stop (/p$^h$/) responses.



Native Korean listeners

| | | | | | | |
|---|---|---|---|---|---|---|
| p' | p' | pʰ | pʰ | pʰ | pʰ | pʰ |
| p' | p' | pʰ | pʰ | pʰ | pʰ | pʰ |
| p' | p' | p | p | p | pʰ | pʰ |
| p | p | p | p | p | p | p |
| p | p | p | p | p | p | p |

| | |
|---|---|
| p' | Fortis |
| p | Lenis |
| pʰ | Aspirated |

Learner C2_1

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Day1 | p' | p' | pʰ | pʰ | pʰ | pʰ | pʰ |
| | p' | p' | pʰ | pʰ | pʰ | pʰ | pʰ/p' |
| | p' | p' | pʰ | pʰ | p | p | p |
| | p' | p' | p | p | p | p | p |
| | p' | p' | p | p | p | p | p |
| Day3 | p' | p' | pʰ | pʰ | pʰ | pʰ | pʰ |
| | p' | p' | pʰ | pʰ | pʰ | pʰ | pʰ |
| | p' | p' | pʰ | p' | pʰ | p | pʰ |
| | p' | p | p | pʰ | p | p | p |
| | p' | p | p | p | p | p | p |

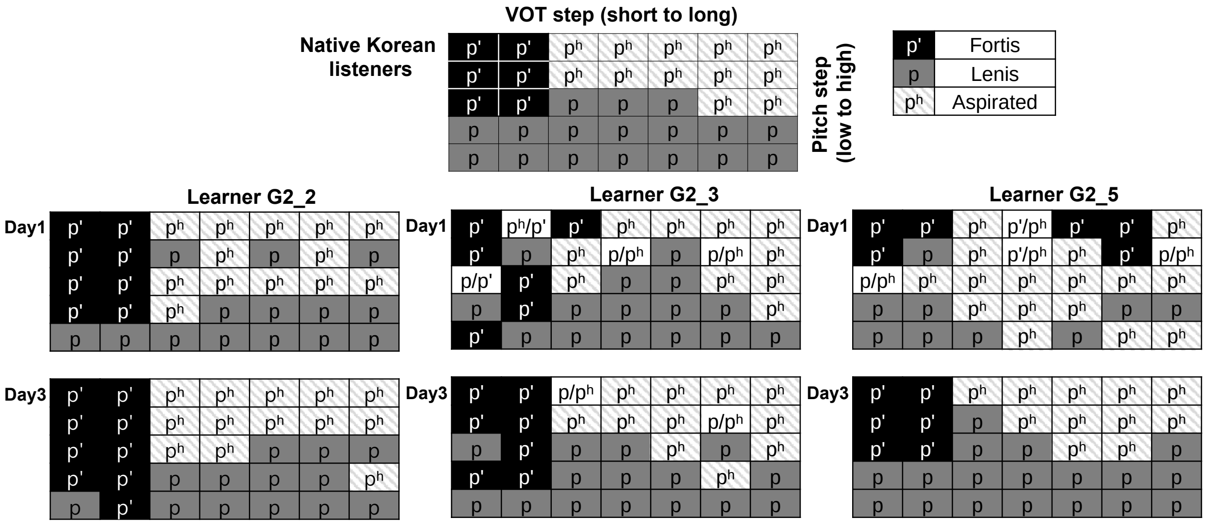
Learner C2_2

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Day1 | p' | p' | pʰ | pʰ | pʰ | pʰ | pʰ |
| | p' | p' | pʰ | pʰ | pʰ | pʰ | p |
| | p' | p | p | p | pʰ | pʰ | pʰ |
| | p' | p | p | p | p | p | p |
| | p' | p' | p | p | p | p | p |
| Day3 | p' | p' | pʰ | pʰ | p' | pʰ | pʰ |
| | p' | p' | pʰ | pʰ | pʰ | pʰ | pʰ |
| | p' | p' | pʰ | pʰ | pʰ | p | pʰ |
| | p | p | pʰ | p | pʰ | pʰ | p |
| | p | p | p | p | p | p | p |

Learner C2_5

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Day1 | p' | p/p' | pʰ | pʰ | pʰ | pʰ | pʰ |
| | p' | p' | p | pʰ | pʰ | pʰ | pʰ |
| | p' | p' | pʰ | pʰ | p | p | pʰ |
| | p | p' | p | p | p | p | p |
| | p' | p' | p | p | p | p | p |
| Day3 | p' | p' | pʰ | pʰ | pʰ | pʰ | pʰ |
| | p' | p' | p | p/pʰ | pʰ | pʰ | pʰ |
| | p' | p' | p | p | p | pʰ | p |
| | p' | p' | p | pʰ | p | p | p |
| | p' | p' | p | p | p | p | p |

Figure 32. Native Korean listeners' patterns in the everyday identification test (top) and mapping plots of responses of everyday ID tests (Day 1 & Day 3) by three participants from the categorical group (C2_1, C2_2, and C2_5).



Native Korean listeners

| | | | | | | |
|---|---|---|---|---|---|---|
| p' | p' | pʰ | pʰ | pʰ | pʰ | pʰ |
| p' | p' | pʰ | pʰ | pʰ | pʰ | pʰ |
| p' | p' | p | p | p | pʰ | pʰ |
| p | p | p | p | p | p | p |
| p | p | p | p | p | p | p |

| | |
|---|---|
| p' | Fortis |
| p | Lenis |
| pʰ | Aspirated |

Learner G2_2

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Day1 | p' | p' | pʰ | pʰ | pʰ | pʰ | pʰ |
| | p' | p' | p | pʰ | p | pʰ | p |
| | p' | p' | pʰ | pʰ | pʰ | pʰ | pʰ |
| | p' | p' | pʰ | p | p | p | p |
| | p | p | p | p | p | p | p |
| Day3 | p' | p' | pʰ | pʰ | pʰ | pʰ | pʰ |
| | p' | p' | pʰ | pʰ | pʰ | pʰ | pʰ |
| | p' | p' | pʰ | pʰ | p | p | p |
| | p' | p' | p | p | p | p | pʰ |
| | p | p' | p | p | p | p | p |

Learner G2_3

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Day1 | p' | pʰ/p' | p' | pʰ | pʰ | pʰ | pʰ |
| | p' | p | pʰ | p/pʰ | p | p/pʰ | pʰ |
| | p/p' | p' | pʰ | p | p | pʰ | pʰ |
| | p | p' | p | p | p | p | pʰ |
| | p' | p | p | p | p | p | p |
| Day3 | p' | p' | p/pʰ | pʰ | pʰ | pʰ | pʰ |
| | p' | p' | pʰ | pʰ | pʰ | p/pʰ | pʰ |
| | p | p' | p | p | pʰ | p | pʰ |
| | p' | p' | p | p | p | pʰ | p |
| | p | p | p | p | p | p | p |

Learner G2_5

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Day1 | p' | p' | pʰ | p'/pʰ | p' | p' | pʰ |
| | p' | p | pʰ | p'/pʰ | pʰ | p' | p/pʰ |
| | p/pʰ | pʰ | pʰ | pʰ | pʰ | pʰ | pʰ |
| | p | p | pʰ | pʰ | pʰ | p | p |
| | p | p | p | pʰ | p | pʰ | pʰ |
| Day3 | p' | p' | pʰ | pʰ | pʰ | pʰ | pʰ |
| | p' | p' | p | pʰ | pʰ | pʰ | pʰ |
| | p' | p' | p | p | pʰ | pʰ | p |
| | p | p | p | p | p | p | p |
| | p | p | p | p | p | p | p |

Figure 33. Native Korean listeners' patterns in the everyday identification test (top) and mapping plots of responses of everyday ID tests (Day 1 & Day 3) by three participants from the gradient group (G2_2, G2_3, and G2_5).

Figure 34 shows individual participants' $\beta_{\mathrm{VOT}}$ and $\beta_{\mathrm{f0}}$ coefficients obtained from a series of logistic regression analyses. Data points are different in color and shape to represent the group information of participants. The overall cue-weighting patterns by group show that both groups used the f0 dimension more than the VOT dimension to identify Korean lenis and aspirated stops. The categorical 2 group's average values of $\beta_{\mathrm{VOT}}$ and $\beta_{\mathrm{f0}}$ were 3.98 (1.67) and 1.01 (0.95), and the gradient 2 groups' average values of $\beta_{\mathrm{VOT}}$ and $\beta_{\mathrm{f0}}$ were 3.17 (1.41) and 0.90 (1.13). The categorical 2 group showed a higher average $\beta_{\mathrm{f0}}$ value than the gradient 2 group. Figure 35 shows the proportion of responses of Korean aspirated stops across f0 steps of test stimuli. Thin lines are logistic curves that fit each participant's response data from the Day 3 ID test, and thick lines are logistic curves that fit the averaged group data from the same test. As in Figure 34, it was observed that participants in both groups showed their primary use of the f0 dimension in the identification of Korean lenis and aspirated stops. After controlling the effect of VOT, Figure 35 shows that there was a more minor degree of variability in individuals' relative use of f0, compared to considerable qualitative variability shown in Figure 21 in experiment 1.

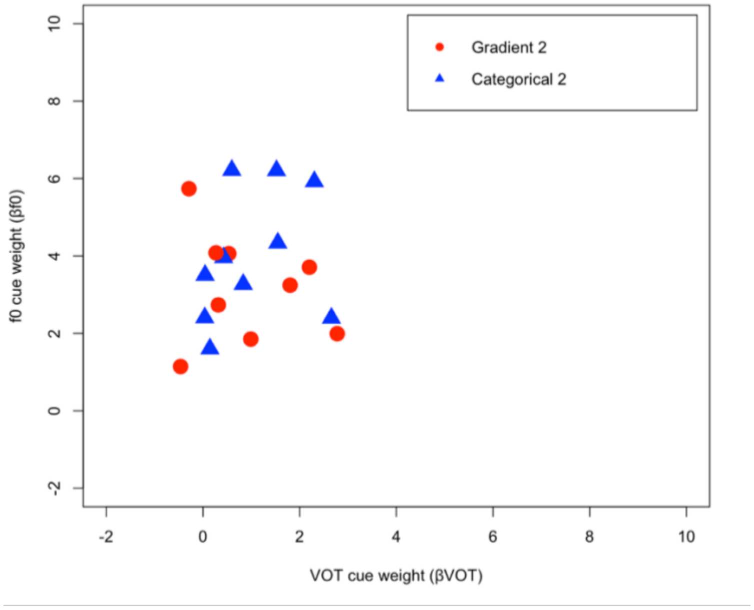


Figure 34. Individual participants' βVOT and βf0 coefficients obtained from a series of logistic regression analysis. Red dots indicate gradient group, and blue dots indicate categorical group.

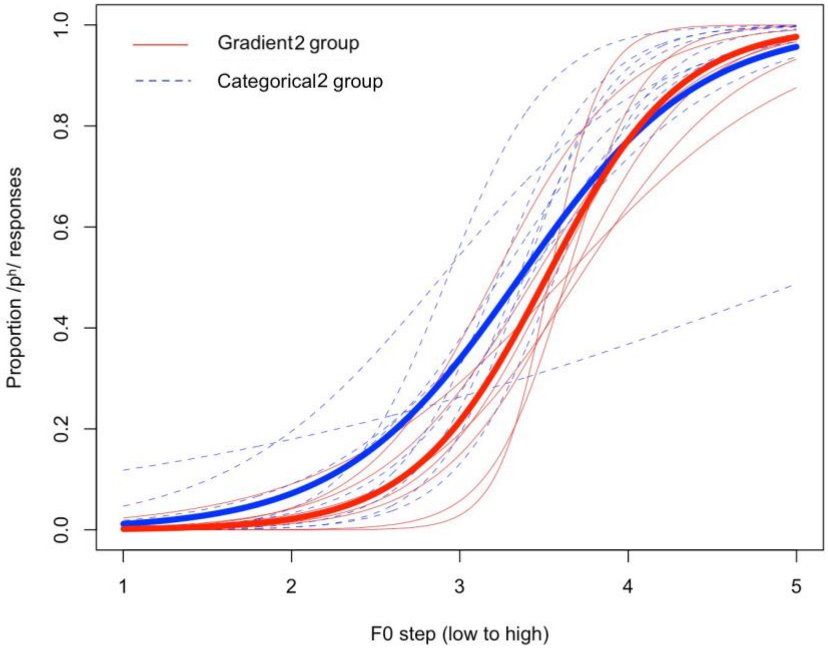


Figure 35. Proportion of /pʰ/ responses along f0 steps of test stimuli. Thin lines are logistic curves fit to each individual participant Day3 ID test data and thick lines are logistic curves fit to group-averaged data. Red lines indicate gradient group, and blue lines indicate categorical group.

Partial correlation tests were conducted to examine whether individual participants' quantified gradiency of the VAS task responses can explain the substantial individual differences in VOT and f0 cue weighting in identifying Korean lenis and aspirated stops in experiment 2. Figure 36 shows the results of correlation analysis, which did not find any significant correlation between individual participants' gradiency and VOT and f0 cue weights ($\beta_{\text{VOT}}$: $r$ = - .05, $p$ = .80, $\beta_{\text{f0}}$: $r$ = .03, $p$ = .89). The results suggest that participants' weights of VOT and f0 dimensions were not associated with how gradient participants' responses in the VAS task were (i.e., how sensitive participants were to the f0 dimension in the perception of English stop voicing contrast). In summary, the quantified gradiency (i.e., coefficients of the quadratic regression curves overlaid on the histograms of the VAS responses) could not account for variability in individuals' relative use of VOT and f0 in nonnative category perception in experiment 2.

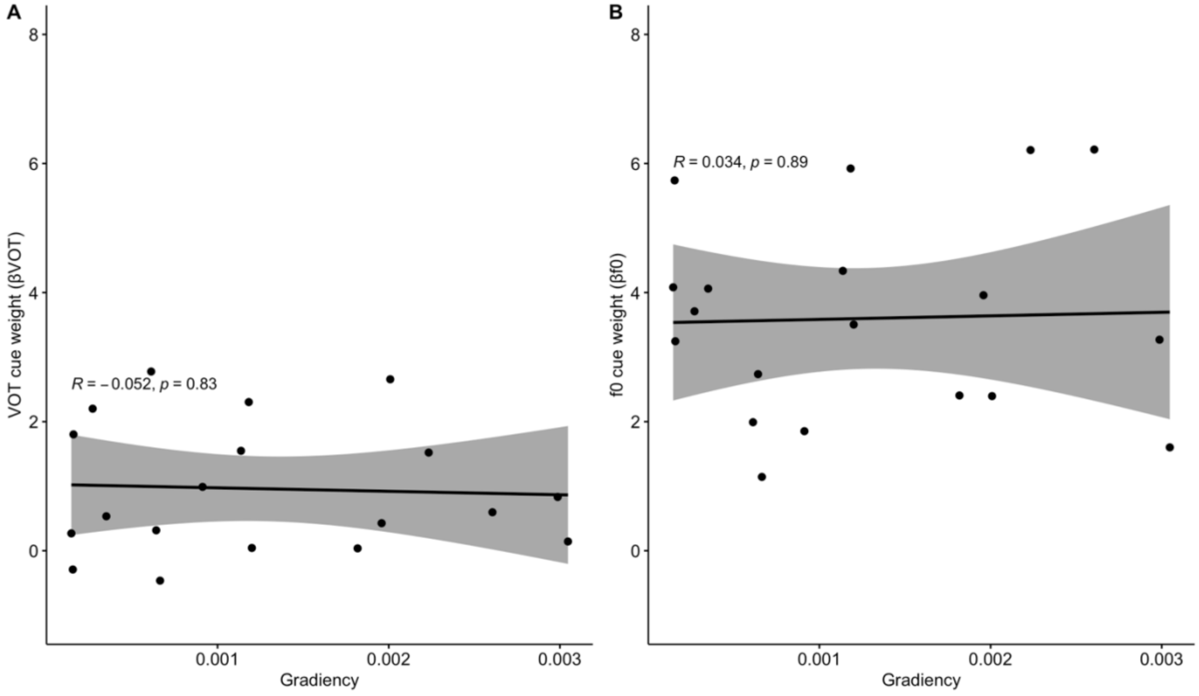


Figure 36. Correlation between individual participants' quantified gradiency of the VAS task and individual participants' beta-coefficients of either VOT (left) or f0 (right).

### 3.2. Summary of experiment 2

In experiment 2, participants received the additional cue-attention switching training before each session of HVPT. Inconsistent with the results in experiment 1, the performances in training between groups were not different from each other. Notably, categorical 2 and gradient 2 groups behaved similarly on the post-test and the new consonant generalization test. The results of everyday ID tests showed that the categorical 2 group demonstrated a similar level of achievement in learning the target Korean stop contrast after receiving three sessions of HPVT with the short and straightforward additional training. The comparison between the categorical 1 and 2 groups confirmed that the categorical 2 group showed greater learning success in HVPT sessions than the categorical 1 group, suggesting the effectiveness of the cue-attention switching training for the categorical group. Considering that the only difference between experiments 1 and 2 was the addition of cue-attention switching training to the HVPT sessions, the results in experiment 2 suggest the effectiveness of the cue-attention switching training method in terms of improving learners' ability to discriminate and identify Korean three-way stop contrast.

The quantified gradiency did not show its association with test scores, indicating that participants with steeper VAS task response patterns did not perform much differently from participants with gradient response patterns. This indicates an additional gain from HPVT with the cue-attention switching training compared to the regular HVPT condition in experiment 1. It seemed that especially participants with higher quantified gradiency values showed a benefit of the cue-attention switching training, in that they also showed an active use of f0 dimension to distinguish test stimuli falling into Korean lenis and aspirated stop categories as participants with lower quantified gradiency values. However, the active use of the f0 dimension was not observed by participants in the categorical 1 group in experiment 1. The correlation analysis between gradiency and $\beta_{f0}$ did not reach its significance level, indicating that substantial individual variability in $\beta_{f0}$ shown in experiment 1 due to participants' different sensitivity to the f0 dimension in the perception of English stop contrast was reduced in experiment 1. Taken together, it can be suggested that the HVPT design with the additional cue-attention switching training aided participants whose f0 cue sensitivity in the perception of native contrast may place them at a disadvantage in the acquisition of L2 cue-weighting strategies.

# CHAPTER 4. DISCUSSION 1

The primary goal of the current study was to examine whether nonnative contrast learning is influenced by learners' individual differences in within-category cue sensitivity in the perception of native speech contrast. The specific question addressed in experiment 1 was, "Are individual differences in the sensitivity to within-category acoustic cues for L1 category perception related to novel phonological contrast learning?" Experiments 1 and 2 examined the case of L1 English speakers learning Korean to investigate whether greater sensitivity to the most informative acoustic cue in L1 results in more attention to that cue in learning L2 speech contrast. The case of native English learners of Korean was particularly interesting because the f0 dimension is a secondary cue to the voicing contrast in English but one of several primary cues, such as VOT, f0 at vowel onset, and closure duration, to the voicing contrast in Korean. In this specific learning condition, English learners of Korean have to more actively engage the f0 dimension than they do in the perception of English voicing contrast to result in the nativelike cue-weighting strategy.

The current study went through several experimental phases. First, the VAS task measured individual participants' sensitivity to within-category acoustic details in L1. The VAS task was administered with stimuli of the English /da/-/ta/ continuum. Then, participants received three consecutive days of HVPT. They were trained with multiple sets of f0 and VOT manipulated stimuli produced by three different female talkers. Pre and post AXB tests, everyday ID tests, and two generalization tests (i.e., new talker generalization ID test & new consonant generalization AXB test) were completed to assess the effectiveness of training and the relationship between the results of the VAS task and participants' improvements during training.

Participants in Experiment 2 received the cue-attention switching training which was additionally added to HVPT paradigm with multiple talkers. The question asked in Experiment 2 was “Can the modified version of HVPT aid learners with less sensitivity to within-category acoustic details in L1 to perceive the target Korean contrast in a nativelike way by shifting their attention to the L2-related acoustic dimension?” The cue-attention switching training presented English stimuli having confined variability of f0 but a wide range of VOT variability to upweight the f0 cue in the perception of learners’ native contrast. This additional training was designed to prompt learners to notice the informativeness of the f0 cue and expected learners to use their increased attention to that cue to identify the target Korean contrast.

Several statistical tests were carried out for the group- and individual-level analysis. For the group-level analysis, participants were divided into two different groups based on their response patterns of the VAS task (i.e., Categorical group vs. Gradient group). For the individual-level analysis, I examined whether individual differences observed in the VAS task predicted learners’ different levels of achievement in learning the target Korean three-way stop contrast and changes in learners’ reliance on f0 and VOT cues before and after training.

The following sections summarize the main findings and discuss them in light of the two research questions.

## 4.1. Individual differences in L2 speech perception and nonnative contrast learning (Research question 1)

The first research question asked, “Are individual differences in the sensitivity to within-category acoustic cues for L1 category perception related to novel phonological contrast learning?” The prediction for this question was that individuals who showed more sensitivity to f0 cues in L1 might have an advantage in mastering the target stop contrast in L2 (i.e., Korean), where the role of f0 is as important as VOT. The results of the VAS indicate that individual L2 learners differ widely in their perception of the native stop contrast. These results confirm the earlier finding that this task shows individual differences in perceiving the voicing contrast by L1 English listeners (e.g., Kapnoula et al., 2017; Kong & Edwards, 2011, 2016). As hypothesized, learners with a more gradient response pattern in the VAS task performed better during the HVPT phase by recording higher scores in everyday ID tests than learners with a more categorical response pattern. In addition, the results provide evidence that only gradient learners (i.e., participants in the gradient 1 group) demonstrated the ability to generalize what they learned to novel and untrained stimuli. Learners who showed categorical response patterns (i.e., categorical 1 group) struggled to learn the target Korean contrast. Compared to the gradient 1 group, the categorical 1 group did not show learning from the HVPT sessions or the generalization effect. The successful L2 learning by learners with gradient response patterns came from their higher accuracy in identifying the Korean lenis and aspirated stops which is more challenging to acquire than Korean fortis stops. Recall that Korean lenis and aspirated stops are primarily distinguished by the f0 dimension while this is the secondary acoustic cue in learners' L1, English. This indicates that gradient learners' higher sensitivity to acoustic details of f0 cues in L1 helped them give more attention to that cue during training and resulted in more nativelike perceptual patterns with primary reliance on f0 cues. However, categorical learners' relatively low sensitivity to f0 cues in the perception of English stop voicing contrast might not aid them to acquire the target Korean by presenting nativelike utilization of cues, especially in terms of f0.

The overall results revealed that individual variability in sensitivity to acoustic details in L1 could function as a mirror of individual variability in perceptual cue weighting in L2. That is, how successfully learners progress to become more nativelike listeners in Korean can be predicted in terms of to what degree they have sensitivity to the L2 informative acoustic cue in L1 speech perception. This also implies that individual differences in the L2-relevant cue sensitivity may determine the initial stage of learning and to what extent learners can benefit from L2 training. Participants in the current study were naïve learners of Korean and received only three training sessions which lasted approximately less than 30 minutes each. Although participants were exposed to a very short training period, individual variability in learning was evident and well-explained by learners' different sensitivity to f0 cues in L1 perception. This suggests that learners' initial starting points in L2 learning were not identical; gradient learners might have started at a better position to actively engage f0 cues in learning the target Korean contrast while categorical learners might have started at an unfavorable position so that acquiring how to properly use f0 cues in learning was more challenging to these learners.

One likely account of the influence of cue sensitivity in L1 on nonnative contrast learning may be the transfer of L1 cue sensitivity to L2 cue utilization. In the current study, individuals who showed more sensitivity to f0 cues in L1 utilized those cues in L2 speech perception, observed in individual participants' identification patterns of training stimuli. This suggests the L1 effect on the perception of L2 sound contrast, which can be attributed to nonnative listeners' transfer of their cue utilization from the L1 to the L2. Previous studies have shown the transfer of L1 cue weighting to the perception of L2 linguistic contrasts. When learners perceive L2 sound contrasts, they weigh cues as a function of their informativeness of signaling sound contrasts in L1 (e.g., Francis et al., 2000; Holt & Lotto, 2006; Iverson et al., 2003; Tremblay et al., 2018). For example, Iverson et al. (2003) showed that the difficulty native Japanese learners of English experience in their perception of the English /l/-/ɹ/ contrast could be attributed to their greater reliance on F2 cue than the F3 cue, with F2 being an essential cue for encoding the Japanese liquid but F3 being the primary cue for the English /l/-/ɹ/ contrast. Trembley et al. (2018) showed the L1-to-L2 transfer on cue weighting and extended the cue-weighting theory to speech segmentation. Importantly, Trembley et al. (2018) showed that some L2 learners have a benefit in L2 learning since acoustic cues that serve one function in the L1 (e.g., to signal word-initial boundaries) can be reallocated to a different function in the L2 (i.e., to signal word-final boundaries). In their study, native Dutch learners of French made earlier/greater use of the F0 rise than native English learners of French to signal word-final boundaries in French. This result came from a stronger functional weight of F0 rise to signal word-initial boundaries in Dutch than in English. It was suggested that as long as a particular cue is important for signaling lexical identity in the L1, learning a new association between that and the function it serves in the L2 should be possible.

Although the present study did not explicitly compare individual participants' L1 and L2 cue weighting strategies, the present results extend previous findings by considering sensitivity to acoustic details in L1 and its individual variability. The current study did not measure participants' cue-weighting strategies with the speech identification task, widely used in cue-weighting research (e.g., Kapnoula et al., 2017; Kong, 2019; Schertz et al., 2015). Instead, this study used the VAS task, which was devised to measure listeners' within-category cue sensitivity, to show how listeners' acoustic cue sensitivity in L1 transfers to the perception of L1 speech contrast. The present study found that the higher sensitivity to f0 cues in L1 was related to the more reliance on those cues in L2, supported by the negative correlation between gradiency of the VAS task responses and beta-coefficients of the f0 dimension in the identification of Korean lenis and aspirated stops (see 2.1.5.3.3). Similar to the understanding of the transfer of L1-to-L2 cue weighting, this finding can thus be understood in regard to the transfer of L1 cue sensitivity to L2 cue utilization. If some learners have a higher sensitivity to the L2 informative acoustic cue (i.e., f0 in this study), learning the importance and function of that cue in L2 speech perception should be easier and happen earlier. To examine L1 effects in the transfer of cue-weighting from L1 to L2, previously mentioned studies either grouped learners based on their L1 backgrounds or examined only a group of learners with the same L1 background. However, the current study did not assume that all learners with the same L1 background would get either advantage or disadvantage from the transfer of L1-to-L2 cue weighting. Instead, I considered that even learners with the same L1 background could present a different advantage or disadvantage from their L1 depending on their utilization of acoustic cues in L1. In this sense, the current findings showed how individual differences in acoustic cue sensitivity in L1 speech perception could predict the learning of nonnative speech contrasts, suggesting the importance of considering individual differences in investigating the transfer of L1-to-L2 cue weighting.

Regarding individual L2 learners' use of acoustic cues in phonetic categorization, Kong and Edwards (2015) provided evidence that L2 learners' individual differences in cue weighting strategies are a by-product of L1-to-L2 transfer on cue weighting and L2 proficiency. Kim et al. (2018) also suggests that individual learners' initial L2 cue weighting patterns contribute to the direction and the rate of development in L2 speech perception over time. Learners' initial L2 cue weighting strategies might indicate baseline differences in their abilities to use acoustic-phonetic information in the speech signal. The present study's findings corroborate these studies and suggest that individual variability in learners' sensitivity to acoustic cues in L1 may represent a baseline in their ability to use acoustic-phonetic information in the L2 speech signal. Therefore, considering these baseline differences might be helpful to predict learners' expected L2 learning outcomes and provide appropriate L2 training to learners whose perceptual abilities may place them at a disadvantage.

## 4.2. The effectiveness of the cue-attention switching training (Research question 2)

The second research question asked, "Can the cue-attention switching which is added to HVPT and administered with L1 speech stimuli decrease the possible learning gap due to individual differences?" The present results showed that the HVPT with the cue-attention switching training aided categorical learners (i.e., Categorical 2 group). The categorical 2 group demonstrated more use of f0 cues in the perception of the target Korean lenis and aspirated stops than the categorical 1 group, which did not receive the additional training.

As discussed earlier, the mechanism of the cue-attention switching training is to increase the within-category variability between stimuli along the relatively uninformative dimension in L2, which was VOT in this study, and consequently, decrease the perceptual distance between those stimuli. Guenther et al. (1999) argued that experience with multiple exemplars encourages listeners to ignore small differences within a single category. Thus, the stimuli representation of the cue-attention switching training was designed to provide multiple same category exemplars largely varying in VOT. This design appears to direct participants in experiment 2 to learn the uninformativeness of the highly variable VOT dimension (i.e., learn to ignore this dimension), resulting in the switch of participants' attention to f0 cues. Through the cue-attention switching training, participants were expected to set up a favorable initial perceptual state for learning the target Korean contrast before initiating the original HVPT sessions. The effectiveness of the cue-attention switching training was reflected in participants' nativelike identification patterns of training stimuli regardless of their different sensitivity to f0 cues in L1.

There are two notable findings on the effectiveness of cue-attention switching training. The first is about the amount of time invested in cue-attention switching training. In this study, participants received only three additional training sessions, and each session lasted less than 10 minutes. This indicates that participants received this training for less than 30 minutes in total. Although this is a very short period, it looks as though it was enough to change L1 cue-weighting strategies in an intended way. This benefit in terms of time efficiency can make this training method applicable to the L2 pedagogy. The second finding is that experiment 2 showed that modified relative attention to multiple acoustic cues could be transferred to nonnative language and bring advantages to language learning circumstances. Participants in experiment 2 purposely changed their attention to VOT and f0 cues and exclusively weighted the f0 dimension in identifying English /b/ and /p/ during the cue-attention switching training. With participants' increased attention to the f0 dimension, they could identify the target contrast in a nativelike way. The present results extend the realm of transfer of L1-to-L2 cue weighting in that even temporally modified L1 cue-weighting strategies can affect L2 cue-weighting. It is possible that the modified L1 strategy will not last long. However, as long as the modified strategy can benefit nonnative language learning, this method can be considered a way to aid disadvantaged or beginner language learners without much knowledge about the target language.

Not all participants showed a benefit of the cue-attention switching training, suggesting individual variability in the effectiveness of the additional training. Some participants did not perform well in the cue-attention switching training and demonstrated lower scores on everyday ID tests compared to others. For example, the participant (C2_5) in experiment 2 scored lower accuracy in the cue-attention switching training sessions than others. In the last cue-attention switching training session, C2_5 reached 78.6% accuracy, lower than the average score (97.1%). This indicates that C2_5 failed to identify English /b/ and /p/ by relying only on the f0 dimension. The Day 3 ID test score of the same participant was also lower than the average score (60.8% vs. 80.3%), suggesting that the failure of the successful learning in the cue-attention switching training resulted in poor identification of training stimuli. In contrast, there was a case in which some participants recorded relatively lower scores in the Day 3 ID test even though they performed well during the cue-attention switching training. For instance, the participant (C2_4) demonstrated successful learning in the cue-attention switching training. However, C2_4 scored 68.7% on the Day 3 ID test, which was lower than the average score across all participants (80.3%), showing no benefit of the additional training in HVPT.

## 4.3. f0 cue sensitivity in the VAS task

This study used the VAS task to measure native English participants' f0 sensitivity and the degree of f0 cue involvement in their native cue-weighting strategies. Even though this study has its ground in the previous research showing the gradiency of responses in the VAS task is related to the f0 sensitivity, the direct measure of f0 sensitivity is missing. This gap should be addressed to reinforce the argument that individual differences in using secondary cues in L1 determine learners' learning achievements. Previous studies with the VAS task often accompanied the identification with stimuli systematically varying in acoustic dimensions to explicitly measure listeners' sensitivity to the secondary cue in the perception of L1 speech contrast (e.g., Kapnoula et al., 2017; Kim et al., 2020; Kong, 2019; Schertz et al., 2015). For example, Kim et al. (2020) measured listeners' proportion of /æ/ responses in the English vowel identification task with the spectrally and durationally manipulated stimuli of the English /ɛ/-/æ/ continuum. By comparing the proportion of /æ/ responses for stimuli with the lowest and the highest duration values, the authors could measure how much listeners attended to duration cues during the categorization of the English vowel contrast.

To fill the gap between previous studies and the current study, I conducted an additional analysis on how gradiency of the VAS task responses relates to participants' use of secondary acoustic cues (i.e., f0) in their responses to stimuli varying in VOT and f0. Inspired by Schertz et al. (2016), stimuli in the reversed condition were particularly examined. Stimuli in the reversed condition were not modeled after the canonical English pattern, such that stimuli with long VOT had high f0 while short VOT had low f0. Instead, in the reversed condition, there were two sets of stimuli in which the relationship between VOT and f0 was not canonical; the first set of stimuli had long VOT (i.e., voiceless stops) with low f0 while the second set of stimuli had short VOT (i.e., voiced stops) with high f0. As described in Figure 37, the covarying stimuli in Quadrants II and IV belong to the Reversed condition, while the stimuli in Quadrant I and Quadrant III belong to the Canonical condition. Schertz et al. (2016) calculated a reliance score to classify native Korean listeners based on their categorization patterns for the covarying English /b/-/p/ stimuli in VOT and f0. Reliance scores were calculated by taking the difference between the ratio of voiceless /p/ response to covarying stimuli in the reversed condition. The authors expected listeners to fall into three groups based on reliance scores: scores clustering near 1 (relying exclusively on VOT), scores clustering near -1 (relying exclusively on f0), and scores clustering around 0 (equal reliance on VOT and f0).

The current study calculated a score for each participant which was inspired by reliance scores in Schertz et al. (2016) with some modifications. Scores were calculated by taking the difference between the mean percentages of the VAS task responses to covarying stimuli in Quadrant II and Quadrant IV. I expected that participants with more reliance on the f0 dimension during the VAS task would result in lower scores. The logic was as follows: if some participants perceived reversed stimuli as bad examples of either English /b/ or /p/ due to their switched relationship between VOT and f0, those participants would click locations closer to the middle of the VAS task scale (Figure 7). Thus, the difference between the mean percentage of the VAS task responses for stimuli in Quadrant II and Quadrant IV would be relatively small. On the other hand, participants who did not rely much on f0 during the VAS task might have relatively higher scores since they might evaluate stimuli in the reversed condition as fairly good examples of either English /b/ or /p/ regardless of the f0 information of stimuli. I conducted a partial correlation analysis between participants' scores and gradiency to examine the relationship between their reliance on the f0 dimension and the response patterns of the VAS task. For experiment 1, the results showed the positive relationship between gradiency and scores ($r = .68$, $p < .001$). This result suggests that participants with categorical responses in the VAS task (i.e., higher gradiency coefficients) paid little attention to the f0 information of stimuli in the reversed condition by evaluating them as either good examples of English /b/ or /p/. A similar pattern was observed in experiment 2. The partial correlation analysis showed the positive correlation between gradiency and scores ($r = .52$, $p < .02$). Together, the results of the correlation analysis suggest the VAS task as a reliable measure of listeners' sensitivity to the secondary acoustic cue (i.e., f0) and their use of that cue in the perception of English stop voicing contrast.

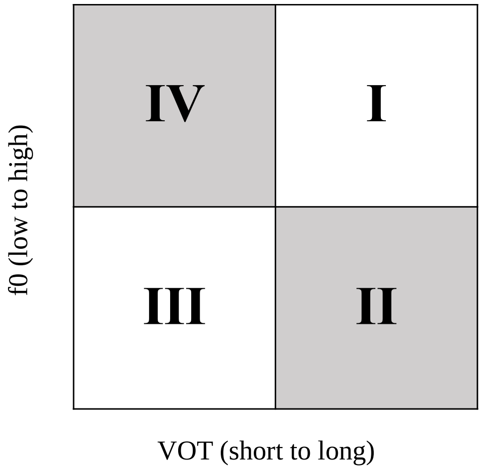


Figure 37. Schematic description in classifying covarying stimuli. The covarying stimuli in Quadrants II and IV belong to the Reversed condition, while the stimuli in Quadrant I and Quadrant III belong to the Canonical condition. (Schertz et al., 2016, p.360).

### 4.4. The VAS task as a pretraining method

Lastly, the present results suggest the possibility of using the VAS task as one of the pretraining methods, which are to predict L2 learners' achievements in training. This suggestion could be understood in line with previous research on the relationship between learners' auditory or cognitive abilities and their L2 learning (e.g., Darcy, Mora, & Daidone, 2016; Lengeris, 2009; Perrachione et al., 2011). For example, Darcy et al. (2016) investigated the role of inhibition in L2 learners' L2 phonological processing. They measured adult L2 learners' inhibition scores with the retrieval-induced inhibition task (Lev-Ari & Peperkamp, 2013). The results of correlation analyses between L2 learners' phonological performances on two tasks (an ABX task and a delayed sentence repetition task) and their inhibition scores showed that learners with higher inhibition scores demonstrated a lower error rate in both tasks. The authors interpreted these results as an indication that learners with higher inhibition ability may have used this ability to inhibit the L1 influence on the acquisition of L2 segmental categories. Perrachione et al. (2011)

showed that successful learning of lexical tones in nonnative phonological contrast depends on individual learners' auditory ability to perceive pitch contours. They found that depending on individual differences in perceptual ability, the HVPT paradigm could be beneficial or detrimental; learners with strong perceptual ability benefitted from globally high stimulus variability, while perceptually weak learners benefited more from low stimulus variability. These studies suggest the importance of pre-instructional assessment and emphasize the need for future work to identify what assessments may best predict the acquisition of unfamiliar phonological contrasts. The relationship between individual variability in the VAS task and participants' learning outcomes evince the possibility of the VAS task as a possible pretraining method. Based on learners' performances on the VAS task, we may be able to foresee the effects of HPVT on the individual learners and tailor training methods accordingly, to overcome disadvantages due to the individual differences in acoustic cue sensitivity.

# CHAPTER 5. EXPERIMENT 3

Experiment 3 focuses on the research question 1 with the prediction 2 targeting native Korean learners of English. The AXB oddity task measures individual participants' within-category sensitivity to the spectral dimension in the perception of Korean /i/ vowel category. Participants complete a total of five-day experimental phase with a five-day HVTP to learn a three English voicing contrasts (/i/-/ɪ/, /ɛ/-/æ/, and /ʊ/-/u/). Experiment 3 is important in that it attempts to extend the results of the results of experiment 1 with a different population of L2 learners, targeting learning English vowel contrasts by Korean learners of English. The results of experiment 3 are expected to suggest whether the role of individual differences in nonnative contrast learning is generalized to two target languages with two different types of segments: English vowel contrasts and Korean consonant contrast.

As a reminder, the prediction 2 states as follows: native Korean learners of English who exhibit more spectral cue sensitivity in the perception of a Korean vowel category outperform in the acquisition of the systematic use of spectral cues to discriminate the challenging English vowel contrasts, /i/-/ɪ/, /ɛ/-/æ/, and /ʊ/-/u/, compared to learners with lower sensitivity to spectral differences.

## 5.1. Methods

### *5.1.1. Participants*

Twenty-five native speakers of Korean were initially signed up for participation, but 24 participants agreed to complete the entire experiment (eleven female, thirteen male: mean age of 22.4 years, range of 19 – 27 years). Participants completed a language background questionnaire asking their demographic and native language backgrounds as well as their English language learning experience and self-reported proficiency.

Table 20 summarizes the relevant demographic information for the Korean L2 learners of English. As determined by a language background questionnaire, two out of 24 participants considered they have a regional accent of Korean dialect (one for Jeju dialect and the other for North Chungcheong dialect). Participants were undergraduate or recently graduated college students at Kangwon National University in South Korea. They graduated elementary, middle, and high schools in South Korea and received a formal English education starting from the age of 10 (i.e., 3rd grade in elementary school) until the age of 18 (i.e., senior in high school). Participants' average length of English education was 14.3 years. None of the participants had an experience of spending or studying abroad in an English-speaking country at the time of the experiment. Participants were self-reported that they relatively did not use English regularly under the circumstances, such as at home or school, with friends, for shopping, a phone call, or social gatherings. The average percentage of English usage in everyday life was 22.8%.

Participants rated their English proficiency for speaking, listening, writing, and reading on a Likert-scale from 1 (Very poor) to 7 (Very good). The average self-rated scores were 2.8 (speaking), 3.5 (listening), 2.8 (writing), and 4.5 (reading). Participants also rated their English accent on a Likert-scale from 1 (Very heavy foreign accent) to 10 (No foreign accent as native speakers), and the average score was 4.8. Participants rated their interest and preference in learning English on a Likert-scale from 1 (Definitely not prefer to learn English) to 10 (Definitely prefer to learn English), and the average rating score was 5.1. All participants were paid for their participation in the experiment under a protocol approved by the Institutional

Review Board (IRB) for the protection of human subjects at the University of Wisconsin-Milwaukee.

Table 20. Demographic information of participants in experiment 3.

| Variable | Mean | *SD* |
|---|---|---|
| Sex | 11 F; 13 M | |
| Age (yrs. old) | 22.4 | 2.0 |
| Birthplace (province, #) | Gyeonggi 10<br>Gangwon 12<br>Jeju 1<br>Chungcheong 1 | |
| Lived out of Korea (yrs.) | 0.0 | 0.0 |
| Father L1 (language, #) | Korean 24 | |
| Mother L1 (language, #) | Korean 24 | |
| L2 (language, #) | English 24 | |
| (1 very poor to 7 very good) | | |
| English Speaking | 2.8 | 1.3 |
| English Listening | 3.5 | 1.5 |
| English Writing | 2.8 | 1.1 |
| English Reading | 4.5 | 1.5 |
| (1 Very heavy foreign accent to 10 No foreign accent) | | |
| Accentedness | 4.8 | 1.5 |
| (1 Not prefer to learn English to 10 Prefer to learn English) | | |
| English preference | 5.1 | 1.6 |

### *5.1.2. Stimuli*

#### 5.1.2.1. Stimuli for the AXB oddity task

A set of 25 Korean /i/ vowel stimuli varying orthogonally in both spectral quality and duration of the vowel /i/ was created for the AXB oddity task. The stimuli were generated according to the following procedure: a 24-year-old male native speaker of Seoul Korean dialect recorded natural production of Korean /i/ vowel with the preceding Korean lenis stop consonant /ti/. All tokens were recorded in the same environment for recording sound files to generate stimuli for the VAS task and HVPT sessions in experiment 1 (see 2.1.2). Among the recorded Korean /ti/ tokens, one token was selected as the baseline token. The selected baseline token had no abrupt changes in formant movement throughout the periodic portion of the signal, no abrupt changes in fundamental frequency, no clicks, and minimal extraneous noise throughout the recording. The baseline token was then manipulated to create two endpoints for creating a continuum of Korean /i/ with multiple spectral steps. The F1 and F2 frequencies of two endpoints were set based on the study by Yang (1998). In Yang (1998), six Korean vowels were synthesized through a formant synthesis method. F1, F2, and F3 formant frequencies of those vowels were modified over and under the original formant frequencies but did not interfere with the formant ranges of the adjacent vowels. Seven native Korean listeners heard two stimuli in a row; after listening to an original stimulus without the formant manipulation followed by the corresponding synthesized stimulus, they judged whether those two stimuli sounded qualitatively 'similar' or 'different.' Based on the results, Yang (1995) presented the formant frequency range for each vowel within which native Korean listeners perceived each pair of stimuli as 'similar.' For the Korean /i/ vowel, the F1 and F2 ranges were 260 – 420 Hz and 2,330 -2,830 Hz, respectively. In this current study, these F1 and F2 ranges were utilized as the target F1 and F2 frequencies of two endpoints of the Korean /i/ vowel.

Two endpoints of the Korean /i/ vowel were created as follows (Yun & Seong, 2013): first, the baseline token was resampled to 10kHz. And then, I created the LPC (berg) object of the resampled recording using "To LPC (berg)…" function in Praat. A new sound file was created by selecting the baseline token and the LPC (berg) together. The new sound file was the estimated source signal, representing everything in the speech signal that cannot be attributed to the resonating cavities. The *FormantGrid* of the new sound file was created, and its original F1 and F2 frequencies were modified to the target formant frequencies using “Formant:Formula (frequencies)…” function in Praat. As the last step, the source signal and the modified *FormantGrid* were selected together and filtered, resulting in the final two endpoint stimuli. The F1 and F2 frequencies of the resulting endpoint stimuli are 268/2,818 Hz and 400/2,332 Hz, respectively. The F1 and F2 frequencies were measured at the mid-point using waveform and spectrographic displays in Praat.

Once two endpoint stimuli were available, a speech continuum between these stimuli was generated by *TANDEM-STRAIGHT* (Kawahara, Takahashi, Morise, & Banno, 2009). More information regarding *TANDEM-STRAIGHT* can be found in section 5.1.2.2. Initially, an eleven-step continuum between two Korean /i/ endpoints was generated. Out of these 11 steps, the author chose five acoustically and auditorily distinct stimuli from each other. Five selected stimuli include two endpoint stimuli. Next, each of the five steps along the natural continuum was manipulated using the “To Manipulation…” function in Praat to create five steps vowel duration continuum ranging from 240 to 400 ms (40 ms/step). Each duration step along the duration continuum was over one just-noticeable difference for native English listeners, namely 25 ms, as reported in Klatt (1976). As a result, a total of 25 stimuli (5 spectral steps × 5 duration steps) were generated. Figure 38 shows the layout of the stimuli and stimulus number from 1 to 25.

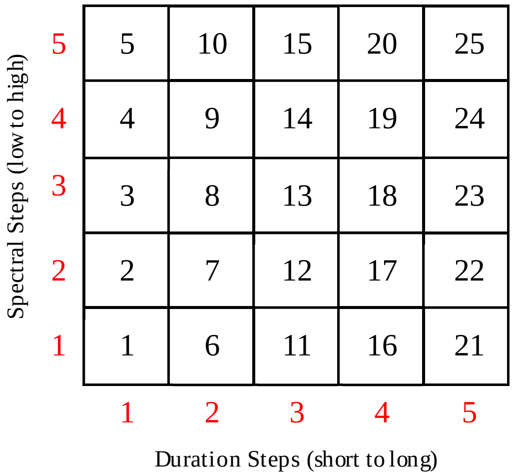


Figure 38. Layout of the stimuli and stimulus number from 1 to 25.

#### 5.1.2.2. Stimuli for the HVPT

In the current study, participants were trained to learn three English vowel contrasts, which are often confused by Korean learners of English. Six English words containing one of the vowels in the contrasts were selected. The selected English words had the same phonetic environment, /hVd/ (Table 21). Six additional English words containing the target English vowels were chosen for the new word generalization test, which were in /bVd/ phonetic environment. /bVd/ was chosen to have the additional words to have a voiced stop consonant in their coda position as six training words in /hVd/ environment. However, it was impossible to find valid English words in /bVd/ for the English /ʊ/-/u/ contrast. Therefore, *nook* and *nuke* were chosen instead.

Table 21. English words for each vowel contrast used in HPVT training and new word generalization test.

| **Target English vowels** | **/ɪ/** | **/i/** | **/ɛ/** | **/æ/** | **/ʊ/** | **/u/** |
|---|---|---|---|---|---|---|
| **Words for HVPT** | hid | heed | head | had | hood | who'd |
| **Words for new word generalization test** | bid | bead | bed | bad | nook | nuke |

Training stimuli consisted of three continua (/i/-/ɪ/, /ɛ/-/æ/, and /ʊ/-/u/) varying in both spectral quality and duration of the vowel. There were three steps involved in the construction of the continua. First, four male monolingual speakers of American English in their early 20s (Talker 1, 2, 3, & 4) were recruited for the recording process. The language background questionnaire confirmed that all talkers spoke a Midwest dialect and never lived outside the Midwest area at the time of the recording. The recording environment was identical as described in 2.1.2.2. Four talkers produced six training words in Table 21 in the context of the carrier sentence ("I say _______ again"). Talkers naturally produced each sentence, including one of six English target words with a normal tone and speech rate, and repeated three times. The best recordings for training words were selected for each talker and modified to generate training stimuli using *TANDEM-STRAIGHT*. The selected recordings of *hid*, *head*, *hood*, *heed, had*, and *who'd* for each talker were set as endpoints to create three sets of resynthesized training stimuli encompassing /i/-/ɪ/, /ɛ/-/æ/, and /ʊ/-/u/ contrasts. *TANDEM-STRAIGHT* is a high-quality vocoder that allows for creating natural-sounding continua between two endpoints with a typical manifestation of perceptual attributes. *TANDEM-STRAIGHT* is a new formulation of its predecessor *STRAIGHT*, which is a contemporary version of a channel of VOCODER. *TANDEM-STRAIGHT* decomposes input speech into three types of positive-values parameters: an interference-free spectrogram, an aperiodicity map, and a fundamental frequency trajectory.

And then, *TANDEM-STRAIGHT* allows for holistic morphing between the two endpoints, rather than just mixing one part of the sound file, so aspects of coarticulation can be better preserved. Alignment between sound files is both temporal in the waveform and also spectral in the frequency domain, allowing for peaks in the spectra to be shifted around, rather than just mixed linearly (McAuliffe, 2017). Figure 39 shows a schematic diagram of *TANDEM-STRAIGHT* (Kawahara et al., 2009, p. 112). The parameter manipulation block independently modifies input speech's source parameters (f0 and aperiodicity) and STRAIGHT spectrogram as well as their coordinates (time and frequency axes) and generates synthesized speech (Kawahara, et al., 2009, p. 112). As mentioned in Kim et al. (2018), it should be noted that the sound quality of the spectral continua generated through *TANDEM-STRAIGHT* represents an improvement over other methods creating synthesized stimuli.

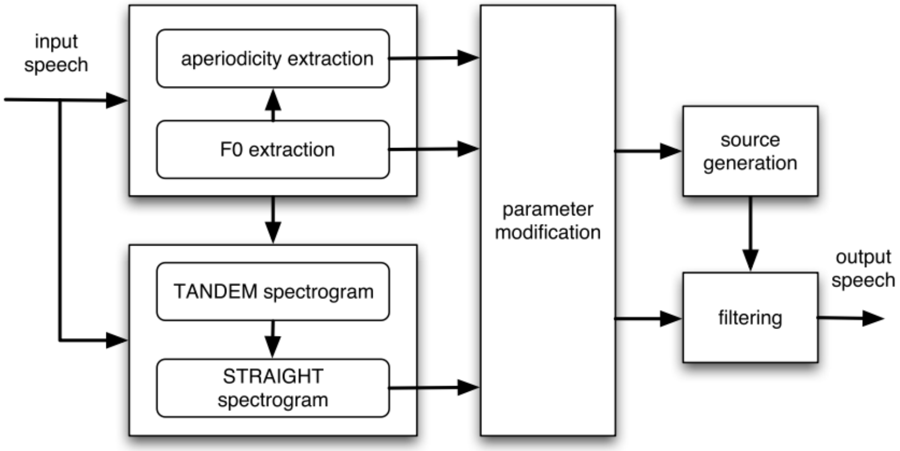


Figure 39. Schematic diagram of TANDEM-STRAIGHT (Kawahara et al., 2009, p. 112).

The identical stimuli generation procedure used for the Korean AXB oddity task was carried out for generating training stimuli sets. There were two sets involved. First, *TANDEM-*

*STRAIGHT* generated three eleven-step continua from *hid*, *head*, and *hood* to *heed*, *had*, and *who'd* were generated. One thing to note here is that the /h/ portion of each continuum was set to neutral between two endpoints in terms of STRAIGHT spectrogram, spectrum level morphing rate, frequency axis morphing rate, aperiodicity morphing rate, f0 morphing rate, and temporal axis morphing rate during the morphing rate manipulation process. The motivation of this setting was to minimize the influence of /h/ portion of stimuli, which may bear some characteristics of the following vowel, in the learning of English vowel contrasts. Among 11 stimuli for each of three continua, five acoustically and auditorily different stimuli were chosen by a native speaker of English, who did not serve as a talker. The /d/ consonant from these stimuli was extracted starting from the beginning of the salient gap to the end of the release of the closure and then was replaced with one of the extracted /d/ to make stimuli have the unified consonant /d/. Since the current study investigates how Korean learners of English utilize spectral and duration cues to identify challenging English vowel contrasts, it was necessary to control other possible acoustic cues in the stimuli. Second, each of five selected stimuli was manipulated in their duration properties using "To manipulation…" function in Praat to generate 5 step vowel duration ranging from 140 ms to 300 ms (40 ms/step). The same criterion for the duration interval was applied as the Korean AXB oddity task. Twenty-five stimuli were generated for each vowel continuum for each talker. As a result, a total of 75 stimuli were created for each talker (25 stimuli × 3 contrasts × 4 talkers = 300 stimuli). The F1 and F2 frequencies in *Hz* and the duration values in *ms* for each vowel continuum from Talker 3 are presented in Table 22. The F1 and F2 frequencies were measured at the mid-point by using waveform and spectrographic displays in Praat. Once again, three sets of stimuli from Talkers 1, 2, and 3 were designated for use only in the HVPT condition, and the set of stimuli from one remaining talker (Talker 4) was used for the new talker generalization test conducted after the entire HVPT sessions.

Table 22. F1 & F2 values and durations of the vowels in the /i/-/ɪ/, /ɛ/-/æ/, and /ʊ/-/u/ continua.

| | Contrasts | | | | | | |
|---|---|---|---|---|---|---|---|
| | /i/-/ɪ/ (hid-heed) | | /ɛ/-/æ/ (head-had) | | /ʊ/-/u/ (hood-who'd) | | |
| Step | F1(*Hz*) | F2(*Hz*) | F1(*Hz*) | F2(*Hz*) | F1(*Hz*) | F2(*Hz*) | Duration (*ms*) |
| 1 | 452 | 1892 | 596 | 1860 | 514 | 1317 | 140 |
| 2 | 396 | 2011 | 687 | 1846 | 416 | 1161 | 180 |
| 3 | 341 | 2114 | 787 | 1834 | 390 | 1094 | 220 |
| 4 | 269 | 2211 | 840 | 1823 | 303 | 985 | 260 |
| 5 | 246 | 2250 | 890 | 1805 | 286 | 948 | 300 |

A two-alternative force choice identification task with the four sets of stimuli was conducted to check the validity of the stimuli, to check whether category boundaries between contrasts were well represented by steps in the continua, and to determine the answers for each stimulus to provide trial-by-trial feedback on participants' performances during the training. The identification task was run online, and it was designed using PennController for Internet Based Experiments (Zehr & Schwarz, 2018). A total of 40 native speakers of American English, who were undergraduates at the University of Wisconsin-Milwaukee at the time of participation, completed the task. Participants were randomly assigned to one of the talkers' /i/-/ɪ/, /ɛ/-/æ/, and /ʊ/-/u/ stimuli sets. After listening to a stimulus, participants were instructed to identify the word they just heard by clicking one of the orthographic words (e.g., hid vs. heed) appearing in the center of the computer screen. Twenty-five stimuli were repeated three times for each vowel continuum in random order, and trials were blocked by contrast (i.e., 25 stimuli × 3 repetitions × 3 vowel contrasts = 225 trials). The task took approximately 20 minutes, and participants received extra credit for their participation. The task results showed native English listeners' overall response patterns and their use of spectral and duration dimensions in the identification of synthesized stimuli. Visual inspection of heat plots in Figure 40, which indicate the proportion of *hid*/*head*/*hood* responses in terms of the colors and darkness of the cells, show that overall, stimuli with lower spectral steps were identified as *hid*, *head*, or *hood* while stimuli with higher spectral steps were identified as *heed*, *had*, or *who'd*. These results suggest that native English listeners predominantly weighted spectral cues while duration cues had a much weaker effect in cue-weighting strategies to categorize English vowels (Baker & Trofimovich, 2005).

One of the training talkers (i.e., Talker 3) additionally recorded his production of new English words for the new word generalization test (Table 21). The recording and stimuli manipulation procedures were identical to the process of generating training stimuli. However, it should be noted that the range for the duration manipulation for two words, *nook* and *nuke*, was set from 70 ms to 230 ms. This range was smaller than the original durational range (140 ms - 300 ms) since previous studies showed that vowel duration is shorter before voiceless consonants (Kluender, Diehl, & Wright, 1988). Additional 15 English native listeners completed a two-alternative force choice identification task under the same condition as previously described.

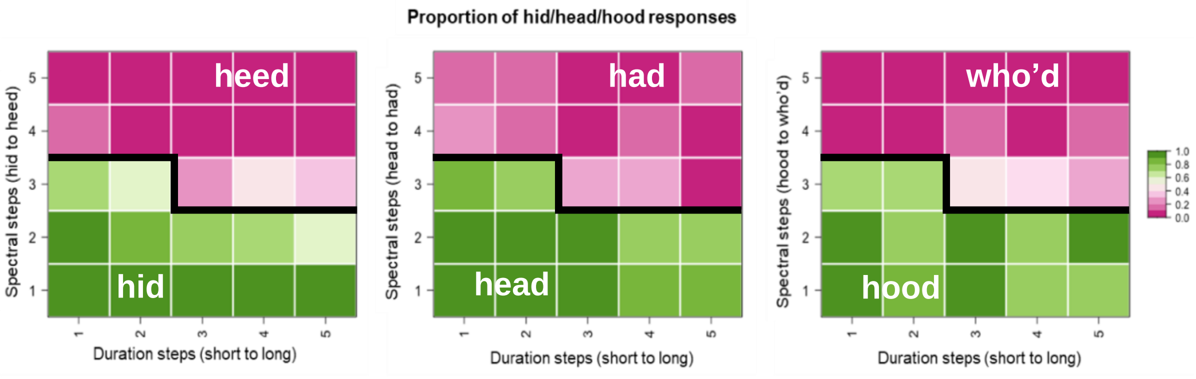


Figure 40. Heat plots of native English listeners' categorization of the three English vowel contrasts, shown plotted across each combination of the two acoustic dimensions. Each cell represents one stimulus, and the darkness of the cell represents the percentage of responses in a forced-choice task; the darkest red cells elicited 100% 'heed/had/who'd responses, while the darkest green cells elicited 100% 'hid/head/hood.'

### *5.1.3. Procedure*

#### 5.1.3.1. The AXB oddity task

The AXB oddity task was administered before participants started the HVPT phase. This task was designed to measure participants' sensitivity to within-category differences induced by spectral and duration cue changes, using a set of stimuli belonging to Korean vowel /i/ but with different spectral and duration properties. The set of stimuli for the AXB oddity task consisted of 25 synthetic Korean /i/ stimuli, systematically varying in 5 spectral and 5 duration steps. The AXB stimuli pair consisted of the compared stimulus (*X*) and two comparing stimuli (*A* and *B*). *A* and *B* stimuli have either a one-step different spectral or duration step from the *X* stimulus, while the other step remains the same as the *X* stimulus. For example, if the *X* is stimulus number 1 in Figure 41, stimulus number 2, which has the same duration step but one step different spectral step, and stimulus number 6, which has the same spectral step but one step different duration step, were selected as either *A* or *B* stimulus as shown in Figure 41. There were two versions for each AXB stimuli pair since each comparing stimulus should appear in both *A* and *B* positions (e.g., 2-1-6 or 6-1-2). The inter-stimulus interval between stimuli in each pair was 300 ms. A total of 128 trials was presented (64 stimuli pairs × 2 versions = 128 trials).

During the AXB oddity task, participants were asked to judge whether the X stimulus is more distinct from either the *A* or *B* stimulus and pick the "most odd one" out. Participants were asked to pay attention to the differences between *A*, *X*, and *B* stimuli since this task was to measure how sensitive the learners are to within-category spectral cue changes. Participants who chose the stimulus with a different spectral step from the *X* stimulus as the odd one was considered to have higher sensitivity in the spectral cue changes than the ones who chose the stimulus with a different duration step from the *X* stimulus. This task started with a short practice session with 12 trials to ensure participants fully understood the procedure and got used to paying attention to small differences between AXB stimuli pairs. This task took about nine minutes with a mid-session break.

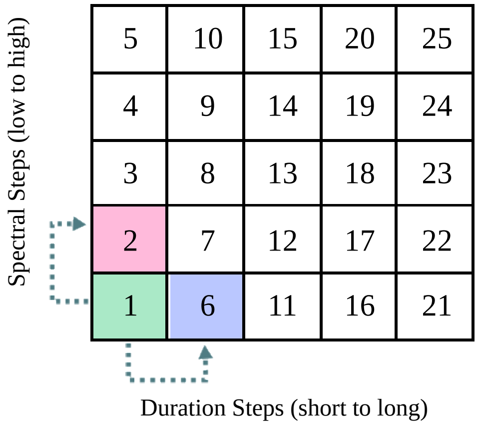


Figure 41. Demonstration of stimuli pairs for the AXB oddity task.

### 5.1.3.2. HVPT

Experiment 3 was managed in the span of five days with five-day online HVPT computer-based sessions. Once participants started their first day of the experiment, no more than a one-day interval was allowed between the training sessions to ensure consistent participation. A total of six English words, two words for each English vowel contrast, served as target words. The HVPT sessions were administered through PsyToolkit (Stoet, 2010, 2017). PsyToolkit is a free-to-use toolkit for demonstrating, programming, and running cognitive-psychological experiments and surveys, including personality tests. Table 23 demonstrates the timeline of the entire experimental phase with five-day of HPVT training.

Table 23. Timeline of the five-day experiment 3 with five-day HVPT.

| | **Participation Day** | | | | |
|---|---|---|---|---|---|
| **Day** | Day 1 | Day 2 | Day 3 | Day 4 | Day 5 |
| **Tasks** | • Korean AXB Task (n = 128)<br>• Daily Familiarization Phase for HVPT<br>• Pre ID Test (n = 450)<br>• HVPT Session 1 (n = 225)<br>• Everyday ID Test (n = 75)<br>• Language Questionnaire | • Daily Familiarization Phase for HVPT<br>• HVPT Session 2 - 4 (n= 225)<br>• Everyday ID Test (n = 75) | | | • Daily Familiarization Phase for HVPT<br>• HVPT Session 5 (n= 225)<br>• Everyday ID Test (n= 75)<br>• Post ID Test (n = 450)<br>• New Talker Generalization Test (n = 225)<br>• New Word Generalization Test (n = 225)<br>• LexTale |
| **Time** | About 1 hour | About 30 minutes | | | About 1 hour |
| **Location** | Laboratory | Online | | | Laboratory or Online |

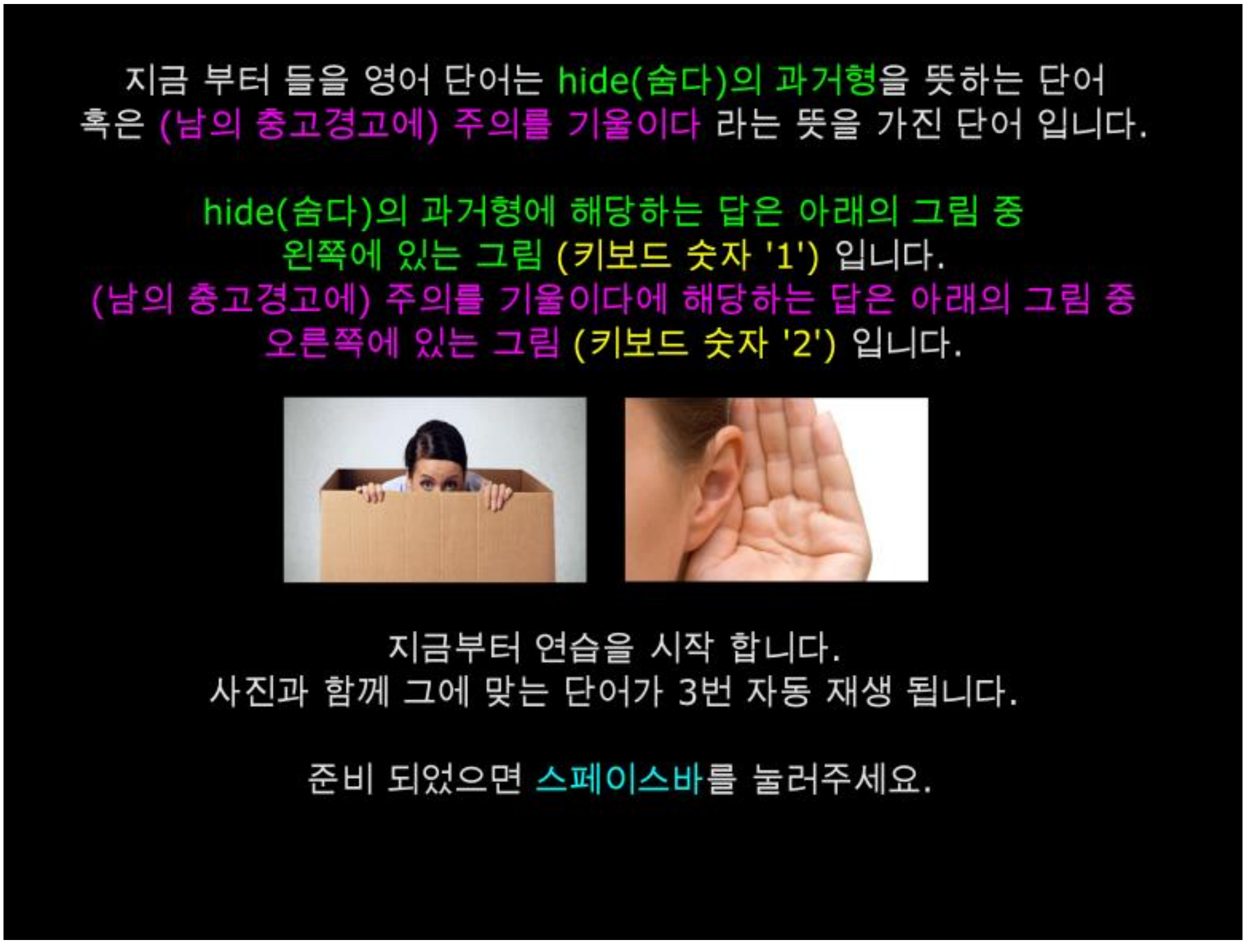
#### *5.1.3.2.1. Day 1*

All participants completed their first day of participation in the Phonetics Lab at Kangwon National University in South Korea. Participants were seated in front of an individual desk equipped with a Samsung Windows computer. Stimuli were presented via headphones at a comfortable listening level. No more than three participants were allowed to participate simultaneously. Participants started Day 1 session with the Korean AXB oddity task, which was followed by a language background questionnaire. The pre-identification test (pretest, henceforth) was completed on the first day to measure participants' pre-training states of using spectral and duration dimensions to identify two words for each English vowel contrast. To avoid orthographic bias, photographs were used to represent target training words (see Appendix C). The sets of training stimuli from Talker 3 and Talker 4 were used in the pretest. Talker 3 was a training talker, and Talker 4 was a new talker who produced stimuli for the new talker generalization test.

Prior to the beginning of the pretest, there was a familiarization phase to ensure participants know all the words with meanings they would be presented with and recognize the pairs of target words with corresponding photographs. The familiarization phase begins with the instruction page in Korean on a computer screen, informing each target word with its meaning in Korean and the corresponding picture (Figure 42, see Appendix D for English translated version). And then, naturally produced tokens of six training words from one of the training talkers (i.e., not manipulated training stimuli) were played three times (6 words x 1 talker x 3 repetitions = 18).

After the familiarization phase, participants heard one stimulus on each trial and were required to press either '1' or '2' on the keyboard. The *hid*, *head*, or *hood* words always corresponded with the '1' button and *heed*, *had*, or *who'd* corresponded with the '2' button. Participants were told that there would be a ten second time limit to respond. Once participants pressed one of the buttons, there was a short pause (500 ms) to indicate the start of a new trial. In total, there were 450 trials, presented in six blocks of 25 trials per block (225 trials from Talker 3

+ 225 trials from Talker 4 = 450 trials). All trials within a block were stimuli from one talker and presented in random order. The order of the presentation of the stimuli sets in the pretest was fixed. First, participants heard the set of stimuli from Talker 3, then the stimuli set from Talker 4. The duration of the pretest was 20 minutes total.



지금 부터 들을 영어 단어는 hide(숨다)의 과거형을 뜻하는 단어
혹은 (남의 충고경고에) 주의를 기울이다 라는 뜻을 가진 단어 입니다.

hide(숨다)의 과거형에 해당하는 답은 아래의 그림 중
왼쪽에 있는 그림 (키보드 숫자 '1') 입니다.
(남의 충고경고에) 주의를 기울이다에 해당하는 답은 아래의 그림 중
오른쪽에 있는 그림 (키보드 숫자 '2') 입니다.

지금부터 연습을 시작 합니다.
사진과 함께 그에 맞는 단어가 3번 자동 재생 됩니다.

준비 되었으면 스페이스바를 눌러주세요.

Figure 42. Examples of one of the instruction pages for the daily familiarization phase (*hid* and *heed* for training English /i/-/ɪ/ contrast).

Upon completion of the pretest, participants began the first HVPT session. Training sessions were always started with a daily familiarization phase, which was the same as the familiarization phase for the pretest. HVPT had the same structure as the pretest but with a few differences. In training, trial-by-trial feedback was provided to the participant about her/his performance. The feedback informed the correctness of participant's every trial. When the participant chose the right picture, a smiley face appeared on a computer screen; however, when the participant chose the wrong picture, he/she again heard the same stimulus with a simultaneous visual presentation of the correct picture/answer. There was a 500 ms interval before a new trial started. Seventy-five stimuli were played during each block of training by talkers; therefore, 225 stimuli were presented per training session in random order (25 stimuli per vowel contrast × 3 training talkers × 3 vowel contrasts = 225). As soon as a training session ended, an everyday ID test started. The procedure was identical to training, except no feedback was given to participants. The set of stimuli from Talker 3 was used for this test, resulting in 25 trials for each vowel contrast (25 trials × 3 contrasts = 75). An everyday ID test was repeated after each training session. Figure 43 shows the steps of training sessions and the overall description of the training process with trial-by-trial feedback. Each training session with everyday ID test averaged 30 minutes with approximately 2 hours and 30 minutes of training in total.

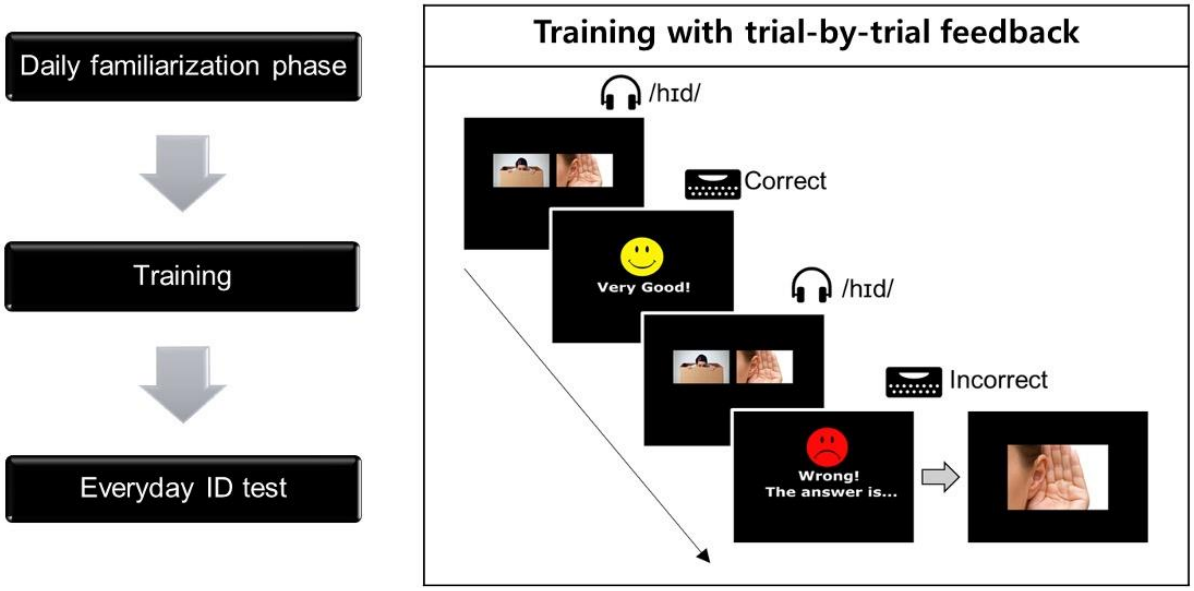


Figure 43. Steps of each training session (left) and description of training with trial-by-trial feedback (right).

*5.1.3.2.2. Day 2, Day 3, & Day 4*

Participants completed one training session each day from Day 2 to Day 4 online when they were available. Participants were given a link directing them to the online webpage for training. They were advised not to take more than a five-minute break while receiving the training online. Once participants finished each training session, the data was automatically saved on the PsyToolkit server. The average time and standard deviation taken to complete one training session were 17.2 (7.78) minutes.

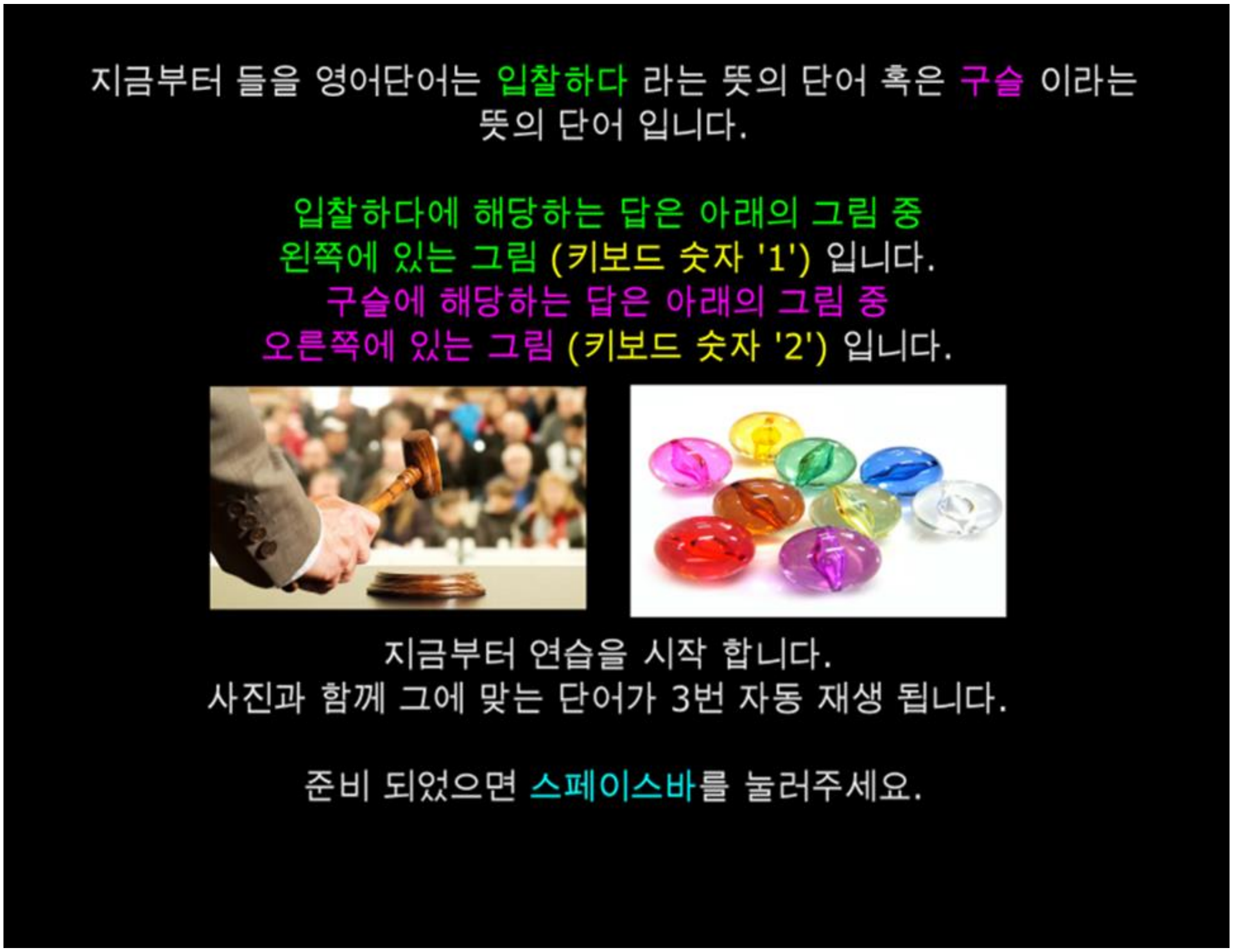
*5.1.3.2.3. Day 5*

For the last day of the experiment, all participants were required to come back to the lab to complete the tasks distributed on Day 5: training session 5, the post-test, two new types of generalization tests, and the LexTale English proficiency test (Lemhöfer & Broersma, 2012).

Participants were given an option to complete the last day of the experiment online through a live Zoom meeting (Zoom Video Communications Inc., 2020) due to the COVID-19 pandemic situation. However, all participants in experiment 3 chose to come back to the lab for Day 5. During the last day of the experiment, participants completed their last session of HVPT, then the post-test (i.e., identical to the pretest), followed by the two generalization tests. The first generalization test was the new talker generalization test. A total of 75 stimuli (25 stimuli × 3 contrasts) from Talker 4, whose stimuli were not included as training stimuli, were presented with three repetitions (75 stimuli × 3 repetitions = 225 trials). The structure of this generalization test was identical to everyday ID tests, except trial-by-trial feedback was not given. The second generalization test was the new word generalization test. As mentioned in section 5.1.2.2, Talker 3 produced six additional English words, which were *bid*, *bead*, *bed*, *bad*, *nook*, and *nuke*, for the generalization test. 225 trials were presented (25 stimuli × 3 contrasts × 3 repetitions = 225 trials). Before the new word generalization test, an instruction informed participants of the meanings of six additional English words in Korean and their corresponding phonographs (Figure 44, see Appendix E for English translated version). Then, the natural production of these words by Talker 3 (i.e., not manipulated stimuli) was played with three repetitions to familiarize participants with these new words.

An English proficiency test was necessary since the English proficiency of Korean learners of English participating in this experiment may vary. Lexical Test for Advanced Learners of English or LexTale English proficiency test (Lemhöfer & Broersma, 2012) was carried out online to check whether participants have similar L2 English proficiency. This test consists of a simple un-speeded visual lexical decision task. During this test, participants were asked to decide whether the presented word is an existing English word or not by clicking a

“Yes” or “No” button on a computer screen. There were 60 trials, which took about 4-5 minutes on average to complete. LexTale test is intended for cognitive researchers studying medium to highly proficient speakers of English as a second language in an experimental setting. It has been shown to give a fair indication of general English proficiency. For example, in a large-scale study (Lemhöfer & Broersma, 2012) on Dutch and Korean advanced learners of English, LexTale scores were found to be good predictors of vocabulary knowledge, to give a fair indication of general English, and to correlate well with experimental word recognition data (from lexical decision and progressive demasking experiments).



지금부터 들을 영어단어는 입찰하다 라는 뜻의 단어 혹은 구슬 이라는
뜻의 단어 입니다.

입찰하다에 해당하는 답은 아래의 그림 중
왼쪽에 있는 그림 (키보드 숫자 '1') 입니다.
구슬에 해당하는 답은 아래의 그림 중
오른쪽에 있는 그림 (키보드 숫자 '2') 입니다.

지금부터 연습을 시작 합니다.
사진과 함께 그에 맞는 단어가 3번 자동 재생 됩니다.

준비 되었으면 스페이스바를 눌러주세요.

Figure 44. Examples of one of the instruction pages for the new consonant generalization test (*bid* and *bead* for testing generalization of learning English /i/-/ɪ/ contrast).

### *5.1.4. Analysis*

#### 5.1.4.1. The AXB oddity task

Participants selected either *A* or *B* stimulus during the oddity task, which sounded more distinct from the *X* stimulus. *A* and *B* stimuli had either a one-step different spectral or duration step from the *X* stimulus, while the step for the other acoustic dimension remained the same as the *X* stimulus. To quantify individual variability in within-category spectral cue changes (i.e., sensitivity to differences in the spectral dimension between *A*, *X*, and *B* stimuli), the proportion of selecting a stimulus with a different spectral step from the *X* stimulus as a more distinct stimulus (i.e., odder one) was calculated for each participant. A higher proportion was considered to indicate a higher sensitivity to changes in the spectral dimension. There were 64 pairs of AXB stimuli with two versions (e.g., 2-1-6 and 6-1-2). To increase the reliability of responses and reduce the possibility to include randomly chosen responses in the analysis, only instances in which a participant consistently chose the stimulus with a different spectral step in both versions of the AXB pair were considered. This indicates that any responses a participant split in her/his responses were not included in the analysis. For example, I counted the cases as “the choice of spectrally different stimuli” only when the participant chose the spectrally different stimuli twice (i.e., in both 2-1-6 and 6-1-2 cases). Therefore, 100% score indicates that a participant selected a spectrally different stimulus during the entire AXB oddity task. For the group analysis, learners were divided into two groups based on the results. Participants who dominantly selected stimuli with different spectral steps (i.e., more sensitive to changes in the spectral dimension) were assigned to the *High Sensitive (HS)-group,* while the other learners were assigned to the *Low Sensitive (LS)-group*. It was done by taking the average of all proportion scores. The participants above the mean were put in the HS group and those below the mean in the LS group.

#### 5.1.4.2. HVPT

The percentage of correctness were collected for all types of tests and used for group and individual analysis: the pretest and post-test, everyday ID tests (Day1 ID test, Day2 ID test, Day3 ID test, Day4 ID test, Day 5 ID test, henceforth), and the new word generalization tests.

#### *5.1.4.2.1. Group analysis (LS-group vs. HS-group)*

As in experiment 1 and 2, linear mixed-effects models using the *lmer*() function from the *lme4* package in *R* were used to analyze participants' scores on tests (i.e., percentage of correctness) in each group. Two separate models were built; one was to compare the scores of the pretest, post-test, and new word generalization tests, and the other was to compare the scores of everyday ID tests. The models included participants, item-level predictors, and their 2-way and 3-way interactions. The same analytical methods and packages in *R* used in experiments 1 and 2 were used here to standardize and center predictors. The participant-level predictor was Group (LS vs. HS), which was centered to -0.5 and 0.5. The item level predictors included Contrast (/i/-/ɪ/, /ɛ/-/æ/, and /ʊ/-/u/ levels), which was centered, and Test, which was dummy coded. For the first model, Test variable was coded to Pre VERSUS Post and Pre VERSUS New Words contrasts, which had the pretest scores as the reference level. For the second model, Test variable was coded to Day1 VERSUS Day2, Day1 VERSUS Day3, Day1 VERSUS Day4, and Day1 VERSUS Day5 contrasts with Day 1 scores as the reference level. Test variable for the second model was coded using the modified version of *lizContrasts4* function (Dong et al., 2019; Wonnacott et al., 2017) to create four dummy variables that stand in place of a five-way factor (condition). The modified version was named *lizContrasts5*. Both models included participants as a random effect (i.e., random intercepts for participants), along with random slopes for Test by participants. The approach for analyses was to inspect models for effects and interactions between the experimental variables where there were clear predictions. For the training data, this is the case for the main effects of Group and Test and the interaction between them. Thus, the analyses began by inspecting these effects in the model. It was then checked to see there was a main effect of

Contrast or interaction between the effects and Contrast. Where this was the case, it was considered to reflect differences in accuracy between three levels of Contrast (/i/-/ɪ/, /ɛ/-/æ/, and /ʊ/-/u/). To break down the effects of Group and Test for each level of Contrast, each level of contrast was investigated by building a separate linear mixed-effect model.

A mixed-effects logistic regression model using the *glmer*() function from the *lme4* package in *R* was implemented to analyze each group's perceptual cue-weighting patterns in their perception of English vowel contrasts as a function of spectral and duration dimensions. To examine how participants changed their weights on spectral and duration cues before and after HVPT, the responses for the pretest and the post-test were analyzed. Three separated logistic models were built, one for each target English vowel contrasts, /i/-/ɪ/, /ɛ/-/æ/, and /ʊ/-/u/. Building three separate models were motivated by the prediction that three vowel contrasts might differ in their acquisition patterns due to their relatively different difficulty (Kim et al., 2018). Participants' responses of *hid*, *head*, and *hood* and responses of *heed*, *had*, and *who'd* were coded as 0/1, respectively. The participant-level variable was Group (LS vs. HS), and the item level predictors were Spectral, Duration, and Test (pretest vs. post-test). The Group and Test variables were centered to -0.5 and 0.5, so that main effects of Spectral and Duration were evaluated as the average effects over all levels of the Group and Test variables. Spectral and Duration were continuous variables consisting of five spectral steps and five duration steps of test stimuli. Both variables were standardized by centering and dividing by 2 standard deviations. All three logistic models included random intercepts for participants to account for participant-specific variability in responses. Random slopes for participants for Spectral, Duration, and Test were included to account for by-participant variability in the effect of each variable on their pre and post-tests responses. The resulting estimated beta-coefficients for a given variable show the change in log odds of the relevant response given a single standard-deviation increase in that variable.

The same logistic regression analysis was employed for participants responses for the new talker generalization test. This analysis was to examine whether participants resulted in the change of weighting of spectral and duration dimensions from the test taken before training (i.e., pretest) to the test taken after training with stimuli produced by a novel talker (i.e., new talker generalization test.

### *5.1.4.2.2. Individual analysis*

Alongside analysis of group patterns of HVPT performances and acoustic cue use, individual participants' different scores in the AXB oddity task allow for a detailed investigation on the relationships between individual variability in the sensitivity to the spectral dimension in the perception of native category and individual variability in the use of acoustic cues in the perception of L2 English vowel contrasts. Therefore, the present study focuses on the extent of individual variability in the successful acquisition of spectral and duration cues for L2 vowel categories by considering a large range of individual differences in the AXB oddity task.

For the individual analysis, the same linear regression mixed-effects analysis for the group analysis was used to analyze how individual differences in the AXB oddity task are related to native English participants' performances in the English vowel contrast training. The identical linear regression models for the group analysis (see 5.1.4.2.1) were built again with the variable `AXB oddity` instead of the `Group` variable included for the group analysis. The `AXB oddity` variable was a continuous variable, which includes individual participants' AXB oddity task scores. The effects of AXB oddity, Test, and their 2-way interactions were checked first, and then the effect of Contrast was discussed by building separate models for each vowel contrast.

The mixed-effects logistic regression analysis employed for the group analysis was utilized for the individual analysis. The purpose of this analysis was to examine the relationship between individual variabilities in the AXB oddity task and the change of weighting of spectral and duration dimensions from the pretest to the post-test and from the pretest to the new talker generalization test. Individual participants' binary responses in the pretest, post-test, and new talker generalization tests were submitted to the logistic regression mixed-effect models. The responses of *hid*, *head*, and *hood* were coded as 0, while the responses of *heed*, *had*, and *who'd* were coded as 1. In the model, the AXB oddity variable was included as a fixed effect with Spectral, Duration, and Test (pretest vs. post-test or pretest vs. new talker generalization test. Three separate models were built for each vowel contrast, taking into account their different learning paths (i.e., different relative difficulties in learning).

Lastly, the relationships between individual variability in the AXB oddity task and participants' use of spectral and duration dimensions in the identification of three English vowel contrasts were examined. Individual participants' cue weights of spectral and duration dimensions were computed via separate logistic regression analysis for each participant with Spectral and Duration as predictors of perceived vowel category. This analysis was performed to investigate whether participants' reliance on spectral dimension after training is related to their sensitivity to the spectral dimension in the native category perception measured by the AXB oddity task. Thus, participants' responses of the post-test with the set of stimuli from the old talker was submitted to the logistic regression analysis. The beta-coefficients ($\beta_{\text{spec}}$ and $\beta_{\text{dur}}$) from each model reflect each participant's perceptual weight assigned to spectral and duration dimension: the greater the coefficient value is, the more weight is given to that acoustic dimension. Once beta-coefficients for each participant were computed, the partial correlation analysis was conducted between participants' AXB oddity task scores and each of the coefficients ($\beta_{\text{spec}}$ or $\beta_{\text{dur}}$). The full results of participants' coefficients are in Appendix J.

### *5.1.5. Results*

#### 5.1.5.1. The AXB oddity task

Table 24 shows the results of the AXB oddity task. Results demonstrated considerable variability in participants' proportion of consistent selection of spectrally different stimulus as the more distinct stimulus from the compared stimulus *X*. Some participants chose the spectrally different stimuli more often during the AXB oddity task than others. The average percentage was 70%. Twelve participants with below the average percentage were assigned to the LS group, and the LS group's average percentage was 58%. The rest of the twelve participants with above average percentage were assigned to the HS group, and the HS group's average percentage was 81%. Table 24 suggests that overall, participants tended to choose the stimulus with a different spectral step from the *X* stimulus in AXB stimuli pairs as an oddball one since no participants chose the spectrally different stimuli lower than 50% of trials in the AXB task.

The English proficiency of participants in the HS and LS groups was controlled based on their LexTale scores. Statistical analysis showed that the HS and LS groups did not differ in their LexTale scores, $p = .19$.

Table 24. Results of the Korean /i/ vowel AXB oddity task by participants in experiment 3.

| Participants | Proportion of selecting a spectrally different stimulus in percentage (%) |
|---|---|
| 1 | 59 |
| 2 | 57 |
| 3 | 50 |
| 4 | 65 |
| 5 | 62 |
| 6 | 69 |
| 7 | 52 |
| 8 | 65 |
| 9 | 30 |
| 10 | 64 |
| 11 | 69 |
| 12 | 62 |
| 13 | 88 |
| 14 | 76 |
| 15 | 94 |
| 16 | 72 |
| 17 | 89 |
| 18 | 82 |
| 19 | 81 |
| 20 | 79 |
| 21 | 95 |
| 22 | 68 |
| 23 | 76 |
| 24 | 79 |
| Average (*SD*) | 70 (*15*) |

#### 5.1.5.2. Results of HVPT by groups (LS group vs. HS group)

##### *5.1.5.2.1. Pre, post, and new words generalization tests*

Figure 45 shows the average percentage of correctness in the pretest, post-test, and new words generalization tests for each of the three English vowel contrasts at group level. Table 25 summarizes the results of linear regression models including Group, Test (Pre VERSUS Post and Pre VERSUS New Words), and Contrast variables. There were main effects of Pre VERSUS Post and Pre VERSUS New Words, reflecting improvement from the pretest to the post-test and that training transferred to new words, which participants did not learn during training. A significant 2-way interaction between Group and Pre VERSUS Post suggests that the HS group found a larger degree of improvement from the pretest to the post-test compared to the LS group. There were reliable 2-way and 3-way interactions with Contrast variable: Contrast and Pre VERSUS Post, Contrast and Pre VERSUS New Words, Group, Contrast, and Pre VERSUS Post, and Group, Contrast, and Pre VERSUS New Words. The interactions with the Contrast variable suggest that a difference in the proportion of correct responses between the pretest and the post-test and the pretest and the new word generalization test were modulated by different English vowel contrasts. In other words, participants demonstrated greater improvement with a certain vowel contrast than others. To break down the interaction effects with Contrast variable, we looked at each target English vowel contrast by building a separate linear regression mixed-effects model.

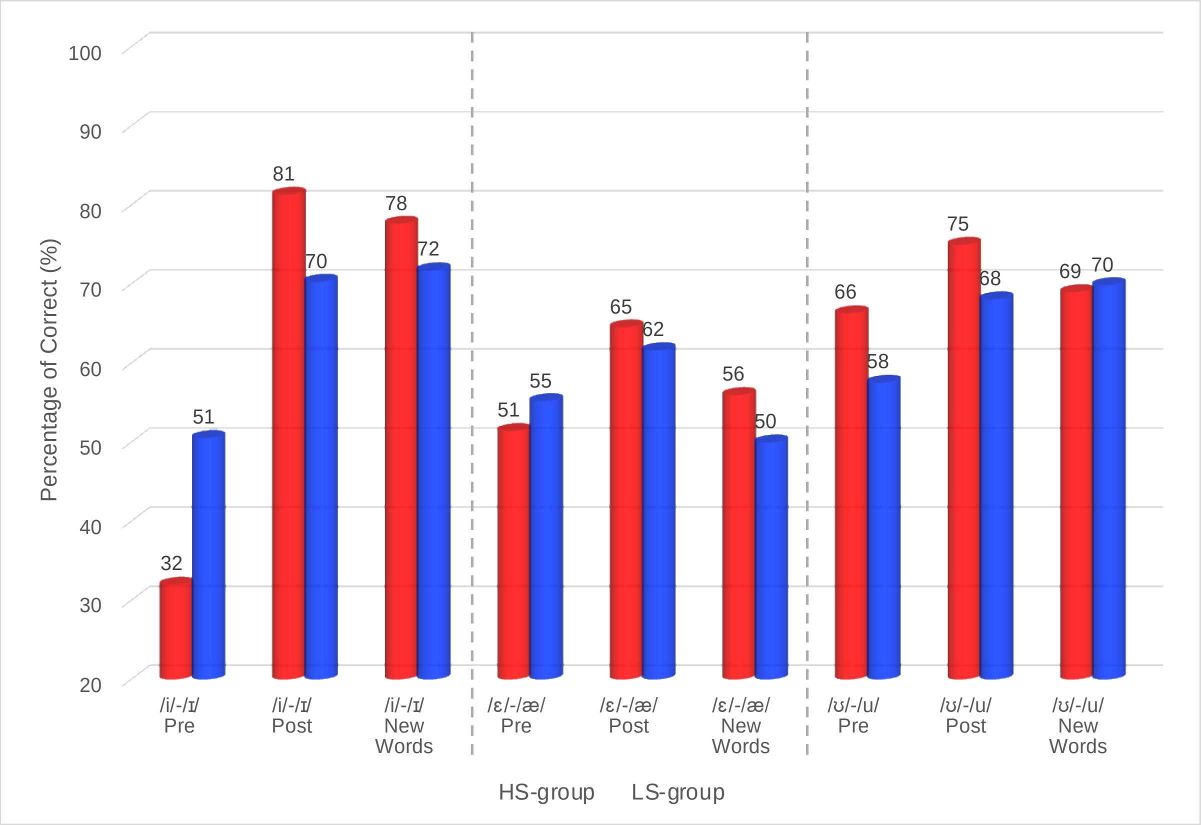


Figure 45. Results of the pretest, the post-test, and the new words generalization test for each of the three English vowel contrasts in percentage correct (%). HS-group: red bar; LS-group: blue bar.

Table 25. Summary of fixed effects for the mixed-effects linear regression model for the pretest, post-test, and new word generalization tests scores.

| Predictor | Estimate ($\beta$) | Std. Error | *t-value* | *p-value* | |
|---|---|---|---|---|---|
| (Intercept) | 62.52 | 1.41 | 44.46 | <0.001 | *** |
| Group | 1.68 | 2.81 | 0.60 | 0.556 | |
| Contrast | 2.61 | 2.04 | 1.28 | 0.204 | |
| Pre_VERSUS_Post | 17.98 | 2.50 | 7.20 | <0.001 | *** |
| Pre_VERSUS_New Words | 13.02 | 2.82 | 4.61 | <0.001 | *** |
| Group:Contrast | 3.64 | 4.09 | 0.89 | 0.375 | |
| Group:Pre_VERSUS_Post | 11.44 | 4.99 | 2.29 | 0.023 | * |
| Group:Pre_VERSUS_New Words | 7.15 | 5.64 | 1.27 | 0.213 | |
| Contrast:Pre_VERSUS_Post | -20.41 | 5.01 | -4.08 | <0.001 | *** |
| Contrast:Pre_VERSUS_New Words | -22.51 | 5.01 | -4.50 | <0.001 | *** |
| Group:Contrast:Pre_VERSUS_Post | -25.73 | 10.01 | -2.57 | 0.011 | * |
| Group:Contrast:Pre_VERSUS_New Words | -30.46 | 10.01 | -3.04 | 0.003 | ** |

Table 26, Table 27, and Table 28 show the results of three linear regression mixed-effects models built based on participants' proportions of correct responses in the pre-test, the post-test, and the new word generalization test for each target vowel contrast. For the /i/-/ɪ/ contrast, the model found significant main effects of Pre VERSUS Post and Pre VERSUS New Words, indicating that participants could identity *hid/heed* and *bid/bead* word pairs better after receiving training than before training. There were significant 2-way interactions between Group and Pre VERSUS Post and Group and Pre VERSUS New Words, suggesting that the HS group showed greater improvements in the identification of /i/-/ɪ/ contrast in the post-test and the new word generalization test compared to their pretest performances than the LS group. For /ɛ/-/æ/, there was a reliable main effect of Pre VERSUS Post, suggesting that participants showed overall better performance in the post-test than the pretest in identifying *head* and *had* words. There was no reliable interaction. For /ʊ/-/u/ contrast, there was a significant main effects of Pre VERSUS Post, reflecting that participants identified *hood/who'd* word pair better in the post-test than they identified *hood/who'd* in the pre-test.

When comparing the three vowel contrasts, the overall group results indicate different patterns of development between /i/-/ɪ/, /ɛ/-/æ/, and /ʊ/-/u/ contrasts throughout training. Participants showed higher scores in the post-test for all three vowel contrasts compared to the scores in the pretest. However, notably, participants did not show their ability in generalization for the /ɛ/-/æ/ and /ʊ/-/u/ contrasts whereas participants resulted in better transfer of learning to new words for the /i/-/ɪ/ contrast. As shown in Figure 45, the group differences were more evident for the /i/-/ɪ/ contrast than /ɛ/-/æ/ and /ʊ/-/u/ contrasts; the HS group showed higher group average scores on the post-test than the LS group for the /i/-/ɪ/ contrast (/i/-/ɪ/: 81% vs. 70%).

Table 26. Summary of fixed effects for the mixed-effects linear regression model for the pretest, post-test, and new word generalization tests scores for the /i/-/ɪ/ contrast.

| Predictor | Estimate ($\beta$) | Std. Error | *t-value* | *p-value* | |
|---|---|---|---|---|---|
| (Intercept) | 63.94 | 2.17 | 29.49 | <0.001 | *** |
| Group | -0.56 | 4.34 | -0.13 | 0.899 | |
| Pre_VERSUS_Post | 34.56 | 4.33 | 7.99 | <0.001 | *** |
| Pre_VERSUS_New Words | 33.44 | 4.78 | 7.00 | <0.001 | *** |
| Group:Pre_VERSUS_Post | 29.56 | 8.65 | 3.42 | 0.002 | ** |
| Group:Pre_VERSUS_New Words | 24.44 | 9.56 | 2.56 | 0.015 | * |

Table 27. Summary of fixed effects for the mixed-effects linear regression model for the pretest, post-test, and new word generalization tests scores for the /ɛ/-/æ/ contrast.

| Predictor | Estimate ($\beta$) | Std. Error | *t-value* | *p-value* | |
|---|---|---|---|---|---|
| (Intercept) | 56.48 | 2.03 | 27.86 | <0.001 | *** |
| Group | 1.70 | 4.05 | 0.42 | 0.678 | |
| Pre_VERSUS_Post | 9.78 | 4.16 | 2.35 | 0.027 | * |
| Pre_VERSUS_New Words | -0.33 | 3.53 | -0.09 | 0.926 | |
| Group:Pre_VERSUS_Post | 6.67 | 8.33 | 0.80 | 0.431 | |
| Group:Pre_VERSUS_New Words | 9.78 | 7.07 | 1.38 | 0.179 | |

Table 28. Summary of fixed effects for the mixed-effects linear regression model for the pretest, post-test, and new word generalization tests scores for the /ʊ/-/u/ contrast.

| Predictor | Estimate ($\beta$) | Std. Error | *t-value* | *p-value* | |
|---|---|---|---|---|---|
| (Intercept) | 67.13 | 1.78 | 37.64 | <0.001 | *** |
| Group | 3.89 | 3.57 | 1.09 | 0.286 | |
| Pre_VERSUS_Post | 9.61 | 2.49 | 3.86 | 0.001 | ** |
| Pre_VERSUS_New Words | 5.94 | 3.31 | 1.80 | 0.082 | . |
| Group:Pre_VERSUS_Post | -1.89 | 4.99 | -0.38 | 0.708 | |
| Group:Pre_VERSUS_New Words | -12.78 | 6.61 | -1.93 | 0.062 | . |

### *5.1.5.2.2. Everyday ID tests*

Figure 46 plots the group results of the percentage of correct responses for everyday ID tests taken from the first to the last training sessions. The results from the different groups are represented in different colors. The visual inspection of Figure 46 demonstrates that the HS group showed better performance in training than the LS group. The HS group could identify the target training words better than the LS group, especially at the later stage of training (i.e., Day 5).

Table 29 summarizes the results of the linear regression mixed-effects analysis with Group, Test (Day1 VERSUS Day2, Day1 VERSUS Day3, Day1 VERSUS Day4, and Day1 VERSUS Day5), and Contrast variables and their interactions. The results of regression analysis in confirmed the visual observation of Figure 46. There were reliable main effects of Group, Day1 VERSUS Day3, Day1 VERSUS Day4, and Day1 VERSUS Day5, showing that the HS group overall showed better performance in everyday ID tests and participants improved in the identification of target training words in Day 3, Day 4, and Day 5 tests compared to their performances in Day 1 test. There were significant and nearly significant 2-way interactions between Group and levels of the Test variable: Group and Day1 VERSUS Day2, Group and Day1 VERSUS Day3, and Group and Day1 VERSUS Day5. These results reflect that improvement across everyday ID tests with Day 1 test scores as a reference was bigger for the HS group than the LS group. The model in Table 29 also found a reliable main effect of Contrast, suggesting that the developmental patterns across everyday ID tests were different for each of three vowel contrasts.

Figure 47 shows the average scores of everyday ID tests by groups for each of three vowel contrasts, and Figure 48 plots the estimated percentage of correctness in everyday ID tests for each vowel contrast. Both Figure 47 and Figure 48 reflect that participants scored greater accuracy in the identification of *hid/heed* than *head/had* and *hood/who'd* word pairs. Critically, Figure 47 shows that on all everyday ID tests, the HS group's average scores for all three English contrasts were higher than the ones of the LS group.

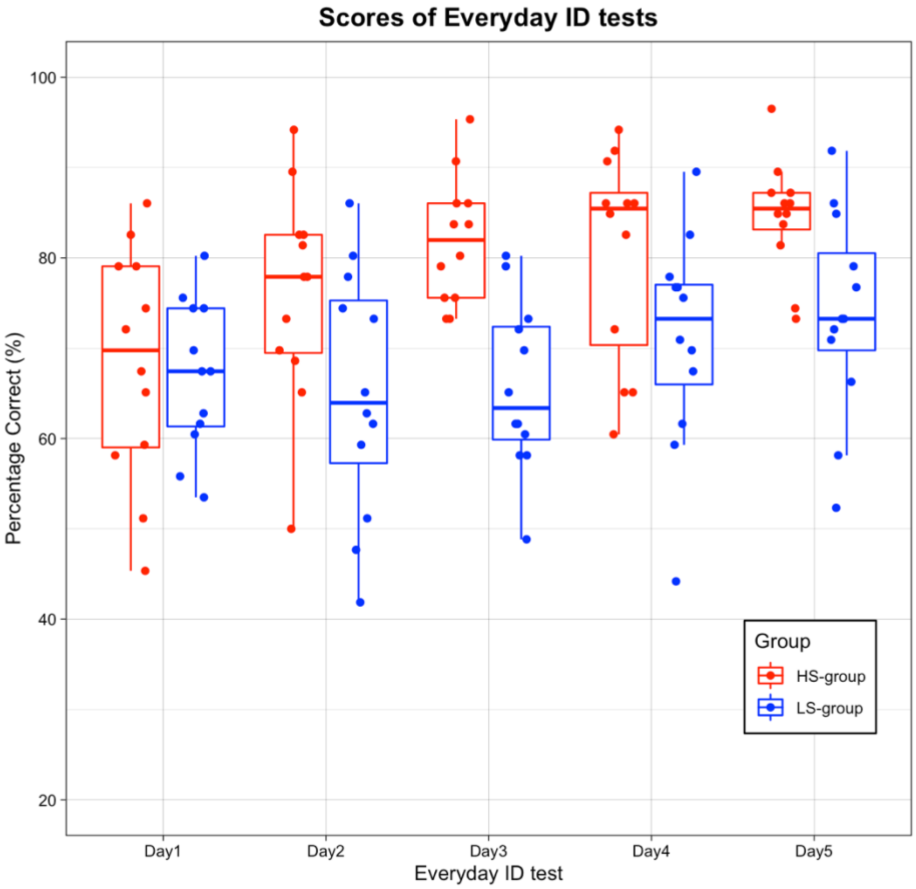


Figure 46. Results of everyday ID tests by group (HS group vs. LS group) in percentage of correct (%).

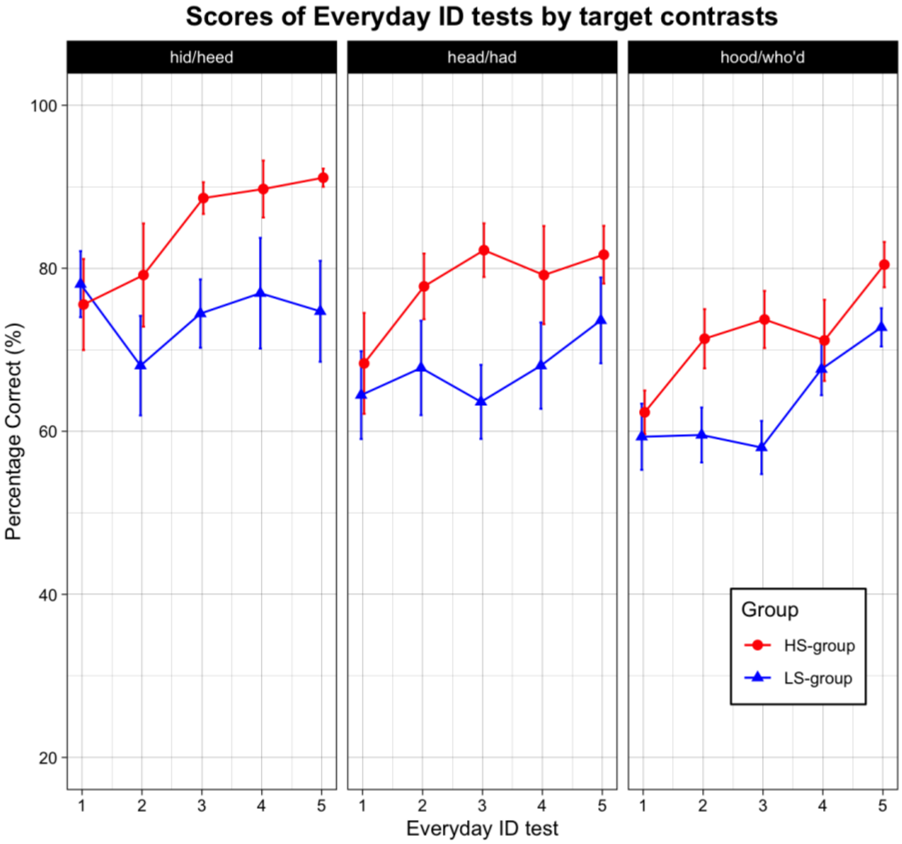


Figure 47. Results of everyday ID tests by three English vowel contrasts in percentage of correct (%). Each line represents either the HS or the LS group.

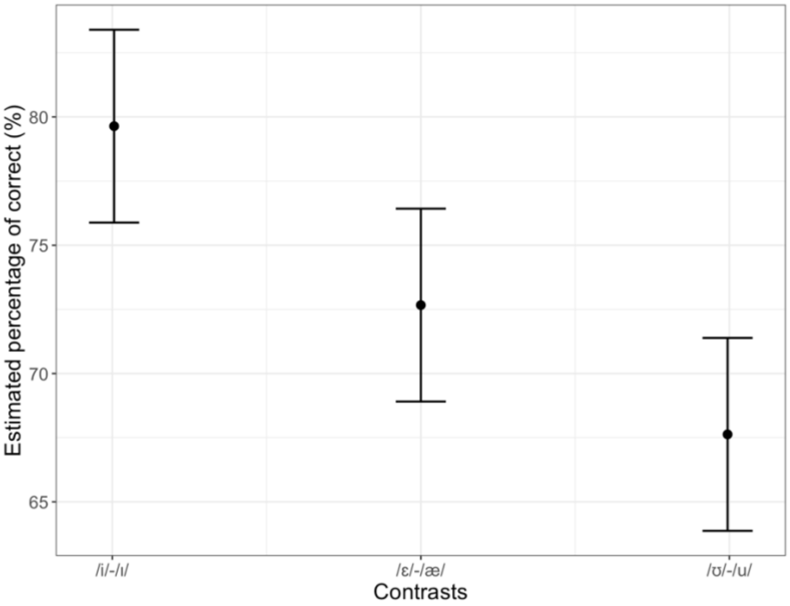


Figure 48. Estimated percentage of correct in everyday ID tests for each of vowel contrasts drawn by the linear regression models with `Contrast`, `Group`, and `Test` variables and their interactions.

Table 29. Summary of fixed effects of the linear regression model for everyday ID test scores.

| Predictor | Estimate ($\beta$) | Std. Error | *t-value* | *p-value* | |
|---|---|---|---|---|---|
| (Intercept) | 73.31 | 1.66 | 44.09 | <0.001 | *** |
| Group | 9.7029 | 3.33 | 2.91 | 0.008 | ** |
| Contrast | -9.82 | 1.36 | -7.22 | <0.001 | *** |
| Day1_VERSUS_Day2 | 2.61 | 2.39 | 1.09 | 0.283 | |
| Day1_VERSUS_Day3 | 5.43 | 2.22 | 2.44 | 0.017 | * |
| Day1_VERSUS_Day4 | 7.44 | 2.65 | 2.80 | 0.010 | ** |
| Day1_VERSUS_Day5 | 11.04 | 2.29 | 4.81 | <0.001 | *** |
| Group:Contrast | -1.67 | 2.73 | -0.61 | 0.540 | |
| Group:Day1_VERSUS_Day2 | 9.52 | 4.79 | 1.99 | 0.054 | . |
| Group:Day1_VERSUS_Day3 | 14.72 | 4.46 | 3.30 | 0.001 | ** |
| Group:Day1_VERSUS_Day4 | 7.69 | 5.31 | 1.45 | 0.160 | |
| Group:Day1_VERSUS_Day5 | 9.26 | 4.60 | 2.02 | 0.048 | * |
| Contrast:Day1_VERSUS_Day2 | 6.39 | 4.30 | 1.49 | 0.139 | |
| Contrast:Day1_VERSUS_Day3 | 0.25 | 4.30 | 0.06 | 0.953 | |
| Contrast:Day1_VERSUS_Day4 | 1.66 | 4.30 | 0.39 | 0.700 | |
| Contrast:Day1_VERSUS_Day5 | 7.90 | 4.30 | 1.84 | 0.067 | . |
| Group:Contrast:Day1_VERSUS_Day2 | -3.93 | 8.62 | -0.46 | 0.648 | |
| Group:Contrast:Day1_VERSUS_Day3 | -3.24 | 8.62 | -0.38 | 0.707 | |
| Group:Contrast:Day1_VERSUS_Day4 | -12.08 | 8.62 | -1.40 | 0.162 | |
| Group:Contrast:Day1_VERSUS_Day5 | -11.62 | 8.62 | -1.35 | 0.178 | |

A follow-up analysis was conducted to confirm the visual inspection of Figure 47, which suggests the group differences in the acquisition of each of the target English vowel contrasts. Specifically, it was examined whether the effects of Group, Test, and their 2-way interactions also hold for the separate analysis for each vowel contrasts. Table 30, Table 31, and Table 32 summarize the results of separate linear regression model analysis for three target English vowel contrasts. For the /i/-/ɪ/ contrast, there was a reliable main effect of Group and nearly significant main effects of Day1 VERSUS Day4 and Day1 VERSUS Day5, suggesting that the HS group recorded a higher overall score averaged across everyday ID tests than the score of the LS group. More importantly, the model found significant 2-way interactions between Group and Test: Group and Day1 VERSUS Day3, Group and Day1 VERSUS Day4, and Group and Day1 VERSUS Day5. These interactions suggest reliable evidence of faster learning in the HS group than in the LS group shown by the comparisons of the LS and the HS groups' subsequent performances in everyday ID tests compared to the Day 1 ID test.

For the /ɛ/-/æ/ contrast, there was a nearly significant main effect of Group and significant main effects of all levels of Test variable, indicating that the HS group showed the overall higher performance and participants in both groups showed an improved skill in the identification of *head/had* word pair as they received more training sessions. The model critically revealed that the HS group demonstrated a larger improvement from the Day1 to the Day 3 ID test than the LS group, as evidenced by the reliable 2-way interaction between Group and Day1 VERSUS Day3.

Lastly, statistical analysis for the /ʊ/-/u/ contrast was conducted. There were main effects of Group and Day1 VERSUS Day3, Day1 VERSUS Day4, and Day1 VERSUS Day5, reflecting improvement across training sessions and overall better performance of the HS group than the

LS group. The reliable interaction between `Group` and `Day1 VERSUS Day3` shows that improvement from the Day1 to Day3 ID tests was larger for the HS group.

Table 30. Summary of fixed effects for the mixed-effects linear regression model for everyday ID test scores for the /i/-/ɪ/ contrast.

| Predictor | Estimate ($\beta$) | Std. Error | *t-value* | *p-value* | |
|---|---|---|---|---|---|
| (Intercept) | 79.64 | 2.45 | 32.49 | <0.001 | *** |
| Group | 10.39 | 4.90 | 2.12 | 0.045 | * |
| Day1_VERSUS_Day2 | -3.19 | 3.89 | -0.82 | 0.416 | |
| Day1_VERSUS_Day3 | 4.72 | 3.22 | 1.47 | 0.149 | |
| Day1_VERSUS_Day4 | 6.53 | 3.47 | 1.88 | 0.067 | . |
| Day1_VERSUS_Day5 | 6.11 | 3.22 | 1.90 | 0.064 | . |
| Group:Day1_VERSUS_Day2 | 13.61 | 7.78 | 1.75 | 0.088 | . |
| Group:Day1_VERSUS_Day3 | 16.67 | 6.44 | 2.59 | 0.013 | * |
| Group:Day1_VERSUS_Day4 | 15.28 | 6.95 | 2.20 | 0.033 | * |
| Group:Day1_VERSUS_Day5 | 18.89 | 6.44 | 2.93 | 0.005 | ** |

Table 31. Summary of fixed effects for the mixed-effects linear regression model for everyday ID test scores for the /ɛ/-/æ/ contrast.

| Predictor | Estimate ($\beta$) | Std. Error | *t-value* | *p-value* | |
|---|---|---|---|---|---|
| (Intercept) | 72.67 | 2.76 | 26.38 | <0.001 | *** |
| Group | 10.33 | 5.51 | 1.88 | 0.073 | . |
| Day1_VERSUS_Day2 | 6.39 | 2.91 | 2.20 | 0.031 | * |
| Day1_VERSUS_Day3 | 6.53 | 2.91 | 2.25 | 0.028 | * |
| Day1_VERSUS_Day4 | 7.22 | 3.52 | 2.05 | 0.046 | * |
| Day1_VERSUS_Day5 | 11.25 | 2.91 | 3.87 | <0.001 | *** |
| Group:Day1_VERSUS_Day2 | 6.11 | 5.81 | 1.05 | 0.296 | |
| Group:Day1_VERSUS_Day3 | 14.72 | 5.81 | 2.53 | 0.013 | * |
| Group:Day1_VERSUS_Day4 | 7.22 | 7.04 | 1.03 | 0.311 | |
| Group:Day1_VERSUS_Day5 | 4.17 | 5.81 | 0.72 | 0.476 | |

Table 32. Summary of fixed effects for the mixed-effects linear regression model for everyday ID test scores for the /ʊ/-/u/ contrast.

| Predictor | Estimate ($\beta$) | Std. Error | *t-value* | *p-value* | |
|---|---|---|---|---|---|
| (Intercept) | 67.63 | 1.59 | 42.46 | < 2e-16 | *** |
| Group | 8.35 | 3.19 | 2.62 | 0.02 | * |
| Day1_VERSUS_Day2 | 4.62 | 2.54 | 1.82 | 0.07 | . |
| Day1_VERSUS_Day3 | 5.03 | 2.38 | 2.11 | 0.04 | * |
| Day1_VERSUS_Day4 | 8.56 | 3.01 | 2.84 | 0.01 | ** |
| Day1_VERSUS_Day5 | 15.77 | 2.38 | 6.63 | 0.00 | *** |
| Group:Day1_VERSUS_Day2 | 8.81 | 5.07 | 1.74 | 0.09 | . |
| Group:Day1_VERSUS_Day3 | 12.71 | 4.76 | 2.67 | 0.01 | * |
| Group:Day1_VERSUS_Day4 | 0.53 | 6.02 | 0.09 | 0.93 | |
| Group:Day1_VERSUS_Day5 | 4.69 | 4.76 | 0.99 | 0.33 | |

#### *5.1.5.2.3. Perceptual cue weighting (Pre vs. Post-test with old talker)*

Heat plots in Figure 49 show participants' overall responses of the pretest and the post-test to spectral and duration cues averaged across three vowel contrasts. Figure 50 shows responses to duration and spectral steps for each vowel contrast in order to focus on each group's changes in weighting of either spectral (i.e., primary) or duration (i.e., secondary) cues from the pretest to the post-test.

Visual inspection of the upper plots in Figure 49 shows that participants used the duration dimension relatively more than the spectral dimension to distinguish two English words in each pair. The HS and LS group participants' cue-weighting strategies in the pretest demonstrate that they did not dominantly use spectral cues. This differs from native English listeners' cue-weighting strategies in that they predominantly use spectral quality and at the same time, duration has a much weaker effect on their vowel categorization (see Figure 40). However, the lower plots in Figure 49 show that both groups differentiated their cue-weighting strategies somewhat closer to native listeners' strategies (i.e., more active use of spectral dimension).

However, there was a notable group difference. In the post-test, the HS group relied on spectral dimension dominantly like native listeners by consistently classifying stimuli with higher spectral steps as *heed/had/who'd*, while stimuli with lower spectral steps as *hid/head/hood*. On the other hand, the LS group's responses did not show this tendency as strong as the HS group, indicating that participants in the LS group used both acoustic dimensions relatively similarly in their categorization of target English vowels.

Figure 50, which demonstrates the extent to which duration and spectral dimensions contribute to participants' vowel categorization for each contrast, shows that both groups used spectral dimension more in the post-test for the /i/-/ɪ/ contrast than /ɛ/-/æ/ and /ʊ/-/u/ contrasts. Thus, three separate logistic regression models were built for each vowel contrast to examine the HS and LS groups' spectral and duration dimensions in vowel categorization and how each group's cue-weighting strategies changed after training.

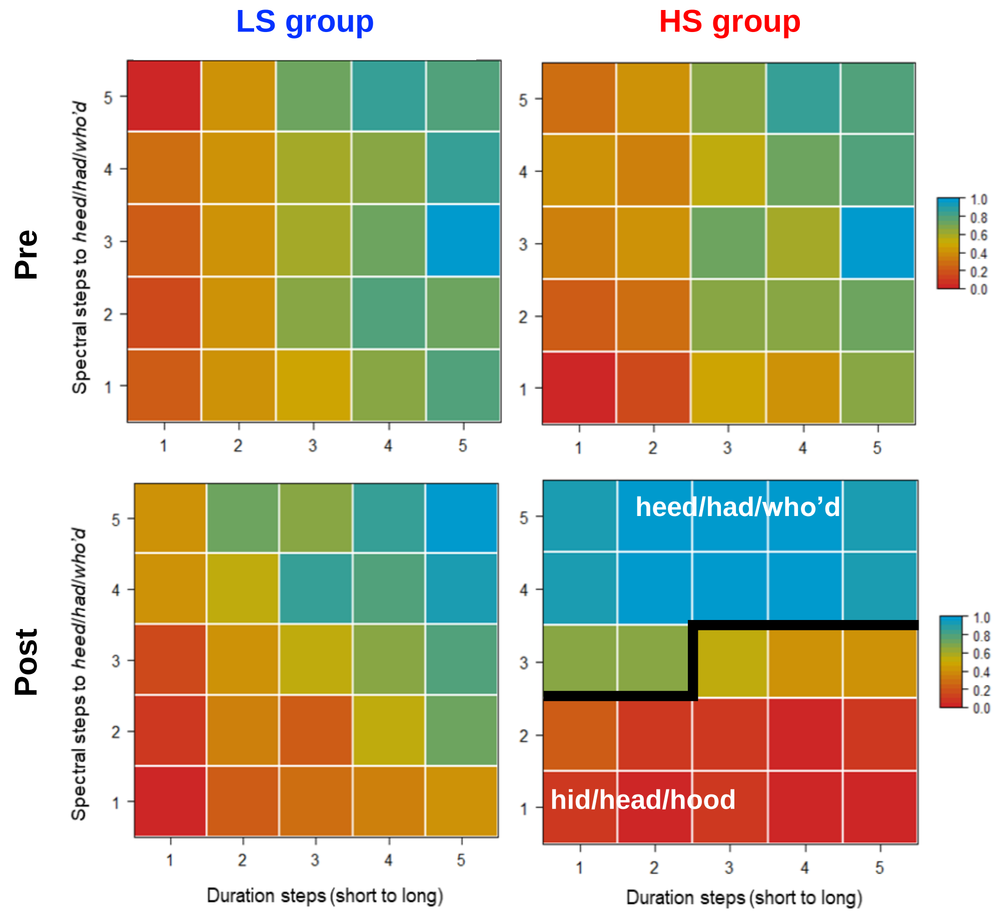


Figure 49. Heat plots of participants' overall responses of the pretest and the post-test to each combination of spectral and duration dimensions averaged across three vowel contrasts. The darkness of the cell represents the percentage of responses in a forced-choice task; the darkest red cells elicited 100% '*hid/head/hood'* responses, while the darkest blue cells elicited 100% '*heed/had/who/d*.'

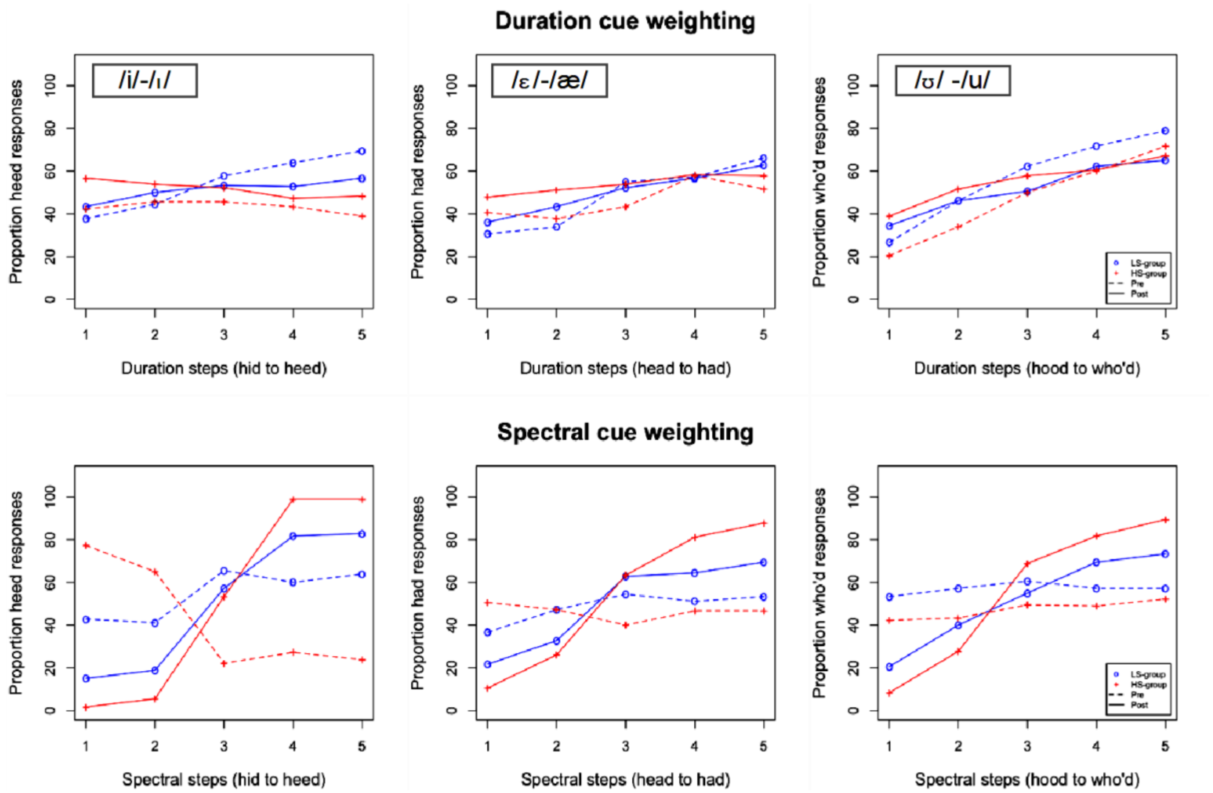


Figure 50. Proportion of heed/had/who'd responses by the LS group (blue) and the HS group (red) in the pretest (dashed lines) and the post-test (solid lines) with the "old" talker.

Figure 51 shows predicted values of /i/ responses as a function of standardized spectral and duration steps for the pretest and the post-test with a set of stimuli from the "old" talker. The left-side graph represents the estimated proportion of /i/ responses for standardized five spectral steps, while the right-side graph shows the estimated proportion of /i/ responses for standardized five duration steps. The logistic regression model results for the /i/-/ɪ/ contrast are summarized in Table 33. The results showed a significant main effect of `Spectral`, indicating that participants gave more /i/ vowel responses (i.e., *heed*) as the vowel spectral steps increased. The model found significant 2-way interactions between each acoustic dimension and `Test` variable, indicating that participants differed their reliance on spectral and duration cues in classifying the /i/-/ɪ/ contrast in the post-test compared to the pretest. Participants increased their reliance on spectral dimension but decreased their reliance on duration dimension, which resembles the perceptual pattern by native listeners of English. The significant 3-way interaction between `Group`, `Spectral`, and `Test` indicates that the effect of spectral dimension in the post-test increased more in the HS group than the LS group.

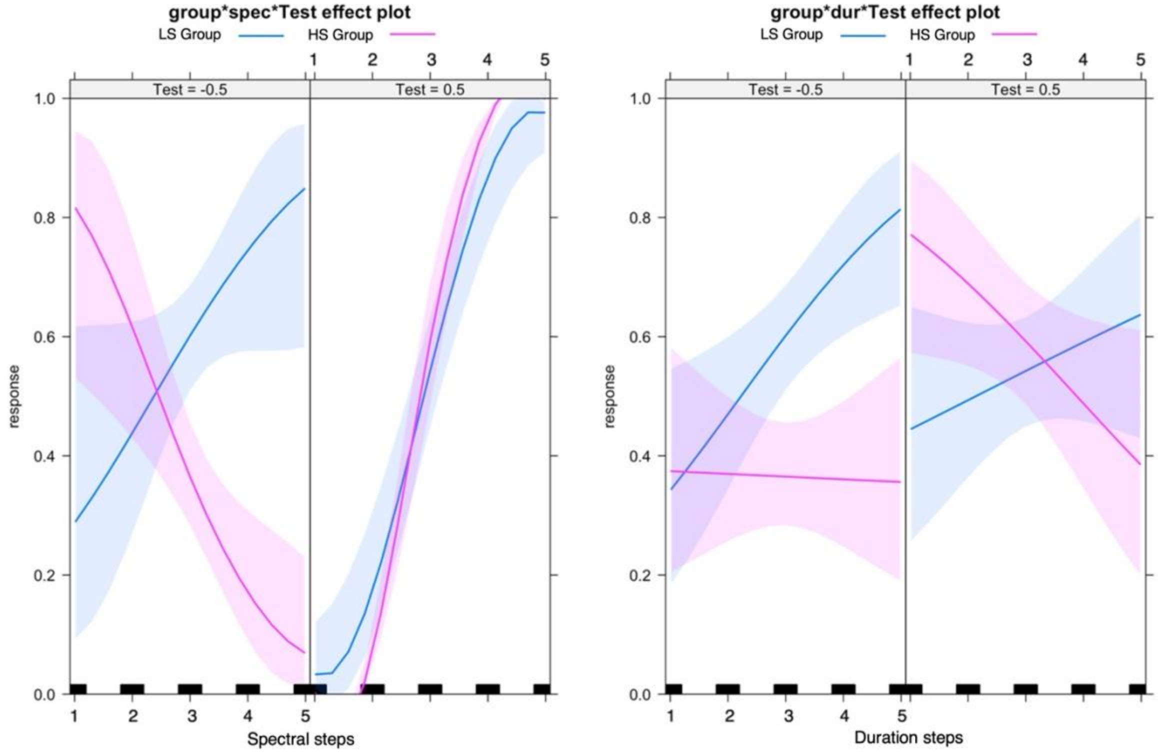


Figure 51. Predicted logit curves for perception of the /i/-/ɪ/ contrast by the LS group (blue) and the HS group (pink). Predictions are drawn from a binary logistic regression model with `Group`, `Spectral`, `Duration`, and `Test` as predictors. Each panel represents the results from the pretest (`Test` = -0.5) and the post-test (`Test` = 0.5) with the "old" talker. The y-axis of each panel represents the predicted probability of a "heed" response.

Table 33. Summary of fixed effects for the mixed-effects logistic regression model predicting changes in spectral and duration cue weights from the pretest to the post-test for the /i/-/ɪ/ contrast.

| Predictor | Estimate ($\beta$) | Std. Error | *z-value* | *p-value* | |
|---|---|---|---|---|---|
| (Intercept) | 0.10 | 0.13 | 0.79 | 0.43 | |
| Group | -0.38 | 0.26 | -1.49 | 0.14 | |
| Duration | 0.20 | 0.38 | 0.54 | 0.59 | |
| Spectral | 3.62 | 0.69 | 5.22 | 0.00 | *** |
| Test | 0.34 | 0.11 | 3.15 | 0.00 | ** |
| Group:Duration | -1.66 | 0.76 | -2.18 | 0.03 | * |
| Group:Spectral | 0.32 | 1.39 | 0.23 | 0.82 | |
| Group:Test | 1.17 | 0.22 | 5.36 | 0.00 | *** |
| Duration:Test | -1.05 | 0.22 | -4.71 | 0.00 | *** |
| Spectral:Test | 8.29 | 0.41 | 20.00 | < .001 | *** |
| Group:Spectral:Test | 10.20 | 0.82 | 12.40 | < .001 | *** |
| Group:Duration:Test | -0.19 | 0.44 | -0.42 | 0.67 | |

Figure 52 shows predicted values of /æ/ responses as a function of standardized spectral and duration steps for the pretest and the post-test with a set of stimuli from the “old” talker. The logistic regression model results for the /ɛ/-/æ/ contrast are summarized in Table 34. The results showed significant main effects of Duration and Spectral, indicating that participants gave more /æ/ responses (i.e., *had*) as the vowel spectral and duration steps increased. The model found a significant 2-way between Spectral and Test and a 3-way interaction between Spectral, Group, and Test, indicating that the use of spectral cues changed from the pretest to the post-test and did so differently for the two groups. The effect of spectral dimension increased at the post-test, which was greater for the HS group participants. As shown in Figure 52, the HS group’s response curve of the post-test as a function of spectral steps was steeper and more canonical (i.e., higher steps of spectral dimension eliciting more of /æ/ responses) than the curve by the LS group. Taken together, although both groups demonstrated the greater reliance on spectral dimension in vowel categorization in the post-test, participants in the HS group depended more heavily on spectral dimension compared to participants in the LS group.

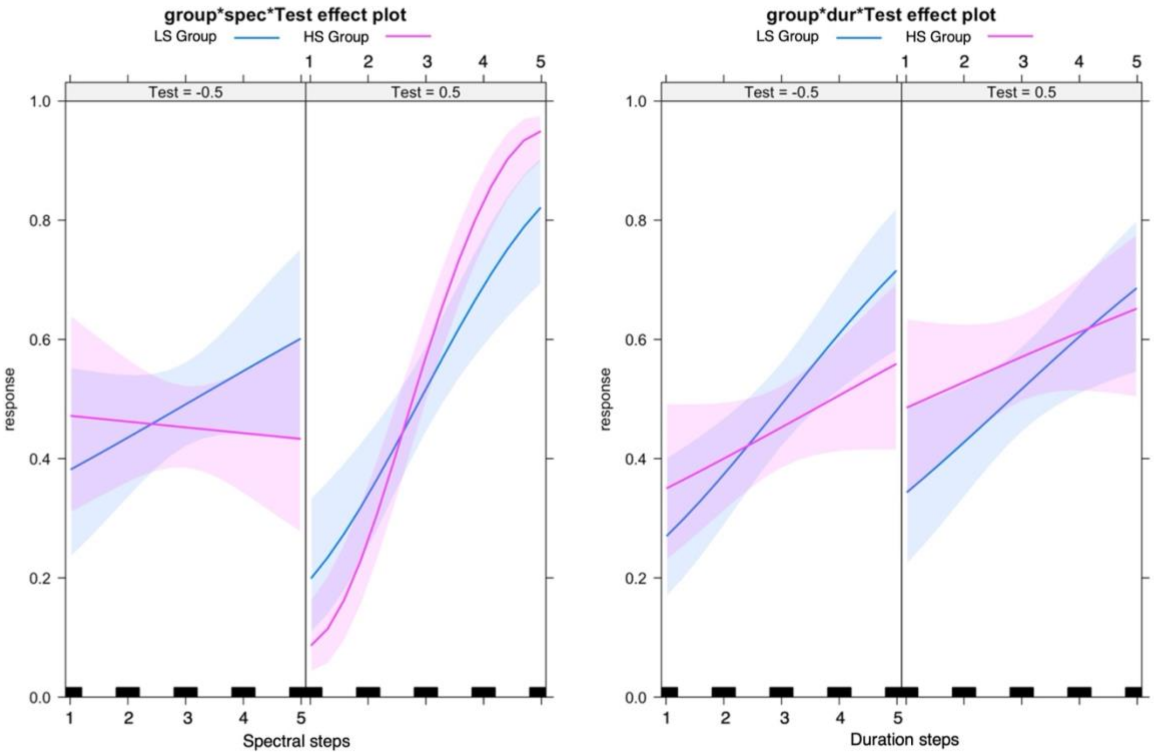


Figure 52. Predicted logit curves for perception of the /ɛ/-/æ/ contrast by the LS group (blue) and the HS group (pink). Predictions are drawn from a binary logistic regression model with `Group`, `Spectral`, `Duration`, and `Test` as predictors. Each panel represents the results from Day 1 ID test (`Test` = -0.5) and Day 3 ID test (`Test` = 0.5). The y-axis of each panel represents the predicted probability of a "had" response.

Table 34. Summary of fixed effects for the mixed-effects logistic regression model predicting changes in spectral and duration cue weights from the pretest to the post-test for the /ɛ/-/æ/ contrast.

| Predictor | Estimate ($\beta$) | Std. Error | *z*-value | *p*-value | |
|---|---|---|---|---|---|
| (Intercept) | 0.03 | 0.10 | 0.30 | 0.765 | |
| Group | 0.04 | 0.20 | 0.20 | 0.839 | |
| Duration | 0.90 | 0.26 | 3.50 | <0.001 | *** |
| Spectral | 1.61 | 0.32 | 4.98 | <0.001 | *** |
| Test | 0.26 | 0.17 | 1.51 | 0.130 | |
| Group:Duration | -0.69 | 0.51 | -1.34 | 0.180 | |
| Group:Spectral | 0.50 | 0.65 | 0.78 | 0.436 | |
| Group:Test | 0.44 | 0.35 | 1.27 | 0.204 | |
| Duration:Test | -0.26 | 0.17 | -1.58 | 0.115 | |
| Spectral:Test | 2.70 | 0.18 | 14.78 | <0.001 | *** |
| Group:Spectral:Test | 2.54 | 0.36 | 6.99 | <0.001 | *** |
| Group:Duration:Test | 0.32 | 0.34 | 0.95 | 0.342 | |

Figure 53 shows predicted values of /u/ responses as a function of standardized spectral and duration steps for the pretest and the post-test with a set of stimuli from the "old" talker. The results of the logistic regression model for the /ʊ/-/u/ contrast are summarized in Table 35. The model found significant main effects of `Duration` and `Spectral`, reflecting that participants assigned more /u/ vowel responses as the vowel spectral and duration steps increased. More importantly, significant 2-way and 3-way interactions between each of acoustic dimensions, `Group`, and `Test` suggest that the use of acoustic dimensions changed before and after training and the magnitude of those changes was different for each group. First, the interaction between `Group` and `Spectral` shows that overall, the effect of spectral dimension was larger for the HS group in vowel classification in the pretest and the post-test. The interactions between `Duration` and `Test` and `Spectral` and `Test` indicate that participants changed their cue weights for spectral and duration cues over time. The effect of `Duration` decreased, and the effect of `Spectral` increased in the post-test. The 3-way interaction between `Group`, `Spectral`, and `Test` shows that the increased effect of spectral in the post-test was greater for the HS group than the LS group. As shown in Figure 53, the response curves predicted by the model, both groups changed their cue-weighting strategies by relying more on spectral dimension than duration dimension. However, the change in participants' reliance on spectral cues was greater for the HS group participants.

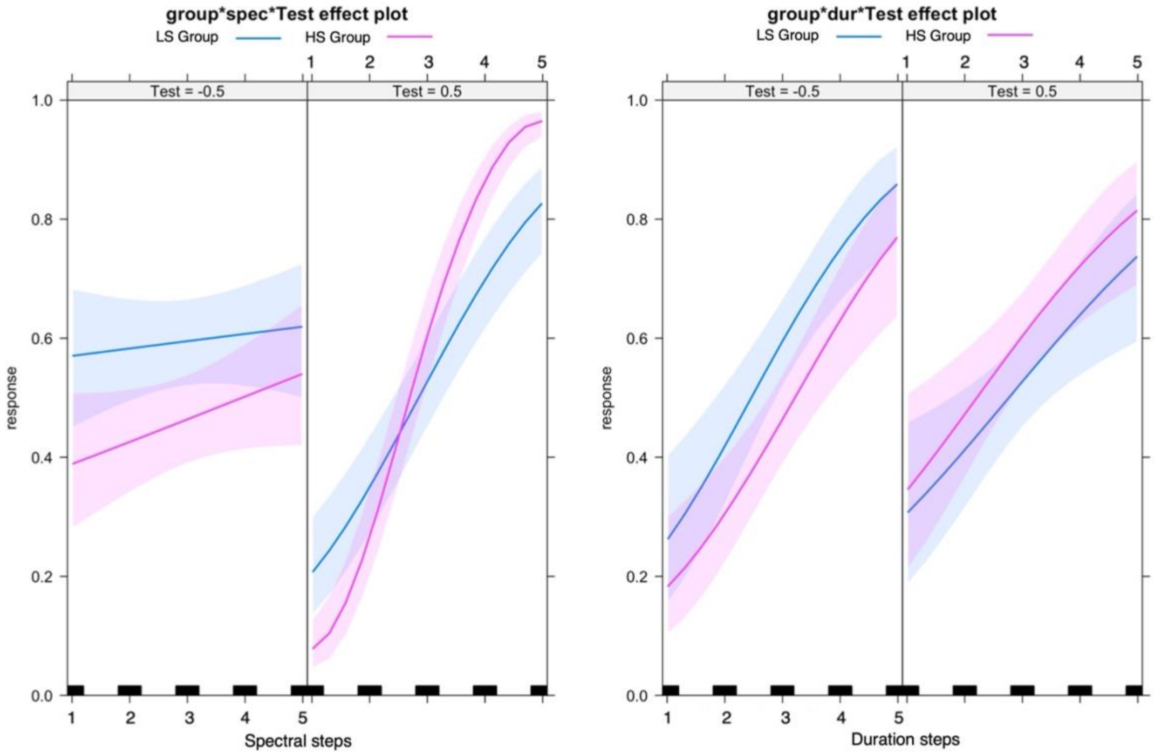


Figure 53. Predicted logit curves for perception of the /ʊ/-/u/ contrast by the LS group (blue) and the HS group (pink). Predictions are drawn from a binary logistic regression model with Group, Spectral, Duration, and Test as predictors. Each panel represents the results from Day 1 ID test (Test = -0.5) and Day 3 ID test (Test = 0.5). The y-axis of each panel represents the predicted probability of a “who’d” response.

Table 35. Summary of fixed effects for the mixed-effects logistic regression model predicting changes in spectral and duration cue weights from the pretest to the post-test for the /ʊ/-/u/ contrast.

| Predictor | Estimate ($\beta$) | Std. Error | *z-value* | *p-value* | |
|---|---|---|---|---|---|
| (Intercept) | 0.18 | 0.11 | 1.62 | 0.106 | |
| Group | -0.16 | 0.22 | -0.72 | 0.474 | |
| Duration | 1.75 | 0.29 | 5.97 | <0.001 | *** |
| Spectral | 1.70 | 0.19 | 9.06 | <0.001 | *** |
| Test | 0.15 | 0.20 | 0.75 | 0.452 | |
| Group:Duration | 0.08 | 0.59 | 0.14 | 0.891 | |
| Group:Spectral | 1.16 | 0.38 | 3.09 | 0.002 | ** |
| Group:Test | 0.93 | 0.41 | 2.29 | 0.022 | * |
| Duration:Test | -0.67 | 0.18 | -3.72 | <0.001 | *** |
| Spectral:Test | 2.83 | 0.19 | 14.89 | <0.001 | *** |
| Group:Spectral:Test | 1.70 | 0.38 | 4.50 | <0.001 | *** |
| Group:Duration:Test | 0.14 | 0.36 | 0.40 | 0.687 | |

#### *5.1.5.2.4. Perceptual cue weighting (new talker generalization test with a novel talker)*

In this section, the results of logistic mixed-effects regression analysis with data from the new talker generalization test are presented. Recall that this generalization test was conducted with the set of stimuli from a novel talker who did not serve as one of the training talkers. As in the previous section, three separate logistic regression models were built for each of the three target English vowel contrasts.

Figure 54 shows the pretest and the post-test responses with a “new” talker as a function of duration and spectral steps for each vowel contrast. Visual inspection of Figure 54 suggests that both groups were influenced by the spectral step changes overall when they classified the /i/-/ɪ/ contrast (*hid/heed*) in the generalization test. However, it seems that the effect of spectral dimension for /ɛ/-/æ/ and /ʊ/-/u/ contrasts (*head/had* and *hood/who'd*) was not as strong as for the /i/-/ɪ/ contrast. The overall group results indicate different patterns of generalization of the use of spectral dimension between three English vowel contrasts.

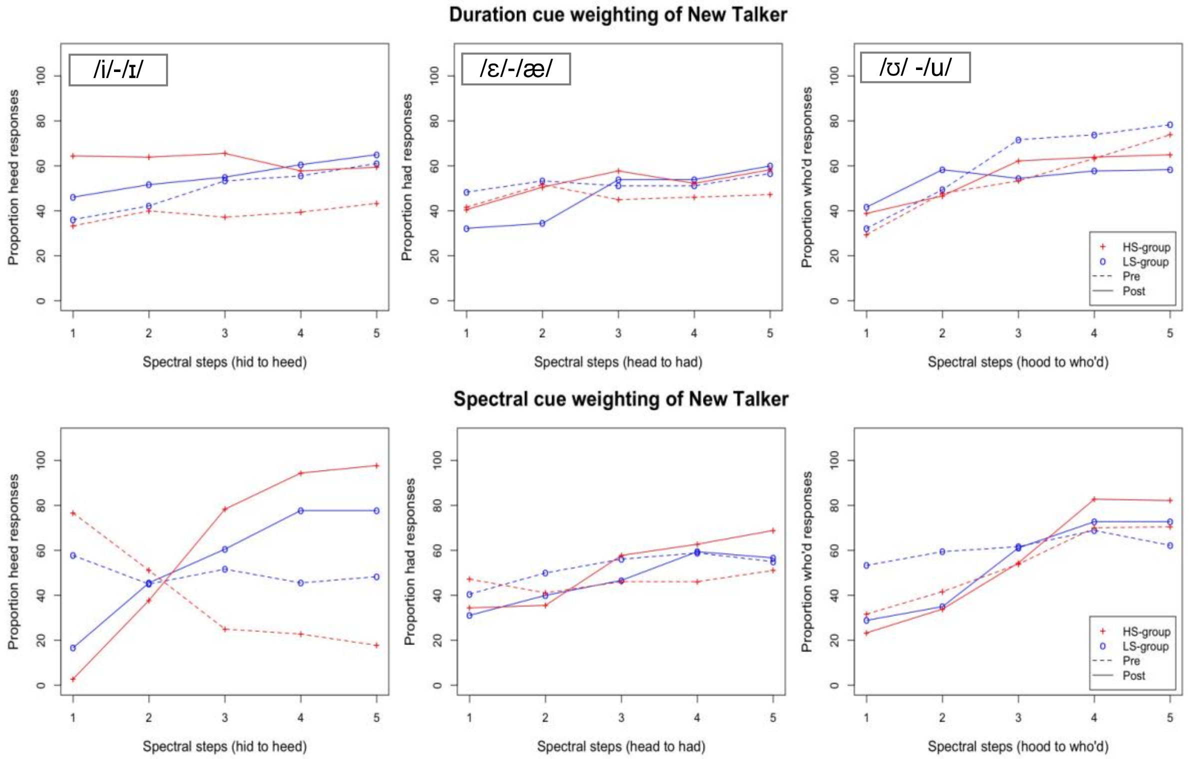


Figure 54. Proportion of heed/had/who'd responses by the LS group (blue) and the HS group (red) in the pretest (dashed lines) and the post-test (solid lines) with a "new" talker.

Figure 55 shows predicted values of /i/ responses as a function of standardized spectral and duration steps for the pretest and the post-test with a set of stimuli from a "novel" talker. As with participants' responses for the /i/-/ɪ/ contrast in the post-test with the "old" talker, novel talker's stimuli with higher spectral steps elicited more /i/ judgments than did those with lower spectral steps. The results of the logistic regression model for the /i/-/ɪ/ contrast are summarized in Table 36. The model found significant main effects of Test and Spectral, indicating that there were more /i/ responses in the post-test, and participants gave more /i/ vowel responses for stimuli with higher spectral steps. The significant 2-way interaction between Duration and Test and Spectral and Test show that participants decreased their reliance on duration dimension while increasing their reliance on spectral dimension to categorize new words of the /i/-/ɪ/ contrast in the post-test. The effect of spectral dimension was greater in the HS group than the LS group, as evidenced by the 3-way interaction between Group, Spectral, and Test.

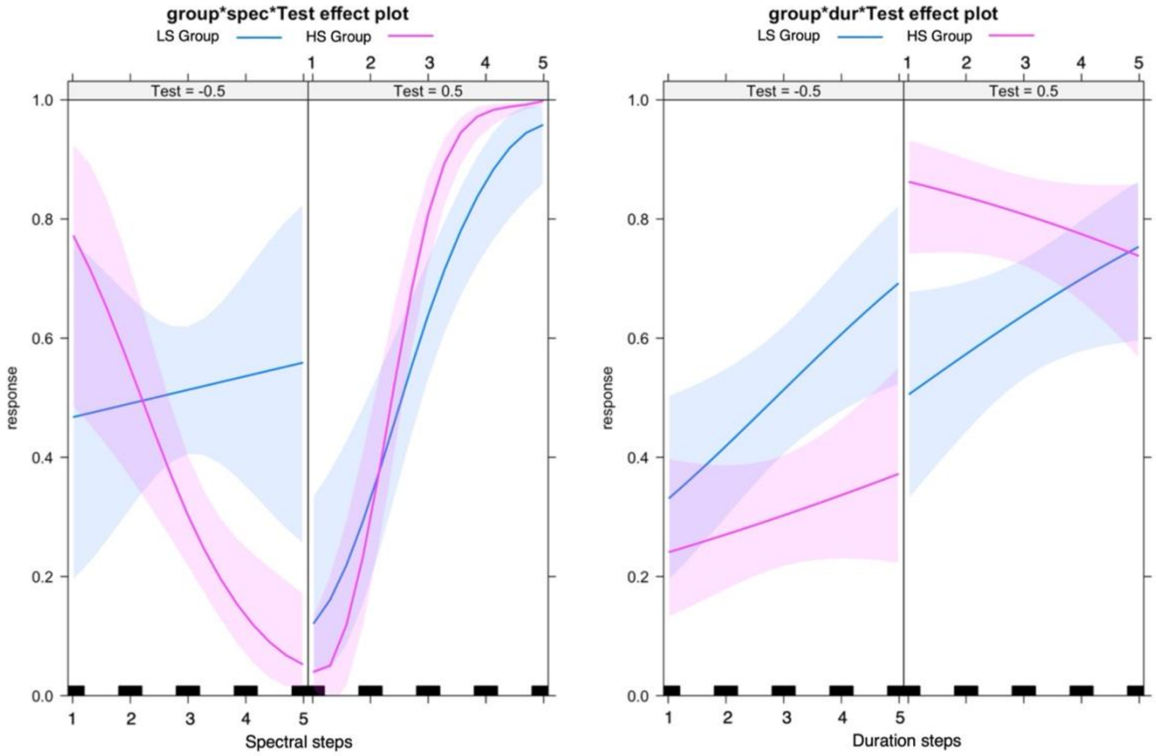


Figure 55. Predicted logit curves for perception of the /i/-/ɪ/ contrast by the LS group (blue) and the HS group (pink). Predictions are drawn from a binary logistic regression model with Group, Spectral, Duration, and Test as predictors. Each panel represents the results from the pretest (Test = -0.5) and the post-test (Test = 0.5) with a “new” talker. The y-axis of each panel represents the predicted probability of a “heed” response.

Table 36. Summary of fixed effects for the mixed-effects logistic regression model predicting changes in spectral and duration cue weights from the pretest to the post-test for the /i/-/ɪ/ contrast.

| Predictor | Estimate ($\beta$) | Std. Error | *z-value* | *p-value* | |
|---|---|---|---|---|---|
| (Intercept) | 0.32 | 0.18 | 1.79 | 0.074 | . |
| Group | -0.07 | 0.36 | -0.18 | 0.855 | |
| Duration | 0.45 | 0.29 | 1.54 | 0.125 | |
| Spectral | 2.16 | 0.71 | 3.05 | 0.002 | ** |
| Test | 1.45 | 0.31 | 4.63 | <0.001 | *** |
| Group:Duration | -1.07 | 0.58 | -1.84 | 0.066 | . |
| Group:Spectral | -0.12 | 1.41 | -0.09 | 0.930 | |
| Group:Test | 2.20 | 0.63 | 3.50 | <0.001 | *** |
| Duration:Test | -0.69 | 0.20 | -3.41 | 0.001 | *** |
| Spectral:Test | 7.24 | 0.33 | 21.93 | <0.001 | *** |
| Group:Spectral:Test | 7.23 | 0.65 | 11.19 | <0.001 | *** |
| Group:Duration:Test | -0.70 | 0.41 | -1.73 | 0.084 | . |

Figure 56 shows predicted values of /æ/ responses as a function of standardized spectral and duration steps for the pretest and the post-test with a set of stimuli from a novel talker. The logistic regression model results for the /ɛ/-/æ/ contrast are summarized in Table 37. The model found significant main effects of Duration and Spectral, suggesting that participants gave more /æ/ responses as spectral and duration steps of new talker generalization test stimuli increased. There were reliable 2-way and 3-way interactions between each acoustic dimension, Group, and Test. The 2-way interactions between Duration and Test and Spectral and Test indicate that the use of duration and spectral steps in the identification of *head* and *had* increased in the post-test. In addition, the 3-way interaction between Group, Spectral, and Test shows that the HS group's increase in the use of spectral dimension was larger than the LS group in the post-test.

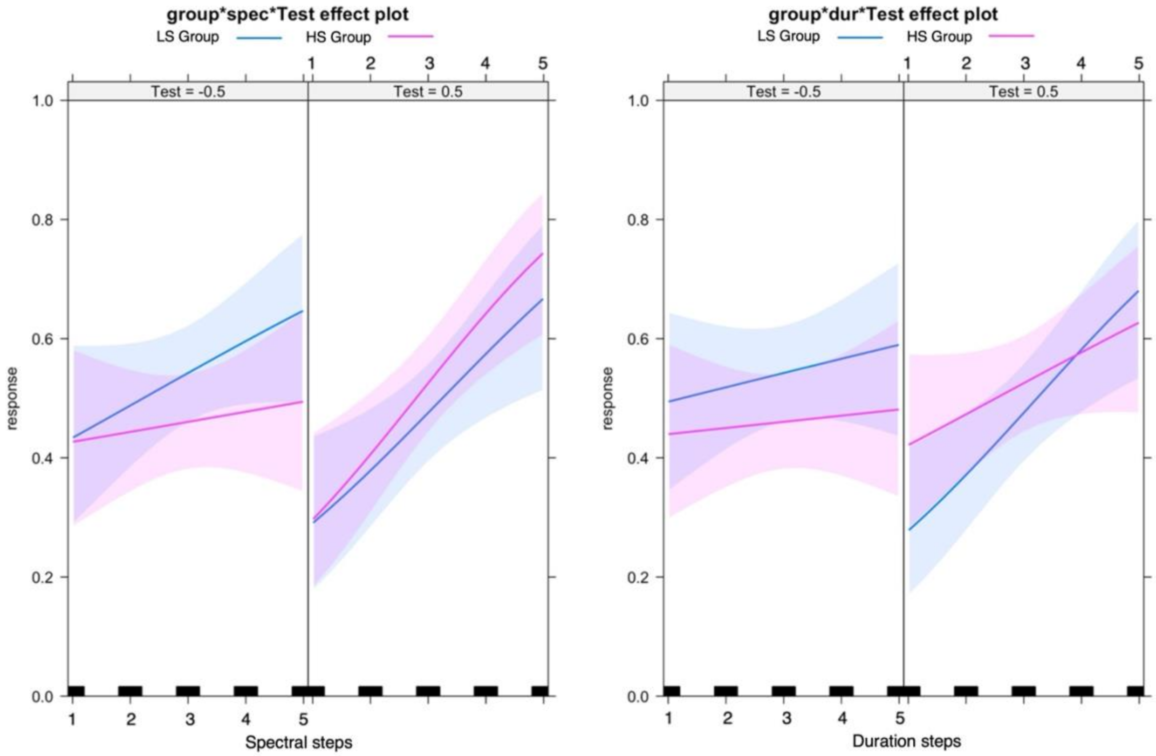


Figure 56. Predicted logit curves for perception of the /ɛ/-/æ/ contrast by the LS group (blue) and the HS group (pink). Predictions are drawn from a binary logistic regression model with Group, Spectral, Duration, and Test as predictors. Each panel represents the results from the pretest (Test = -0.5) and the post-test (Test = 0.5) with a "new" talker. The y-axis of each panel represents the predicted probability of a "had" response.

Table 37. Summary of fixed effects for the mixed-effects logistic regression model predicting changes in spectral and duration cue weights from the pretest to the post-test for the /ɛ/-/æ/ contrast.

| Predictor | Estimate ($\beta$) | Std. Error | *z-value* | *p-value* | |
|---|---|---|---|---|---|
| (Intercept) | 0.00 | 0.11 | 0.03 | 0.974 | |
| Group | -0.06 | 0.22 | -0.29 | 0.770 | |
| Duration | 0.55 | 0.26 | 2.15 | 0.032 | * |
| Spectral | 0.83 | 0.26 | 3.13 | 0.002 | ** |
| Test | 0.00 | 0.07 | -0.05 | 0.958 | |
| Group:Duration | -0.39 | 0.51 | -0.76 | 0.448 | |
| Group:Spectral | -0.09 | 0.53 | -0.17 | 0.862 | |
| Group:Test | 0.53 | 0.15 | 3.55 | <0.001 | *** |
| Duration:Test | 0.71 | 0.15 | 4.65 | <0.001 | *** |
| Spectral:Test | 0.84 | 0.15 | 5.50 | <0.001 | *** |
| Group:Spectral:Test | 0.67 | 0.31 | 2.19 | 0.029 | * |
| Group:Duration:Test | -0.47 | 0.30 | -1.53 | 0.125 | |

Figure 57 shows predicted values of /u/ responses as a function of standardized spectral and duration steps for the pretest and the post-test with a set of stimuli from a novel talker. The results of the logistic regression model for the /ʊ/-/u/ contrast are summarized in Table 38. The model found significant main effects of `Duration` and `Spectral`, indicating that both acoustic dimensions were used more in the post-test with a set of stimuli from a novel talker. The significant 2-way interaction between `Group` and `Spectral` shows that the HS group overall utilized spectral dimension more than the LS group in identifying *hood* and *who'd* words. The reliable 2-way interactions between each acoustic dimension and `Test` variable suggest that participants decreased use of duration dimension while increased spectral dimension in the post-test. The model for the /ʊ/-/u/ contrast did not find a reliable 3-way interaction between `Group`, `Spectral`, and `Test`; however, found a significant interaction between `Group`, `Duration`, and `Test`, indicating that the effect of duration dimension decreased more for the LS group in the post-test.

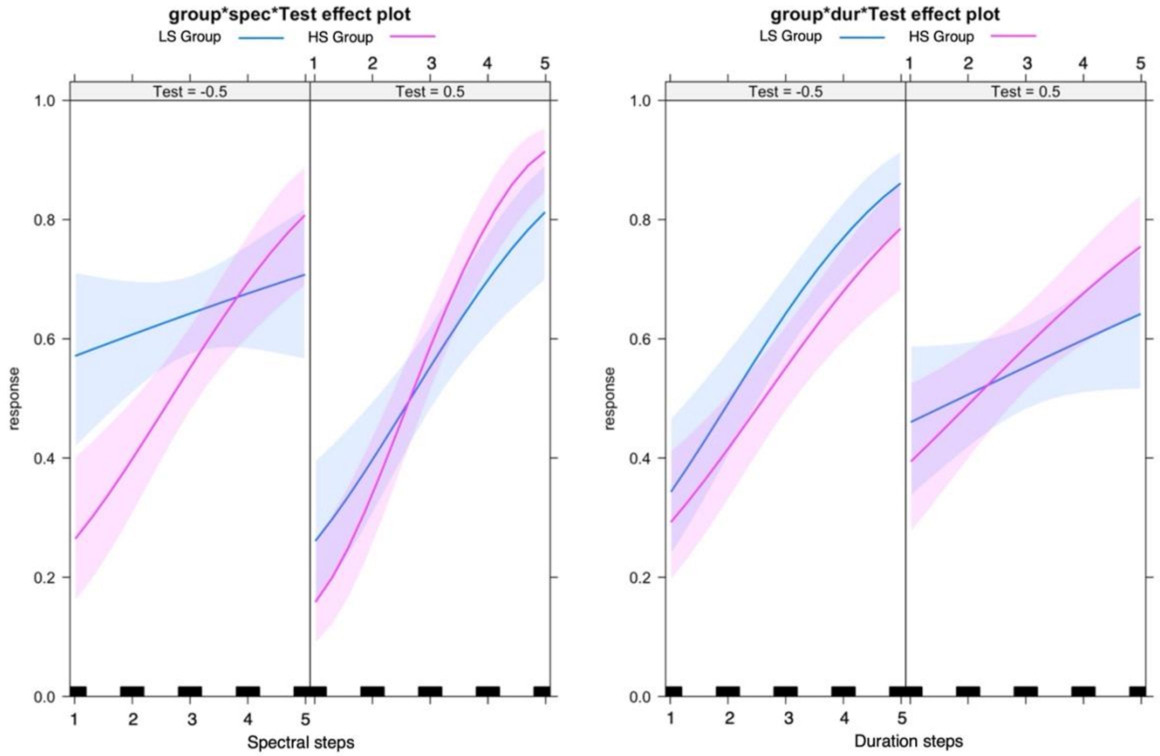


Figure 57. Predicted logit curves for perception of the /ʊ/-/u/ contrast by the LS group (blue) and the HS group (pink). Predictions are drawn from a binary logistic regression model with Group, Spectral, Duration, and Test as predictors. Each panel represents the results from the pretest (Test = -0.5) and the post-test (Test = 0.5) with a "new" talker. The y-axis of each panel represents the predicted probability of a "who'd" response.

Table 38. Summary of fixed effects for the mixed-effects logistic regression model predicting changes in spectral and duration cue weights from the pretest to the post-test for the /ʊ/-/u/ contrast.

| Predictor | Estimate ($\beta$) | Std. Error | *z-value* | *p-value* | |
|---|---|---|---|---|---|
| (Intercept) | 0.35 | 0.10 | 3.48 | <0.001 | *** |
| Group | -0.14 | 0.20 | -0.70 | 0.482 | |
| Duration | 1.29 | 0.22 | 5.84 | <0.001 | *** |
| Spectral | 1.74 | 0.28 | 6.28 | <0.001 | *** |
| Test | -0.15 | 0.17 | -0.86 | 0.393 | |
| Group:Duration | 0.22 | 0.44 | 0.50 | 0.619 | |
| Group:Spectral | 1.21 | 0.55 | 2.20 | 0.028 | * |
| Group:Test | 0.51 | 0.34 | 1.50 | 0.134 | |
| Duration:Test | -0.95 | 0.17 | -5.60 | <0.001 | *** |
| Spectral:Test | 1.24 | 0.17 | 7.21 | <0.001 | *** |
| Group:Spectral:Test | -0.33 | 0.34 | -0.95 | 0.342 | |
| Group:Duration:Test | 0.75 | 0.34 | 2.23 | 0.026 | * |

### 5.1.5.3. Results of HVPT by individuals

The same linear regression mixed-effects analysis previously employed for the examination of group results was used to analyze individual results as well. The fixed and random factors included in the models were identical to group analysis except AXB oddity variable (i.e., participants' AXB oddity test scores) was included instead of Group variable. The effect of the AXB oddity variable was considered to examine whether individual variability in HVPT training can be explained by individual differences in the AXB oddity task. Since only one fixed-effect variable, AXB oddity, is newly inserted into models while Test and Contrast variables are intact, our primary interest was to investigate a main effect of AXB oddity and interaction effects with other variables. The following section, therefore, begins by inspecting these effects in the models.

#### *5.1.5.3.1. Pre, post, and new word generalization test*

Table 39 summarizes the results of linear regression models including AXB oddity, Test (Pre VERSUS Post and Pre VERSUS New Words), and Contrast variables. There was no reliable main effects of predictors or interaction effect with AXB oddity.

Table 39. Summary of fixed effects for the mixed-effects linear regression model of the pretest, post-test, and new word generalization tests scores.

| Predictor | Estimate ($\beta$) | Std. Error | *t*-value | *p*-value | |
|---|---|---|---|---|---|
| (Intercept) | 62.52 | 1.41 | 44.22 | <0.001 | *** |
| AXB oddity | -0.86 | 2.83 | -0.30 | 0.765 | |
| Contrast | 2.61 | 2.11 | 1.24 | 0.218 | |
| Pre_VERSUS_Post | 17.98 | 2.58 | 6.98 | <0.001 | *** |
| Pre_VERSUS_New Words | 13.02 | 2.84 | 4.58 | <0.001 | *** |
| AXB oddity:Contrast | 2.11 | 4.23 | 0.50 | 0.619 | |
| AXB oddity:Pre_VERSUS_Post | 6.88 | 5.17 | 1.33 | 0.185 | |
| AXB oddity:Pre_VERSUS_New Words | 4.57 | 5.69 | 0.80 | 0.427 | |
| Contrast:Pre_VERSUS_Post | -20.41 | 5.17 | -3.95 | <0.001 | *** |
| Contrast:Pre_VERSUS_New Words | -22.51 | 5.17 | -4.36 | <0.001 | *** |
| AXB oddity:Contrast:Pre_VERSUS_Post | -15.74 | 10.36 | -1.52 | 0.130 | |
| AXB oddity:Contrast:Pre_VERSUS_New Words | -16.80 | 10.36 | -1.62 | 0.107 | |

Table 40, Table 41, and Table 42 summarize the results of linear regression analyses run separately for each of vowel contrasts. As shown in Table 40, there was a nearly significant 2-way interaction between AXB oddity and Pre VERSUS Post, indicating that the higher AXB oddity task scores become, the greater improvement from the pretest to the post-test was found in the classification of English /i/-/ɪ/ contrast.

Table 40. Summary of fixed effects for the mixed-effects linear regression model of the pretest, post-test, and new word generalization tests scores for the /i/-/ɪ/ contrast.

| Predictor | Estimate ($\beta$) | Std. Error | *t-value* | *p-value* | |
|---|---|---|---|---|---|
| (Intercept) | 67.71 | 10.59 | 6.39 | <0.001 | *** |
| AXB oddity | -0.05 | 0.15 | -0.36 | 0.720 | |
| Pre_VERSUS_Post | -13.14 | 23.84 | -0.55 | 0.587 | |
| Pre_VERSUS_New Words | 1.81 | 24.47 | 0.07 | 0.942 | |
| AXB oddity:Pre_VERSUS_Post | 0.68 | 0.33 | 2.04 | 0.052 | . |
| AXB oddity:Pre_VERSUS_New Words | 0.45 | 0.34 | 1.32 | 0.195 | |

Table 41. Summary of fixed effects for the mixed-effects linear regression model of the pretest, post-test, and new word generalization tests scores for the /ɛ/-/æ/ contrast.

| Predictor | Estimate ($\beta$) | Std. Error | *t-value* | *p-value* | |
|---|---|---|---|---|---|
| (Intercept) | 61.24 | 9.92 | 6.18 | <0.001 | *** |
| AXB oddity | -0.07 | 0.14 | -0.49 | 0.628 | |
| Pre_VERSUS_Post | 9.79 | 20.66 | 0.47 | 0.640 | |
| Pre_VERSUS_New Words | -18.96 | 17.56 | -1.08 | 0.291 | |
| AXB oddity:Pre_VERSUS_Post | 0.00 | 0.29 | 0.00 | 0.999 | |
| AXB oddity:Pre_VERSUS_New Words | 0.26 | 0.24 | 1.08 | 0.289 | |

Table 42. Summary of fixed effects for the mixed-effects linear regression model of the pretest, post-test, and new word generalization tests scores for the /ʊ/-/u/ contrast.

| Predictor | Estimate ($\beta$) | Std. Error | *t-value* | *p-value* | |
|---|---|---|---|---|---|
| (Intercept) | 58.50 | 8.42 | 6.95 | <0.001 | *** |
| AXB oddity | 0.14 | 0.12 | 1.23 | 0.229 | |
| Pre_VERSUS_Post | 7.93 | 12.24 | 0.65 | 0.524 | |
| Pre_VERSUS_New Words | 4.68 | 13.74 | 0.34 | 0.736 | |
| AXB oddity:Pre_VERSUS_Post | 0.02 | 0.17 | 0.14 | 0.889 | |
| AXB oddity:Pre_VERSUS_New Words | 0.08 | 0.19 | 0.44 | 0.665 | |

### *5.1.5.3.2. Everyday ID tests*

Table 43 summarizes the results of linear regression analysis for scores of everyday ID tests considering individual variability in the AXB oddity task. The model found a reliable main effect of AXB oddity, suggesting that overall scores of everyday ID tests were higher for participants who showed higher scores in the AXB oddity task. There also was a significant 2-way interaction between AXB oddity and Day1 VERSUS Day3, indicating that as AXB oddity scores became higher, an improvement from Day1 ID test to Day3 ID test became larger.

Table 43. Summary of fixed effects for the mixed-effects linear regression model of everyday ID test scores.

| Predictor | Estimate ($\beta$) | Std. Error | *t-value* | *p-value* | |
|---|---|---|---|---|---|
| (Intercept) | 73.31 | 1.70 | 43.01 | <0.001 | *** |
| AXB oddity | 8.97 | 3.41 | 2.63 | 0.015 | * |
| Contrast | -9.82 | 1.37 | -7.17 | <0.001 | *** |
| Day1_VERSUS_Day2 | 2.61 | 2.52 | 1.03 | 0.309 | |
| Day1_VERSUS_Day3 | 5.43 | 2.39 | 2.27 | 0.028 | * |
| Day1_VERSUS_Day4 | 7.44 | 2.76 | 2.69 | 0.013 | * |
| Day1_VERSUS_Day5 | 11.04 | 2.40 | 4.60 | <0.001 | *** |
| AXB oddity:Contrast | -0.58 | 2.74 | -0.21 | 0.831 | |
| AXB oddity:Day1_VERSUS_Day2 | 2.73 | 5.05 | 0.54 | 0.593 | |
| AXB oddity:Day1_VERSUS_Day3 | 9.20 | 4.79 | 1.92 | 0.061 | . |
| AXB oddity:Day1_VERSUS_Day4 | -1.07 | 5.54 | -0.19 | 0.848 | |
| AXB oddity:Day1_VERSUS_Day5 | 1.96 | 4.81 | 0.41 | 0.686 | |
| Contrast:Day1_VERSUS_Day2 | 6.39 | 4.33 | 1.48 | 0.141 | |
| Contrast:Day1_VERSUS_Day3 | 0.25 | 4.33 | 0.06 | 0.953 | |
| Contrast:Day1_VERSUS_Day4 | 1.66 | 4.33 | 0.38 | 0.702 | |
| Contrast:Day1_VERSUS_Day5 | 7.90 | 4.33 | 1.82 | 0.069 | . |
| AXB oddity:Contrast:Day1_VERSUS_Day2 | -3.82 | 8.67 | -0.44 | 0.660 | |
| AXB oddity:Contrast:Day1_VERSUS_Day3 | -2.82 | 8.67 | -0.33 | 0.745 | |
| AXB oddity:Contrast:Day1_VERSUS_Day4 | -10.18 | 8.67 | -1.17 | 0.241 | |
| AXB oddity:Contrast:Day1_VERSUS_Day5 | -3.58 | 8.67 | -0.41 | 0.680 | |

As for group analysis, separate linear regression models were built for each of vowel contrasts. Table 44 summarizes the results of linear regression model for the /i/-/ɪ/ contrast. There was a significant 2-way interaction between AXB oddity and Day1 VERSUS Day3, indicating that as the continuous variable, AXB oddity, got higher, the positive difference between Day1 and Day3 ID test scores got higher, as well. The results suggest that the perception of English /i/-/ɪ/ contrast became more accurate in Day3 ID test for participants who chose spectrally different stimuli as more odd ones in the AXB oddity task.

Table 45 summarizes the results of the linear regression model for the /ɛ/-/æ/ contrast. A trend effect of AXB oddity showed that overall scores of everyday ID tests were higher for those who had higher AXB oddity scores; however, the model did not find any reliable interaction between AXB oddity and Test. Thus, there was no evidence that the degree of improvement across everyday ID tests with Day 1 scores as the reference level was affected by individual variability in participants' AXB oddity task scores.

Finally, Table 46 summarizes the results of the linear regression model for the /ʊ/-/u/ contrast. There was a main effect of AXB oddity, reflecting that as one standardized unit increases in AXB oddity, the estimated increase in average score of everyday ID tests was 7.86. The model also found an interaction between AXB oddity and Day1 VERSUS Day3, indicating that improvement in the proportion of correct responses across Day1 and Day3 tests increases as AXB oddity scores increase.

Table 44. Summary of fixed effects for the mixed-effects linear regression model of everyday ID test scores for the /i/-/ɪ/ contrast.

| Predictor | Estimate ($\beta$) | Std. Error | *t-value* | *p-value* | |
|---|---|---|---|---|---|
| (Intercept) | 79.64 | 2.53 | 31.55 | <0.001 | *** |
| AXB oddity | 8.57 | 5.07 | 1.69 | 0.104 | |
| Day1_VERSUS_Day2 | -3.19 | 3.95 | -0.81 | 0.423 | |
| Day1_VERSUS_Day3 | 4.72 | 3.40 | 1.39 | 0.171 | |
| Day1_VERSUS_Day4 | 6.53 | 3.55 | 1.84 | 0.072 | . |
| Day1_VERSUS_Day5 | 6.11 | 3.40 | 1.80 | 0.078 | . |
| AXB oddity:Day1_VERSUS_Day2 | 6.83 | 7.93 | 0.86 | 0.394 | |
| AXB oddity:Day1_VERSUS_Day3 | 14.10 | 6.83 | 2.07 | 0.044 | * |
| AXB oddity:Day1_VERSUS_Day4 | 6.26 | 7.13 | 0.88 | 0.384 | |
| AXB oddity:Day1_VERSUS_Day5 | 6.35 | 6.83 | 0.93 | 0.357 | |

Table 45. Summary of fixed effects for the mixed-effects linear regression model of everyday ID test scores for the /ɛ/-/æ/ contrast.

| Predictor | Estimate ($\beta$) | Std. Error | *t-value* | *p-value* | |
|---|---|---|---|---|---|
| (Intercept) | 72.67 | 2.75 | 26.45 | <0.001 | *** |
| AXB oddity | 10.57 | 5.52 | 1.92 | 0.067 | . |
| Day1_VERSUS_Day2 | 6.39 | 2.97 | 2.15 | 0.037 | * |
| Day1_VERSUS_Day3 | 6.53 | 3.07 | 2.13 | 0.038 | * |
| Day1_VERSUS_Day4 | 7.22 | 3.53 | 2.04 | 0.047 | * |
| Day1_VERSUS_Day5 | 11.25 | 2.97 | 3.79 | <0.001 | *** |
| AXB oddity:Day1_VERSUS_Day2 | -0.78 | 5.96 | -0.13 | 0.897 | |
| AXB oddity:Day1_VERSUS_Day3 | 2.96 | 6.16 | 0.48 | 0.633 | |
| AXB oddity:Day1_VERSUS_Day4 | -3.26 | 7.10 | -0.46 | 0.648 | |
| AXB oddity:Day1_VERSUS_Day5 | -2.40 | 5.96 | -0.40 | 0.689 | |

Table 46. Summary of fixed effects for the mixed-effects linear regression model of everyday ID test scores for the /ʊ/-/u/ contrast.

| Predictor | Estimate ($\beta$) | Std. Error | *t-value* | *p-value* | |
|---|---|---|---|---|---|
| (Intercept) | 67.63 | 1.62 | 41.74 | <0.001 | *** |
| AXB oddity | 7.86 | 3.25 | 2.41 | 0.024 | * |
| Day1_VERSUS_Day2 | 4.62 | 2.57 | 1.80 | 0.079 | . |
| Day1_VERSUS_Day3 | 5.03 | 2.41 | 2.09 | 0.042 | * |
| Day1_VERSUS_Day4 | 8.56 | 2.91 | 2.94 | 0.005 | ** |
| Day1_VERSUS_Day5 | 15.77 | 2.41 | 6.54 | <0.001 | *** |
| AXB oddity:Day1_VERSUS_Day2 | 2.14 | 5.16 | 0.42 | 0.680 | |
| AXB oddity:Day1_VERSUS_Day3 | 10.64 | 4.84 | 2.20 | 0.033 | * |
| AXB oddity:Day1_VERSUS_Day4 | -6.23 | 5.84 | -1.07 | 0.292 | |
| AXB oddity:Day1_VERSUS_Day5 | 1.95 | 4.84 | 0.40 | 0.688 | |

#### *5.1.5.3.3. Perceptual cue weighting (Pre vs. Post-test with old talker)*

The estimated response curves in Figure 58, Figure 59, and Figure 60 were drawn from the logistic regression models investigating the effect of individual differences in scores of the AXB oddity task on the proportion of /i/, /æ/, and /u/ responses respectively for the pretest and the post-test with a set of stimuli from the "old" talker as a function of standardized spectral and duration steps. The visual comparisons between Figure 58, Figure 59, and Figure 60 suggest that the /i/-/ɪ/ contrast would be easier to learn than /ɛ/-/æ/ and /ʊ/-/u/ contrasts. As shown in Figure

58 for the /i/-/ɪ/ contrast, regardless of AXB oddity scores of individual participants, higher spectral steps prompted more /i/ responses as indicated by relatively gradient estimated curved (see the pink line or the post-test result in the figure). However, this pattern is not visually observed in Figure 59 and Figure 60 as steeper positive slopes were estimated for higher spectral steps (i.e., steps 4 and 5).

Visual inspection of Figure 58 for the /i/-/ɪ/ contrast shows that spectral dimension influenced participants' decisions in the pretest. However, some participants, especially those with higher AXB oddity scores, used spectral dimension oppositely, assigning more /i/ responses to stimuli with lower spectral steps. This trend indicates that overall, participants paid attention to spectral differences between stimuli in classifying the /i/-/ɪ/ contrast, although some had to learn how to use spectral cues appropriately. After training, all participants successfully classified stimuli by predominantly relying on spectral dimension. In the post-test, participants who properly used spectral dimension even in the pretest increased their use of that dimension by mainly assigning /i/ responses to stimuli with higher spectral steps. In contrast, participants who used spectral dimension in the pretest but in an opposite way changed their direction of using spectral dimension. Thus, all participants benefited from training in learning the /i/-/ɪ/ contrast.

Table 47 summarizes the logistic regression analysis results for the /i/-/ɪ/ contrast. There was an interaction between AXB oddity and Duration, indicating that the effect of duration dimension decreased as AXB oddity score decreased. The 3-way interaction between AXB oddity, Spectral, and Test shows that the increased use of spectral dimension in the post-test was larger as the AXB oddity scores increased. However, it should be noted that this result came from the fact that participants with higher AXB scores initially oppositely utilized spectral dimension; therefore, their use of spectral dimension in the post-test inevitably increased in a larger degree than the use of the same acoustic dimension by participants with relatively lower AXB oddity scores.

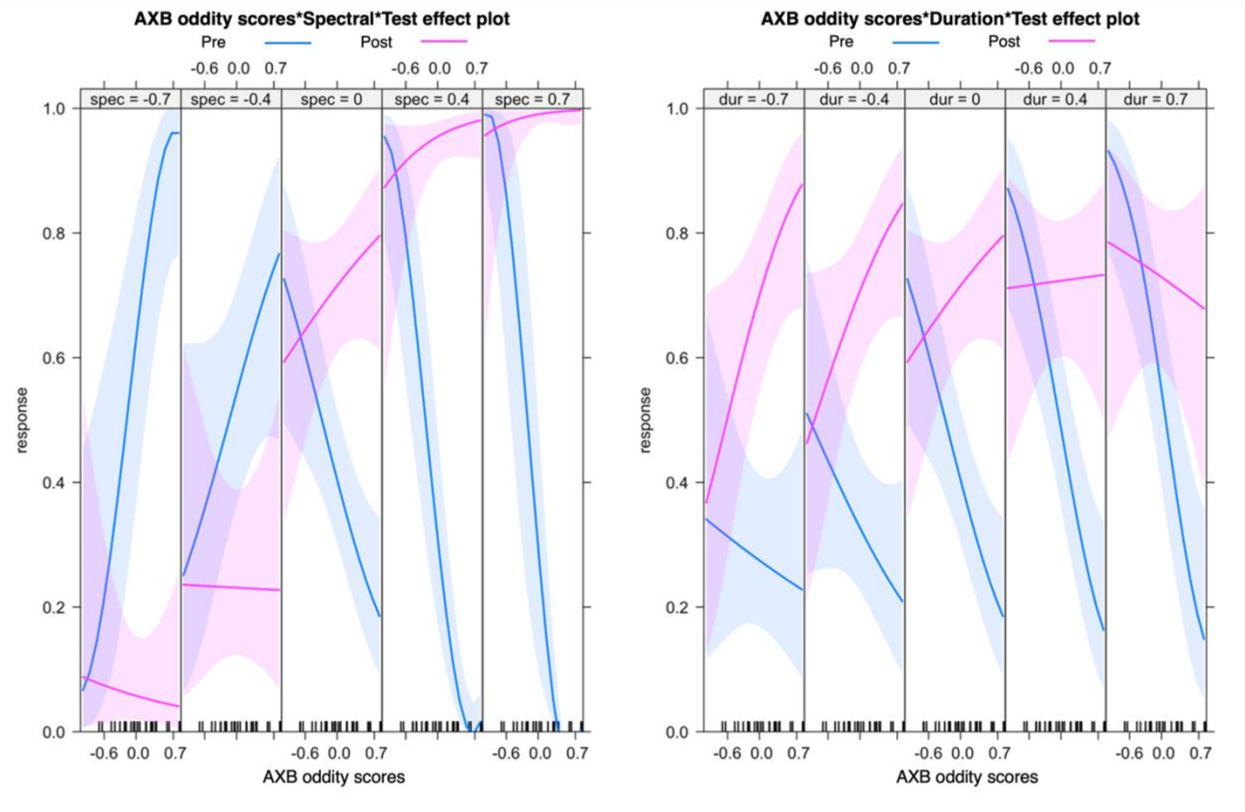


Figure 58. Predicted logit curves for perception of the /i/-/ɪ/ contrast by individual participants' AXB oddity scores (x-axis). Predictions are drawn from a binary logistic regression model with AXB oddity, Spectral, Duration, and Test as predictors. Each panel represents the results from the pretest (Blue: Test = -0.5) and the post-test (Pink: Test = 0.5) with the "old" talker. The y-axis of each panel represents the predicted probability of a "heed" response.

Table 47. Summary of fixed effects for the mixed-effects logistic regression model predicting changes in spectral and duration cue weights from the pretest to the post-test with the “old” talker for the /i/-/ɪ/ contrast.

| Predictor | Estimate ($\beta$) | Std. Error | *z*-value | *p*-value | |
|---|---|---|---|---|---|
| (Intercept) | 0.08 | 0.12 | 0.71 | 0.476 | |
| AXB oddity | -0.42 | 0.24 | -1.73 | 0.084 | . |
| Duration | 0.34 | 0.34 | 1.00 | 0.316 | |
| Spectral | 3.13 | 0.74 | 4.23 | <0.001 | *** |
| Test | 0.09 | 0.22 | 0.43 | 0.665 | |
| AXB oddity:Duration | -1.64 | 0.68 | -2.42 | 0.016 | * |
| AXB oddity:Spectral | -1.47 | 1.48 | -0.99 | 0.320 | |
| AXB oddity:Test | 1.00 | 0.45 | 2.24 | 0.025 | * |
| Duration:Test | -0.97 | 0.21 | -4.67 | <0.001 | *** |
| Spectral:Test | 6.92 | 0.30 | 22.74 | <0.001 | *** |
| AXB oddity:Spectral:Test | 6.35 | 0.67 | 9.41 | <0.001 | *** |
| AXB oddity:Duration:Test | 0.96 | 0.44 | 2.15 | 0.032 | * |

Figure 59, demonstrating the estimated proportion of /æ/ responses in the pretest and the post-test, shows that the effect of spectral dimension was more evident in the post-test than the pretest, and the predominant use of spectral dimension in the post-test was observed as AXB oddity scores increased. The results of the logistic regression analysis in Table 48 confirmed this observation. The 3-way interaction between AXB oddity, Spectral, and Test indicates that the effect of spectral dimension increased in the post-test, and the increase was larger as AXB oddity scores increased as well. The 2-way interaction between AXB oddity and Duration suggests that overall use of the duration dimension decreased as AXB oddity scores increased. The 3-way interaction between AXB oddity, Duration, and Test further shows that the effect of duration dimension decreased in the post-test, and the decrease was larger as AXB oddity scores increased.

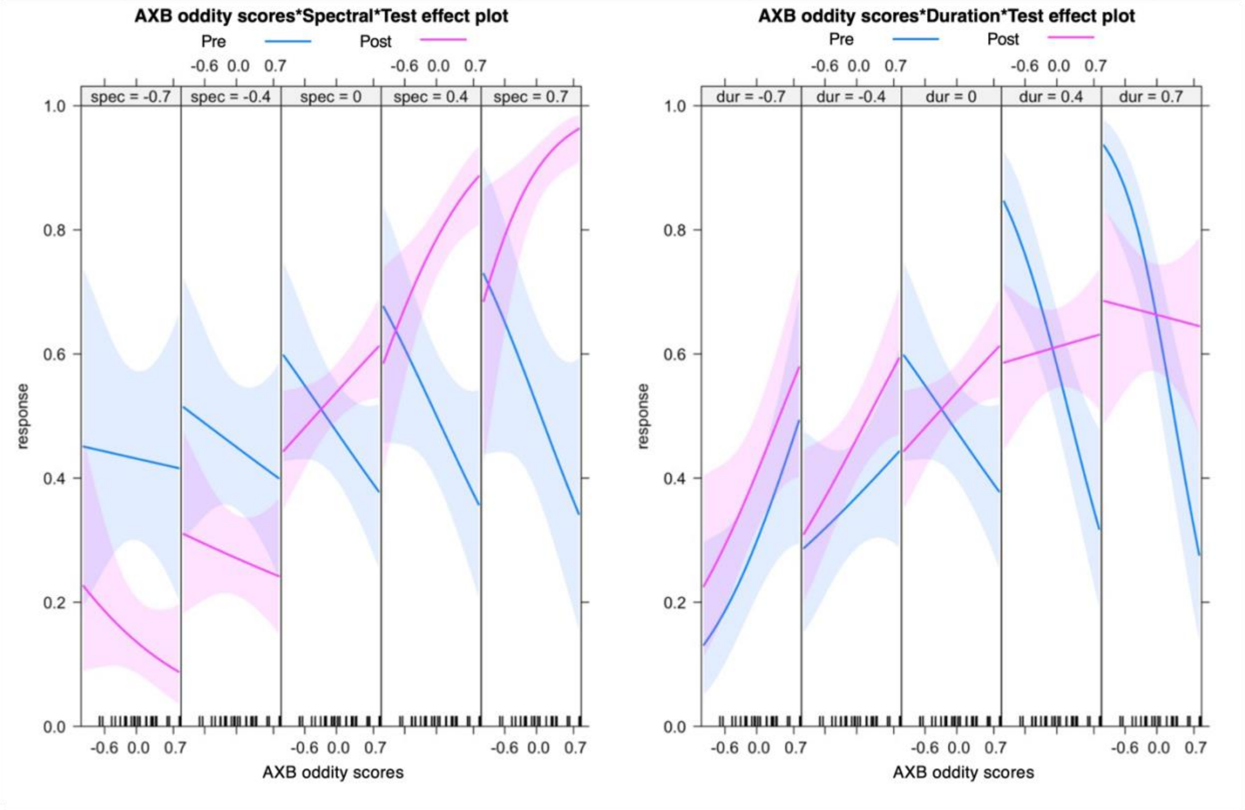


Figure 59. Predicted logit curves for perception of the /ɛ/-/æ/ contrast by individual participants' AXB oddity scores (x-axis). Predictions are drawn from a binary logistic regression model with AXB oddity, Spectral, Duration, and Test as predictors. Each panel represents the results from the pretest (Blue: Test = -0.5) and the post-test (Pink: Test = 0.5) with the "old" talker. The y-axis of each panel represents the predicted probability of a "had" response.

Table 48. Summary of fixed effects for the mixed-effects logistic regression model predicting changes in spectral and duration cue weights from the pretest to the post-test with the “old” talker for the /ɛ/-/æ/ contrast.

| Predictor | Estimate | Std. Error | *z*-value | *p*-value | |
|---|---|---|---|---|---|
| (Intercept) | 0.03 | 0.10 | 0.30 | 0.761 | |
| AXB oddity | -0.06 | 0.20 | -0.28 | 0.779 | |
| Duration | 0.91 | 0.23 | 4.04 | <0.001 | *** |
| Spectral | 1.57 | 0.32 | 4.90 | <0.001 | *** |
| Test | 0.26 | 0.16 | 1.63 | 0.102 | |
| AXB oddity:Duration | -1.43 | 0.46 | -3.13 | 0.002 | ** |
| AXB oddity:Spectral | 0.43 | 0.64 | 0.68 | 0.496 | |
| AXB oddity:Test | 0.86 | 0.32 | 2.66 | 0.008 | ** |
| Duration:Test | -0.31 | 0.17 | -1.84 | 0.066 | . |
| Spectral:Test | 2.63 | 0.18 | 14.59 | <0.001 | *** |
| AXB oddity:Spectral:Test | 2.04 | 0.36 | 5.67 | <0.001 | *** |
| AXB oddity:Duration:Test | 1.47 | 0.37 | 3.98 | <0.001 | *** |

Figure 60 for the /ʊ/-/u/ contrast demonstrates a very similar pattern as the /ɛ/-/æ/ contrast in that the effect of spectral dimension was larger in the post-test and as AXB oddity scores increased. The results of the regression analysis in Table 49 supported this observation. There was a nearly significant interaction between AXB oddity and Spectral, indicating that overall, the reliance on spectral dimension in the pretest and the post-test was larger as AXB scores became higher. In addition, the 3-way interaction between AXB oddity, Duration, and Test revealed that the increased use of spectral dimension in the post-test for the /ʊ/-/u/ contrast was larger as AXB oddity scores increased.

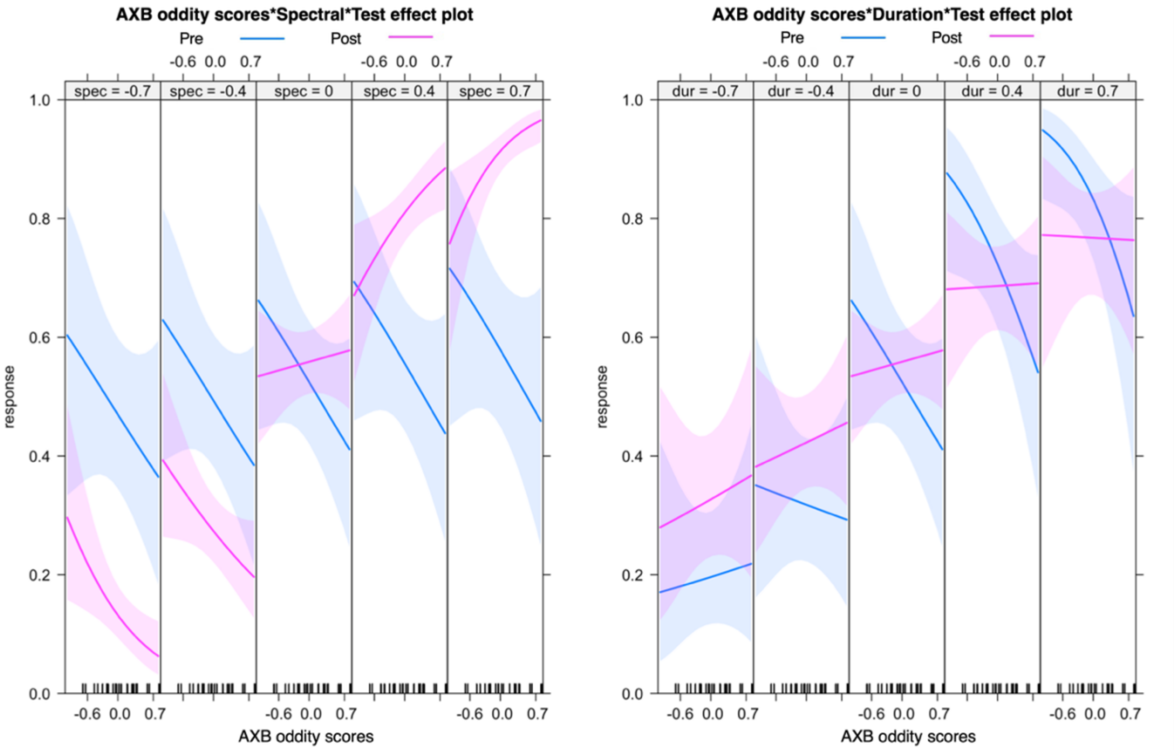


Figure 60. Predicted logit curves for perception of the /ʊ/-/u/ contrast by individual participants' AXB oddity scores (x-axis). Predictions are drawn from a binary logistic regression model with AXB oddity, Spectral, Duration, and Test as predictors. Each panel represents the results from the pretest (Blue: Test = -0.5) and the post-test (Pink: Test = 0.5) with the "old" talker. The y-axis of each panel represents the predicted probability of a "who'd" response.

Table 49. Summary of fixed effects for the mixed-effects logistic regression model predicting changes in spectral and duration cue weights from the pretest to the post-test with the "old" talker for the /ʊ/-/u/ contrast.

| Predictor | Estimate ($\beta$) | Std. Error | *z*-value | *p*-value | |
|---|---|---|---|---|---|
| (Intercept) | 0.18 | 0.11 | 1.65 | 0.098 | . |
| AXB oddity | -0.22 | 0.22 | -1.01 | 0.315 | |
| Duration | 1.75 | 0.28 | 6.15 | <0.001 | *** |
| Spectral | 1.69 | 0.20 | 8.24 | <0.001 | *** |
| Test | 0.12 | 0.21 | 0.58 | 0.565 | |
| AXB oddity:Duration | -0.64 | 0.57 | -1.12 | 0.261 | |
| AXB oddity:Spectral | 0.77 | 0.41 | 1.89 | 0.059 | . |
| AXB oddity:Test | 0.64 | 0.42 | 1.51 | 0.132 | |
| Duration:Test | -0.73 | 0.18 | -4.09 | <0.001 | *** |
| Spectral:Test | 2.75 | 0.19 | 14.76 | <0.001 | *** |
| AXB oddity:Spectral:Test | 1.64 | 0.38 | 4.38 | <0.001 | *** |
| AXB oddity:Duration:Test | 0.88 | 0.37 | 2.39 | 0.017 | * |

#### *5.1.5.3.4. Perceptual cue weighting (new talker generalization test with a novel talker)*

Figure 61 shows predicted values of /i/ responses as a function of standardized spectral and duration steps for the pretest and the post-test with a set of stimuli from a novel talker. The pattern of the opposite use of spectral dimension by participants with higher scores of AXB oddity was visually confirmed again via the positive slopes for stimuli with lower spectral steps (i.e., step 1 and 2) and the negative slopes for stimuli with higher spectral steps (i.e., step 4 and 5). However, after training, spectral dimension was used in a nativelike way by all participants regardless of participants' different AXB oddity scores. The result of the logistic regression in Table 50 found a reliable 3-way interaction between AXB oddity, Spectral, and Test, suggesting that the effect of spectral dimension increased in the post-test with a larger degree as AXB oddity scores became higher. However, as mentioned earlier, this result should be taken cautiously due to some participants' non-standard strategies in the use of spectral dimension. In terms of duration dimension, a 2-way interaction between AXB oddity and Duration suggests that higher AXB scores were related to lesser use of duration dimension in pretest and the post-test.

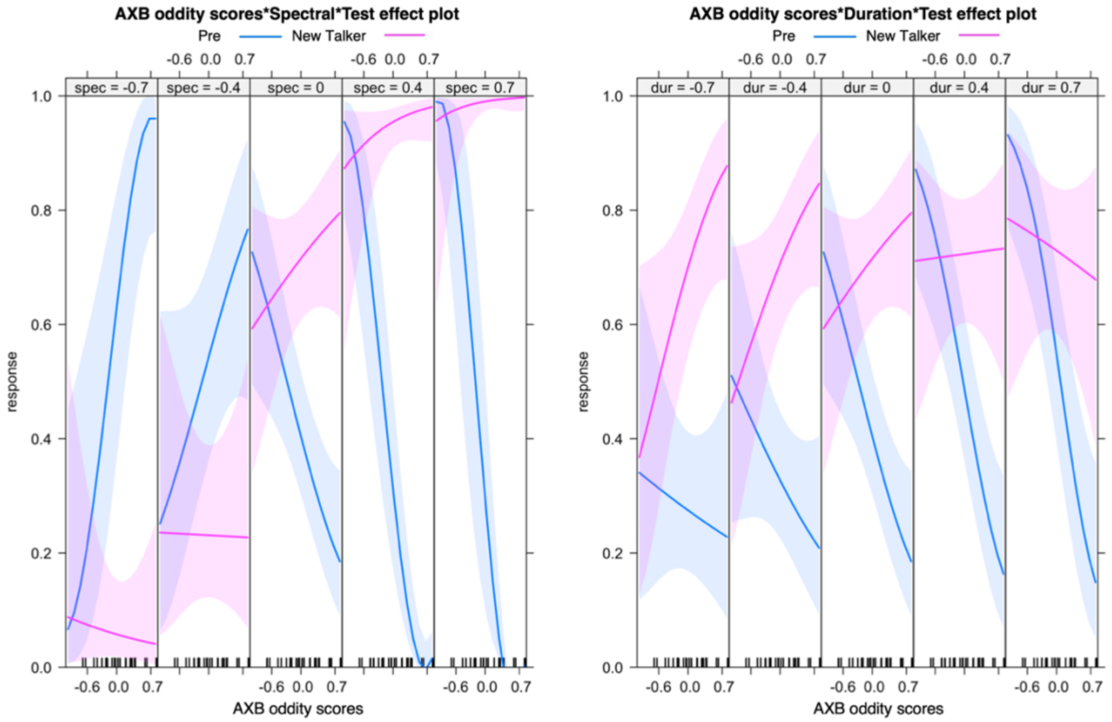


Figure 61. Predicted logit curves for perception of the /i/-/ɪ/ contrast by individual participants' AXB oddity scores (x-axis). Predictions are drawn from a binary logistic regression model with AXB oddity, Spectral, Duration, and Test as predictors. Each panel represents the results from the pretest (Blue: Test = -0.5) and the post-test (Pink: Test = 0.5) with a "new" talker. The y-axis of each panel represents the predicted probability of a "heed" response.

Table 50. Summary of fixed effects for the mixed-effects logistic regression model predicting changes in spectral and duration cue weights from the pretest to the post-test with a "new" talker for the /i/-/ɪ/ contrast.

| Predictor | Estimate ($\beta$) | Std. Error | *z-value* | *p-value* | |
|---|---|---|---|---|---|
| (Intercept) | 0.26 | 0.17 | 1.57 | 0.116 | |
| AXB oddity | -0.42 | 0.35 | -1.21 | 0.227 | |
| Duration | 0.47 | 0.28 | 1.66 | 0.098 | . |
| Spectral | 1.95 | 0.70 | 2.76 | 0.006 | ** |
| Test | 1.31 | 0.31 | 4.24 | 0.000 | *** |
| AXB oddity:Duration | -1.36 | 0.58 | -2.36 | 0.018 | * |
| AXB oddity:Spectral | -2.60 | 1.48 | -1.76 | 0.079 | . |
| AXB oddity:Test | 1.92 | 0.64 | 3.01 | 0.003 | ** |
| Duration:Test | -0.73 | 0.20 | -3.65 | 0.000 | *** |
| Spectral:Test | 6.72 | 0.30 | 22.67 | < 2e-16 | *** |
| AXB oddity:Spectral:Test | 8.01 | 0.68 | 11.75 | < 2e-16 | *** |
| AXB oddity:Duration:Test | 0.30 | 0.46 | 0.65 | 0.515 | |

As shown in Figure 62, stimuli with higher spectral steps prompt more /æ/ responses in the post-test. And, the effect of spectral dimension increases as AXB oddity scores increase, as evidenced by the larger difference between the estimated proportion of /æ/ responses for stimuli with the lowest spectral step and the estimated proportion for stimuli with the highest spectral step. The results of the logistic regression analysis agree with this pattern. The model in Table 51 found a significant 3-way interaction between `AXB oddity`, `Spectral`, and `Test`, suggesting that the effect of spectral dimension increased in the post-test with a larger degree for participants with higher AXB oddity scores.

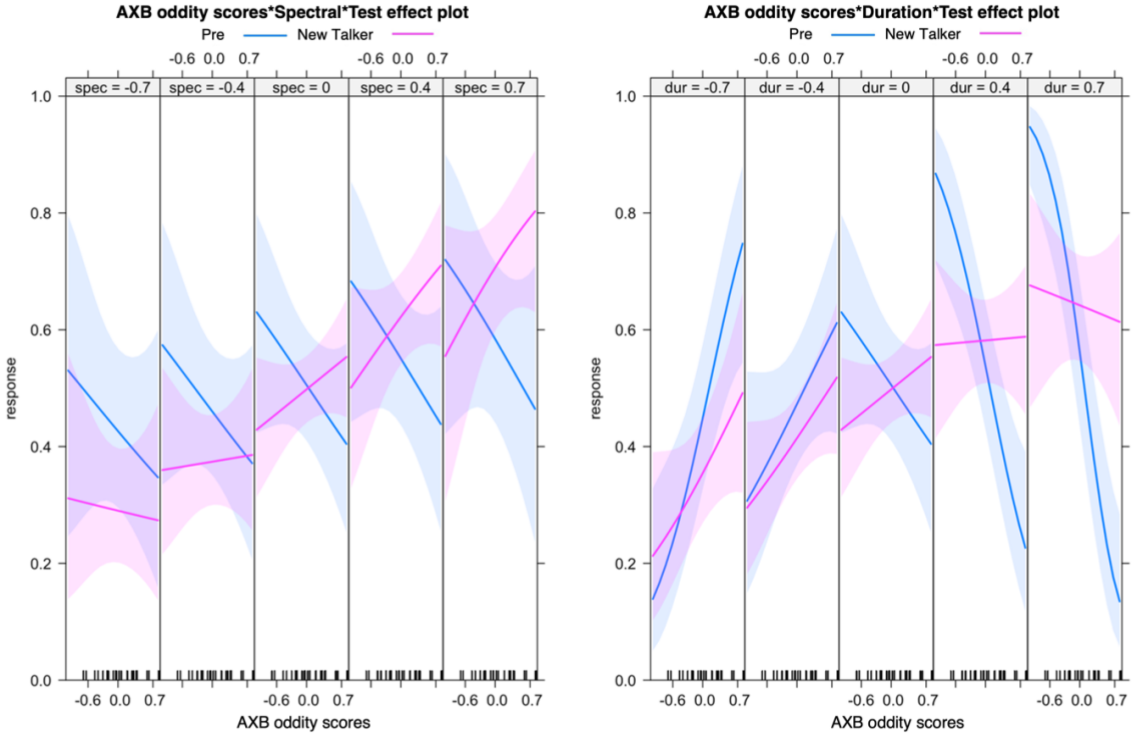


Figure 62. Predicted logit curves for perception of the /ɛ/-/æ/ contrast by individual participants' AXB oddity scores (x-axis). Predictions are drawn from a binary logistic regression model with `AXB oddity`, `Spectral`, `Duration`, and `Test` as predictors. Each panel represents the results from the pretest (Blue: `Test` = -0.5) and the post-test (Pink: `Test` = 0.5) with a "new" talker. The y-axis of each panel represents the predicted probability of a "had" response.

Table 51. Summary of fixed effects for the mixed-effects logistic regression model predicting changes in spectral and duration cue weights from the pretest to the post-test with a “new” talker for the /ɛ/-/æ/ contrast.

| Predictor | Estimate ($\beta$) | Std. Error | *z-value* | *p-value* | |
|---|---|---|---|---|---|
| (Intercept) | 0.01 | 0.12 | 0.05 | 0.960 | |
| AXB oddity | -0.11 | 0.24 | -0.44 | 0.662 | |
| Duration | 0.58 | 0.21 | 2.75 | 0.006 | ** |
| Spectral | 0.86 | 0.28 | 3.06 | 0.002 | ** |
| Test | -0.02 | 0.18 | -0.12 | 0.902 | |
| AXB oddity:Duration | -1.80 | 0.44 | -4.11 | 0.000 | *** |
| AXB oddity:Spectral | 0.22 | 0.56 | 0.40 | 0.690 | |
| AXB oddity:Test | 0.76 | 0.37 | 2.06 | 0.039 | * |
| Duration:Test | 0.54 | 0.16 | 3.39 | 0.001 | *** |
| Spectral:Test | 0.83 | 0.16 | 5.25 | 0.000 | *** |
| AXB oddity:Spectral:Test | 0.67 | 0.33 | 1.99 | 0.046 | * |
| AXB oddity:Duration:Test | 2.36 | 0.37 | 6.45 | 0.000 | *** |

Visual inspection of Figure 63 for the /ʊ/-/u/ contrast shows that even in the pretest, the effect of spectral dimension increased as AXB oddity scores increased. In Table 52, the nearly significant 2-way interaction between `AXB oddity` and `Spectral` confirmed this visual observation. This result suggests that the proportion of participants’ /u/ responses was predicted by the `Spectral` variable and the effect of spectral dimension was larger for participants with higher AXB oddity scores.

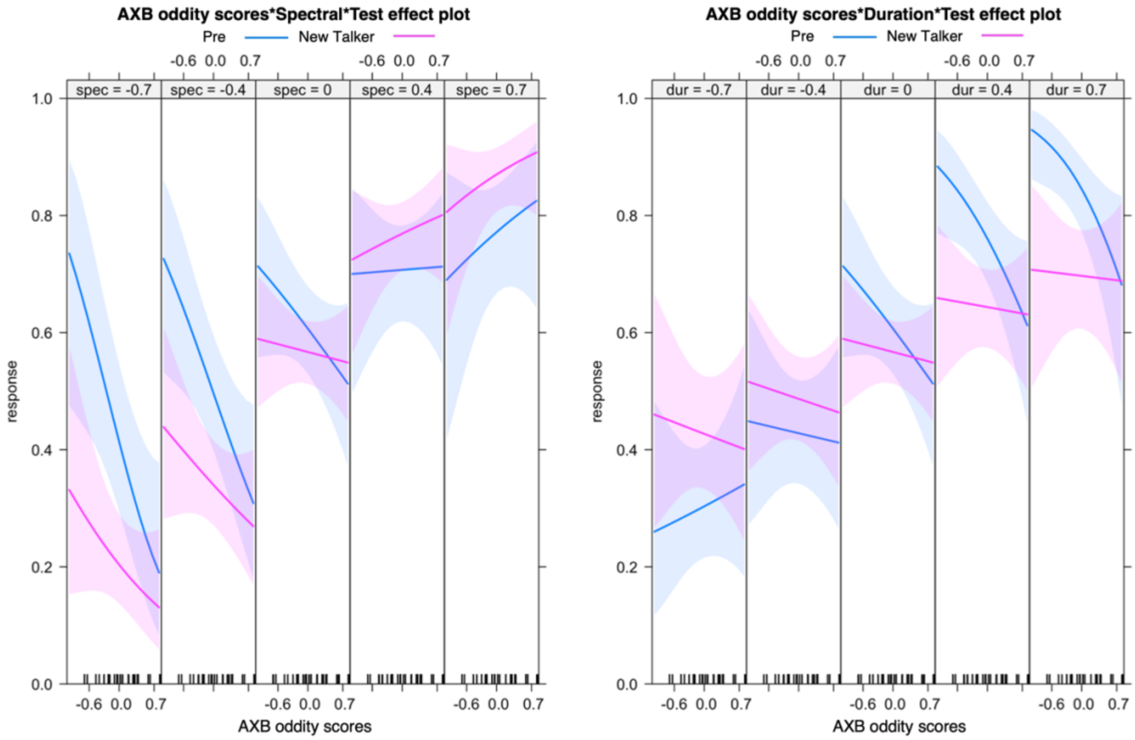


Figure 63. Predicted logit curves for perception of the /ʊ/-/u/ contrast by individual participants' AXB oddity scores (x-axis). Predictions are drawn from a binary logistic regression model with AXB oddity, Spectral, Duration, and Test as predictors. Each panel represents the results from the pretest (Blue: Test = -0.5) and the post-test (Pink: Test = 0.5) with a "new" talker. The y-axis of each panel represents the predicted probability of a "who'd" response.

Table 52. Summary of fixed effects for the mixed-effects logistic regression model predicting changes in spectral and duration cue weights from the pretest to the post-test with a "new" talker for the /ʊ/-/u/ contrast.

| Predictor | Estimate ($\beta$) | Std. Error | *z*-value | *p*-value | |
|---|---|---|---|---|---|
| (Intercept) | 0.35 | 0.10 | 3.60 | 0.000 | *** |
| AXB oddity | -0.28 | 0.20 | -1.44 | 0.150 | |
| Duration | 1.30 | 0.22 | 5.95 | 0.000 | *** |
| Spectral | 1.73 | 0.28 | 6.12 | 0.000 | *** |
| Test | -0.16 | 0.18 | -0.90 | 0.367 | |
| AXB oddity:Duration | -0.46 | 0.44 | -1.05 | 0.292 | |
| AXB oddity:Spectral | 1.04 | 0.57 | 1.84 | 0.066 | . |
| AXB oddity:Test | 0.38 | 0.35 | 1.08 | 0.279 | |
| Duration:Test | -0.98 | 0.17 | -5.77 | 0.000 | *** |
| Spectral:Test | 1.23 | 0.17 | 7.18 | 0.000 | *** |
| AXB oddity:Spectral:Test | -0.46 | 0.36 | -1.29 | 0.198 | |
| AXB oddity:Duration:Test | 1.04 | 0.35 | 2.98 | 0.003 | ** |

#### *5.1.5.3.5. Individual trajectories in spectral and duration cue weights*

Individual participants' cue weights were computed via separate logistic regression analysis for each participant with Spectral and Duration as predictors of perceived either *heed, had,* or *who'd* responses for the pretest and the post-test with a set of stimuli from either old talker or novel talker. Each model generated two beta-coefficients, $\beta_{\text{spec}}$ and $\beta_{\text{dur}}$, for one of the vowel contrasts. Thus, each participant had six beta-coefficients for three vowel contrasts for the old talker and also for the new talker. Recall that negative coefficients indicate that an acoustic dimension influenced participants in the opposite direction from native listeners.

Figure 64 and Figure 65 illustrate group trends of participants' changes in spectral and duration cue weights in the pretest and the post-test with old and novel talkers. The plots on the left side present the differences between the pretest $\beta_{\text{spec}}$ and the post-test $\beta_{\text{spec}}$ values (i.e., $\beta_{\text{pretest spec}} - \beta_{\text{post-test spec}}$); and the plots on the right side present the difference between the pretest $\beta_{\text{dur}}$ and the post-test $\beta_{\text{dur}}$ values (i.e., $\beta_{\text{pretest dur}} - \beta_{\text{post-test dur}}$). Note that values over zero indicate more use of the corresponding acoustic dimension in the post-test than the pretest. The higher the values were, the more the acoustic dimension affected in the post-test. As we can see in the figures, participants in the HS group increased their reliance on spectral dimension in the post-tests than participants in the LS group, and this trend is commonly observed in both post-tests with old and novel talkers. The $\beta_{\text{spec}}$ difference between groups was more evident in the test results with the old talker, especially for the /i/-/ɪ/ contrast than /ɛ/-/æ/ and /ʊ/-/u/ contrasts. These results suggest that participants in the HS group could enhance the importance of spectral dimension in classifying the /i/-/ɪ/ contrast since this contrast was relatively easier to learn than the other two contrasts.

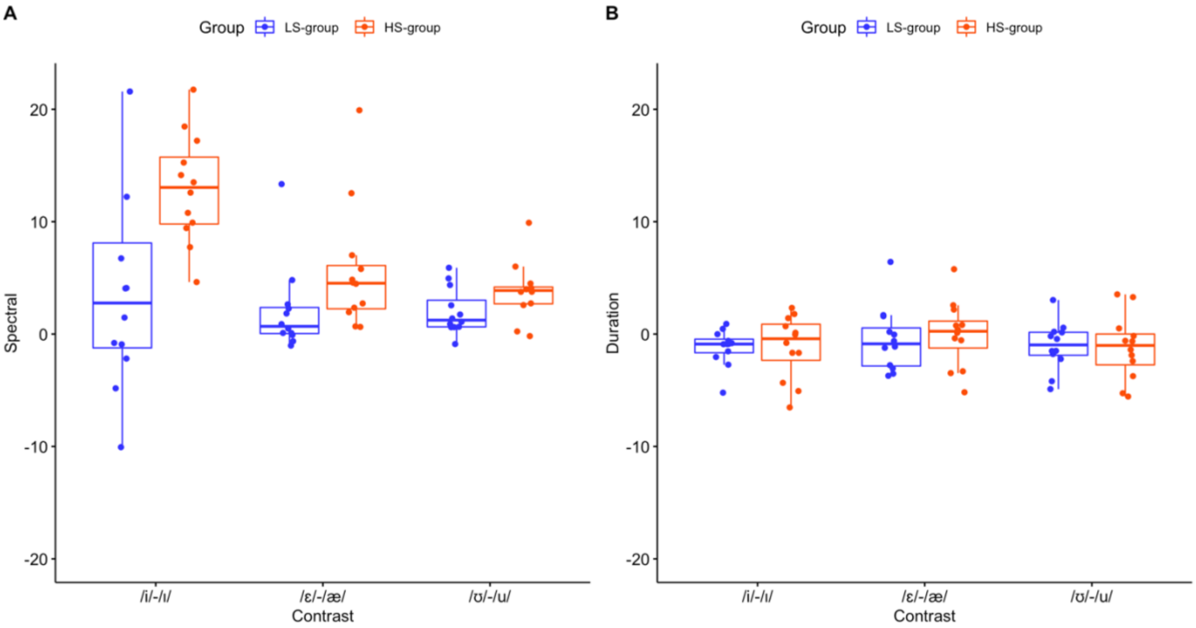


Figure 64. Group trends of changes in spectral and duration cue weights in the pretest and the post-test with the "old" talker. The x-axis represents each vowel contrast. The y-axis represents the differences between each group participants' pre-test and post-test beta-coefficients of spectral dimension (left) and duration dimension (right). Higher values indicate more use of the corresponding acoustic dimension in the post-test.

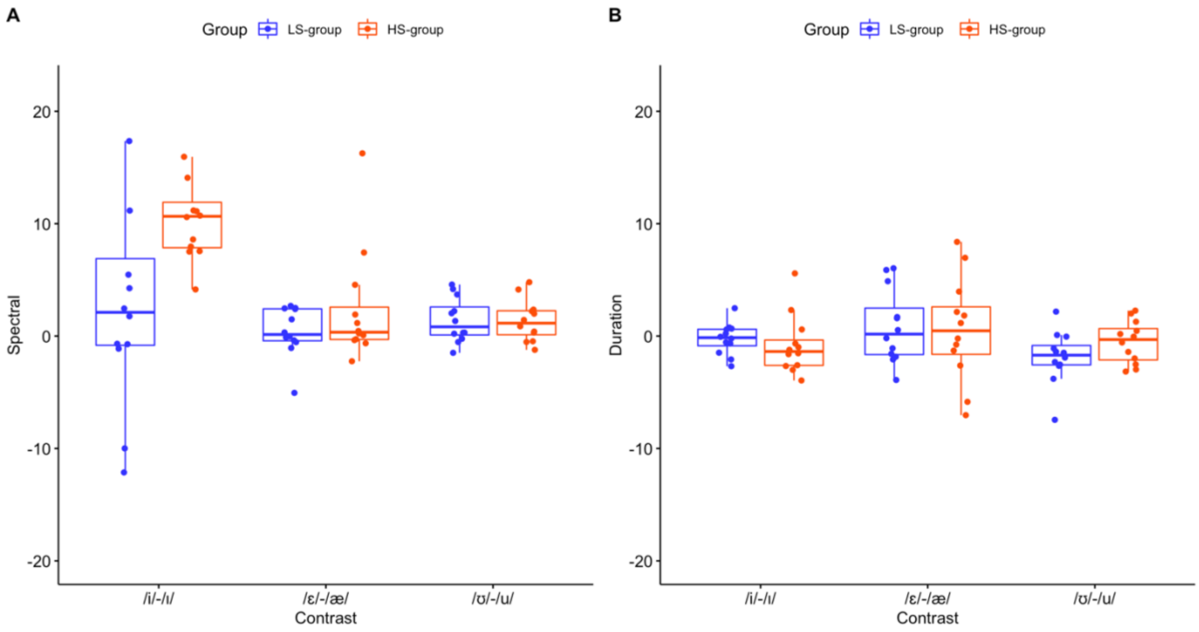


Figure 65. Group trends of changes in spectral and duration cue weights in the pretest and the post-test with a “new” talker. The x-axis represents each vowel contrast. The y-axis represents the differences between each group participants’ pre-test and post-test beta-coefficients of spectral dimension (left) and duration dimension (right). Higher values indicate more use of the corresponding acoustic dimension in the post-test.

Figure 66 displays individual participants’ developmental trajectories in spectral and duration cue weighting for each target vowel contrasts over time (i.e., before and after training). The x-axis shows trajectories of duration cue weights from the pretest to the post-test, and the y-axis show trajectories of spectral cue weights from the pretest to the post-test. The panels in Figure 66 represent each participant with their subject number; the letters before the subject number indicate their group, either the HS or the LS group. This analysis was inspired by Kim et al. (2018) for investigating English vowel acquisition developmental trajectories across time points for Korean adult and child learners of English.

Individual participants’ direction and length of arrows show how much they changed their cue-weighting strategies in the perception of English vowel contrasts. Participants could be categorized into several groups where each group shows a similar pattern in changes in spectral and duration cue weights based on visual inspection. Roughly, six different groups were generated based on the directions of arrows, which show changes in the logistic regression coefficients for each dimension in the classification of the /i/-/ɪ/ contrast. As mentioned earlier, participants showed more evident changes in their cue-weighting strategies in the identification of words for the /i/-/ɪ/ contrast. For the clearer examination of individual trajectories from the pretest to the post-test, participants' cue-weighting change patterns were initially determined based on the /i/-/ɪ/ contrast. Following Kim et al. (2018), the characteristics of six groups for the /i/-/ɪ/ contrast are described as follows: the negative sign (−) indicates that the corresponding acoustic dimension was used to distinguish the vowel contrast in the opposite direction from native listeners; the positive sign (+) indicates that the acoustic dimension was used in a nativelike way. For example, "–Spec, +Spec" group indicates that spectral dimension was used to distinguish the vowel contrast by with negative $\beta_{\text{spec}}$ values in the pretest (i.e., fewer /i/ responses were made as spectral steps of stimuli increased), but the same dimension was used with positive $\beta_{\text{spec}}$ values in the post-test (i.e., more /i/ responses were made as spectral steps of stimuli increased). The six groups were "–Spec, +Spec", +Spec, +Spec", "+Dur, +Spec", "+Dur, –Dur", "–Dur, +Spec", and "+Spec, –Spec".

The "–Spec, +Spec" group initially interpreted spectral cues in the opposite direction from native listeners with negative values of $\beta_{\text{spec}}$; however, in the post-test, they altered their way of using spectral cues and showed the primary reliance on spectral cues like native listeners (HS_1, HS_5, HS_6,HS_7, HS_8, HS_9, HS_12, LS_4, LS_10). This result suggests that some participants might be initially sensitive to spectral cues, but they might have mistakenly thought that the concept *hid* sounded like /hid/, and the concept *heed* sounded like /hɪd/. Due to their higher sensitivity to spectral cues even before training, participants in this group could shift to

"+Spec" at the later stage of training relatively easier than other participants who did have less sensitivity to spectral dimension. The fact that this pattern was found more often by participants in the HS group leads us to speculate that the HS group participants had higher sensitivity to spectral cues than the LS group even before training, although they did not know how to utilize spectral cues appropriately. The "+Spec, +Spec" group (LS_9, LS_11) showed spectral reliance from earlier time points and increased the reliance on spectral cues after training.

The "–Dur, +Spec" (HS_3, SHS_4, HS_11) and "+Dur, +Spec" (HS10, LS_2, LS_3, LS_5) groups initially distinguished the /i/-/ɪ/ contrast by duration information exclusively, but in the post-test, these groups showed a native-like use of spectral cues with down-weighting duration cues. The overall direction of change in primary weighting from duration to spectral cues is consistent with the developmental stage hypothesis (Escudero, 2000) that in the initial stage, learners are not able to identify tokens of /i/ versus /ɪ/. However, at the final stage of learning, learners show English-like use of both spectral and duration cues, with spectral cues receiving primary weighting. The "±Dur, +Spec" groups show a benefit of training and suggest that even participants who gave priority to duration cues can change their cue-weighting strategies to classify challenging vowel contrasts after a short laboratory training. Together with the "±Spec, +Spec" pattern, "±Dur, +Spec" patterns starting with the dominant use of duration cues were the most common patterns of cue-weighting change in the current study.

Some participants relied dominantly on duration dimension in the pretest but reduced their use of duration cues at the later stage of learning (i.e., the post-test). One participant (LS_7) belongs to the "+Dur, –Dur" group. This pattern suggests that some participants used their sensitivity to duration cues initially. However, after the training, they might realize the less informativeness of the duration dimension, but their learning still did not reach the level enough for them to prioritize spectral cues. This pattern seems to be the middle stage of the developmental stage hypothesis, which suggests that the participants in this group might change their patterns of cue-weighting to "–Dur, +Spec" if they received further training.

Lastly, some participants showed the "+Spec, –Spec" pattern in their cue-weighting strategies. For example, LS_6 and LS_8 first showed higher reliance on spectral dimension for the /i/-/ɪ/ contrast; however, they reduced their reliance on spectral dimension in the post-test. It is also noticeable that these participants somewhat decrease the effect of duration dimension in the post-test for /ɛ/-/æ/ and /ʊ/-/u/ contrasts. One possible reason for this pattern is that some participants might have tried different cue-weighting strategies and fluctuated between them in the process of learning to apply a unified strategy to classify all three vowel contrasts. However, they might have failed to systematically adapt one relevant strategy, thus resulting in reducing the reliance on both duration and spectral dimensions in the post-test. In other words, this pattern may show a much more considerable confusion in learning and, consequently, a possible disadvantage of highly variable training with spectrally and durationally manipulated stimuli.

Figure 66 shows things to note regarding participants' trajectories in perceptual cue weighting for /ɛ/-/æ/ and /ʊ/-/u/ contrasts. A large number of participants in the HS group (9 participants for the /ɛ/-/æ/contrast and 8 participants for the /ʊ/-/u/ contrast) moved toward more nativelike perception with positive $\beta_{\text{spec}}$ values. Some of participants in the LS group (4 participants for the /ɛ/-/æ/contrast and 5 participants for the /ʊ/-/u/ contrast) also showed their change in $\beta_{\text{spec}}$ values in the post-test toward more nativelike perception. These patterns suggest that as a whole, majority of participants could perceive /ɛ/-/æ/ and /ʊ/-/u/ contrasts by increasing their use of spectral cues at the later time of the experiment compared to the earlier time of the experiment. However, there are two noticeable differences between the three vowel contrasts' patterns of trajectories. The first difference is that the degrees of increased spectral cue use in the post-test were lower for /ɛ/-/æ/ and /ʊ/-/u/ contrasts than the /i/-/ɪ/ contrast except for the three participants (HS_6, HS_7, and LS_1). In conjunction with the results of test scores presented earlier, this pattern suggests that the acquisition of /ɛ/-/æ/ and /ʊ/-/u/ contrasts might be harder than the /i/-/ɪ/ contrast. The second difference is that with the exception of one participant (LS_1), no participants showed an increase of spectral dimension in the post-test for either the /ɛ/-/æ/ or the /ʊ/-/u/ contrast without increasing spectral dimension for the /i/-/ɪ/ contrast. This suggests that learners may start to receive benefit from HVPT in the acquisition of spectral cue weighting for the /i/-/ɪ/ contrast, and then the acquisition of /ɛ/-/æ/ and /ʊ/-/u/ contrasts may or may not follows. Although individual learners may take different rates to show their learning in English vowel contrasts, once they start to learn how to successfully utilize spectral cues, their learning outcome may be shown in the acquisition of the /i/-/ɪ/ contrast first.

Figure 67 shows participants' changes in cue-weightings of duration and spectral dimensions shown in the results of the new talker generalization test. Figure 67 demonstrated whether individual trajectories in the results of the pretest and post-test with the "old" talker are consistent with their trajectories in the generalization test. Overall, it is shown that the lengths of arrows are shorter than the ones in Figure 66, indicating that participants did not much increase or decrease the effects of acoustic dimensions in the generalization test. However, it seems that participants' trajectories in the generalization test followed a similar pattern shown in Figure 66. For instance, participants who were categorized as the "+Spec, −Spec" group, still belong to the same group based on the results of the generalization test. These results suggest that individuals' developmental patterns in the use of duration and spectral cues transfer to a novel set of stimuli produced by a "new" talker.

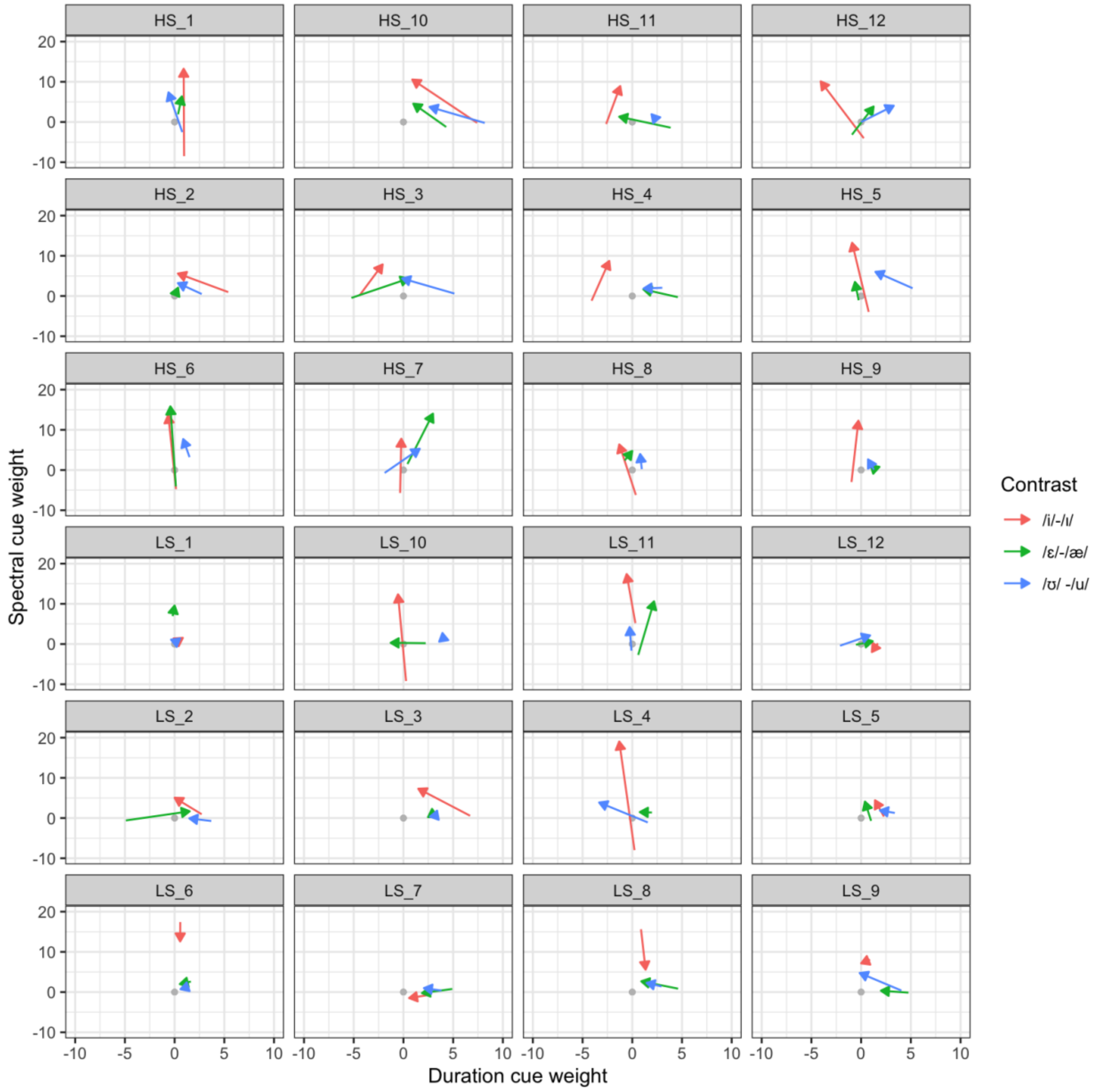


Figure 66. Developmental trajectories from the pretest to the post-test for individual participants. Each arrow indicates the changes in beta-coefficients of spectral and duration dimensions drawn by the logistic regression analysis fitted to each participant’s response data from the pretest to the post-test with the “old” talker. The vowel contrasts are displayed in different colors.

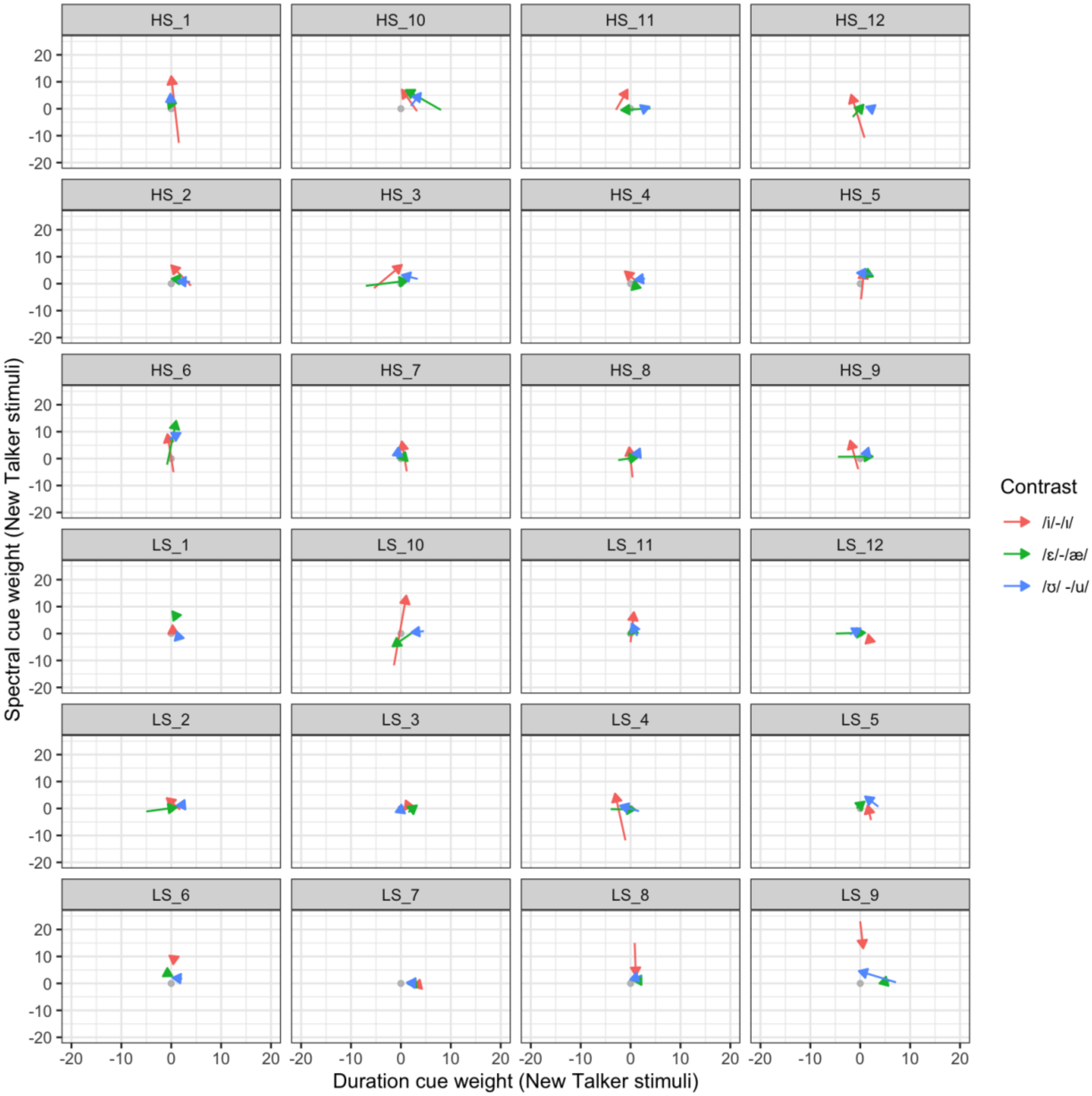


Figure 67. Developmental trajectories from the pretest to the post-test for individual participants. Each arrow indicates the changes in beta-coefficients of spectral and duration dimensions drawn by the logistic regression analysis fitted to each participant's response data from the pretest to the post-test with a "new" talker. The vowel contrasts are displayed in different colors.

#### *5.1.5.3.6. Correlation between AXB oddity task scores and spectral cue uses*

Three partial correlation tests were conducted to examine whether individual participants' AXB oddity task scores can explain substantial individual differences in duration and spectral weighting after training in classifying each of three English vowel contrasts. Since the primary interest here is to investigate the relationship between AXB oddity task scores and participants' successful use of spectral cues as a primary dimension after training, the correlation tests were run between AXB task scores and participants' $\beta_{\text{spec}}$ values obtained from the post-test with the "old" talker and the new talker generalization test with a novel talker. As shown in Figure 68 and Figure 69 with the results of the post-test with the "old" talker, the analysis revealed that there were significant positive correlations between individual participants' AXB oddity task scores and spectral cue weights for /i/-/ɪ/ and /ʊ/-/u/ contrasts ($r$ = .43, $p$ = .04; $r$ = .46, $p$ = .02), indicating that the higher AXB oddity task scores are associated with higher use of spectral dimension after training. The correlation between AXB oddity task scores and spectral cue weights for /ɛ/-/æ/ did not reach its significance ($r$ = .35, $p$ = .09). There was no significant correlation between AXB oddity task scores and their $\beta_{\text{spec}}$ values from the new talker generalization test. In summary, to some extent, AXB oddity task scores could account for variability in Korean learners' use of spectral cues in the perception of three challenging English vowel contrasts.

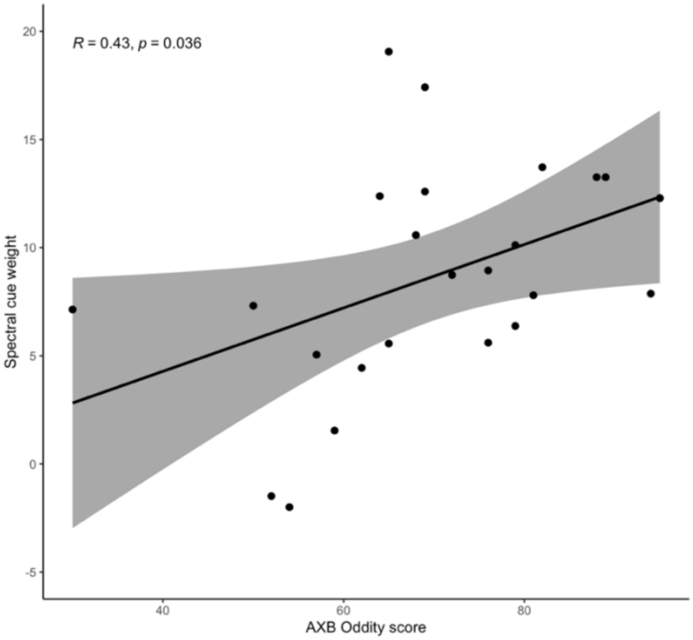


Figure 68. Correlation between individual participants' AXB oddity task scores and their beta-coefficients of spectral dimension for the /i/-/ɪ/ contrast from the logistic regression analysis fitted to each participants' response data of the post-test with the "old" talker.

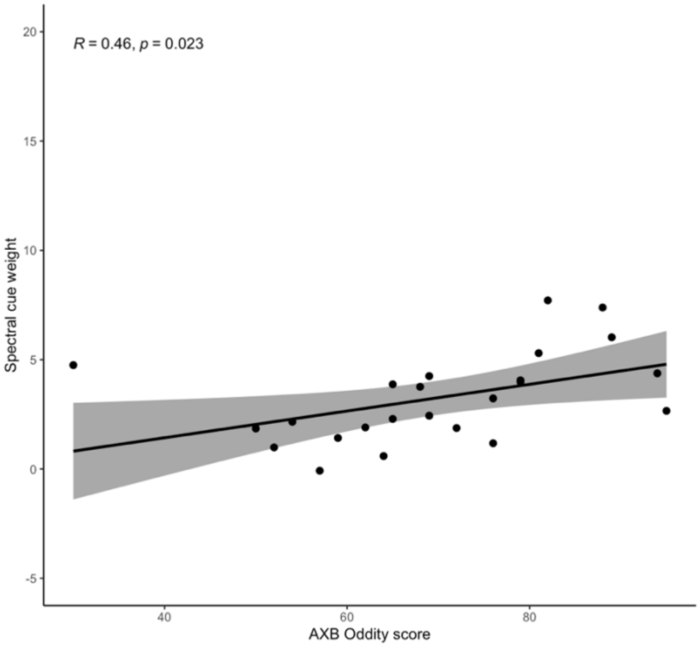


Figure 69. Correlation between individual participants' AXB oddity task scores and their beta-coefficients of spectral dimension for the /ʊ/-/u/ contrast from the logistic regression analysis fitted to each participants' response data of the post-test with the "old" talker.

## 5.2. Summary of experiment 3

In experiment 3, participants were assigned to two groups depending on their performances in the AXB oddity task: the high sensitivity (HS) and the low sensitivity (LS) groups. The AXB oddity task was intended to measure how much participants were sensitive to spectral differences between stimuli falling into a Korean /i/ category. The learning outcomes of the HS group were better than the outcomes of the LS group under the HVPT environment. The HS group demonstrated a larger benefit of HVPT than the LS group as evidenced by the HS group's better performances in all types of tests, including the pretest and the post-test, everyday ID tests, and new talker/words generalization tests. To compare the acquisition of the three vowel contrasts, the statistical analysis was separately conducted for each vowel contrast. The results showed that both groups' improvement was especially noticeable in learning the /i/-/ɪ/ contrast than the other two contrasts, the /ɛ/-/æ/ and the /ʊ/-/u/ contrasts. In terms of group trends in use of the cues changed from the pretest to the post-test, the effect of spectral dimension increased more for the HS group than the LS group. This result demonstrates that the HS group more actively engaged the spectral dimension than the LS group in classifying each vowel contrast stimuli into two vowel categories.

The analysis examining the relationships between individual variability in the AXB oddity task and their achievements in HVPT suggests that participants with higher AXB oddity task scores improved more in HVPT. Moreover, the significant positive correlation between AXB oddity task scores and participants' spectral cue weights in the post test suggest that the systematical and successful use of primary acoustic dimension (i.e., spectral cues) in the perception of English vowel contrasts can be explained by the degree of variation in performance in the AXB oddity task. However, the AXB oddity task scores failed to show their relationship with participants' ability to generalize their learning in how to weigh duration and spectral cues to novel talkers' stimuli.

# CHAPTER 6. EXPERIMENT 4

Experiment 3 showed that learners with a relatively less sensitivity to spectral differences in the perception of Korean /i/ vowel category (i.e., LS group) could not successfully learn the target English vowel contrasts. More importantly, they could not primarily rely on the spectral dimension to distinguish two English vowels for each target vowel contrasts. The LS group did not perform in the generalization tests enough to show their ability to generalize their learning in perceiving sets of stimuli either produced by a novel talker or consisted of new words which were not trained during HVPT.

The primary purpose of experiment 4 is to test the effectiveness of the cue-attention switching training with a different population of L2 learners, English learners of Korean. In experiment 1, learners with less sensitivity to the f0 dimension in the perception of English stop voicing contrast did not result in successful learning compared to learners with higher sensitivity to f0. Experiment 2 adapted the cue-attention switching training, which was designed to reallocate learners' attention to the primary acoustic dimension of the target language (i.e., f0 in experiment 2) in the perception of native language. The results of experiment 2 showed the effectiveness of the HVPT paradigm with the cue-attention switching training. Learners with less sensitivity to the f0 dimension resulted in a similar level of achievements as learners with higher sensitivity to the f0 dimension and presented nativelike cue-weighting strategy in the perception of Korean three-way stop contrast.

Focusing on the second research question, experiment 4 tests whether the cue-attention switching training acts positively or negatively in learning English vowel contrasts by Korean learners of English. The cue-attention switching training in experiment 4 exposes learners with stimuli which highly vary in the less informative acoustic dimension in L2 (i.e., duration) but with limited variability in the informative acoustic dimension in L2 (i.e., spectral). Same as experiment 2, the set of stimuli for this additional training is in learners' L1 (i.e., Korean in experiment 4). I expected that the modified HVPT parading with the cue-attention switching training would decrease the possible difficulty in placing the primary attention to the spectral dimension during training. I also expected the cue-attention switching training to specifically aid learners with a relatively less sensitivity to the spectral dimension to realize the informativeness of that dimension in perceiving the target language contrasts.

## 6.1. Methods

### *6.1.1. Participants*

A new group of 25 Korean learners of English (fourteen female, eleven male: mean age of 23.6 years, range of 19 – 33 years) participated in experiment 4. All participants were undergraduate or recently graduate students from Kangwon National University in South Korea. As in experiment 3, participants completed a language background questionnaire asking their demographic and native language backgrounds as well as their English language learning experience and self-reported proficiency. Table 53 summarizes the relevant demographic information for the Korean L2 learners of English. As determined by a language background questionnaire, participants had a similar English language learning background. Briefly, they graduated elementary, middle, and high schools in South Korea, learned English as their second language at schools, and did not have experience living in an English-speaking country at the time of participation. Participants self-reported that they use English in everyday life, about 20% on average. Participants' average self-rated scores of their English proficiency were 3.3 (speaking), 4.3 (listening), 3.2 (writing), and 4.5 (reading) on a 7-point Likert scale, and their self-reported English accentedness score was 5.2 on a 10-point Likert scale. None of the participants had a history of speech and hearing problems and got monetary compensation after each experimental session.

Table 53. Demographic information of participants in experiment 4.

| Variable | Mean | *SD* |
|---|---|---|
| Sex | 14 F; 11 M | |
| Age (yrs. old) | 23.6 | 3.4 |
| Birthplace (province, #) | Gyeonggi 11<br>Gangwon 13<br>Chungcheong 1 | |
| Lived out of Korea (yrs.) | 0.0 | 0.0 |
| Father L1 (language, #) | Korean 25 | |
| Mother L1 (language, #) | Korean 25 | |
| L2 (language, #) | English 25 | |
| (1 very poor to 7 very good) | | |
| English Speaking | 3.3 | 1.1 |
| English Listening | 4.3 | 1.4 |
| English Writing | 3.2 | 1.1 |
| English Reading | 4.5 | 1.5 |
| (1 Very heavy foreign accent to 10 No foreign accent) | | |
| Accentedness | 5.2 | 1.6 |
| (1 Not prefer to learn English to 10 Prefer to learn English) | | |
| English preference | 5.6 | 1.7 |

### *6.1.2. Stimuli*

Stimuli for all the tests and HVTP training were identical to the stimuli used in experiment 3.

#### 6.1.2.1. Stimuli for the cue-attention switching training

In experiment 4, the cue-attention switching training was added to the original HVPT implemented in experiment 3. The stimuli for this training were a subset of stimuli for the AXB oddity task. Recall that stimuli for the AXB oddity task consist of spectrally and durationally manipulated Korean /i/ vowel stimuli with five spectral steps and five duration steps (see 5.1.2.1). Among 25 manipulated stimuli, only ten stimuli with the highest and lowest steps of spectral dimension were used in the cue-attention switching training (i.e., stimuli marked as blue in Figure 70). The detailed procedure of the cue-attention switching training is provided in section 6.1.3.2.

### *6.1.3. Procedure*

The apparatus and procedures for the AXB oddity task and the HVPT were identical to those used in experiment 3.

#### 6.1.3.1. HVPT

The overall procedure of HVPT for learning English vowel contrasts was identical to experiment 3, except that participants received the cue-attention switching training before the initiation of each training session. As in experiment 3, experiment 4 consisted of five days, as shown in Table 54. On the first day of the experiment, participants were asked to come to the lab and complete the tasks. Participants were given an option to complete the last day of the experiment online through a live Zoom meeting (Zoom Video Communications Inc., 2020), considering the COVID-19 pandemic. If they could not come back to the lab, the last day experiment was administered through a live one-to-one Zoom meeting between a participant and the experimenter. Only one participant was allowed in each Zoom meeting to enable the experimenter to guide and monitor the participant’s performance. During the meeting, participants were guided verbally by the experimenter, and the experimenter monitored their performances through the video chat function on Zoom in real-time. All participants in experiment 4 completed their Day 5 session at home through Zoom.

Table 54. Timeline of the five-day experiment 4 with five-day HVPT.

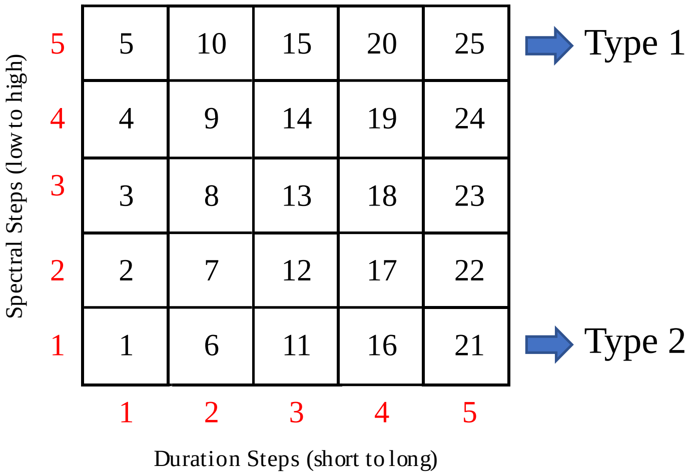
| | Participation Day | | | | |
|---|---|---|---|---|---|
| **Day** | Day 1 | Day 2 | Day 3 | Day 4 | Day 5 |
| **Tasks** | • Korean AXB Task (n = 128)<br>• Daily Familiarization Phase for HVPT<br>• Pre ID Test (n = 450)<br>• Cue-Attention Switching Training<br>• HVPT Session 1 (n = 225)<br>• Everyday ID Test (n = 75)<br>• Language Questionnaire | • Daily Familiarization Phase for HVPT<br>• Cue-Attention Switching Training<br>• HVPT Session 2 - 4 (n= 225)<br>• Everyday ID Test (n = 75) | | | • Daily Familiarization Phase for HVPT<br>• HVPT Session 5 (n= 225)<br>• Everyday ID Test (n= 75)<br>• Post ID Test (n = 450)<br>• New Talker Generalization Test (n = 225)<br>• New Word Generalization Test (n = 225)<br>• LexTale |
| **Time** | About 1 hour | About 30 minutes | | | About 1 hour |
| **Location** | Laboratory | Online | | | Laboratory or Online |

### 6.1.3.2. The cue-attention switching training

Before initiating each session of HPVT, participants in experiment 4 received the additional cue-attention switching training. This training took the 2FAC paradigm, giving participants two response choices, either ‘Type 1’ or ‘Type 2’. Participants were instructed that they will hear sound files from a native speaker of Korean who may speak a different dialect of Korean (e.g., unfamiliar regional dialects or dialects spoken by immigrants or second-generation immigrants), and there are two types of Korean /i/ pronunciation in that dialect, which are Type 1 and Type 2. Participants were told that they have to decide whether the sound they hear is either ‘Type 1’ or ‘Type 2’ /i/ sound. After each stimulus was played, a screen with two choices, ‘Type

1' or 'Type 2', appeared, and participants had to click one of the options based on their judgments. Before the next stimulus, a burst of white noise (1,250ms) was presented to mask the echoic memory of the previously presented token.

As the cue-attention switching training adopted in experiment 2, the primary purpose of this training is to switch participants' attention to spectral cues in the perception of the Korean vowel. To accomplish this, stimuli varying along with five duration steps but only with the minimum and maximum spectral steps (1 and 5) were presented. Therefore, only ten stimuli, five with the minimum spectral step and five with the maximum spectral step, were used. The feedback on participants' performances was always 'Type 1' when the spectral step was minimum and 'Type 2' when the spectral step was maximum. In other words, for a stimulus, the 'Type 1' response was correct when the spectral step of that stimulus was minimum, regardless of the duration step. This feedback design was to make spectral dimension as informative and predictive cues for dividing the single Korean /i/ vowel into two categories while informing participants that they do not need to pay attention to duration cues to be successful in the cue-attention switching training (Figure 70). The session took approximately 5 minutes. The ten stimuli were presented with eight times repetition (i.e., total of 80 trials).



| Spectral Steps (low to high) | | | | | | |
|---|---|---|---|---|---|---|
| 5 | 5 | 10 | 15 | 20 | 25 | Type 1 |
| 4 | 4 | 9 | 14 | 19 | 24 | |
| 3 | 3 | 8 | 13 | 18 | 23 | |
| 2 | 2 | 7 | 12 | 17 | 22 | |
| 1 | 1 | 6 | 11 | 16 | 21 | Type 2 |
| | 1 | 2 | 3 | 4 | 5 | |

Figure 70. Stimuli for the cue-attention switching training and their answers.

### 6.1.4. *Analysis*

The analytical methods for the AXB oddity task and the HVPT were identical to those used in experiment 3. The full results of participants' coefficients ($\beta_{spec}$ or $\beta_{dur}$) from individual logistic regression analyses can be found in Appendix I.

#### 6.1.4.1. The cue-attention switching training

The percentage of correctness was calculated for each participant. Figure 71 shows the results of the cue-attention switching training in the percentage of correct. Participants' accuracy was relatively high in the first session of the cue-attention switching training (*M* =77.3); however, there was a wide range of individual differences in the first session (*SD* =20.6), indicating that some participants had hard time classifying stimuli into two Korean /i/ categories by focusing only on spectral differences between stimuli. However, participants' average score increased to 92.6 with decreased standard deviation of 9.7. These results indicate that participants could learn how to divide a single Korean /i/ category into two categories with increased use of spectral dimension, which was what the cue-attention switching training intended to train.

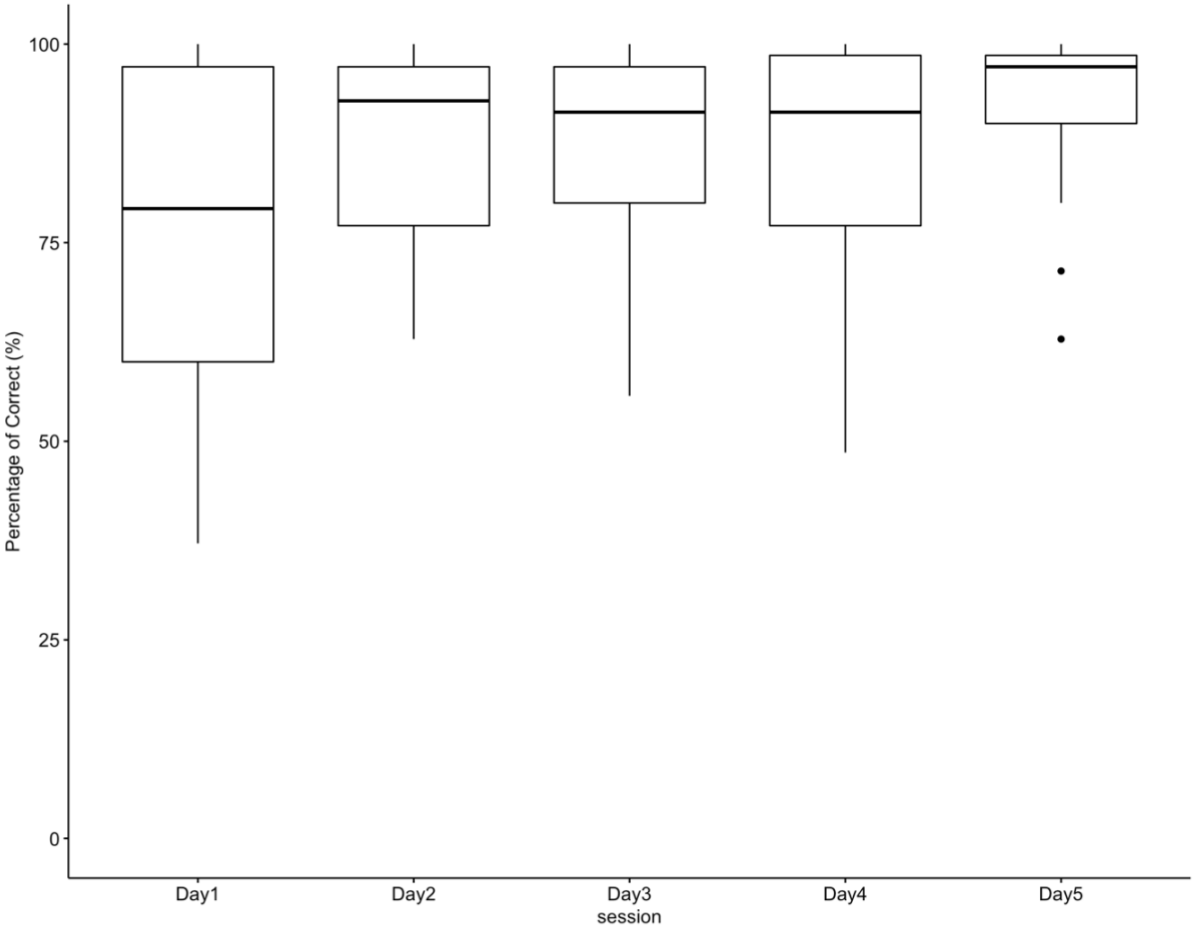


Figure 71. Results of the cue-attention switching training in the percentage of correct (%).

### *6.1.5. Results*

#### 6.1.5.1. The AXB oddity task

Table 55 shows the results of the AXB oddity task. To examine whether the cue-attention switching training effectively aids participants in the LS group to learn how to distinguish English vowel contrasts, only participants who recorded similar AXB oddity task scores as the LS group in experiment 3 were included. The criteria to divide participants into the HS and LS groups in experiment 3 was 70%. Thus, all participants in experiment 4 did not exceed 5% above from the mean score of 70% in their scores of the AXB oddity task. Out of 32 participants who showed their initial interests to participate in experiment 4, 25 participants were chosen for further training based on their AXB oddity task results (i.e., those who did not exceed 75% in thier AXB oddity task scores). The average AXB task score in experiment 4 was 63%. The group of participants in experiment 4 was named the LS 2 group, and their performances in HVPT were compared to those of the LS group in experiment 1. The independent *t*-test showed that the AXB task scores between the LS and LS 2 groups were not significantly different ($p$ = .50).

Table 55. Results of the Korean /i/ vowel AXB oddity task by participants in experiment 4.

| **Participants** | **Proportion of selecting a spectrally different stimulus in percentage (%)** |
|---|---|
| 1 | 56 |
| 2 | 61 |
| 3 | 62 |
| 4 | 52 |
| 5 | 60 |
| 6 | 62 |
| 7 | 57 |
| 8 | 66 |
| 9 | 63 |
| 10 | 66 |
| 11 | 70 |
| 12 | 67 |
| 13 | 66 |
| 14 | 68 |
| 15 | 66 |
| 16 | 58 |
| 17 | 66 |
| 18 | 71 |
| 19 | 73 |
| 20 | 52 |
| 21 | 62 |
| 22 | 70 |
| 23 | 67 |
| 24 | 59 |
| 25 | 54 |
| **Average (*SD*)** | **63 (*5.57*)** |

#### 6.1.5.2. Results of HVPT by group (LS group vs. LS 2 group)

For the group-level analysis, the LS and the LS 2 groups were compared on all types of tests in HVPT. This group comparison was a major part of the analysis in experiment 4 since it was designed to test the effectiveness of the cue-attention switching training and examine whether the LS 2 group resulted in better performances in HVPT than the LS group. Two separate linear regression mixed-effects models were built with participants' proportion of correct responses to the pretest, post-test, new word generalization test, and five everyday ID tests.

The English proficiency of participants in the LS 2 and the LS groups was controlled based on their LexTale scores. Statistical analysis showed that the LS 2 group and LS group participants did not differ in their LexTale scores ($t = 1.5$, $p = .14$).

#### *6.1.5.2.1. Pre, post, and new word generalization tests*

Table 56 summarizes the results of the linear regression analysis with participants' percentage of correct answers for the pretest, post-test, and new talker generalization tests across three vowel contrasts. The model found significant main effects of Pre VERSUS Post and Pre VERSUS New Words, indicating that overall participants improved in the post-test and the new word generalization test. The main effect of Contrast suggests that participants' overall performance in tests were modulated by different vowel contrasts with higher test scores on the /i/-/ɪ/ contrast followed by /ɛ/-/æ/ and /ʊ/-/u/ contrasts. These results are consistent with the results in experiment 3, which showed that the participants had a relatively harder time classifying the target words in /ɛ/-/æ/ and /ʊ/-/u/ contrasts than the words in the /i/-/ɪ/ contrast. The model also found a significant 2-way interaction between Group and Pre VERSUS Post, indicating that the LS 2 group's improvement from the pretest to the post-test was larger than the LS group's improvement. Since there was a significant main effect of Contrast, three separate linear regression models were built for each vowel contrast.

Table 56. Summary of fixed effects for the mixed-effects linear regression model for the pretest, post-test, and new word generalization tests scores.

| Predictor | Estimate ($\beta$) | Std. Error | *t*-value | *p*-value | |
|---|---|---|---|---|---|
| (Intercept) | 61.87 | 1.33 | 46.58 | <0.001 | *** |
| Group | 0.28 | 2.84 | 0.10 | 0.923 | |
| Contrast | -3.53 | 1.68 | -2.10 | 0.037 | * |
| Pre_VERSUS_Post | 5.83 | 2.05 | 2.84 | 0.005 | ** |
| Pre_VERSUS_New Words | 6.04 | 2.05 | 2.94 | 0.004 | ** |
| Group:Contrast | -6.38 | 3.59 | -1.78 | 0.076 | . |
| Group:Pre_VERSUS_Post | -9.52 | 4.38 | -2.17 | 0.031 | * |
| Group:Pre_VERSUS_New Words | -5.04 | 4.38 | -1.15 | 0.252 | |
| Contrast:Pre_VERSUS_Post | 0.06 | 4.11 | 0.01 | 0.989 | |
| Contrast:Pre_VERSUS_New Words | -8.07 | 4.11 | -1.96 | 0.050 | . |
| Group:Contrast:Pre_VERSUS_Post | 11.25 | 8.78 | 1.28 | 0.201 | |
| Group:Contrast:Pre_VERSUS_New Words | -1.19 | 8.78 | -0.14 | 0.892 | |

Table 57, Table 58, and Table 59 show the results of three linear regression mixed-effects models built based on participants' proportions of correct responses in the pretest, post-test, and new word generalization tests for each target vowel contrast. For the /i/-/ɪ/ contrast, the model found a significant main effect of Pre VERSUS New Words, indicating that participants could identity *hid/heed* and *bid/bead* word pairs better after receiving training than before training. There was a significant 2-way interaction between Group and of Pre VERSUS Post, suggesting that the LS 2 group showed greater improvement in the classification of /i/-/ɪ/ contrast in the post-test than the LS group. For /ɛ/-/æ/, only Pre VERSUS Post was significant, suggesting that participants showed overall better performance in the post-test than the pretest in identifying *head* and *had* words. There was no reliable interaction. For /ʊ/-/u/ contrast, there were reliable main effects of Pre VERSUS Post and Pre VERSUS New Words, reflecting that participants identified *hood/who'd* and *nook/nuke* word pairs better in the post-test and the new word generalization test, respectively, than they identified *hood/who'd* in the pre-test.

Table 57. Summary of fixed effects for the mixed-effects linear regression model for the pretest, post-test, and new word generalization tests scores for the /i/-/ɪ/ contrast.

| Predictor | Estimate ($\beta$) | Std. Error | *t-value* | *p-value* | |
|---|---|---|---|---|---|
| (Intercept) | 68.17 | 2.20 | 30.94 | <0.001 | *** |
| Group | 5.84 | 4.71 | 1.24 | 0.222 | |
| Pre_VERSUS_Post | 6.02 | 3.74 | 1.61 | 0.116 | |
| Pre_VERSUS_New Words | 15.68 | 3.81 | 4.11 | <0.001 | *** |
| Group:Pre_VERSUS_Post | -20.36 | 7.99 | -2.55 | 0.015 | * |
| Group:Pre_VERSUS_New Words | -8.21 | 8.15 | -1.01 | 0.318 | |

Table 58. Summary of fixed effects for the mixed-effects linear regression model for the pretest, post-test, and new word generalization tests scores for the /ɛ/-/æ/ contrast.

| Predictor | Estimate ($\beta$) | Std. Error | *t-value* | *p-value* | |
|---|---|---|---|---|---|
| (Intercept) | 53.57 | 1.34 | 40.05 | <0.001 | *** |
| Group | -3.04 | 2.86 | -1.07 | 0.294 | |
| Pre_VERSUS_Post | 5.37 | 2.15 | 2.50 | 0.017 | * |
| Pre_VERSUS_New Words | -3.35 | 2.44 | -1.38 | 0.177 | |
| Group:Pre_VERSUS_Post | -1.59 | 4.59 | -0.35 | 0.731 | |
| Group:Pre_VERSUS_New Words | 2.77 | 5.20 | 0.53 | 0.598 | |

Table 59. Summary of fixed effects for the mixed-effects linear regression model for the pretest, post-test, and new word generalization tests scores for the /ʊ/-/u/ contrast.

| Predictor | Estimate ($\beta$) | Std. Error | *t-value* | *p-value* | |
|---|---|---|---|---|---|
| (Intercept) | 63.86 | 1.59 | 40.19 | <0.001 | *** |
| Group | -1.97 | 3.39 | -0.58 | 0.566 | |
| Pre_VERSUS_Post | 6.09 | 2.53 | 2.41 | 0.021 | * |
| Pre_VERSUS_New Words | 5.80 | 2.68 | 2.17 | 0.037 | * |
| Group:Pre_VERSUS_Post | -6.61 | 5.41 | -1.22 | 0.229 | |
| Group:Pre_VERSUS_New Words | -9.67 | 5.72 | -1.69 | 0.100 | . |

### *6.1.5.2.2. Everyday ID tests*

Figure 72 plots the group results of the percentage of correct responses for everyday ID tests from Day 1 to Day 5. The results from the different groups are represented in different colors. Table 60 summarizes the results of the linear regression mixed-effects analysis with

Group, Test (Day1 VERSUS Day2, Day1 VERSUS Day3, Day1 VERSUS Day4, and Day1 VERSUS Day5) variables, and their interactions.

The visual inspection of Figure 72 demonstrates that the LS and the LS 2 groups did not perform much differently in the earlier stage of training (i.e., Day 1) while the LS 2 group outperformed the LS group at the later stage of learning (i.e., Day 2, Day 3 & Day5). However, the group differences were not evident compared to the group differences between the HS and LS groups in experiment 3.

The closer observation for each of the vowel contrasts in Figure 73 suggests that the LS 2 group showed higher scores in the classification of the /i/-/ɪ/ contrast than /ɛ/-/æ/ and /ʊ/-/u/ contrasts, which indicates that the LS 2 group achieved a higher level of learning for a certain contrast than others as shown in experiment 3. In terms of group differences, the LS 2 group outperformed the LS group in identifying *hid/heed* words of the /i/-/ɪ/ contrast. For the other two contrasts, both groups performed similarly.

The results of the regression analysis in Table 60 confirmed the visual inspection of Figure 72. The model found reliable main effects of Day1 VERSUS Day3, Day1 VERSUS Day4, and Day1 VERSUS Day5, suggesting that participants in both groups improved across training sessions. More importantly, there were significant 2-way interactions between Group and levels of Test variables. The interactions between Group, Day1 VERSUS Day2, Day1 VERSUS Day3, and Day1 VERSUS Day5 show that the LS 2 group improved more across everyday ID tests with Day 1 test scores as a reference than the LS group. In sum, the results suggest that the LS 2 group performed better than the LS group in training, indicating a benefit of the cue-attention switching training.

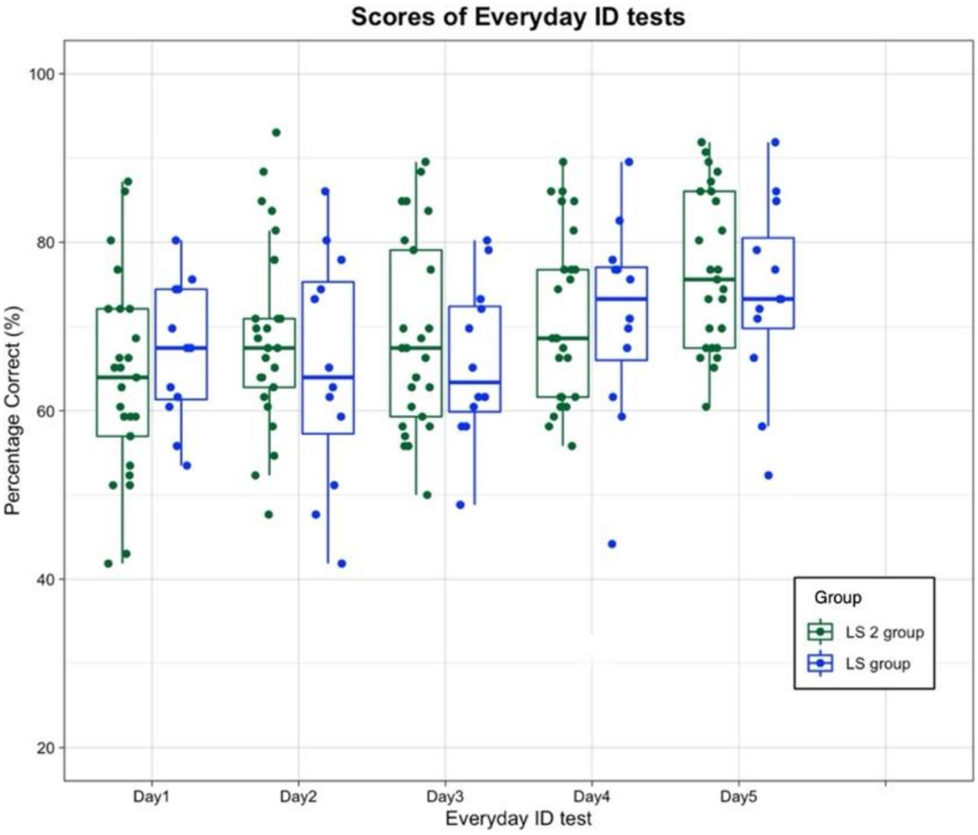


Figure 72. Results of everyday ID tests by group (LS 2 group vs. LS group) in percentage of correct (%).

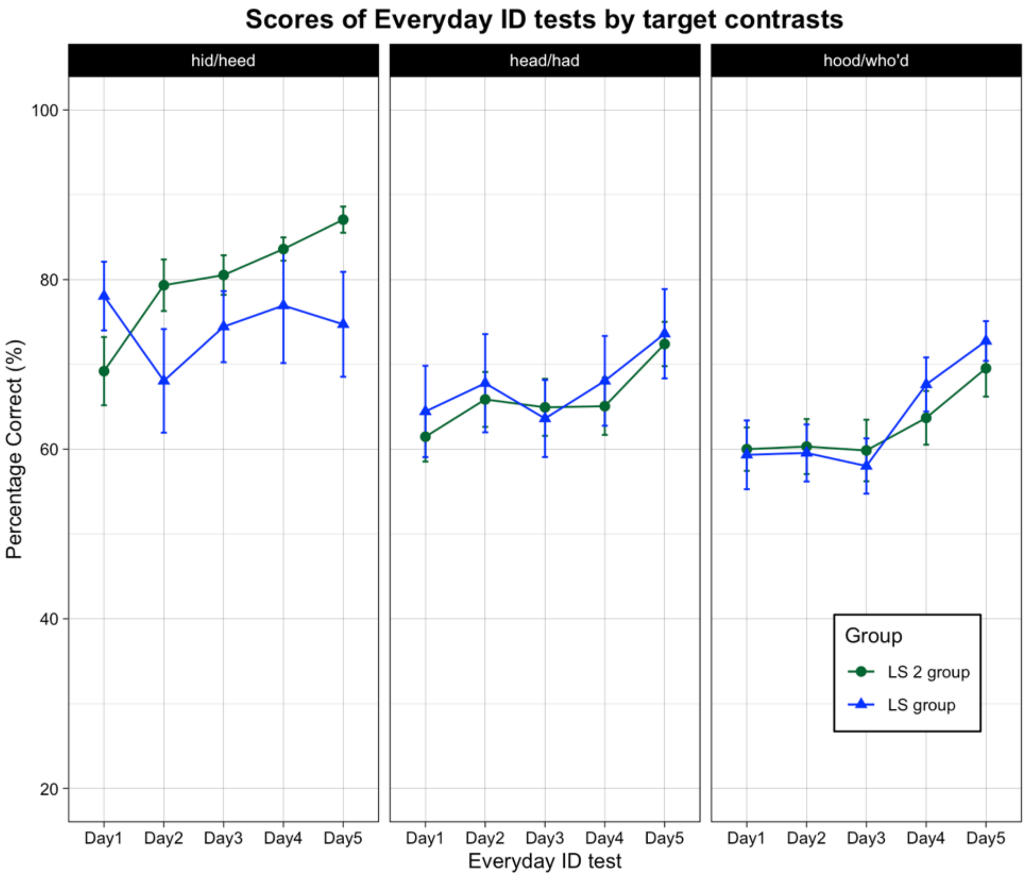


Figure 73. Results of everyday ID tests by three English vowel contrasts in percentage of correct (%). Each line represents either the LS 2 or the LS groups.

Table 60. Summary of fixed effects for the mixed-effect linear regression model for everyday ID test scores.

| Predictor | Estimate ($\beta$) | Std. Error | *t-value* | *p-value* | Pr(>\|t\|) |
|---|---|---|---|---|---|
| (Intercept) | 69.41 | 1.51 | 46.03 | <0.001 | *** |
| Group | 1.33 | 3.22 | 0.41 | 0.681 | |
| Day1_VERSUS_Day2 | 2.89 | 1.46 | 1.98 | 0.052 | . |
| Day1_VERSUS_Day3 | 3.05 | 1.39 | 2.20 | 0.031 | * |
| Day1_VERSUS_Day4 | 6.32 | 1.26 | 5.01 | <0.001 | *** |
| Day1_VERSUS_Day5 | 10.94 | 1.26 | 8.67 | <0.001 | *** |
| Group:Day1_VERSUS_Day2 | 7.00 | 3.11 | 2.25 | 0.028 | * |
| Group:Day1_VERSUS_Day3 | 6.38 | 2.96 | 2.16 | 0.035 | * |
| Group:Day1_VERSUS_Day4 | 3.33 | 2.70 | 1.23 | 0.221 | |
| Group:Day1_VERSUS_Day5 | 6.15 | 2.70 | 2.28 | 0.025 | * |

Three separate linear regression models were built to examine whether the interactions between Group and levels of Test variable presented in Table 60 are intact for each of the three target vowel contrasts. Table 61, Table 62, and Table 63 summarize regression analysis results. For the /i/-/ɪ/ contrast, the model found significant 2-way interactions between Group and all levels of Test variable. These results reflect that the LS 2 group improved more in their classification of the /i/-/ɪ/ contrast on Day2 to Day5 ID tests than the LS group. However, the results of regression models for /ɛ/-/æ/ and /ʊ/-/u/ contrasts did not find either the main effect of Group or 2-way interactions between Group and Test. When comparing the three vowel contrasts, the LS 2 group showed the cue-attention switching training effect only in their learning of the /i/-/ɪ/ contrast. This result suggests that the LS 2 group's improvement was limited for the /i/-/ɪ/ contrast since this contrast might be a relatively easier vowel contrast to learn than the other remaining contrasts for Korean learners of English.

Table 61. Summary of fixed effects for the mixed-effect linear regression model for everyday ID test scores for the /i/-/ɪ/ contrast.

| Predictor | Estimate ($\beta$) | Std. Error | *t*-value | *p*-value | |
|---|---|---|---|---|---|
| (Intercept) | 78.16 | 1.91 | 40.98 | <0.001 | *** |
| Group | 5.50 | 4.08 | 1.35 | 0.185 | |
| Day1_VERSUS_Day2 | 3.60 | 2.61 | 1.38 | 0.172 | |
| Day1_VERSUS_Day3 | 6.49 | 2.37 | 2.74 | 0.007 | ** |
| Day1_VERSUS_Day4 | 9.37 | 2.37 | 3.96 | <0.001 | *** |
| Day1_VERSUS_Day5 | 10.99 | 2.37 | 4.64 | <0.001 | *** |
| Group:Day1_VERSUS_Day2 | 20.13 | 5.58 | 3.61 | <0.001 | *** |
| Group:Day1_VERSUS_Day3 | 14.94 | 5.06 | 2.95 | 0.004 | ** |
| Group:Day1_VERSUS_Day4 | 15.51 | 5.06 | 3.07 | 0.003 | ** |
| Group:Day1_VERSUS_Day5 | 21.20 | 5.06 | 4.19 | <0.001 | *** |

Table 62. Summary of fixed effects for the mixed-effect linear regression model for everyday ID test scores for the /ɛ/-/æ/ contrast.

| Predictor | Estimate ($\beta$) | Std. Error | *t-value* | *p-value* | |
|---|---|---|---|---|---|
| (Intercept) | 66.45 | 2.21 | 30.07 | <0.001 | *** |
| Group | -1.55 | 4.72 | -0.33 | 0.744 | |
| Day1_VERSUS_Day2 | 4.05 | 2.31 | 1.75 | 0.084 | . |
| Day1_VERSUS_Day3 | 2.07 | 2.23 | 0.93 | 0.356 | |
| Day1_VERSUS_Day4 | 3.60 | 2.24 | 1.61 | 0.112 | |
| Day1_VERSUS_Day5 | 10.36 | 2.23 | 4.65 | <0.001 | *** |
| Group:Day1_VERSUS_Day2 | 1.07 | 4.94 | 0.22 | 0.830 | |
| Group:Day1_VERSUS_Day3 | 4.30 | 4.76 | 0.90 | 0.370 | |
| Group:Day1_VERSUS_Day4 | -0.01 | 4.79 | 0.00 | 0.998 | |
| Group:Day1_VERSUS_Day5 | 1.77 | 4.76 | 0.37 | 0.712 | |

Table 63. Summary of fixed effects for the mixed-effect linear regression model for everyday ID test scores the /ʊ/-/u/ contrast.

| Predictor | Estimate ($\beta$) | Std. Error | *t-value* | *p-value* | |
|---|---|---|---|---|---|
| (Intercept) | 62.93 | 1.74 | 36.21 | <0.001 | *** |
| Group | -0.78 | 3.71 | -0.21 | 0.835 | |
| Day1_VERSUS_Day2 | 0.28 | 2.58 | 0.11 | 0.914 | |
| Day1_VERSUS_Day3 | -0.53 | 2.59 | -0.21 | 0.838 | |
| Day1_VERSUS_Day4 | 5.19 | 2.27 | 2.29 | 0.026 | * |
| Day1_VERSUS_Day5 | 10.80 | 2.27 | 4.77 | <0.001 | *** |
| Group:Day1_VERSUS_Day2 | 0.09 | 5.52 | 0.02 | 0.987 | |
| Group:Day1_VERSUS_Day3 | 1.17 | 5.53 | 0.21 | 0.834 | |
| Group:Day1_VERSUS_Day4 | -4.60 | 4.84 | -0.95 | 0.346 | |
| Group:Day1_VERSUS_Day5 | -3.88 | 4.84 | -0.80 | 0.425 | |

Although the primary interest of experiment 4 is to compare learning outcomes between the LS group and the LS 2 group, I would like to point out that the LS 2 group did not reach to the level of learning by the HS group in experiment 3, especially for /ɛ/-/æ/ and /ʊ/-/u/ contrasts. Figure 82 and Figure 83 in Appendix F show that except for the /i/-/ɪ/ contrast, the HS group identified /ɛ/-/æ/ and /ʊ/-/u/ contrasts in everyday ID tests better than the LS 2 group. Table 70 in Appendix G summarizes the result of the linear regression analysis, which compares the scores of everyday ID tests for the HS group and the LS 2 group. The model found a significant main effect of `Group`, indicating that the HS group overall performed better than the LS 2 group in everyday ID tests (β = - 8.41, *t* = 2.68, *p* = 0.01). Three separate linear regression analyses for each of the target vowels show that the main effect of Group in Table 70 in Appendix G came from the group differences for /ɛ/-/æ/ and /ʊ/-/u/ contrasts in everyday ID tests. The LS 2 group achieved a similar level of learning for the /i/-/ɪ/ contrast, which was shown by the insignificant main effect of Group in Table 71 (β = - 4.89, *t* = 3.16, *p* = 0.13). On the other hand, the LS 2 group did not perform similarly as the HS group for /ɛ/-/æ/ and /ʊ/-/u/ contrasts. As we can see in Table 72 and Table 73, there were main effects of Group for both regression models (β = - 11.89, *t* = 4.37, *p* = 0.01 and β = - 9.13, *t* = 3.86, *p* = 0.02, respectively). Therefore, the comparison between the LS 2 group and the HS group suggests that although the LS 2 group showed the effectiveness of the cue-attention switching training, their achievement failed to reach the same level of achievement of the HS group which showed a relatively higher sensitivity to the spectral dimension in the AXB oddity task.

#### *6.1.5.2.3. Perceptual cue weighting (Pre vs. Post-test with old talker)*

The primary purpose of the cue-attention switching training was to direct participants' attention to the most relevant and informative acoustic cue before training and ultimately aid them to use spectral dimension in the perception of three target English vowel contrasts. Thus, to examine the effectiveness of the cue-attention switching training in Korean English learners' learning in English vowel contrasts, it is essential to analyze how participants in the LS 2 group used spectral and duration cues differently from the LS group in experiment 3. In this section, the logistic regression analysis results along with data from the pretest and the post-test with the "old" talker are presented. The statistical methods used in this section were identical to the methods used in experiment 3 (see 5.1.5.3.3).

In the previous section, the results of everyday ID tests showed that the LS and the LS 2 groups showed the biggest difference in their learning for the /i/-/ɪ/ contrast than other contrasts. Based on these results, the interest in this section was set to examine whether the LS and the LS 2 groups behaved differently in their changes in the use of spectral and duration dimensions from the pretest to the post-test depending on the type of vowel contrasts. Figure 74, Figure 75, and Figure 76 are heat plots showing participants' responses of the pretest and the post-test with the "old" talker to spectral and duration cues for each vowel contrast. These figures show the response patterns of the LS (top) and the LS 2 (bottom) groups to the pre and post-test stimuli and compare how these two groups changed their use of duration and spectral dimensions before and after training. The plots on the left side demonstrate responses of the pretest as a function of spectral and duration steps of test stimuli, while the plots on the right side demonstrate responses of the post-test as a function of spectral and duration steps of test stimuli. The proportion of *heed/had/who'd* responses are presented in terms of the blueness of the cells, and the proportion of *hid/head/hood* responses are presented in terms of the redness of the cells. In the following section, the results of logistic regression analysis are presented with a visual inspection of figures to check whether visually observed response patterns are in line with the statistical analysis.

Figure 74 shows the pretest and the post-test responses for the /i/-/ɪ/ contrast. Confirming that the /i/-/ɪ/ contrast is relatively easy to learn, both groups changed their use of spectral and duration dimensions as English native listeners. The test stimuli with higher spectral steps were consistently identified as *heed,* while test stimuli with lower spectral steps were consistently identified as *hid* by both groups. However, the LS 2 group involved the use of spectral dimension to a larger degree as visually confirmed by the clear distinction between red and blue color regions. The results of the logistic regression confirmed these observations (Table 64). There were significant main effects of Duration and Spectral. Participants gave more /i/ responses as the vowel spectral and duration steps shift to the /i/ vowel, indicating that they are sensitive to spectral and duration changes. Notably, the beta-coefficients of Duration and Spectral variables show that participants relied more on the spectral dimension than the duration dimension when they identified the test stimuli ($\beta_{Dur}$:0.76, $\beta_{Spec}$: 3.31). The model also found significant interactions between each acoustic dimension and Test, indicating that participants increased their use of spectral dimension in the post-test and at the same time, decreased their use of duration dimension. The reliable 3-way interaction between Group, Test, and Spectral suggests that the LS 2 group increased the effect of spectral dimension in the post-test more than the LS group did.

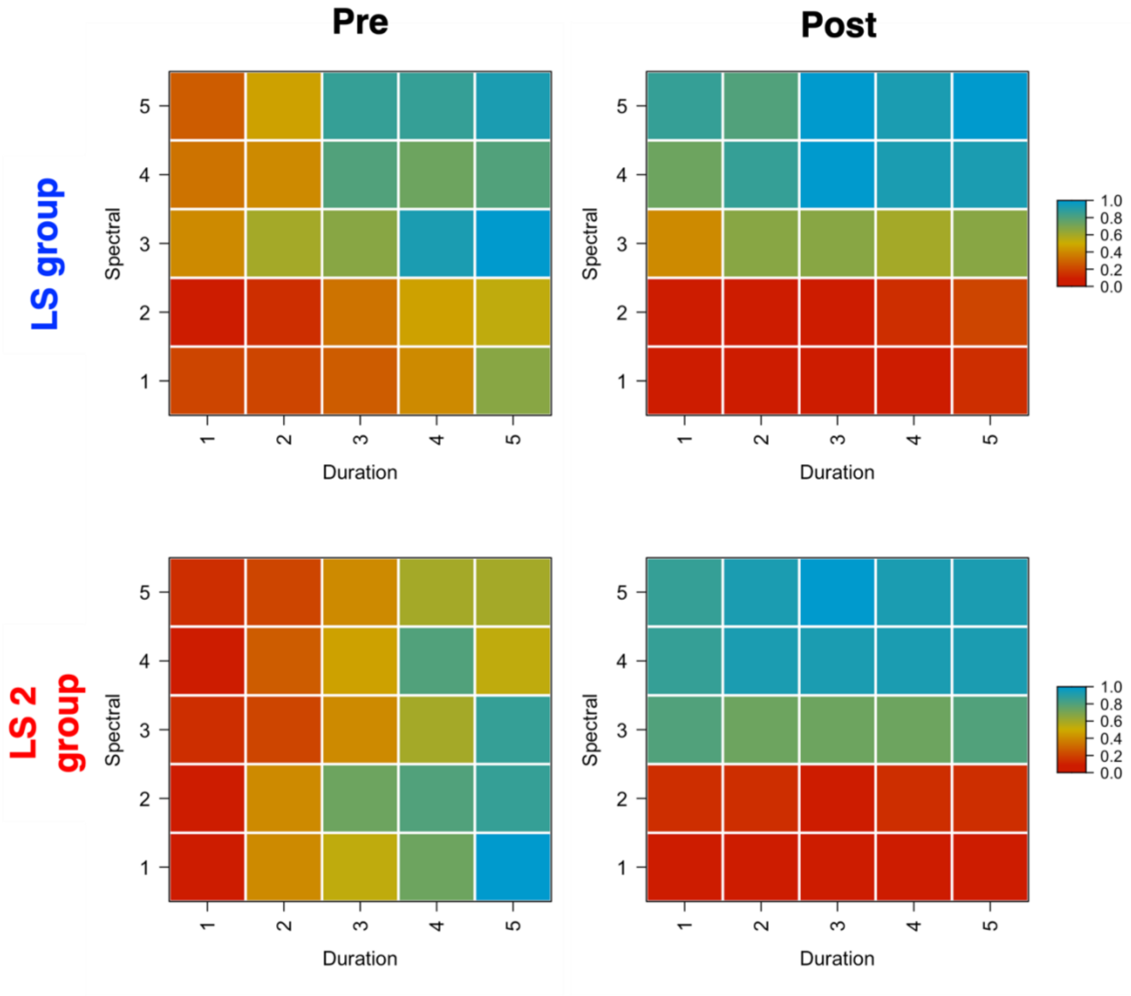


Figure 74. Heat plots of participants' overall responses of the pretest and the post-test with the "old" talker to each combination of spectral and duration dimensions for the /i/-/ɪ/ contrast. The darkness of the cell represents the percentage of responses in a forced-choice task; the darkest red cells elicited 100% '*hid'* responses, while the darkest blue cells elicited 100% '*heed*.'

Table 64. Summary of fixed effects for the mixed-effects logistic regression model predicting changes in spectral and duration cue weights from the pretest to the post-test with the “old” talker for the /i/-/ɪ/ contrast (LS group vs. LS 2 group).

| Predictor | Estimate ($\beta$) | Std. Error | *z-value* | *p-value* | |
|---|---|---|---|---|---|
| (Intercept) | 0.37 | 0.10 | 3.57 | <0.001 | *** |
| Group | 0.12 | 0.22 | 0.53 | 0.595 | |
| Duration | 0.76 | 0.18 | 4.19 | <0.001 | *** |
| Spectral | 3.31 | 0.57 | 5.82 | <0.001 | *** |
| Test | 0.08 | 0.17 | 0.48 | 0.632 | |
| Group:Duration | -0.45 | 0.39 | -1.16 | 0.245 | |
| Group:Spectral | -0.47 | 1.22 | -0.39 | 0.698 | |
| Group:Test | 0.81 | 0.37 | 2.19 | 0.029 | * |
| Duration:Test | -0.93 | 0.16 | -5.98 | <0.001 | *** |
| Spectral:Test | 5.42 | 0.21 | 25.43 | <0.001 | *** |
| Group:Spectral:Test | 3.29 | 0.41 | 8.02 | <0.001 | *** |
| Group:Duration:Test | 0.11 | 0.33 | 0.33 | 0.741 | |

Figure 75 shows the pretest and the post-test responses for the /ɛ/-/æ/ contrast. This figure shows much confusion in identifying two English words, *head* and *had*. In the pretest, the LS group showed their trend to assign *head* responses to stimuli with lower (i.e., shorter) duration steps and *had* responses to stimuli with higher (i.e., longer) duration steps. The LS group’s initial attention to duration was slightly shifted to the spectral dimension in the post-test. The plot on the right side shows that stimuli with lower spectral steps were more likely to be identified as *head* tokens; on the other hand, stimuli with higher spectral steps were mainly identified as *had* tokens. The LS 2 group showed a higher degree of confusion in identifying stimuli into two categories, resulting in a seemingly random response pattern. The LS 2 group failed to systematically use both spectral and duration dimensions to distinguish the contrast.

Table 65 summarizes the results of the logistic regression model. The model found significant main effect of `Spectral` and nearly significant effect of `Duration`. Participants gave more /æ/ responses as the vowel spectral and duration steps shift to the /æ/ vowel. Although participants overall were influenced by the spectral changes in the identification of the /ɛ/-/æ/, they were more sensitive to the spectral changes in the identification of the /i/-/ɪ/ contrast. This was shown as a much larger coefficient of Spectral in Table 64 ($\beta_{\text{Spec}}$: 3.31) than in Table 65 ($\beta_{\text{Sepc}}$: 0.87). There was a significant 2-way interaction between Spectral and Test, indicating that spectral dimension was used more in the post-test. The 3-way interaction between Group, Test, and Spectral suggests that the LS group increased the effect of spectral dimension in the post-test more than the LS 2 group. The noticeable difference between the /i/-/ɪ/ and /ɛ/-/æ/ models was that coefficients of the Spectral and Test interaction and the Group, Test, and Spectral interaction were larger for the /i/-/ɪ/ model ($\beta_{\text{Spec and Test}}$:5.42, $\beta_{\text{Group, Sepc, and Test}}$: 3.29) than the /ɛ/-/æ/ model ($\beta_{\text{Spec and Test}}$:0.90, $\beta_{\text{Group, Sepc, and Test}}$: -0.76). This indicates that participants could primarily rely on spectral cues for the /i/-/ɪ/ contrast in a more nativelike way than the /ɛ/-/æ/ contrast and that the learning of the /ɛ/-/æ/ contrast lagged behind when compared to the /i/-/ɪ/ contrast.

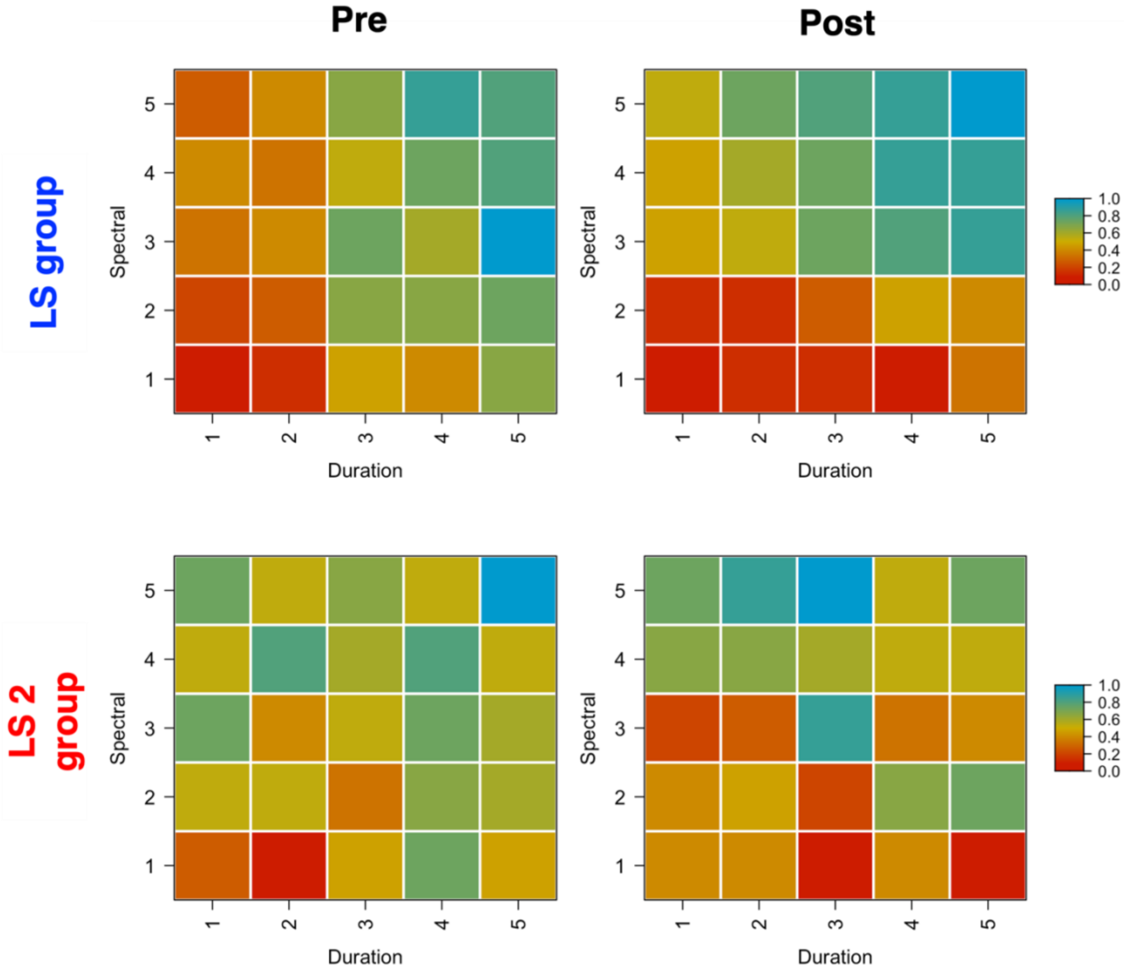


Figure 75. Heat plots of participants' overall responses of the pretest and the post-test with the "old" talker to each combination of spectral and duration dimensions for the /ɛ/-/æ/ contrast. The darkness of the cell represents the percentage of responses in a forced-choice task; the darkest red cells elicited 100% '*head'* responses, while the darkest blue cells elicited 100% '*had*.'

Table 65. Summary of fixed effects for the mixed-effects logistic regression model predicting changes in spectral and duration cue weights from the pretest to the post-test with the "old" talker for the /ɛ/-/æ/ contrast (LS group vs. LS 2 group).

| Predictor | Estimate ($\beta$) | Std. Error | *z-value* | *p-value* | |
|---|---|---|---|---|---|
| (Intercept) | -0.09 | 0.09 | -0.99 | 0.323 | |
| Group | -0.14 | 0.20 | -0.71 | 0.480 | |
| Duration | 0.45 | 0.23 | 1.92 | 0.055 | . |
| Spectral | 0.87 | 0.18 | 4.82 | <0.001 | *** |
| Test | 0.16 | 0.16 | 0.99 | 0.325 | |
| Group:Duration | -1.20 | 0.50 | -2.41 | 0.016 | * |
| Group:Spectral | -0.65 | 0.39 | -1.67 | 0.096 | . |
| Group:Test | 0.17 | 0.34 | 0.49 | 0.624 | |
| Duration:Test | -0.24 | 0.13 | -1.89 | 0.059 | . |
| Spectral:Test | 0.90 | 0.13 | 7.15 | <0.001 | *** |
| Group:Spectral:Test | -0.76 | 0.28 | -2.74 | 0.006 | ** |
| Group:Duration:Test | 0.29 | 0.28 | 1.02 | 0.306 | |

Lastly, Figure 76 shows the pretest and the post-test responses for the /ʊ/-/u/ contrast. In the pretest, both groups relied on duration dimension in a similar pattern, in which stimuli with lower duration steps were associated with *hood* responses and stimuli with higher duration steps were associated with *who'd* responses. The LS group still showed their ability to reliably distinguish the contrast by relying on duration cues even after HVPT (i.e., the post-test). On the contrary, the LS 2 group showed a different pattern in the post-test by presenting their ability to utilize spectral dimension as a primary cue. This pattern is consistent with the LS 2 group's pattern for classifying the /i/-/ɪ/ contrast. Compared to Figure 74, Figure 76 shows that the LS 2 group relied less on spectral dimension for the /ʊ/-/u/ contrast than the /i/-/ɪ/ contrast. The logistic regression analysis results in Table 66 show main effects of Spectral and Duration. There were significant 2-way interactions between each of acoustic dimensions and Test, indicating that the effect of spectral dimension was increased in the post-test while the effect of duration dimension was decreased. The 3-way interaction between Group, Test, and Spectral suggests that the increased effect of spectral dimension in the post-test was larger for the LS 2 group than the

LS group. The model comparison between the /i/-/ɪ/ and /ʊ/-/u/ contrasts is very similar to the one between the /i/-/ɪ/ and /ɛ/-/æ/ contrasts, suggesting a relatively slower learning process for the /ʊ/-/u/ contrast compared to the learning of the /i/-/ɪ/ contrast. The coefficients of the Spectral and Test interaction and the Group, Test, and Spectral interaction for the /ʊ/-/u/ model were smaller than the coefficients in the model for the /i/-/ɪ/ contrast: $\beta_{\text{Spec and Test}}$:1.53, $\beta_{\text{Group, Sepc, and Test}}$: 1.84 (/ʊ/-/u/ contrast) vs. $\beta_{\text{Spec and Test}}$:5.42, $\beta_{\text{Group, Sepc, and Test}}$: 3.29 (/i/-/ɪ/ contrast).

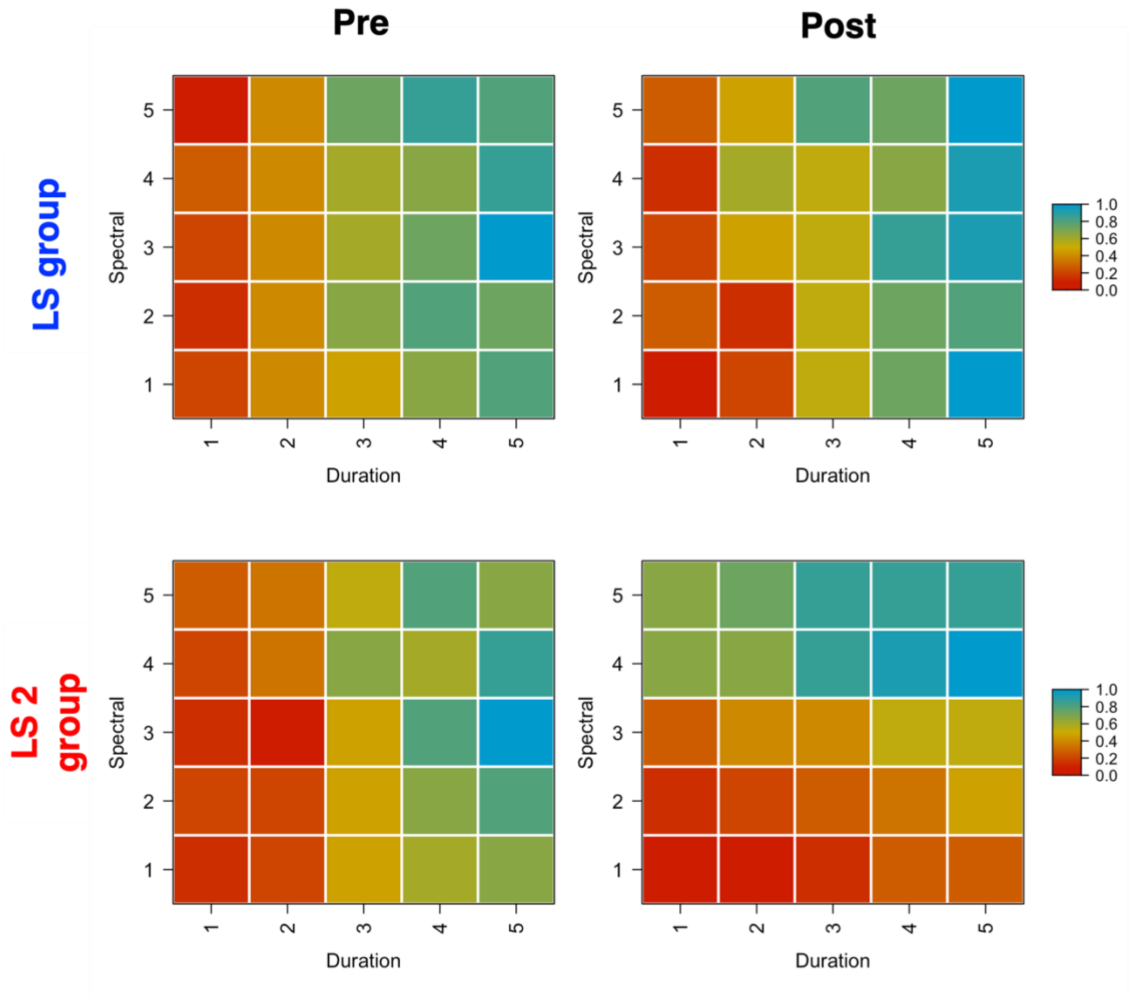


Figure 76. Heat plots of participants' overall responses of the pretest and the post-test with the "old" talker to each combination of spectral and duration dimensions for the /ʊ/-/u/ contrast. The darkness of the cell represents the percentage of responses in a forced-choice task; the darkest red cells elicited 100% '*hood'* responses, while the darkest blue cells elicited 100% '*who'd*.'

Table 66. Summary of fixed effects for the mixed-effects logistic regression model predicting changes in spectral and duration cue weights from the pretest to the post-test with the “old” talker for the /ʊ/-/u/ contrast (LS group vs. LS 2 group).

| Predictor | Estimate ($\beta$) | Std. Error | *z-value* | *p-value* | |
|---|---|---|---|---|---|
| (Intercept) | -0.03 | 0.08 | -0.39 | 0.701 | |
| Group | -0.21 | 0.18 | -1.20 | 0.232 | |
| Duration | 1.44 | 0.29 | 4.97 | <0.001 | *** |
| Spectral | 0.98 | 0.16 | 6.19 | <0.001 | *** |
| Test | 0.06 | 0.18 | 0.31 | 0.760 | |
| Group:Duration | -0.89 | 0.62 | -1.44 | 0.149 | |
| Group:Spectral | 0.97 | 0.34 | 2.87 | 0.004 | ** |
| Group:Test | 0.92 | 0.39 | 2.38 | 0.017 | * |
| Duration:Test | -0.32 | 0.14 | -2.30 | 0.021 | * |
| Spectral:Test | 1.53 | 0.14 | 11.30 | <0.001 | *** |
| Group:Spectral:Test | 1.84 | 0.28 | 6.51 | <0.001 | *** |
| Group:Duration:Test | -0.18 | 0.30 | -0.61 | 0.541 | |

Figure 77 illustrates group trends of participants’ changes in spectral and duration cue weights in the pretest and the post-test with the “old” talker. The plots on the left side present the differences between the pretest $\beta_{spec}$ and the post-test $\beta_{spec}$ values (i.e., $\beta_{pretest\ spec} - \beta_{post\text{-}test\ spec}$); and the plots on the right side present the difference between the pretest βdur and the post-test $\beta_{dur}$ values (i.e., $\beta_{pretest\ dur} - \beta_{post\text{-}test\ dur}$). Note that values over zero indicate more use of the corresponding acoustic dimension in the post-test than the pretest. The higher the values were, the more the acoustic dimension affected in the post-test. The right-side plot is in line with the statistical results presented above in that the LS 2 group increased their reliance on spectral dimension more than the LS group in the post-test for the /i/-/ɪ/ contrast. The other two vowel contrasts did not show much group difference, indicating that the LS and the LS 2 groups did not differ in their changes in spectral cue weights from the pretest to the post-test.

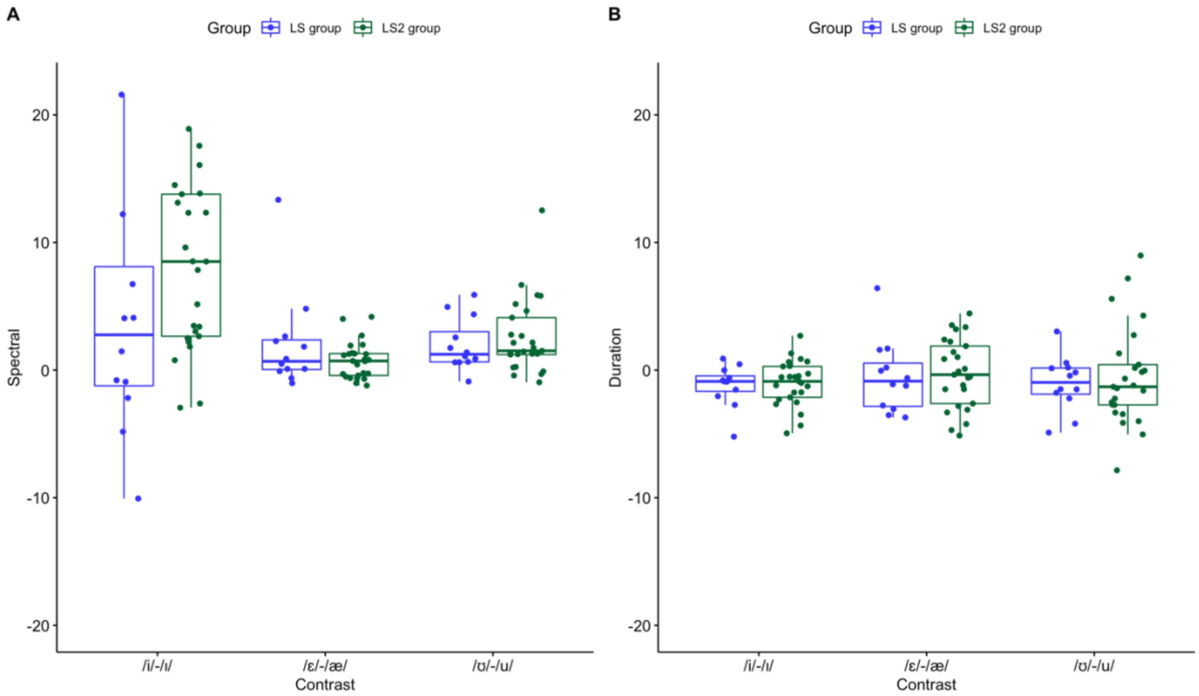


Figure 77. Group trends of changes in spectral and duration cue weights in the pretest and the post-test with the "old" talker. The x-axis represents each vowel contrast. The y-axis represents the differences between each group participants' pre-test and post-test beta-coefficients of spectral dimension (left) and duration dimension (right). Higher values indicate more use of the corresponding acoustic dimension in the post-test.

#### *6.1.5.2.4. Perceptual cue weighting (new talker generalization test with a novel talker)*

This section presents the results of mixed-effects logistic regression analyses for the new talker generalization data. The logistic regression models were built for each vowel contrast. The major purpose of this section is to examine whether the LS and the LS 2 groups show similar or different patterns of changing their cue weightings from the pretest to the post-test with a novel talker compared to their previous presented patterns in tests with the "old" talker (see 6.1.5.2.3).

Figure 78, Figure 79, and Figure 80 are heat plots showing participants' responses of the pretest and the post-test with a new talker to spectral and duration cues for each vowel contrast. These figures show the response patterns of the LS and the LS 2 groups to the test stimuli and compare how these two groups changed their use of duration and spectral dimensions before and after training. The plots on the left side demonstrate responses of the pretest as a function of spectral and duration steps of test stimuli, while the plots on the right side demonstrate responses of the post-test as a function of spectral and duration steps of test stimuli. The proportion of *heed/had/who'd* responses are presented in terms of the blueness of the cells, and the proportion of *hid/head/hood* responses are presented in terms of the redness of the cells. In the following section, the results of logistic regression analysis are presented with a visual inspection of figures to check whether visually observed response patterns are in line with the statistical analysis.

Figure 78 shows the new talker generalization test responses for the /i/-/ɪ/ contrast. Similar to the pattern shown in the post-test with old talker stimuli, both groups showed a pattern closer to native listeners in the post-test with novel talker stimuli. However, the boundary between the red and blue area is clearer in the plot of the LS 2 group's post-test results. The results of the logistic regression summarized in Table 67 confirmed the visual observation. The significant 2-way interaction between Spectral and Test suggests that participants increased their use of spectral dimension in classifying the /i/-/ɪ/ contrast in the post-test. The significant 3-way interaction between Group, Spectral, and Test further suggests that the increase of spectral cue reliance in the post-test was greater for the LS 2 group.

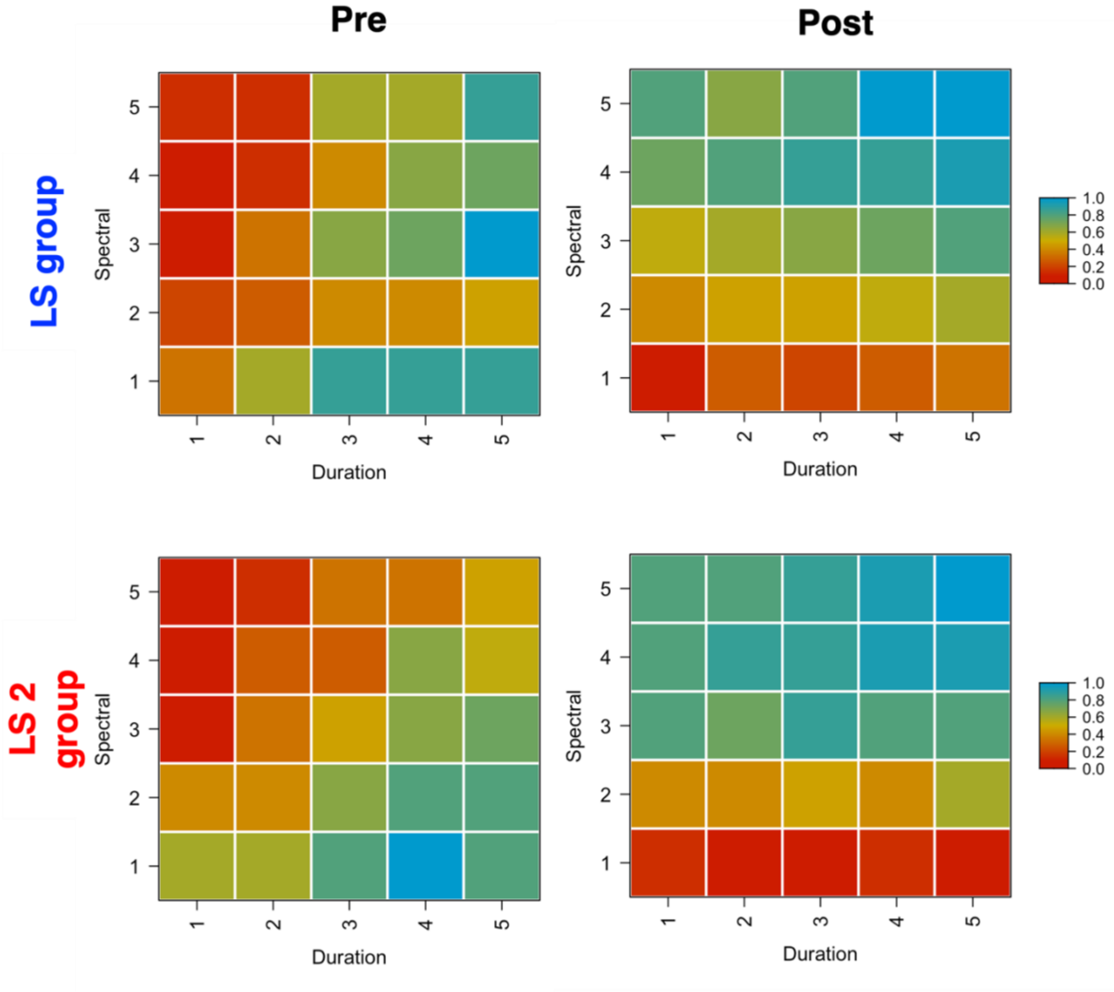


Figure 78. Heat plots of participants' overall responses of the pretest and the post-test with a "new" talker to each combination of spectral and duration dimensions for the /i/-/ɪ/ contrast. The darkness of the cell represents the percentage of responses in a forced-choice task; the darkest red cells elicited 100% '*hid'* responses, while the darkest blue cells elicited 100% '*heed*.'

Table 67. Summary of fixed effects for the mixed-effects logistic regression model predicting changes in spectral and duration cue weights from the pretest to the post-test with a “new” talker for the /i/-/ɪ/ contrast (LS group vs. LS 2 group).

| Predictor | Estimate ($\beta$) | Std. Error | *z-value* | *p-value* | |
|---|---|---|---|---|---|
| (Intercept) | 0.40 | 0.14 | 2.77 | 0.006 | ** |
| Group | 0.08 | 0.31 | 0.25 | 0.805 | |
| Duration | 0.84 | 0.18 | 4.59 | 0.000 | *** |
| Spectral | 1.86 | 0.51 | 3.66 | 0.000 | *** |
| Test | 0.80 | 0.19 | 4.13 | 0.000 | *** |
| Group:Duration | -0.20 | 0.39 | -0.52 | 0.603 | |
| Group:Spectral | -0.39 | 1.09 | -0.36 | 0.717 | |
| Group:Test | 0.63 | 0.41 | 1.52 | 0.127 | |
| Duration:Test | -0.45 | 0.15 | -3.09 | 0.002 | ** |
| Spectral:Test | 4.76 | 0.18 | 26.08 | < 2e-16 | *** |
| Group:Spectral:Test | 1.76 | 0.36 | 4.84 | 0.000 | *** |
| Group:Duration:Test | -0.18 | 0.31 | -0.58 | 0.560 | |

Figure 79 displays new talker generalization responses for the /ɛ/-/æ/ contrast. Before receiving HVPT, both groups showed a great deal of confusion in identifying *head* and *had* during the pre-test with new talker stimuli. However, after training, the LS 2 group showed their ability in generalizing their learning to a novel set of stimuli by relying more on spectral dimension in the post-test. The 3-way interaction between `Group`, `Spectral`, and `Test` (Table 68) indicates that the effect of spectral dimension in the post-test increased more for the LS 2 group than the LS group. These results are different from the results of the pretest and the post-test with old talker stimuli, which showed that the LS group showed more reliance on spectral dimension in the post-test than the LS 2 group.

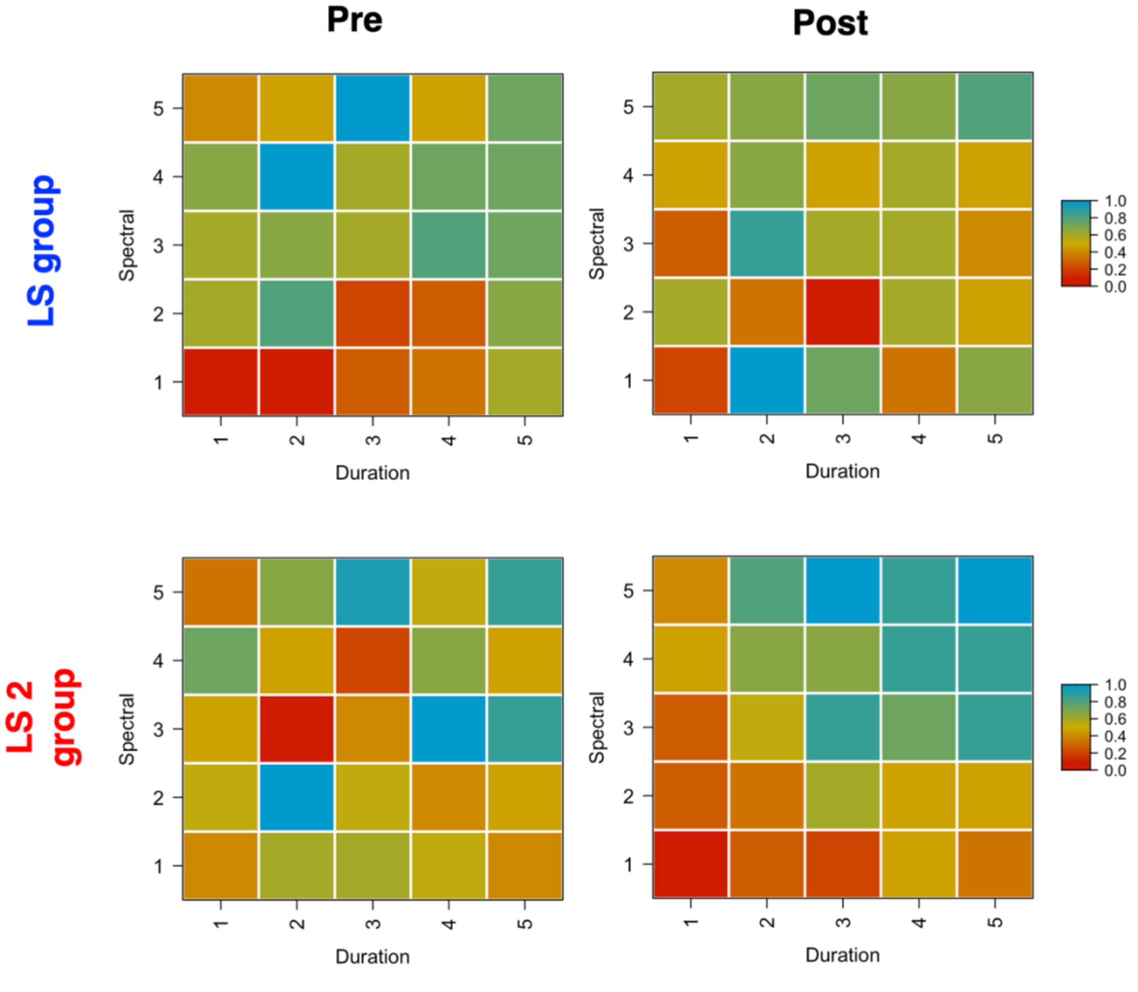


Figure 79. Heat plots of participants' overall responses of the pretest and the post-test with a novel talker to each combination of spectral and duration dimensions for the /ɛ/-/æ/ contrast. The darkness of the cell represents the percentage of responses in a forced-choice task; the darkest red cells elicited 100% '*head'* responses, while the darkest blue cells elicited 100% '*had*.'

Table 68. Summary of fixed effects for the mixed-effects logistic regression model predicting changes in spectral and duration cue weights from the pretest to the post-test with a "new" talker for the /ɛ/-/æ/ contrast (LS group vs. LS 2 group).

| Predictor | Estimate ($\beta$) | Std. Error | *z-value* | *p-value* | |
|---|---|---|---|---|---|
| (Intercept) | -0.05 | 0.10 | -0.48 | 0.630 | |
| Group | 0.01 | 0.21 | 0.04 | 0.966 | |
| Duration | 0.36 | 0.33 | 1.08 | 0.278 | |
| Spectral | 0.43 | 0.12 | 3.74 | 0.000 | *** |
| Test | 0.09 | 0.17 | 0.51 | 0.610 | |
| Group:Duration | 0.35 | 0.71 | 0.49 | 0.624 | |
| Group:Spectral | 0.14 | 0.25 | 0.56 | 0.574 | |
| Group:Test | 0.62 | 0.36 | 1.70 | 0.089 | . |
| Duration:Test | 0.29 | 0.13 | 2.29 | 0.022 | * |
| Spectral:Test | 0.40 | 0.12 | 3.27 | 0.001 | ** |
| Group:Spectral:Test | 1.21 | 0.26 | 4.65 | 0.000 | *** |
| Group:Duration:Test | 0.76 | 0.27 | 2.84 | 0.005 | ** |

Figure 80 displays new talker generalization responses for the /ʊ/-/u/ contrast. This figure reflects that both groups relied primarily on duration information in categorizing the target vowels for the /ʊ/-/u/ contrast in the pretest. In the post-test, however, both groups changed their primary reliance from duration dimension to spectral dimension. Table 69 confirms this visual inspection with a significant 2-way interaction between Spectral and Test and non-significant 3-way interaction between Group, Spectral, and Test.

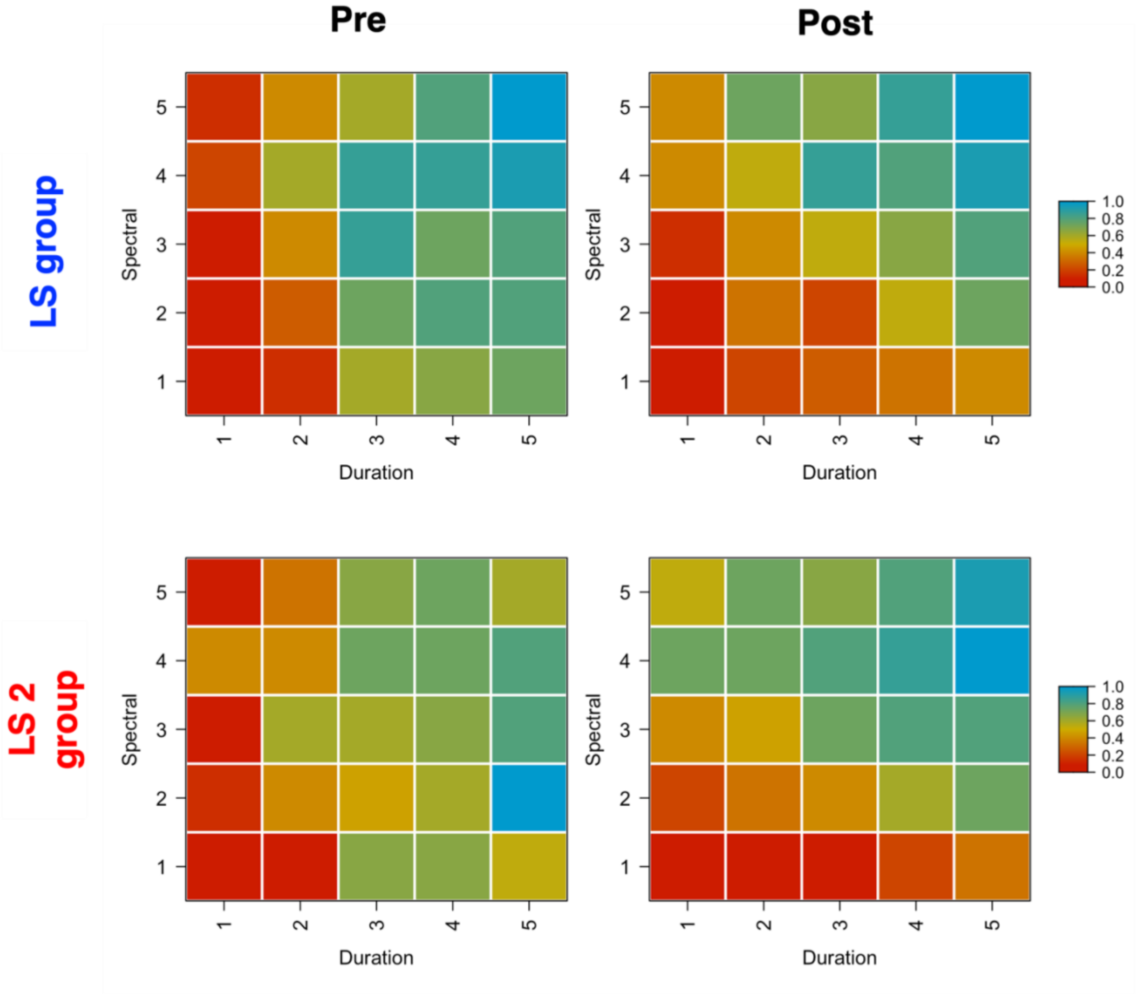


Figure 80. Heat plots of participants' overall responses of the pretest and the post-test with a "new" talker to each combination of spectral and duration dimensions for the /ʊ/-/u/ contrast. The darkness of the cell represents the percentage of responses in a forced-choice task; the darkest red cells elicited 100% '*hood'* responses, while the darkest blue cells elicited 100% '*who'd*.'

Table 69. Summary of fixed effects for the mixed-effects logistic regression model predicting changes in spectral and duration cue weights from the pretest to the post-test with a “new” talker for the /ʊ/-/u/ contrast (LS group vs. LS 2 group).

| Predictor | Estimate (*β*) | Std. Error | *z-value* | *p-value* | |
|---|---|---|---|---|---|
| (Intercept) | 0.33 | 0.09 | 3.79 | 0.000 | *** |
| Group | -0.14 | 0.18 | -0.74 | 0.457 | |
| Duration | 1.33 | 0.25 | 5.31 | 0.000 | *** |
| Spectral | 1.02 | 0.19 | 5.32 | 0.000 | *** |
| Test | 0.12 | 0.18 | 0.66 | 0.513 | |
| Group:Duration | -0.70 | 0.53 | -1.31 | 0.191 | |
| Group:Spectral | -0.16 | 0.41 | -0.39 | 0.700 | |
| Group:Test | 0.73 | 0.39 | 1.88 | 0.061 | . |
| Duration:Test | -0.03 | 0.13 | -0.25 | 0.805 | |
| Spectral:Test | 1.44 | 0.13 | 10.84 | < 2e-16 | *** |
| Group:Spectral:Test | 0.29 | 0.28 | 1.00 | 0.316 | |
| Group:Duration:Test | 0.16 | 0.29 | 0.56 | 0.575 | |

Figure 81 illustrates group trends of participants’ changes in spectral and duration cue weights in the pretest and the post-test with a “new” talker. The plots present the same information as Figure 77, which means that on the left side present the differences between the pretest $\beta_{spec}$ and the post-test $\beta_{spec}$ values (i.e., $\beta_{\text{pretest spec}} - \beta_{\text{post-test spec}}$); and the plots on the right side present the difference between the pretest βdur and the post-test $\beta_{dur}$ values (i.e., $\beta_{\text{pretest dur}} - \beta_{\text{post-test dur}}$). Once again, the higher the values were, the more the acoustic dimension affected in the post-test with new talker stimuli. Notably, the group difference in the left-side plot was most for the /i/-/ɪ/ contrast, suggesting that the LS 2 group not only increased the effect of spectral dimension in classifying the /i/-/ɪ/ contrast in the post-test with the “old” taker (Figure 77), but also presented better ability to generalize their learning from HVPT to a new environment created by a novel talker.

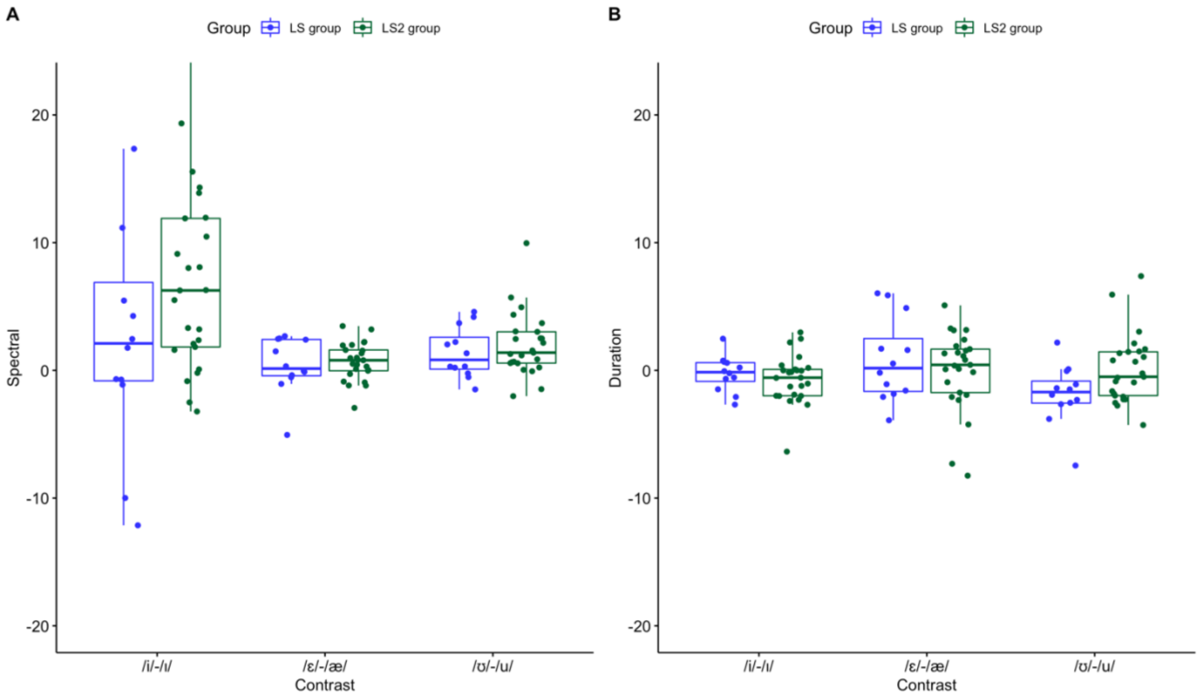


Figure 81. Group trends of changes in spectral and duration cue weights in the pretest and the post-test with a "new" talker. The x-axis represents each vowel contrast. The y-axis represents the differences between each group participants' pre-test and post-test beta-coefficients of spectral dimension (left) and duration dimension (right). Higher values indicate more use of the corresponding acoustic dimension in the post-test.

## 6.2. Summary of experiment 4

As a whole, a group of participants in experiment 4 (i.e., the LS 2 group) was comparable to the LS group in experiment 3 in that the LS 2 group's AXB oddity scores fall into the range of scores by the LS group. The LS group additionally received the cue-attention switching training before starting each HVPT session, and their performances were compared to the ones of the LS group to examine the effectiveness of the short additional training in nonnative phonological contrast learning. The results showed that the LS 2 group performed better than the LS group in Day 2, Day 3, and Day 5 everyday ID tests. The follow-up statistical analysis separately conducted to each vowel contrast revealed that the LS 2 group exceeded the

LS group' learning especially for the /i/-/ɪ/ contrast. The higher achievement in the perception of the /i/-/ɪ/ contrast over the other two contrasts suggest the difficulty hierarchy among three English vowel contrast with the /i/-/ɪ/ contrast as the easiest contrast to acquire.

The logistic regression analyses showed a group difference in how the LS and LS 2 groups handled two acoustic cues in the identification of English words of the target vowel contrasts. In the post-test with the "old" talker, the LS 2 group showed greater use of spectral dimension to classify stimuli into two vowel categories for /i/-/ɪ/ and /ʊ/-/u/ contrasts. The patterns of using acoustic cues in the post-test with a novel talker resembled this tendency in that the LS 2 group increased their reliance on spectral cues more than the LS group to classify /i/-/ɪ/ and /ɛ/-/æ/ contrasts. Taken together, experiment 4 provided empirical evidence of the effectiveness of the cue-attention switching training accompanied with a traditional HVPT paradigm. This short additional training can aid learners to improve their learning by switching their attention to the most relevant acoustic dimension in the perception of nonnative contrasts. However, learners may receive the benefits of this additional training, starting with a relatively easy nonnative contrast (in our study, /i/-/ɪ/ contrast) and then experiencing the benefits with a relative harder contrast to learn (in our study, /ɛ/-/æ/ and /ʊ/-/u/ contrasts).

# CHAPTER 7. DISCUSSION 2

Experiment 3 and Experiment 4 examined L1 Korean speakers learning English to quantify individual variability in sensitivity to fine-grained differences of acoustic cues in the perception of the L1 category and investigate its relation to L2 learning success. Korean adult learners of English were trained to learn challenging English vowel contrasts /i/-/ɪ/, /ɛ/-/æ/, and /ʊ/-/u/. As argued in Bohn (1995), Korean learners of English were assumed to have difficulties in learning the target English vowel contrasts because of their lack of sensitivity to spectral properties to distinguish English vowels of target contrasts due to the absence of those contrasts in learners' L1, Korean. In addition, Korean learners of English tend to perceptually assimilate these pairs of English vowels to Korean vowels, /i/, /e/, and /u/, respectively. Thus, I hypothesized that learners with higher sensitivity to spectral properties would master the distinctions between English vowels of target contrasts better than learners with comparably lower sensitivity to spectral properties.

As in experiment 1 and experiment 2, Korean adult learners of English went through several experimental phases. First, the AXB oddity task quantified learners' sensitivity to within-category differences induced by spectral and duration cue changes, using a set of stimuli belonging to Korean vowel /i/ varying in 5 spectral and 5 duration steps. The AXB stimuli pair consisted of the compared stimulus (*X*) and two comparing stimuli. *A* and *B* stimuli have either a one-step different spectral or duration step from the *X* stimulus, while the other step remains the same as the *X* stimulus. During the AXB oddity task, the learners were asked to judge whether the *X* stimulus is more distinct from either the *A* or *B* stimulus and pick the "most odd one" out. We required the learners to pay attention to the differences between *A*, *X*, and *B* stimuli since this task was to measure how sensitive the learners are to within-category spectral cue changes. We divided learners into two groups (HS groups vs. LS groups) based on the results. Second, participants received five-day computer-based auditory training (i.e., HVPT) with feedback. The training stimuli were six English words which were synthesized to comprise minimal pairs from the endpoints of a five-step spectral and duration continuum of either /i/-/ɪ/, /ɛ/- /æ/, or /ʊ/-/u/: hid/heed, head/had, and hood/who'd. Pre- and post-ID tests, everyday ID tests, and two generalization tests (i.e., new talker generalization ID test & new words generalization ID test) measured improvements during training and the ability to generalize learning to a novel set of stimuli. Participants in experiment 4 additionally received the cue-attention switching training administered with systematically varying Korean /i/ vowel stimuli. Only the subset of stimuli with highest and lowest values in the spectral dimension was presented to upweight spectral cues in the perception of the Korean vowel category.

Several statistical tests were performed for the group- and individual-level analysis. The group-level analysis compared HS and LS groups' performances on multiple tests conducted during and after HVPT. For the individual-level analysis, I examined individual participants' developmental trajectories in perceptual cue weighting across the vowel contrasts over time and investigated whether individual variability in the AXB oddity task predicted participants' changes in their spectral cue weights from before training to after training.

The following sections summarize the main findings and discuss them in light of the two research questions.

## 7.1. Individual differences in nonnative contrast learning and developmental trajectories (Research question 1)

The prediction for the first research question was that native Korean learners of English who exhibit more spectral cue sensitivity in the perception of a Korean vowel category outperform in the acquisition of the systematic use of spectral cues to discriminate the challenging English vowel contrasts, /i/-/ɪ/, /ɛ/-/æ/, and /ʊ/-/u/, compared to learners with lower sensitivity to spectral differences. The results of experiment 3 showed that spectral cue weight changed before and after the training and did so differently for both the HS group and the LS group learners. The HS group learners increased their use of spectral cues to identify the challenging English vowel contrasts more than the LS group learners. Notably, the HS group had advantages in generalizing their learning on utilizing spectral cues to the stimuli produced by a novel talker. Experiment 3 suggests how sensitive learners are to the L2 relevant acoustic cues in the L1 within-category perception could affect their L2 learning outcomes. On the one hand, higher sensitivity to the "right" acoustic cues (e.g., spectral cues), which the AXB oddity task can measure, benefits learners to identify the target L2 sounds in a more native-like way. On the other hand, learners' less sensitivity to L2-relevant cues may hinder their ability to detect the informativeness of those cues.

The results of experiment 3 shed important light on individual differences in within-category cue sensitivity in L1 and its relation to L2 speech perception. As mentioned in the introduction section, Korean adult learners of English do not have much sensitivity to the spectral property of vowels since the Korean vowel system is missing English-like vowel contrasts. The AXB oddity task showed that Korean listeners showed individual variability in within-category cue sensitivity to spectral cues in the perception of single vowel contrast. This suggests that some Korean listeners have better auditory abilities than others, although Korean listeners do not have experience in tuning to small-scale spectral differences in L1. It should be noted that Korean listeners' perceptual abilities to perceive spectral differences may be weaker than listeners whose L1 has a higher functional weight of spectral cues to signal speech contrasts, such as English (Tremblay et al., 2018). The results of experiment 3 suggest a point regarding the L1 influence. In a larger sense, Korea (L1) affects the learning of English (L2) vowel contrasts negatively in general due to the differences between Korean and English vowel systems. However, experiment 3 showed that some learners actually benefited from their higher sensitivity to the L2-relevant acoustic cue in the perception of native vowel contrast. This indicates that the assumption that the discrepancy between native and target language systems is an obstacle in L2 learning may be misleading. Considering that there is a wide range of individual differences in cue sensitivity, the key to successfully mastering the L2 cue-weighting strategy is not exclusively affected by learners' L1 background but also affected by individual learners' sensitivity to the "right" (i.e., L2-relevant) acoustic cues in L1.

Section 5.1.5.3.5 analyzed individual participants' developmental trajectories for each target vowel contrast. The results suggest that seemingly random individual variability in L2 perceptual cue weighting can be understood as distinctive patterns. There were roughly six different patterns of changes spectral cue weights before and after training to identify English target vowel contrasts: "–Spec, +Spec", +Spec, +Spec", "+Dur, +Spec", "+Dur, –Dur", "–Dur, +Spec", and "+Spec, –Spec". Recall that "–" indicates the opposite use of the cue from native listeners, and "+" indicates a nativelike use of acoustic cue. Except for one pattern (+Spec, -Spec), all patterns showed that participants developed their cue-weighting strategy during training in a more nativelike way. Five patterns suggest that participants either increased their reliance on the spectral dimension or reduced their reliance on the duration dimension.

Considering that participants in the current study received only five training sessions, this result is very inspiring. Previous studies suggested that Korean learners of English frequently recognize the English vowel contrasts in terms of duration quality due to a duration-focused EFL (English as a foreign language) instruction. For example, Lee (2009) argued that Korean learners of English tend to assume that there are only length differences between English vowels in contrast because of Korean classroom instructions and dictionaries. This indicates that Korean learners of English need to detect and utilize non-instructed acoustic cues (e.g., spectral cues) in learning English vowel contrasts. Participants in the current study could achieve this process, especially participants showing "−Spec, +Spec", +Spec, +Spec", "+Dur, +Spec", or "−Dur, +Spec" developmental patterns. There was no explicit instruction during the training sessions that mentioned the informativeness of spectral dimension or directed participants' attention to spectral property between training stimuli. Thus, the current study provided additional evidence of the effectiveness of HVPT with highly variable stimuli induced by multiple talkers. However, it should be noted that HVPT is not always effective to all types of learners, indicating that a benefit of HVPT does not always emerge (e.g., Giannakopoulou et al., 2017; Perrachione et al., 2011). For example, Giannakopoulou et al. (2017) showed that both native Greek adults and child learners of English showed consistently stronger performance following low- rather than high-variability phonetic training. In the current study, participants showing the "+Spec, -Spec" pattern might be a group that showed a possible disadvantage in HVPT. Highly phonetically variable multi-talker stimuli possibly impose a burden on some learners acquiring their knowledge of L2 phonological contrast.

## 7.2. The cue-attention switching training and its limitation (Research question 2)

This study assumed Korean English learners' ability to ignore duration cues during the cue-attention switching training based on Kondaurova and Francis (2010). Kondaurova and Francis (2010) argued that the relative importance of acoustic cues might play a significant role in determining cue weighting, and that this importance may be different across individuals as well as across languages (both L1 and L2). The authors stated that since the role of vowel duration in Spanish is considerably less important than in other languages (e.g., English), Spanish learners of English may have an easier experience in reducing the weight they give to duration during the inhibition training as compared to listeners whose L1 extensively use the duration cues. The case of Spanish learners of English applies to the case of Korean learners of English. Even though some native Korean speakers have vowel length distinction, the duration contrasts have been lost, especially among younger Korean speakers. This means that the role of vowel duration in Korean is considerably less important than in English. Therefore, it was predicted that Korean listeners' limited experience with duration cues within the Korean linguistic system might permit inhibition of attention toward this less important acoustic cue.

As an answer to the second research question, the findings of experiment 4 are partly in line with the hypothesized effects of the cue-attention switching training in learning English vowel contrasts by Korean learners of English. It was found that the LS 2 group performed better than the LS group in experiment 3 in identifying the target English vowel contrast with primary reliance on the spectral dimension. However, the LS 2 group did not reach the level of achievement of the HS group in experiment 3. The HS group outperformed the LS 2 group on all three target vowel contrasts. The effectiveness of the cue-attention switching training was shown differently depending on the target contrasts. The effectiveness was most evident in the English

/i/-/ɪ/ contrast (Figure 73). The LS 2 group did not perform better than the LS group in experiment 3 in identifying /ɛ/-/æ/ and /ʊ/-/u/ contrasts.

One potential explanation of the different effects of the cue-attention switching training for each target contrast might be the selection of the Korean vowel category for the additional training. The additional training in this study includes a set of stimuli falling into the Korean vowel /i/ but did not include stimuli falling into Korean /e/ or /u/ vowels. Because of this design, the cue-attention switching training might have its direct influence only on the perception of English /i/-/ɪ/ contrast. The training might help learners pay exclusive attention to spectral cues to perceive the Korean vowel /i/, and consequently, the increased learners' attention to spectral cues might result in better identification of English *hid* and *heed* in a nativelike way. However, this effect was not observed for the other two English target contrasts. This result is congruent with Nishi and Kewley-Port's (2007) results on training English vowels to Japanese learners of English. Training results from Japanese learners of English indicated that learners who were trained with the subset of English vowels which are more difficult than other vowels, showed rapid improvement on the trained subset vowels. However, the learners did not show improvement for the untrained English vowels. In contrast, the learners trained using the full set of English vowels improved gradually on all vowels.

The findings of experiment 4 suggest future work which must examine the selection of vowels for the cue-attention switching training. Specifically, future work should address two questions. The first question is whether all Korean vowels should be included in the cue-attention switching training to observe its benefit in learning all three English vowel contrasts. Secondly, it should be asked whether including a Korean vowel that assimilates to the most difficult English vowel contrast to the cue-attention switching training can increase the effectiveness of this training. Based on the study by Nishi and Kewley-Port (2007, 2008), I predict that providing the cue-attention switching training using all three Korean vowels /i, e, u/ (i.e., full set) would be more beneficial to Korean learners of English than the training only with a set of stimuli of Korean vowel /e/ or /u/ (i.e., subset) which are mapped onto relatively more difficult English /i/-/ɪ/ contrast than /ɛ/-/æ/ and /ʊ/-/u/ contrasts. However, the biggest possible disadvantage of training with the full set of stimuli would be a longer training period. The primary purpose of modifying the training method is to maximize learning outcomes while minimizing the training time span. Therefore, the efficacy of the hybrid protocol, which is a combination of the full set and subset training methods, should be tested in comparison to the full set training method. If the hybrid protocol is found more effective than the full set method in terms of time, it will support the potential of the cue-attention switching training to be improved as a better version. According to Nishi and Kewley-Port (2008), the hybrid protocol may not be effective as much as training with the full set of stimuli, it is still unknown which training method is more beneficial as the cue-attention switching training.

## 7.3. Relative levels of difficulty of the target English vowel contrasts

The target nonnative sounds for Korean learners of English are six English vowels in three contrasts. Therefore, this study could help suggest the relative levels of difficulty in learning three English vowel contrasts and propose a more effective training timeline for Korean learners of English (i.e., assigning more time to relatively tricky contrasts during the training phase). The results of the present study showed that Korean learners had more difficulty in learning appropriate cue utilization for /ɛ/-/æ/ and /ʊ/-/u/ contrasts than the /i/-/ɪ/ contrast. These results confirm previous studies reporting higher levels of difficulty in learning English /ɛ/-/æ/

and /ʊ/-/u/ contrasts than the /i/-/ɪ/. For example, Tsukada et al. (2005) reported that both adult and child Korean learners of English obtained lower discrimination test scores for the English /ɛ/-/æ/ contrast than /i/-/ɪ/ contrast. Kim et al. (2018) suggested the relative difficulty of the acquisition of English /ɛ/-/æ/ contrast than the English /i/-/ɪ/ contrast. Results showed that Korean learners of English used both spectral and duration cues to distinguish the /i/-/ɪ/ contrast like English listeners, while learners primarily used duration cues for the English /ɛ/-/æ/ contrast. Lee (2009) found that after receiving a pronunciation instruction, Korean EFL teachers improved their perception scores slightly better in the English /i/-/ɪ/ contrast than the /ʊ/-/u/ contrast, and they improved their English vowel production test scores significantly only in the English /i/-/ɪ/ contrast.

Kim et al. (2018) proposed two possible accounts for the different levels of difficulty of English vowel contrasts: perceptual mapping patterns between Korean and English vowels and the acoustic distinctiveness of the spectral and duration cues to the two vowels in each English contrast. Baker et al. (2002) have reported trends in the perceptual assimilation patterns between English /i/-/ɪ/ and /ʊ/-/u/ contrasts and Korean /i/ and /u/ vowels in that one of English vowels in each contrast assimilates more to the Korean vowel. In their perceptual mapping study, Korean listeners assimilated English /i/ to Korean /i/ more than English /ɪ/ and assimilated English /u/ to Korean /u/ than English /ʊ/. These patterns are shown in other studies (e.g., Lee & Cho, 2018; Tsukada et al., 2005). The results of Baker et al. (2002) showed that Korean adult listeners chose Korean /i/ for English /i/ 92% of the time and for /ɪ/ 68% of the time and chose Korean /u/ for English /u/ for 83% of the time and /ʊ/61% of the time. The results indicate that English /i/ and /u/ are better matched to Korean /i/ and /u/, respectively, than English /ɪ/ and /ʊ/. These two cases can be considered as *category goodness assimilation* (Best, 1995; Best & Tyler, 2007) when two

L2 sounds are assimilated to one L1 category, but one of the L2 sounds is a better exemplar of the L1 category than the other. On the other hand, neither English /ɛ/ nor /æ/ vowel is well matched to Korean /e/ vowel. Korean adult listeners in Baker et al. (2002) chose Korean /e/ for English /ɛ/ 57% of the time and /æ/ 67% of the time, indicating that both English /ɛ/ and /æ/ vowels poorly correspond to Korean /e/. This case can be considered as *single-category assimilation* when both L2 sounds are assimilated to the same L1 category equally well or poorly. According to the PAM/PAM-L2 (Best, 1995; Best & Tyler, 2007), difficulty levels in discrimination of L2 contrasts are higher for single-category assimilation cases than category goodness assimilation cases. Two different assimilation cases (English /i/-/ɪ/ and /ʊ/-/u/ contrasts vs. English /ɛ/-/æ/ contrast) can explain why learning of English /ɛ/-/æ/ contrast was more problematic for Korean learners of English in the current study than the English /i/-/ɪ/ contrast. More importantly, the learners generalized their learning to the new words only for the English /i/-/ɪ/ contrast. However, it is still unclear why the English /ʊ/-/u/ contrast was also relatively more difficult to learn than the English /i/-/ɪ/ contrast. Although the exact reason cannot be provided in this study, the results of the current study agree with Nishi and Kewley-Port (2008), showing that Korean learners of English identified English /ʊ, ʌ/ least accurately (39% and 31%, respectively).

The acoustic distinctiveness of the spectral and duration cues to the two vowels in each English contrast could also account for the relatively greater difficulty of the English /ɛ/-/æ/ contrast in learning (Kim et al., 2018). As pointed out in the introduction (see 1.3.2), compared to the English /i/-/ɪ/ and /ʊ/-/u/ contrasts, native English listeners relatively relied more on duration cues in identifying the English /ɛ/-/æ/ contrast. Hillenbrand et al. (2000) suggested that this phenomenon is because spectral differences between /ɛ/ and /æ/ are not enough to distinguish due to their greater amount of spectral overlap. To test the relatively smaller spectral differences between /ɛ/ and /æ/, I calculated the F1 and F2 differences between training stimuli with the highest spectral step and the lowest spectral step for each English vowel contrast. For the /i/-/ɪ/ contrast, the F1 and F2 differences were 206 Hz and 358 Hz; for the /ʊ/-/u/ contrasts, the F1 and F2 differences were 228 Hz and 369 Hz; and for the /ɛ/-/æ/ contrast, the F1 and F2 differences were 294 Hz and 55 Hz. Therefore, Korean English learners' relative difficulty with using spectral differences for the English /ɛ/-/æ/ contrast could be attributed to the relatively smaller acoustic distinctiveness in terms of spectral cues, especially F2.

### 7.4. Applicability of the AXB oddity task

The results of the AXB oddity task in the present study successfully predicted participants' changes in spectral cue weights before and after HVPT, as shown by the correlation analysis (see 5.1.5.3.6). Thus, the current study suggests the possibility of the AXB oddity task as a measure of sensitivity to within-category acoustic cues. Together with the VAS task, the AXB oddity task may be used as a pretraining method to predict learning outcomes. Especially, when learners' within-category cue sensitivity is required in L2 learning since multiple L2 target sounds are mapped onto learners' single L1 category, the AXB oddity task is expected to measure learners' ability to perceive small-scale differences of the L2-relevant acoustic cue.

The AXB oddity task has several different characteristics in comparison to the traditional AXB and oddity tasks. First, either *A* or *B* stimulus is not the same as the reference stimulus *X*; thus, there is no correct answer for listeners' choices. Second, unlike the traditional oddity task, which asks whether a stimuli pair (*AX*) is the same or different, the AXB oddity task presents three stimuli and specifies a reference stimulus (*X*) to decide which one of *A* or *B* stimulus sounds more distinct from the *X* stimulus. Third, the AXB oddity task instructs listeners to pay attention to the magnitude of the differences between *AX* and *XB* pairs so that listeners enable to base their decisions on the pairwise comparison. The AXB oddity task is similar to the 4IAX task (Pisoni & Lazarus, 1974) in that both tasks direct listeners to respond to the magnitude of differences between pairs of stimuli. However, unlike the AXB oddity task, the 4IAX task always has the correct answer since two pairs of stimuli are always *AA*, which is the same, and *AX*, which is different. The AXB oddity task would be better than the 4IAX task in terms of testing time since the AXB oddity task presents three stimuli for each trial rather than four stimuli as in the 4IAX task. However, the 4IAX task may have its advantage over the AXB oddity task in that the 4IAX task plays the reference stimuli for each pair (*AA* and *AX*). This design may reduce the perceptual burden on listeners more than the AXB oddity task in which listeners should compare the magnitude of differences between pairs with listening to the reference stimuli only one time. However, the AXB oddity task attempted to overcome this limitation through its way of calculating within-category cue sensitivity scores. Listeners' choices were counted as valid ones only when they were agreed in two pairs for stimuli presentation, where one pair was *AXB*, and the other pair was *BXA*. If a listener chose *A* as a more distinct stimulus from A in the *AXB* pair, the same *A* stimulus should be chosen in the *BXA* pair to be considered for the score. This calculation method considers listeners' consistency in responses and expects to increase the reliability of calculated scores.

# CHAPTER 8. CONCLUSION

The present study contributes to existing literature with an in-depth understanding of how individual differences in L1 speech perception mediate L2 speech perception (e.g., Francis et al., 2000; Holt & Lotto, 2006; Iverson et al., 2003; Tremblay et al., 2018). More importantly, finding evidence for different patterns of phonetic cue utilization in L2 learning by individual learners with the same L1 background suggests that the L1 influence is not always something to overcome for successful L2 learning. L1 influence can be either beneficial or detrimental depending on learners' individual differences in how sensitive they are to sub-phonemic acoustic details of the L2-relevant acoustic dimension. Thus, the results of L2 training experiments in this study suggest two points: (1) Considering individual differences is essential to understand why L2 learners' achievements are highly variable, and (2) Understanding the sources of learning variabilities must be precedent to predict learners' difficulties and provide an appropriate training paradigm to enhance their learning outcomes.

The current study helps us understand how individual differences appear in the course of L2 learning by observing individual learners' developmental processes (i.e., trajectories) of multiple acoustic cue utilization in L2 perception. Instead of relating individual differences with one-time observation of L2 learning performances, the present study closely looked into learners' performances throughout multiple days. This approach provided more insightful evidence for individual differences in cue sensitivity as a possible source of L2 learning variations. However, compared to the previous L2 training studies, the span of training was relatively short enough to be considered as a longitudinal study. Future longitudinal research would benefit from examining how the role of within-category cue sensitivity in L2 learning changes as learners receive more L2 exposure. In a similar sense, the present work opens up a possibility for future research in the role of individual L2 proficiency in modulating the degrees of the transfer of L1 cue sensitivity to L2 cue utilization. The current study targeted relatively less proficient groups of L2 learners and showed that their L2 learning were heavily influenced by their sensitivity to acoustic details in L1 perception. This result suggests the existence of the transfer of L1 cue sensitivity to L2 cue utilization. However, it is unknown that whether learners' L2 proficiency plays an important role how much learners are influenced by their acoustic cue sensitivities, especially in the situation such as learning the reversed cue-weighting is essential (e.g., English learners of Korean). If advanced learners utilize acoustic cues in a nativelike way regardless of their acoustic cue sensitivity in L1 perception, it may suggest that separate long-term memory representations (categories) for L2 phonation-type contrast can be built as learners' proficiency is enhanced. Kong and Edward (2015) examined whether L2 cue-weighting strategies are influenced by individual learners' L2 proficiency. Targeting English learners of Korean, the authors found that as L2 proficiency in Korean increased, VOT became less important and f0 became more important in differentiating the Korean phonation-type contrast relative to how leaners processed the voicing contrast in their native language. Kong and Edward (2015) suggested the L1 to L2 transfer effect modulated by L2 proficiency based on the results of the identification tests of English and Korean stop contrasts. Involving the VAS task as a measure of the degrees of sensitivity to acoustic details as in the present study, future research can suggest the effect of the transfer of L1 cue sensitivity to L2 cue utilization modulated by L2 proficiency.

This study generalizes the effect of individual differences in cue sensitivity in L2 learning in two senses. First, the results of phonetic training targeting two groups of L2 learners with different L1 backgrounds provide cross-language evidence for to what extent for individual differences in cue sensitivity in L1 perception can account for patterns of the acquisition of nativelike cue utilization in L2 perception. Second, this study includes both nonnative consonant and vowel contrast learning contexts to observe whether the relation between individual differences in within-category cue sensitivity and L2 contrast learning can be generalized in the different types of target segments.

Regarding the results of the VAS and AXB oddity tasks, considerable individual variability in L2 learners' response patterns were found. These patterns were expected to represent to what extent L2 learners are sensitive to a certain acoustic dimension in L1 perception which is informative in L2 perception (i.e., in this study, f0 or spectral). Previous research has demonstrated the validity of the VAS task in measuring listeners' within-category cue sensitivity in both consonant and vowel category perception (e.g., Kapnoula et al., 2017; Kim et al., 2020; Kong, 2019; Kong & Edwards, 2011, 2016). However, the AXB oddity task has not been tested its reliability in measuring within-category cue sensitivity. By targeting to measure Korean listeners' sensitivity to spectral and duration cues in Korean vowel perception, this study could suggest the applicability of the AXB oddity task as a measurement of cue sensitivity. One limitation in designing the AXB oddity tasks is the lack of considering psychoacoustic salience of acoustic dimensions. In the AXB oddity tasks, spectral and duration dimensions were manipulated varying in several steps. The interval between duration steps were set to 40 ms, which was less than two just-noticeable differences, namely 50 ms as reported in Klatt (1976). The problem was that one just-noticeable difference (i.e., 25 ms) was based on native English listeners' L1 speech perception. The results of the AXB oddity tasks in experiments 3 and 4 showed that most participants selected the spectrally different stimulus as a more distinct stimulus from the reference stimulus (*X*). This may be due to the loss of the vowel length contrasts in Seoul Korean, especially among younger Korean speakers (Kang, Yoon, &

Han, 2015). Thus, for Korean listeners, a larger duration interval, supposably more than 40 ms, is needed. Future work is needed to explore psychoacoustic salience of the duration dimension in the Korean vowel perception to determine the proper duration interval between stimuli for the AXB oddity task. In terms of the spectral dimension, this study conducted a small-scale of speech perception experiment to determine an interval between spectral steps. Seven native Korean listeners were presented with pairs of spectrally manipulated Korean /i/ vowel stimuli and asked if they could detect the difference between stimuli in each pair. However, a larger scale of experiment is required to determine a noticeable difference of the spectral dimension in the perception of Korean vowels.

The target nonnative sounds for Korean learners of English are six English vowels in three contrasts. Therefore, this study suggested the relative levels of difficulty in learning three English vowel contrasts and propose a more effective training timeline for Korean learners of English (i.e., assigning more time to relatively tricky contrasts during the training phase). In accordance with the previous findings on differences in difficulty between three English vowel contrasts, the current study showed that Korean learners had more difficulty in learning appropriate cue utilization for /ɛ/-/æ/ and /ʊ/-/u/ contrasts than the /i/-/ɪ/ contrast.

This study did not stop at finding a possible source of individual variability in L2 learning. This study is meaningful in that it took one step further and suggested the modified version of HVPT to aid learners whose perceptual abilities in L1 may place them at a disadvantage. The successful application of this training showed that the addition of perceptual adaptation period (i.e., the cue-attention switching training) which requires noncanonical use of acoustic cues can result in nativelike perceptual patterns in L2 perception regardless of learners' initial sensitivity to the L2-relevant acoustic cue. Interestingly, when there are multiple L2 target contrasts, the cue-attention switching training improved learners' performance only for a certain target contrast if the training involved a subset of L1 categories on which the target contrasts are mapped. Additional work is necessary to pull apart the benefits of diverse combinations of training stimuli for the effectiveness of the cue-attention switching training to reach a broader range in L2 learning.

# REFERENCES

Ahn, S.-W. (2004). An Acoustic Study of English and Korean Vowels Revisited. *Korean Journal of Linguistics*, *29*(1), 45–62. Retrieved from https://www.kci.go.kr/kciportal/ci/sereArticleSearch/ciSereArtiView.kci?sereArticleSearchBean.artiId=ART001101387

Andruski, J. E., Blumstein, S. E., & Burton, M. (1994). The effect of subphonetic differences on lexical access. *Cognition*, *52*(3), 163–187.

Baker, W. & Trofimovich, P. (2005). Interaction of native-and second-language vowel system(s) in early and late bilinguals. *Language and Speech*, *48*(1), 1–27. https://doi.org/10.1177/00238309050480010101

Baker, W., Trofimovich, P., Mack, M., & Flege, J. E. (2002). The effect of perceived phonetic similarity on non-native sound learning by children and adults. *Proceedings of the Annual Boston University Conference on Language Development*, *26*(1), 36–47. Cascadilla Press.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Best, C. T. (1995). A direct realist view of cross-language speech perception. *Speech Perception and Linguistic Experience*, 171–206.

Best, C. T. & Tyler, M. D. (2007). *Nonnative and second-language speech perception: Commonalities and complementarities* (M. J. Munro & O.-S. Bohn, Eds.). Amsterdam: John Benjamins. Retrieved from https://www.academia.edu/48127265/Nonnative_and_second_language_speech_perception

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International*, *5*(9/10), 341–345. Retrieved from https://ci.nii.ac.jp/naid/10026090047/

Bohn, O.-S. (1995). Cross language speech perception in adults: First language transfer doesn't tell it all. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 279–304). Baltimore: York Press.

Bohn, O.-S. & Flege, J. E. (1990). Interlingual identification and the role of foreign language experience in L2 vowel perception. *Applied Psycholinguistics*, *11*(3), 303–328.

Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. (1999). Training Japanese listeners to identify English /r/and /l/: Long-term retention of learning in perception and production. *Perception & Psychophysics*, *61*(5), 977–985. https://doi.org/10.3758/BF03206911

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *The Journal of the Acoustical Society of America*, *101*(4), 2299–2310. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3507383/

Carney, A. E., Widin, G. P., & Viemeister, N. F. (1977). Noncategorical perception of stop consonants differing in VOT. *The Journal of the Acoustical Society of America*, *62*(4), 961–970.

Cheng, K. & Chen, X. (2020). The Effects of Intrinsic Acoustic Cues on Categorical Perception in Children with Cochlear Implants. *International Journal of English Linguistics*, *10*(5), 110–124.

Crystal, T. H. & House, A. S. (1988). The duration of American-English vowels: An overview. *Journal of Phonetics*, *16*(3), 263–284.

Darcy, I., Mora, J. C., & Daidone, D. (2016). The role of inhibitory control in second language phonological processing: Inhibitory control and L2 phonology. *Language Learning*, *66*(4), 741–773. https://doi.org/10.1111/lang.12161

Dong, H., Clayards, M., Brown, H., & Wonnacott, E. (2019). The effects of high versus low talker variability and individual aptitude on phonetic training of Mandarin lexical tones. *PeerJ*, *7*, e7191. https://doi.org/10.7717/peerj.7191

Escudero, P. (2000). *Developmental patterns in the adult L2 acquisition of new contrasts: The acoustic cue weighting in the perception of Scottish tense/lax vowels by Spanish speakers* (M.Sc. dissertation). University of Edinburgh, Edinburgh.

Escudero, P. & Boersma, P. (2004). Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition*, *26*(4), 515–585. https://doi.org/10.1017/S0272263104040021

Eychenne, J. & Jang, T.-Y. (2015). On the merger of Korean mid front vowels: Phonetic and phonological evidence. *Phonetics and Speech Sciences*, *7*(2), 119–129. https://doi.org/10.13064/KSSS.2015.7.2.119

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–272). Baltimore, MD: York Press.

Flege, J. E., Bohn, O.-S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, *25*(4), 437–470.

Francis, A. L., Baldwin, K., & Nusbaum, H. C. (2000). Effects of training on attention to acoustic cues. *Perception & Psychophysics*, *62*(8), 1668–1680. https://doi.org/10.3758/BF03212164

Francis, A. L., Kaganovich, N., & Driscoll-Huber, C. (2008). Cue-specific effects of categorization training on the relative weighting of acoustic cues to consonant voicing in English. *The Journal of the Acoustical Society of America*, *124*(2), 1234–1251.

Gelman, A., Su, Y.-S., Yajima, M., Hill, J., Pittau, M. G., Kerman, J., & Dorie, V. (2021). arm: Data Analysis Using Regression and Multilevel/Hierarchical Models (Version 1.12-2). Retrieved from https://CRAN.R-project.org/package=arm

Gerrits, E. & Schouten, M. E. H. (2004). Categorical perception depends on the discrimination task. *Perception & Psychophysics*, *66*(3), 363–376. https://doi.org/10.3758/BF03194885

Giannakopoulou, A., Brown, H., Clayards, M., & Wonnacott, E. (2017). High or low? Comparing high and low-variability phonetic training in adult and child second language learners. *PeerJ*, *5*, e3209.

Grenon, I., Kubota, M., & Sheppard, C. (2019). The creation of a new vowel category by adult learners after adaptive phonetic training. *Journal of Phonetics*, *72*, 17–34.

Guenther, F. H., Husain, F. T., Cohen, M. A., & Shinn-Cunningham, B. G. (1999). Effects of categorization and discrimination training on auditory perceptual space. *The Journal of the Acoustical Society of America*, *106*(5), 2900–2912. https://doi.org/10.1121/1.428112

Harmon, Z., Idemaru, K., & Kapatsinski, V. (2019). Learning mechanisms in cue reweighting. *Cognition*, *189*, 76–88. https://doi.org/10.1016/j.cognition.2019.03.011

Harnad, S. (2003). Categorical perception. In L. Nadel (Ed.), *Encyclopedia of Cognitive Science* (pp. 169–177). NY: Nature Publishing Group, Macmillan.

Hazan, V. & Barrett, S. (2000). The development of phonemic categorization in children aged 6–12. *Journal of Phonetics*, *28*(4), 377–396.

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *The Journal of the Acoustical Society of America*, *97*(5), 3099–3111.

Hillenbrand, J. M., Clark, M. J., & Houde, R. A. (2000). Some effects of duration on vowel recognition. *The Journal of the Acoustical Society of America*, *108*(6), 3013–3022.

Holliday, J. J. (2015). A longitudinal study of the second language acquisition of a three-way stop contrast. *Journal of Phonetics*, *50*, 1–14. https://doi.org/10.1016/j.wocn.2015.01.004

Holt, L. L. & Lotto, A. J. (2006). Cue weighting in auditory categorization: Implications for first and second language acquisition. *The Journal of the Acoustical Society of America*, *119*(5), 3059–3071. https://doi.org/10.1121/1.2188377

Hwang, H. & Moon, S.-J. (2005). Hankwukini palumhan hankwuke /에, 애/wa yenge /ɛ, æ/ moum (An acoustic comparative study of Korean /에, 애/ and English /ɛ, æ/ pronounced by Korean young male speakers). *Malsori*, *56*, 29–47.

Idemaru, K. & Holt, L. L. (2011). Word recognition reflects dimension-based statistical learning. *Journal of Experimental Psychology. Human Perception and Performance*, *37*(6), 1939–1956. https://doi.org/10.1037/a0025641

Idemaru, K. & Holt, L. L. (2014). Specificity of dimension-based statistical learning in word recognition. *Journal of Experimental Psychology. Human Perception and Performance*, *40*(3), 1009–1021. https://doi.org/10.1037/a0035269

Ingram, J. C. & Park, S.-G. (1997). Cross-language vowel perception and production by Japanese and Korean learners of English. *Journal of Phonetics*, *25*(3), 343–370.

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for nonnative phonemes. *Cognition*, *87*(1), B47–B57. https://doi.org/10.1016/S0010-0277(02)00198-1

Joanisse, M. F., Manis, F. R., Keating, P., & Seidenberg, M. S. (2000). Language deficits in dyslexic children: Speech perception, phonology, and morphology. *Journal of Experimental Child Psychology*, *77*(1), 30–60.

Kabakoff, H., Go, G., & Levi, S. V. (2020). Training a non-native vowel contrast with a distributional learning paradigm results in improved perception and production. *Journal of Phonetics*, *78*, 100940. https://doi.org/10.1016/j.wocn.2019.100940

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

Kang, Y., Yoon, T. J., & Han, S. (2015). Frequency effects on the vowel length contrast merger in Seoul Korean. *Laboratory Phonology*, *6*(3-4), 469-503.

Kapnoula, E. C., Winn, M. B., Kong, E. J., Edwards, J., & McMurray, B. (2017). Evaluating the sources and functions of gradiency in phoneme categorization: An individual differences approach. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(9), 1594.

Kawahara, H., Takahashi, T., Morise, M., & Banno, H. (2009). Development of exploratory research tools based on TANDEM-STRAIGHT. *Proceedings: APSIPA ASC 2009 : Asia-Pacific Signal and Information Processing Association, 2009 Annual Summit and Conference*, 111–120. Retrieved from https://eprints.lib.hokudai.ac.jp/dspace/handle/2115/39651

Kim, D., Clayards, M., & Goad, H. (2017). Individual differences in second language speech perception across tasks and contrasts: The case of English vowel contrasts by Korean learners. *Linguistics Vanguard*, *3*(1), 1–11.

Kim, D., Clayards, M., & Goad, H. (2018). A longitudinal study of individual differences in the acquisition of new vowel contrasts. *Journal of Phonetics*, *67*, 1–20. https://doi.org/10.1016/j.wocn.2017.11.003

Kim, D., Clayards, M., & Kong, E. J. (2020). Individual differences in perceptual adaptation to unfamiliar phonetic categories. *Journal of Phonetics*, *81*, 100984. https://doi.org/10.1016/j.wocn.2020.100984.

Kim, M. (2004, October 4). *Correlation between VOT and F0 in the perception of Korean stops and affricates*. 1–6. Jeju Island, Korea.

Klatt, D. H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *The Journal of the Acoustical Society of America*, *59*(5), 1208–1221. https://doi.org/10.1121/1.380986

Kluender, K. R., Diehl, R. L., & Wright, B. A. (1988). Vowel-length differences before voiced and voiceless consonants: An auditory explanation. *Journal of Phonetics*, *16*(2), 153–169. https://doi.org/10.1016/S0095-4470(19)30480-2

Kondaurova, M. V. & Francis, A. L. (2008). The relationship between native allophonic experience with vowel duration and perception of the English tense/lax vowel contrast by Spanish and Russian listeners. *The Journal of the Acoustical Society of America*, *124*(6), 3959–3971.

Kondaurova, M. V. & Francis, A. L. (2010). The role of selective attention in the acquisition of English tense and lax vowels by native Spanish listeners: Comparison of three training methods. *Journal of Phonetics*, *38*(4), 569–587.

Kong, E. J. (2019). Individual differences in categorical perception: L1 English learners' L2 perception of Korean stops. *Phonetics and Speech Sciences*, *11*(4), 63–70. https://doi.org/10.13064/KSSS.2019.11.4.063

Kong, E. J. & Edwards, J. (2011). Individual differences in speech perception: Evidence from visual analogue scaling and eye-tracking. *Proceedings of the 17th International Congress of Phonetic Sciences*, 1126–1129. Hong Kong, China.

Kong, E. J. & Edwards, J. (2015). Individual differences in L2 learners' perceptual cue weighting patterns. *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, U.K.

Kong, E. J. & Edwards, J. (2016). Individual differences in categorical perception of speech: Cue weighting and executive function. *Journal of Phonetics*, *59*, 40–57. https://doi.org/10.1016/j.wocn.2016.08.006

Kong, E. J. & Kang, S. (in press). Individual differences in categorical judgment of L2 stops: A link to proficiency and acoustic cue-weighting. *Language and Speech*.

Kong, E. J. & Kang, S. (2017). Individual variability in processing English source input into Korean stops. *Studies in Phonetics, Phonology, and Morphology*, *23*(2), 201–219. https://doi.org/10.17959/SPPM.2017.23.2.201

Ladefoged, P. & Johnson, K. (2014). *A course in phonetics*. Boston, MA, USA: Cengage learning.

Lee, H., Politzer-Ahles, S., & 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*(2), 117–132. https://doi.org/10.1016/j.wocn.2012.12.002

Lee, J.-Y. (2008). Perception of English high vowels: Duration as a cue by Korean speakers of English. *Kansas Working Papers in Linguistics*, *30*, 195–204. https://doi.org/10.17161/KWPL.1808.3915

Lee, J.-Y. (2009). *The effects of pronunciation instruction using duration manipulation on the acquisition of English vowel sounds by pre-service Korean EFL teachers* (Ph.D., University of Kansas). University of Kansas, United States -- Kansas. Retrieved from https://www.proquest.com/docview/304918596/abstract/8CBA6E1BFAB942F4PQ/1

Lee, S. & Cho, M.-H. (2018). Predicting L2 vowel identification accuracy from cross-language mappings between L2 English and L1 Korean. *Language Sciences*, *66*, 183–198. https://doi.org/10.1016/j.langsci.2017.09.006

Lemhöfer, K. & Broersma, M. (2012). Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods*, *44*(2), 325–343. https://doi.org/10.3758/s13428-011-0146-0

Lengeris, A. (2009). *Individual differences in second-language vowel learning* (Doctoral dissertation). UCL (University College London), London.

Lev-Ari, S. & Peperkamp, S. (2013). Low inhibitory skill leads to non-native perception and production in bilinguals' native language. *Journal of Phonetics*, *41*(5), 320–331. https://doi.org/10.1016/j.wocn.2013.06.002

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, *74*(6), 431.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, *54*(5), 358–368.

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

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English/r/and/l: A first report. *The Journal of the Acoustical Society of America*, *89*(2), 874–886.

Massaro, D. W. & Cohen, M. M. (1983). Phonological context in speech perception. *Perception & Psychophysics*, *34*(4), 338–348.

McAuliffe, M. (2017, February 23). How to create synthetic speech continua using STRAIGHT. Retrieved from https://memcauliffe.com/how-to-create-synthetic-speech-continua-using-straight.html

McClelland, J. L., Fiez, J. A., & McCandliss, B. D. (2002). Teaching the/r/–/l/discrimination to Japanese adults: Behavioral and neural aspects. *Physiology & Behavior*, *77*(4–5), 657–662.

McMurray, B., Aslin, R. N., Tanenhaus, M. K., Spivey, M. J., & Subik, D. (2008). Gradient sensitivity to within-category variation in words and syllables. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(6), 1609–1631. https://doi.org/10.1037/a0011747

McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition*, *86*(2), B33–B42.

Miller, J. L. (1994). On the internal structure of phonetic categories: A progress report. *Cognition*, *50*(1–3), 271–285.

Moon, S.-J. (2007). Hankwuke tanmouuy umsenghakcek kipanyenkwu (A fundamental phonetics investigation of Korean Monophthongs). *Malsori*, *62*, 1–17.

Nishi, K. & Kewley-Port, D. (2007). Training Japanese listeners to perceive American English vowels: Influence of training set. *Journal of Speech, Language, and Hearing Research : JSLHR*, *50*, 1496–1509. https://doi.org/10.1044/1092-4388(2007/103)

Nishi, K. & Kewley-Port, D. (2008). Non-native speech perception training using vowel subsets: Effects of vowels in sets and order of training. *Journal of Speech, Language, and Hearing Research : JSLHR*, *51*(6), 1480–1493. https://doi.org/10.1044/1092-4388(2008/07-0109)

Paradigm Stimulus Presentation. (2007). Perception Research Systems. Retrieved from Retrieved from http://www.paradigmexperiments.com

Perrachione, T. K., Lee, J., Ha, L. Y., & Wong, P. C. M. (2011). Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. *The Journal of the Acoustical Society of America*, *130*(1), 461–472. https://doi.org/10.1121/1.3593366

Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics*, *13*(2), 253–260.

Pisoni, D. B., Aslin, R. N., Perey, A. J., & Hennessy, B. L. (1982). Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants. *Journal of Experimental Psychology: Human Perception and Performance*, *8*(2), 297.

Pisoni, D. B. & Lazarus, J. H. (1974). Categorical and noncategorical modes of speech perception along the voicing continuum. *The Journal of the Acoustical Society of America*, *55*(2), 328–333.

Powell, M. J. (2009). The BOBYQA algorithm for bound constrained optimization without derivatives. *Cambridge NA Report NA2009/06, University of Cambridge, Cambridge*, 26–46.

R Core Team. (2020). *R Core Team R: a language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing.

Reetz, H., & Jongman, A. (2020). *Phonetics: Transcription, production, acoustics, and perception*. Hoboken, NJ: John Wiley & Sons.

Ryu, N.-Y. (2018). Korean vowel identification by English and Mandarin listeners: Effects of L1-L2 vowel inventory size and acoustic relationship. *Toronto Working Papers in Linguistics*, *40*. Retrieved from https://twpl.library.utoronto.ca/index.php/twpl/article/view/29209

Samuel, A. G. (1977). The effect of discrimination training on speech perception: Noncategorical perception. *Perception & Psychophysics*, *22*(4), 321–330.

Schertz, J., Cho, T., Lotto, A., & Warner, N. (2015). Individual differences in phonetic cue use in production and perception of a non-native sound contrast. *Journal of Phonetics*, *52*, 183–204. https://doi.org/10.1016/j.wocn.2015.07.003

Schertz, J., Cho, T., Lotto, A., & Warner, N. (2016). Individual differences in perceptual adaptability of foreign sound categories. *Attention, Perception, & Psychophysics*, *78*(1), 355–367. https://doi.org/10.3758/s13414-015-0987-1

Schmidt, A. M. (2007). Cross-language consonant identification. *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*, *17*, 185–200.

Schouten, B., Gerrits, E., & Van Hessen, A. (2003). The end of categorical perception as we know it. *Speech Communication*, *41*(1), 71–80.

Shin, J. (2015). Vowels and consonants. In L. Brown & J. Yeon (Eds.), *The Handbook of Korean Linguistics* (pp. 1–21). Chichester, U.K.: John Wiley & Sons, Inc.

Shin, J., Kiaer, J., & Cha, J. (2012). *The sounds of Korean*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139342858

Sohn, H.-M. (2001). *The Korean language*. Cambridge: Cambridge University Press.

Souza, H. K., Carlet, A., Jułkowska, I. A., & Rato, A. (2017). Vowel inventory size matters: Assessing cue-weighting in L2 vowel perception. *Ilha Do Desterro*, *70*, 33–46. https://doi.org/10.5007/2175-8026.2017v70n3p33

Stoet, G. (2010). PsyToolkit: A software package for programming psychological experiments using Linux. *Behavior Research Methods*, *42*(4), 1096–1104. https://doi.org/10.3758/BRM.42.4.1096

Stoet, G. (2017). PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology*, *44*(1), 24–31. https://doi.org/10.1177/0098628316677643

Strange, W. & Dittmann, S. (1984). Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English. *Perception & Psychophysics*, *36*(2), 131–145.

Studdert-Kennedy, M., Liberman, A. M., Harris, K. S., & Cooper, F. S. (1970). *Motor theory of speech perception: A reply to Lane's critical review.*

Toscano, J. C. & Lansing, C. R. (2019). Age-related changes in temporal and spectral cue weights in speech. *Language and Speech*, *62*(1), 61–79. https://doi.org/10.1177/0023830917737112

Tremblay, A., Broersma, M., & Coughlin, C. E. (2018). The functional weight of a prosodic cue in the native language predicts the learning of speech segmentation in a second language. *Bilingualism: Language and Cognition*, *21*(3), 640–652. https://doi.org/10.1017/S136672891700030X

Tsukada, K., Birdsong, D., Bialystok, E., Mack, M., Sung, H., & Flege, J. E. (2005). A developmental study of English vowel production and perception by native Korean adults and children. *Journal of Phonetics*, *33*(3), 263–290.

Utman, J. A., Blumstein, S. E., & Burton, M. W. (2000). Effects of subphonetic and syllable structure variation on word recognition. *Perception & Psychophysics*, *62*(6), 1297–1311.

Wonnacott, E., Brown, H., & Nation, K. (2017). Skewing the evidence: The effect of input structure on child and adult learning of lexically based patterns in an artificial language. *Journal of Memory and Language*, *95*, 36–48. https://doi.org/10.1016/j.jml.2017.01.005

Yamada, R. A. & Tohkura, Y. (1992). The effects of experimental variables on the perception of American English /r/ and /l/ by Japanese listeners. *Perception & Psychophysics*, *52*(4), 376–392. https://doi.org/10.3758/BF03206698

Yang, B. (1998). Vowel perception by formant variation. *The Journal of the Acoustical Society of America*, *103*(5), 3093–3093. https://doi.org/10.1121/1.422952

Yoon, T.-J. & Kang, Y. (2014). Monophthong analysis on a large-scale speech corpus of read-style Korean. *Phonetics and Speech Sciences*, *6*(3), 139–145.

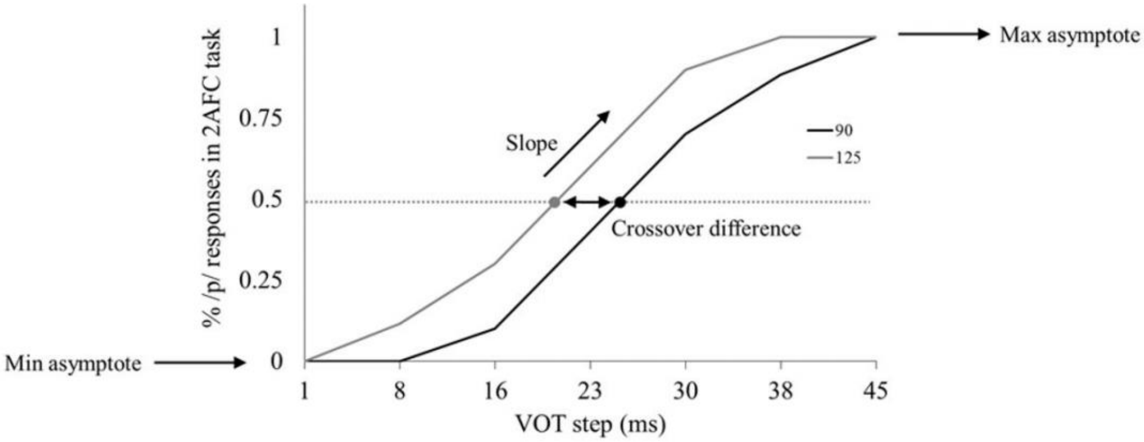
Zehr, J. & Schwarz, F. (2018). *PennController for Internet Based Experiments (IBEX)*. Retrieved from https://doi.org/10.17605/OSF.IO/MD832

Zhang, H., Chen, F., Yan, N., Wang, L., Shi, F., & Ng, M. L. (2016). The influence of language experience on the categorical perception of vowels: Evidence from Mandarin and Korean. *Proc. Interspeech 2016*, 873–877. San Francisco, USA.

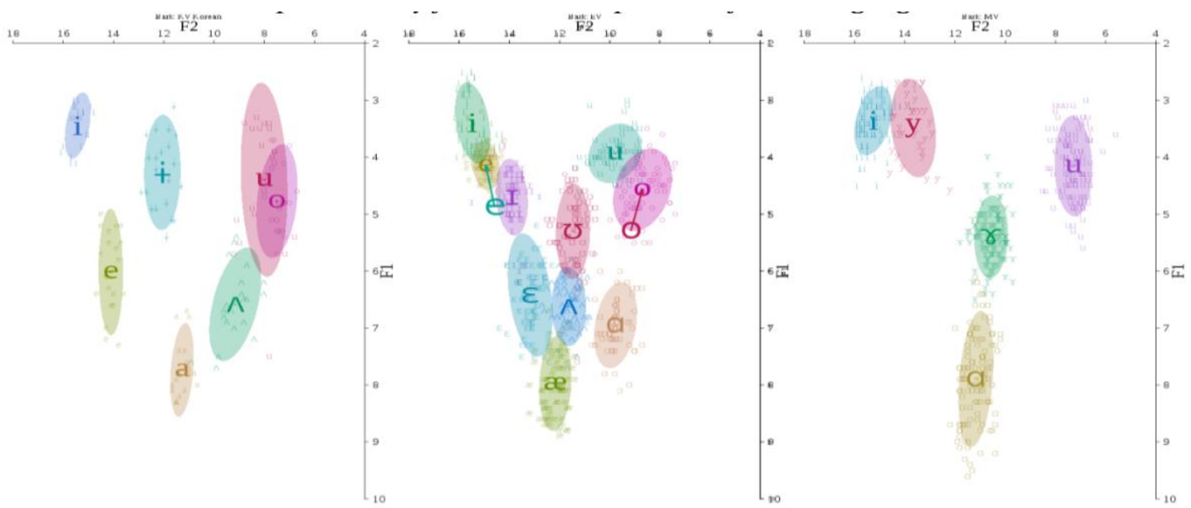
Zoom Video Communications Inc. (2020). *Security guide. Zoom Video Communications Inc.* Retrieved from https://zoom.us/docs/doc/ZoomSecurity-White-Paper.pdf

# APPENDICES

**Appendix A:** Hypothetical illustration of f0 cue use measured by the difference in 2AFC crossover points between low (90 *Hz*) and high (125 *Hz*) f0 steps (Kapnoula et al. 2017, p. 32).



**Appendix B:** First formant (F1) and second formant (F2) (bark) plots of Korean, English1 and Mandarin vowels produced by female native speakers of each language (Ryu, 2018, p. 4).



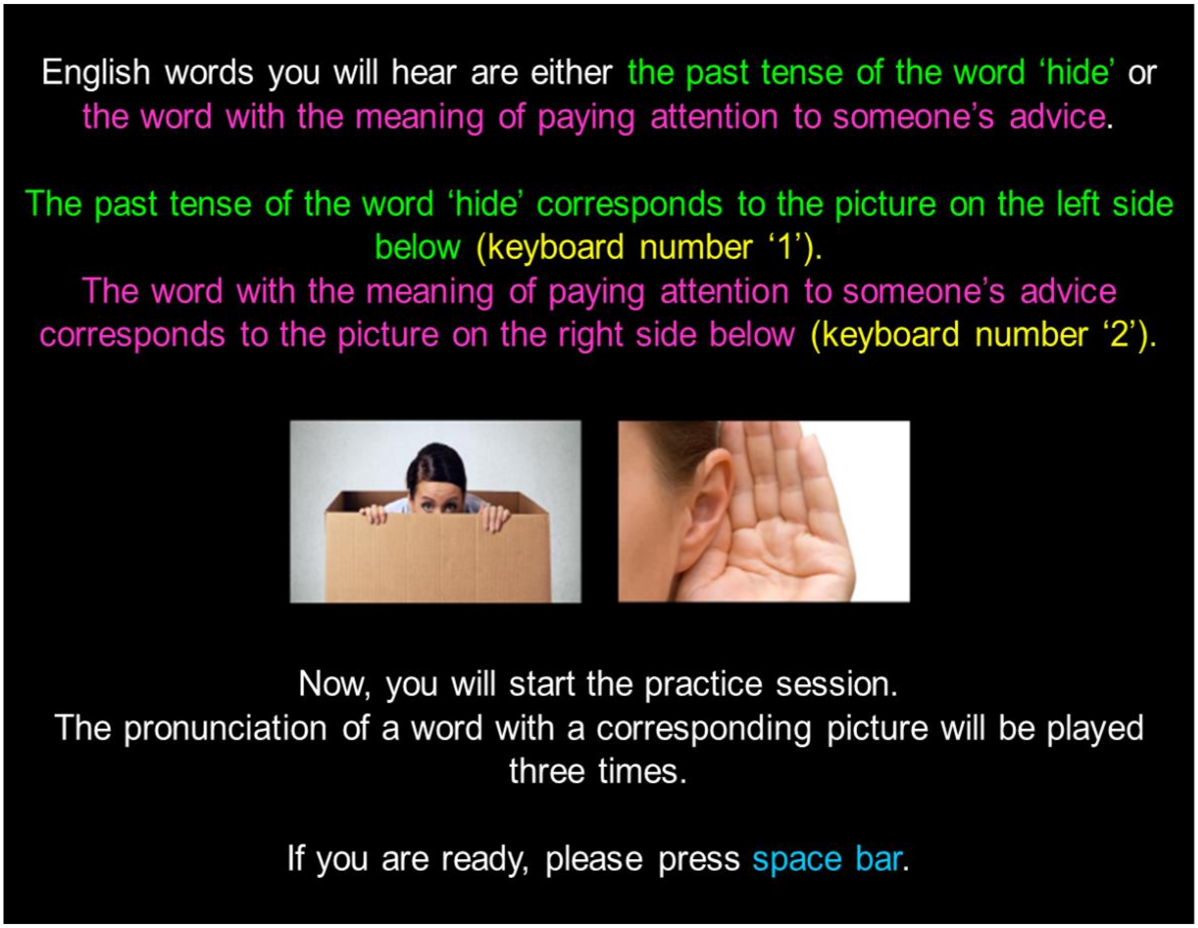
## **Appendix C:** Photographs associated with the target English words

1. hid/heed

2. head/had

3. hood/who’d

**Appendix D:** English translated version of Figure 42.



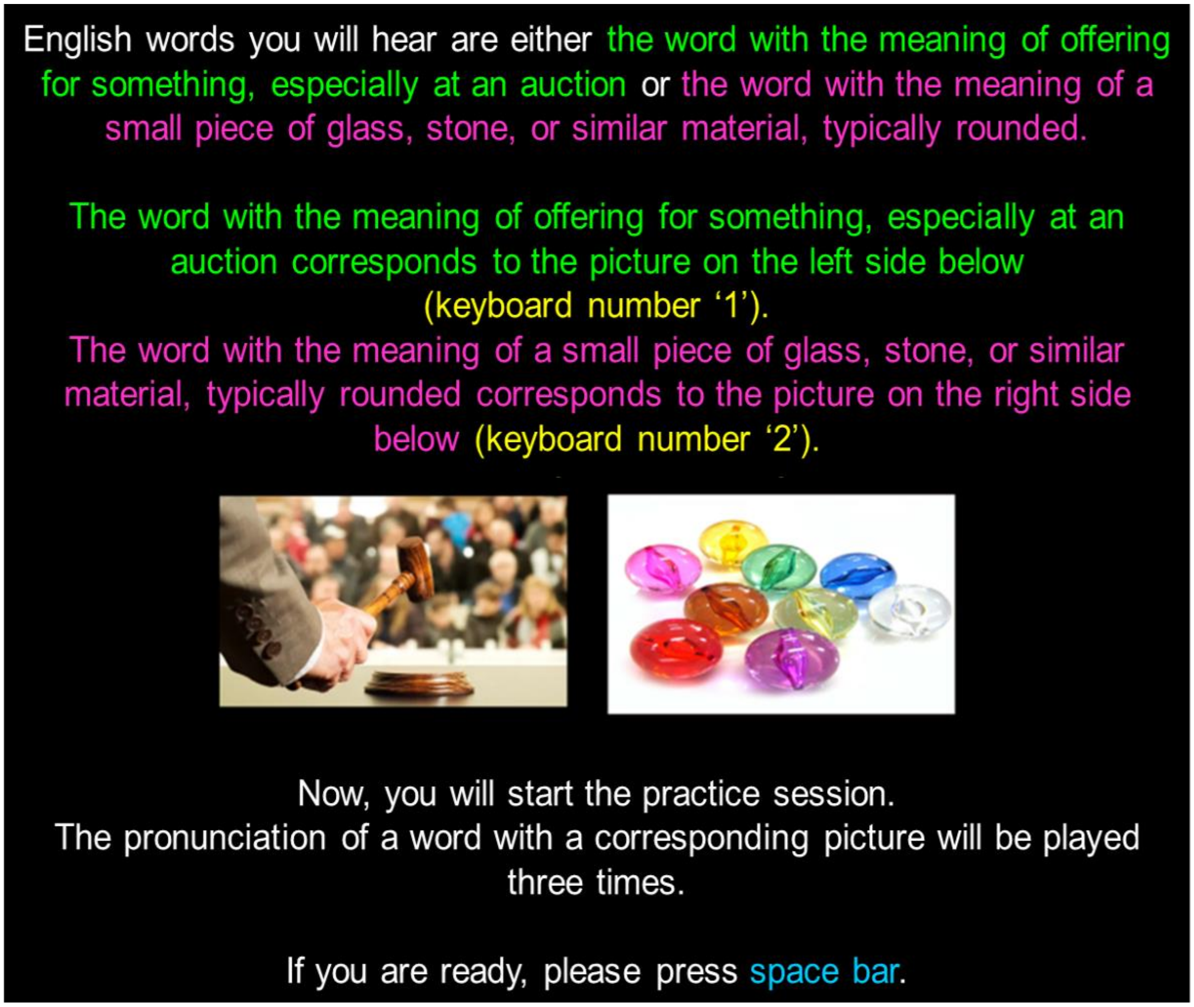
English words you will hear are either the past tense of the word ‘hide’ or the word with the meaning of paying attention to someone’s advice.

The past tense of the word ‘hide’ corresponds to the picture on the left side below (keyboard number ‘1’).
The word with the meaning of paying attention to someone’s advice corresponds to the picture on the right side below (keyboard number ‘2’).

Now, you will start the practice session.
The pronunciation of a word with a corresponding picture will be played three times.

If you are ready, please press space bar.

**Appendix E:** English translated version of Figure 44.



English words you will hear are either the word with the meaning of offering for something, especially at an auction or the word with the meaning of a small piece of glass, stone, or similar material, typically rounded.

The word with the meaning of offering for something, especially at an auction corresponds to the picture on the left side below (keyboard number ‘1’).

The word with the meaning of a small piece of glass, stone, or similar material, typically rounded corresponds to the picture on the right side below (keyboard number ‘2’).

Now, you will start the practice session.

The pronunciation of a word with a corresponding picture will be played three times.

If you are ready, please press space bar.

**Appendix F:** Results of everyday ID tests by group (LS 2 group vs. HS group).

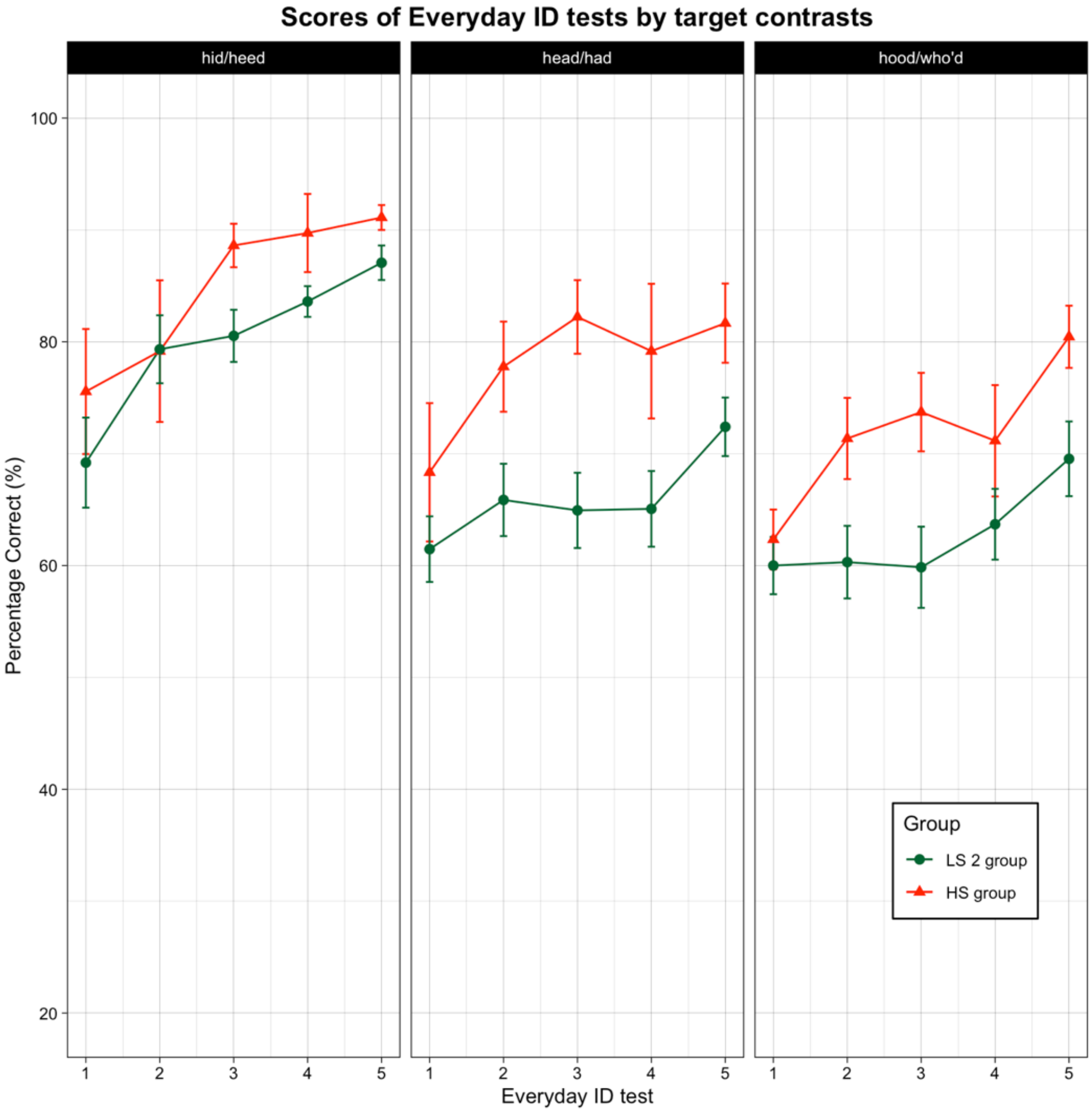


Figure 82. Results of everyday ID tests by group (LS 2 group vs. HS group) in percentage of correct (%).

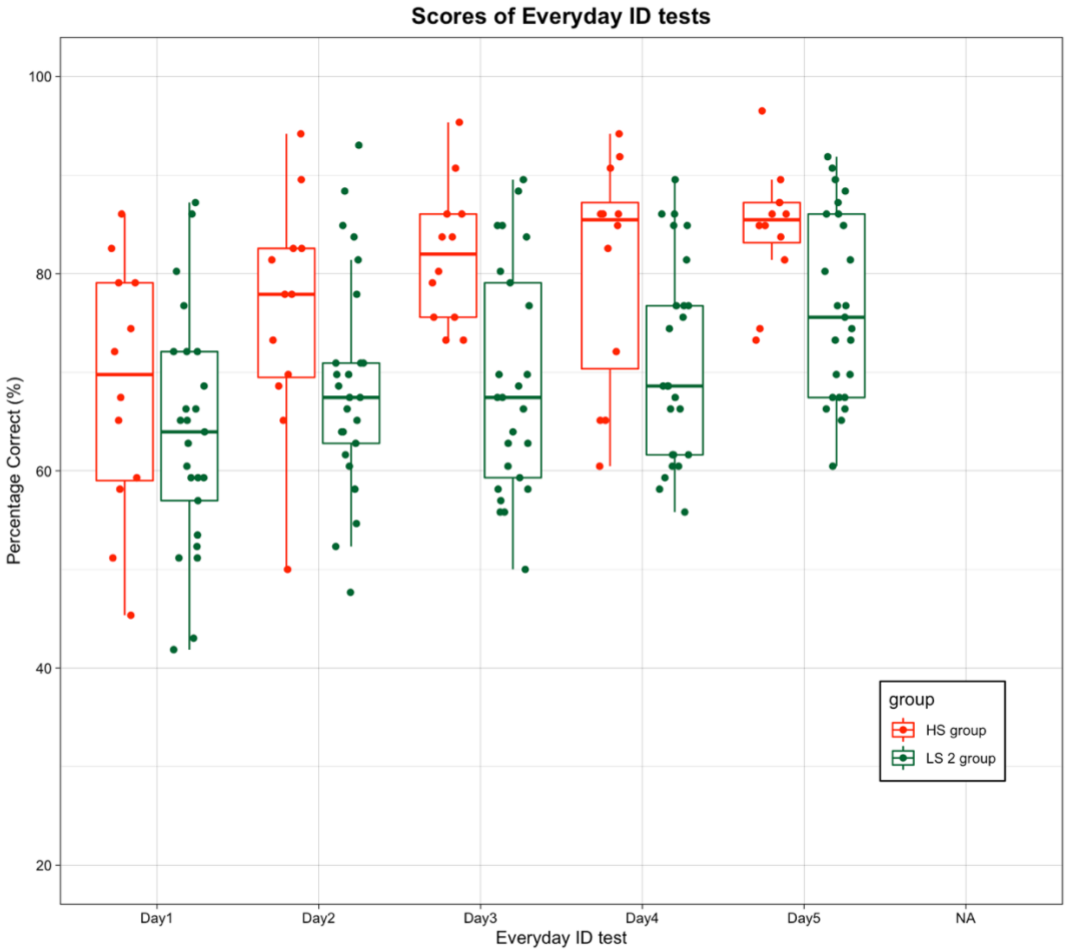


Figure 83. Results of everyday ID tests by three English vowel contrasts in percentage of correct (%). Each line represents either the LS 2 or the HS groups.

**Appendix G:** Results of the linear regression analysis, comparing the scores of everyday ID tests for the HS group and the LS 2 group.

Table 70. Summary of fixed effects for the mixed-effect linear regression model for everyday ID test scores.

| Predictor | Estimate ($\beta$) | Std. Error | *t-value* | *p-value* | Pr(>\|t\|) |
|---|---|---|---|---|---|
| (Intercept) | 72.57 | 1.47 | 49.32 | < .001 | *** |
| Group | -8.41 | 3.14 | -2.68 | 0.011 | * |
| Day1_VERSUS_Day2 | 6.00 | 1.36 | 4.43 | 0.000 | *** |
| Day1_VERSUS_Day3 | 7.86 | 1.24 | 6.34 | 0.000 | *** |
| Day1_VERSUS_Day4 | 8.93 | 1.42 | 6.28 | 0.000 | *** |
| Day1_VERSUS_Day5 | 14.02 | 1.24 | 11.32 | < .001 | *** |
| Group:Day1_VERSUS_Day2 | -2.59 | 2.90 | -0.89 | 0.374 | |
| Group:Day1_VERSUS_Day3 | -8.45 | 2.65 | -3.19 | 0.002 | ** |
| Group:Day1_VERSUS_Day4 | -4.72 | 3.04 | -1.55 | 0.125 | |
| Group:Day1_VERSUS_Day5 | -3.35 | 2.65 | -1.27 | 0.209 | |

Table 71. Summary of fixed effects for the mixed-effect linear regression model for everyday ID test scores for the /i/-/ɪ/ contrast.

| Predictor | Estimate ($\beta$) | Std. Error | *t*-value | *p*-value | |
|---|---|---|---|---|---|
| (Intercept) | 81.53 | 1.48 | 55.11 | < .001 | *** |
| Group | -4.89 | 3.16 | -1.55 | 0.13 | |
| Day1_VERSUS_Day2 | 8.02 | 2.82 | 2.84 | 0.01 | ** |
| Day1_VERSUS_Day3 | 11.89 | 2.35 | 5.07 | 0.00 | *** |
| Day1_VERSUS_Day4 | 14.32 | 2.35 | 6.10 | 0.00 | *** |
| Day1_VERSUS_Day5 | 17.12 | 2.35 | 7.29 | 0.00 | *** |
| Group:Day1_VERSUS_Day2 | 6.52 | 6.03 | 1.08 | 0.28 | |
| Group:Day1_VERSUS_Day3 | -1.72 | 5.02 | -0.34 | 0.73 | |
| Group:Day1_VERSUS_Day4 | 0.23 | 5.02 | 0.05 | 0.96 | |
| Group:Day1_VERSUS_Day5 | 2.31 | 5.02 | 0.46 | 0.65 | |

Table 72. Summary of fixed effects for the mixed-effect linear regression model for everyday ID test scores for the /ɛ/-/æ/ contrast.

| Predictor | Estimate ($\beta$) | Std. Error | *t-value* | *p-value* | |
|---|---|---|---|---|---|
| (Intercept) | 69.80 | 2.04 | 34.16 | < .001 | *** |
| Group | -11.89 | 4.37 | -2.72 | 0.01 | ** |
| Day1_VERSUS_Day2 | 6.04 | 2.18 | 2.77 | 0.01 | ** |
| Day1_VERSUS_Day3 | 6.85 | 2.18 | 3.14 | 0.00 | ** |
| Day1_VERSUS_Day4 | 5.95 | 2.68 | 2.22 | 0.03 | * |
| Day1_VERSUS_Day5 | 11.71 | 2.18 | 5.38 | 0.00 | *** |
| Group:Day1_VERSUS_Day2 | -5.04 | 4.65 | -1.08 | 0.28 | |
| Group:Day1_VERSUS_Day3 | -10.42 | 4.65 | -2.24 | 0.03 | * |
| Group:Day1_VERSUS_Day4 | -7.23 | 5.73 | -1.26 | 0.21 | |
| Group:Day1_VERSUS_Day5 | -2.40 | 4.65 | -0.52 | 0.61 | |

Table 73. Summary of fixed effects for the mixed-effect linear regression model for everyday ID test scores the /ʊ/-/u/ contrast.

| Predictor | Estimate ($\beta$) | Std. Error | *t-value* | *p-value* | |
|---|---|---|---|---|---|
| (Intercept) | 65.64 | 1.81 | 36.33 | < .001 | *** |
| Group | -9.13 | 3.86 | -2.36 | 0.02 | * |
| Day1_VERSUS_Day2 | 3.14 | 2.32 | 1.35 | 0.18 | |
| Day1_VERSUS_Day3 | 3.59 | 2.38 | 1.51 | 0.14 | |
| Day1_VERSUS_Day4 | 5.36 | 2.33 | 2.30 | 0.03 | * |
| Day1_VERSUS_Day5 | 12.32 | 2.25 | 5.47 | 0.00 | *** |
| Group:Day1_VERSUS_Day2 | -8.72 | 4.97 | -1.76 | 0.09 | . |
| Group:Day1_VERSUS_Day3 | -11.54 | 5.07 | -2.27 | 0.03 | * |
| Group:Day1_VERSUS_Day4 | -5.13 | 4.97 | -1.03 | 0.31 | |
| Group:Day1_VERSUS_Day5 | -8.58 | 4.81 | -1.78 | 0.08 | . |

## **Appendix H:** Distributions of the VAS task click locations of all participants in experiment 1 & 2.

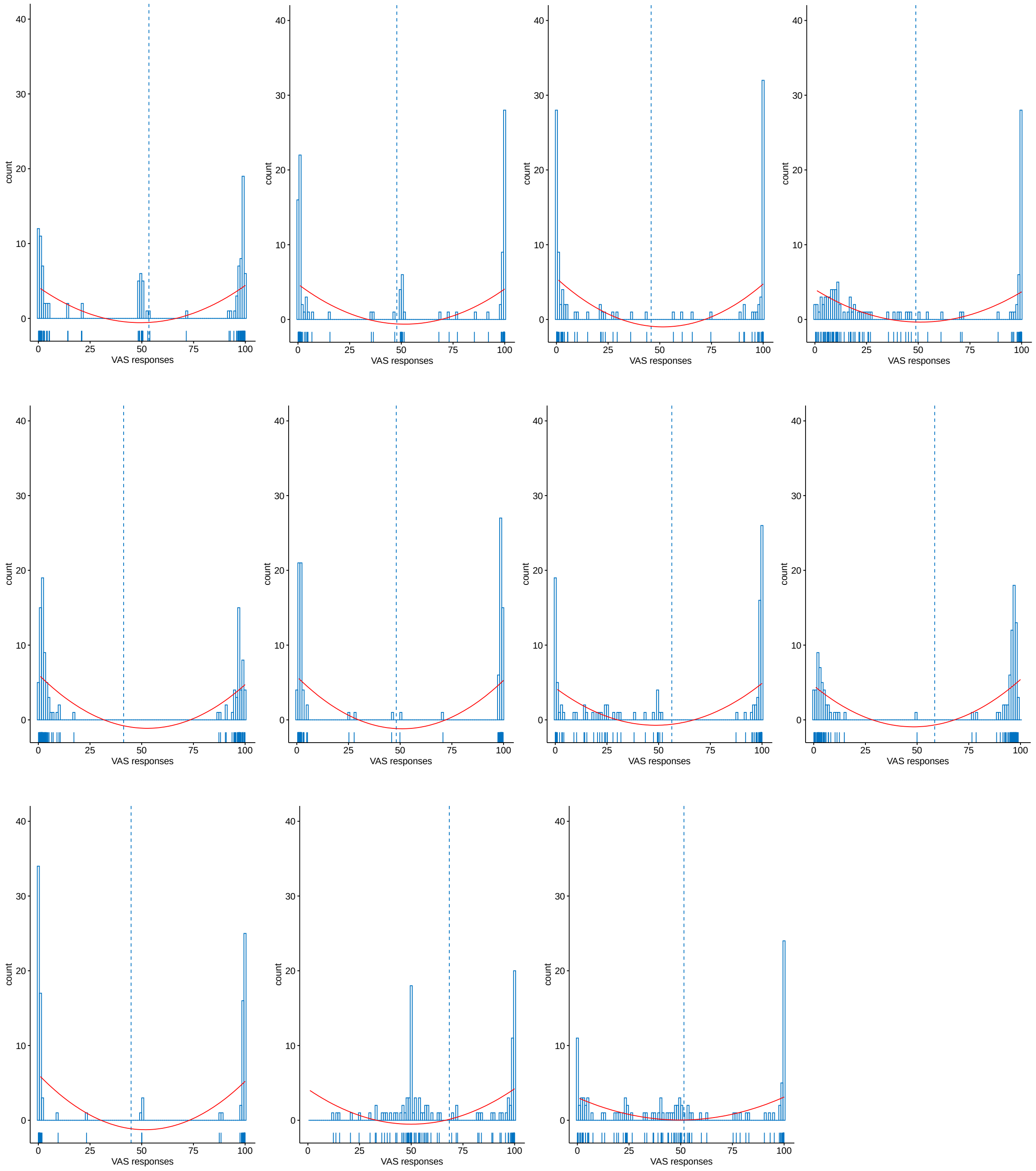


Figure 84. Distributions of the VAS task click locations of participants who showed categorical responses in experiment 1.

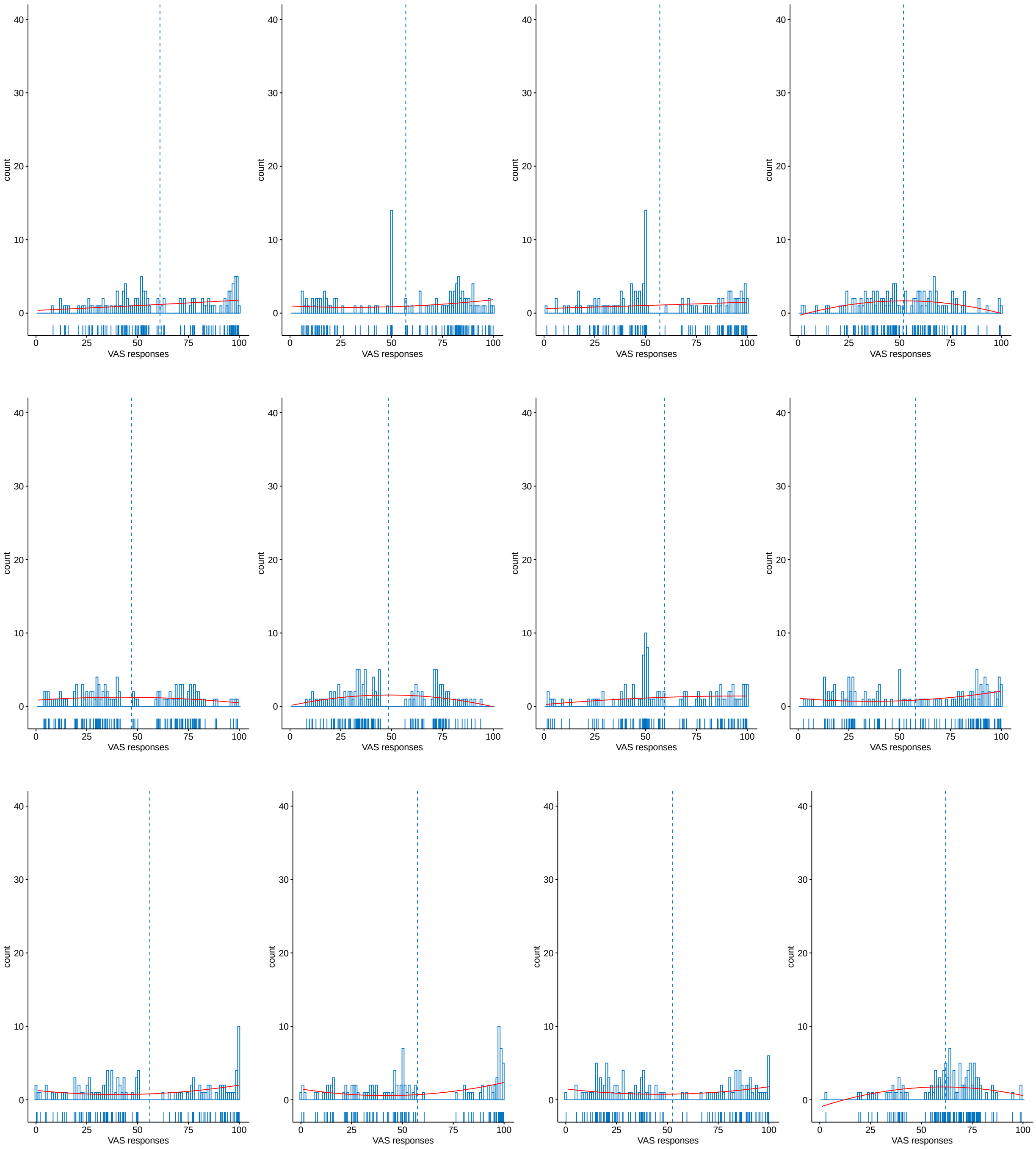


Figure 85. Distributions of the VAS task click locations of participants who showed gradient responses in experiment 1.

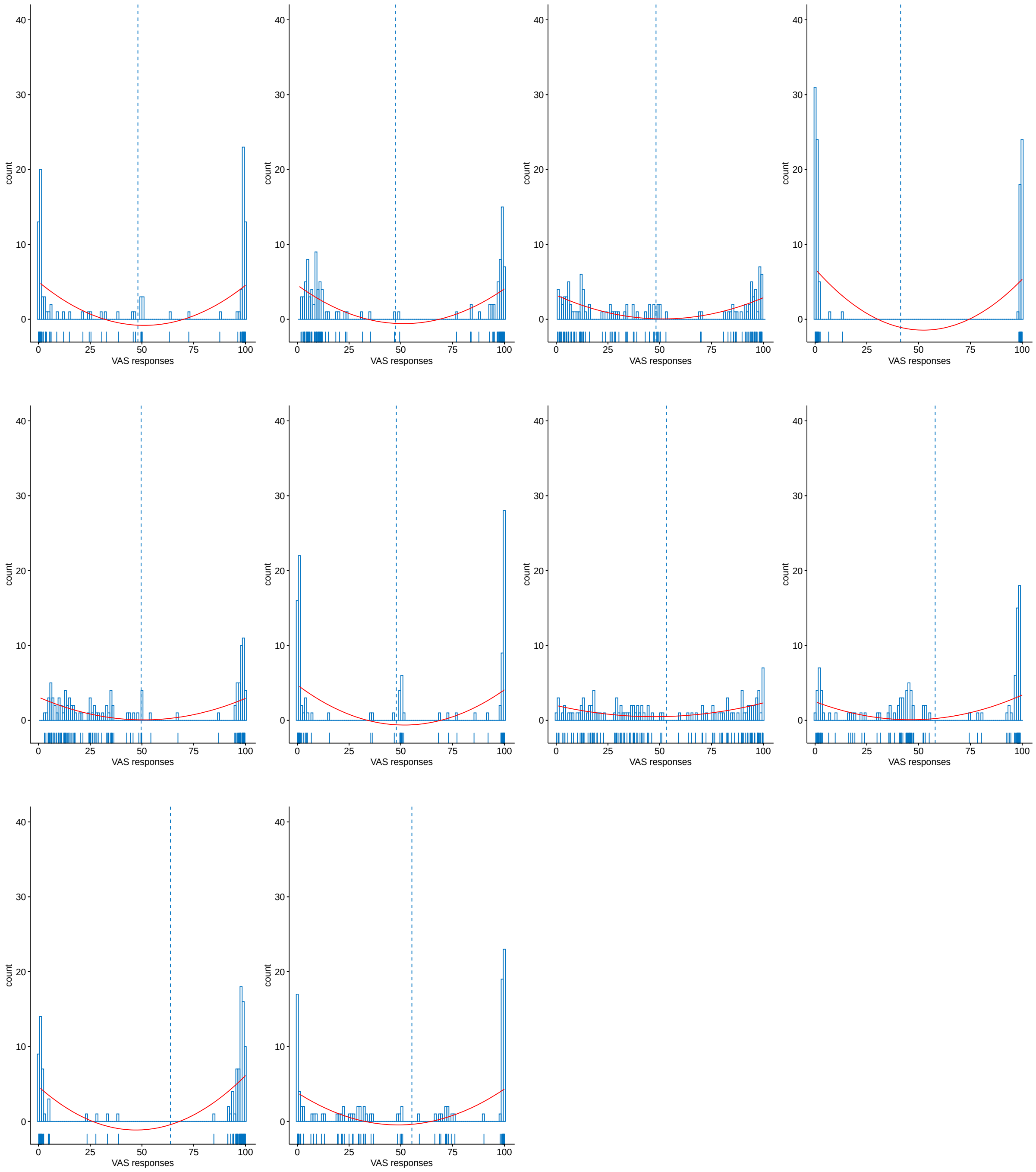


Figure 86. Distributions of the VAS task click locations of participants who showed categorical responses in experiment 2.

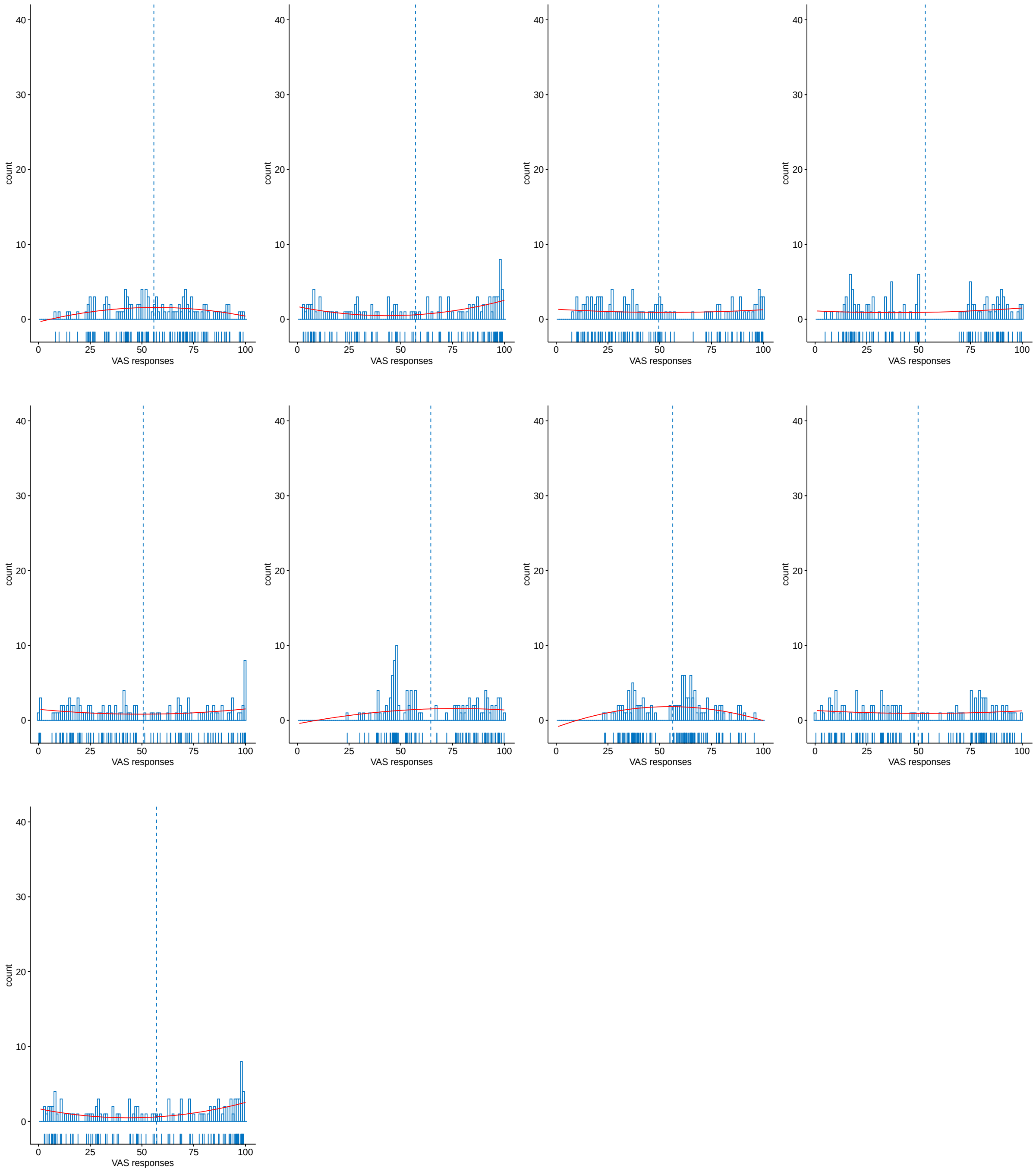


Figure 87. Distributions of the VAS task click locations of participants who showed gradient responses in experiment 2.

**Appendix I:** F0 and VOT cue weights of participants in experiment 1 & 2. Cue weights were measured with the results of the Day 3 ID test (excluding /p'/ responses).

Table 74. F0 and VOT cue weights of participants in experiment 1.

| Group | Participants | F0 weight | VOT weight |
|---|---|---|---|
| Categorical | C1_1 | 4.2617 | 6.8018 |
| Categorical | C1_2 | -0.0479 | 0.2706 |
| Categorical | C1_3 | -0.1788 | 1.0763 |
| Categorical | C1_4 | 0.6401 | 0.5140 |
| Categorical | C1_5 | 3.0914 | -0.5413 |
| Categorical | C1_6 | -0.1670 | -1.0025 |
| Categorical | C1_7 | 0.6956 | 0.0087 |
| Categorical | C1_8 | -0.4482 | 0.6127 |
| Categorical | C1_9 | 5.5037 | 0.7353 |
| Categorical | C1_11 | 1.8214 | 1.4643 |
| Categorical | C1_12 | -0.1726 | 2.1894 |
| Categorical | C1_13 | 2.2174 | 0.1464 |
| Gradient | G1_1 | 3.1093 | 2.1457 |
| Gradient | G1_2 | 3.8421 | -0.3038 |
| Gradient | G1_3 |  | 0.7680 |
| Gradient | G1_4 | 3.0088 | 1.8229 |
| Gradient | G1_5 | 5.0930 | 1.3569 |
| Gradient | G1_6 | 1.8223 | 1.7426 |
| Gradient | G1_7 | 2.0978 | 4.2665 |
| Gradient | G1_8 | 2.1357 | 3.3572 |
| Gradient | G1_9 | 8.2267 | 2.5759 |
| Gradient | G1_10 | 2.5401 | 1.1655 |
| Gradient | G1_11 | 5.5267 | 2.1822 |
| Gradient | G1_12 | 5.5572 | 1.0877 |

Table 75. F0 and VOT cue weights of participants in experiment 2.

| Group | Participants | F0 weight | VOT weight |
|---|---|---|---|
| Categorical | C2_1 | 4.3365 | 1.5482 |
| Categorical | C2_2 | 3.2696 | 0.8327 |
| Categorical | C2_3 | 5.9224 | 2.3034 |
| Categorical | C2_4 | 2.3957 | 2.6566 |
| Categorical | C2_5 | 1.6016 | 0.1432 |
| Categorical | C2_6 | 6.2077 | 1.5210 |
| Categorical | C2_7 | 6.2158 | 0.5964 |
| Categorical | C2_8 | 3.9586 | 0.4268 |
| Categorical | C2_9 | 2.4058 | 0.0367 |
| Categorical | C2_10 | 3.5046 | 0.0431 |
| Gradient | G2_1 | 5.7378 | -0.2907 |
| Gradient | G2_2 | 2.7359 | 0.3167 |
| Gradient | G2_3 | 3.2455 | 1.8026 |
| Gradient | G2_4 | 3.7101 | 2.2010 |
| Gradient | G2_5 | 1.1437 | -0.4634 |
| Gradient | G2_6 | 4.0610 | 0.5338 |
| Gradient | G2_7 | 1.9914 | 2.7762 |
| Gradient | G2_8 | 1.8523 | 0.9895 |
| Gradient | G2_9 | 4.0815 | 0.2690 |

**Appendix J:** Duration and spectral cue weights of participants in experiment 3 & 4. Cue weights were measured with the results of the post-test with old talker and new talker stimuli.

Table 76. Duration and spectral cue weights of participants in experiment 3 with old talker stimuli.

| Group | Participants | i_duration weight | i_spec weight | e_dur weight | e_spec weight | u_duration weight | u_spec weight |
|---|---|---|---|---|---|---|---|
| LS | LS_1 | 0.8300 | 1.5421 | 0.0000 | 9.5910 | 0.5646 | 1.4186 |
| LS | LS_2 | 0.0000 | 5.0500 | 1.5200 | 1.6840 | 1.4759 | -0.0801 |
| LS | LS_3 | 1.4900 | 7.3119 | 2.4700 | 0.2091 | 3.3837 | 1.8492 |
| LS | LS_4 | -1.3100 | 19.0582 | 0.7540 | 1.4808 | -3.3344 | 3.8713 |
| LS | LS_5 | 1.4000 | 4.4371 | 0.3850 | 4.1062 | 1.8970 | 1.8970 |
| LS | LS_6 | 0.5590 | 12.5903 | 0.5210 | 1.9821 | 1.2157 | 2.4342 |
| LS | LS_7 | 0.5640 | -1.4895 | 1.8600 | -0.2540 | 2.0561 | 0.9860 |
| LS | LS_8 | 1.3400 | 5.5628 | 0.9010 | 2.6825 | 1.4334 | 2.2879 |
| LS | LS_9 | 0.0000 | 7.1385 | 1.9900 | 0.3487 | -0.1445 | 4.7496 |
| LS | LS_10 | -0.5460 | 12.3799 | -1.2900 | 0.3110 | 3.5926 | 0.5914 |
| LS | LS_11 | -0.5680 | 17.4161 | 2.1900 | 10.6344 | -0.2656 | 4.2475 |
| LS | LS_12 | 1.0800 | -2.0010 | 1.1700 | 0.7128 | 0.9081 | 2.1546 |
| HS | HS_1 | 0.9090 | 13.2541 | 0.7320 | 6.3885 | -0.6156 | 7.3824 |
| HS | HS_2 | 0.3260 | 5.6019 | 0.3520 | 2.0808 | 0.3275 | 3.2278 |
| HS | HS_3 | -2.0900 | 7.8712 | 0.5420 | 4.3702 | -0.1340 | 4.3750 |
| HS | HS_4 | -2.3100 | 8.7347 | 1.1100 | 1.6668 | 1.1405 | 1.8706 |
| HS | HS_5 | -0.9090 | 13.2541 | -0.6220 | 3.5126 | 1.4390 | 6.0230 |
| HS | HS_6 | -0.6390 | 13.7159 | -0.4260 | 15.8487 | 0.8620 | 7.7089 |
| HS | HS_7 | -0.2150 | 7.7989 | 2.9700 | 14.0316 | 1.6424 | 5.2973 |
| HS | HS_8 | -1.3100 | 6.3756 | 0.0000 | 4.7451 | 0.7934 | 3.9857 |
| HS | HS_9 | -0.2690 | 12.2851 | 1.8500 | 0.7859 | 0.6939 | 2.6572 |
| HS | HS_10 | 0.8840 | 10.5749 | 0.9860 | 4.5622 | 2.6245 | 3.7584 |
| HS | HS_11 | -1.2100 | 8.9431 | -1.3200 | 1.3226 | 2.8162 | 1.1749 |
| HS | HS_12 | -4.0700 | 10.1143 | 1.2600 | 3.8562 | 3.2872 | 4.0558 |

Table 77. Duration and spectral cue weights of participants in experiment 3 with new talker stimuli.

| Group | Participants | i_duration weight | i_spec weight | e_dur weight | e_spec weight | u_duration weight | u_spec weight |
|---|---|---|---|---|---|---|---|
| LS | LS_1 | 1.2900 | 0.1640 | 0.2278 | 8.2300 | 1.1302 | 0.6980 |
| LS | LS_2 | -0.9000 | 3.8371 | 1.0476 | 0.3790 | 1.1338 | 1.1338 |
| LS | LS_3 | 0.9740 | 2.9543 | 3.1966 | 1.1300 | 0.0941 | 1.4142 |
| LS | LS_4 | -3.1100 | 5.5232 | 0.8962 | -0.3770 | -2.1109 | 1.0361 |
| LS | LS_5 | 1.4900 | 1.1731 | 0.8648 | 2.5500 | 0.9724 | 4.4157 |
| LS | LS_6 | 0.3940 | 7.1199 | 0.1994 | 2.7400 | 0.2676 | 2.0421 |
| LS | LS_7 | 2.4300 | -1.0560 | 1.4073 | -0.1600 | 1.2169 | 0.3103 |
| LS | LS_8 | 1.0500 | 2.8844 | 0.5397 | 1.1400 | -0.0784 | 1.3718 |
| LS | LS_9 | 0.5880 | 13.0247 | 3.8443 | 0.0000 | -0.2809 | 4.5828 |
| LS | LS_10 | 1.0900 | 14.0835 | -1.6734 | -4.7800 | 2.0736 | 0.3728 |
| LS | LS_11 | 0.6620 | 7.9293 | 0.3202 | 2.9700 | 0.2416 | 3.7012 |
| LS | LS_12 | 1.6100 | -0.4148 | 0.9574 | 0.3210 | 0.0851 | 1.8689 |
| HS | HS_1 | 0.0000 | 12.0758 | 0.9879 | 2.0600 | -0.1621 | 5.5442 |
| HS | HS_2 | 0.0000 | 6.7996 | 0.0849 | 1.7900 | 1.2674 | 0.9601 |
| HS | HS_3 | 0.1960 | 6.9474 | 1.3535 | 1.1100 | 0.2161 | 3.1909 |
| HS | HS_4 | -1.1400 | 4.5637 | 0.1935 | -2.3200 | 0.8920 | 1.4329 |
| HS | HS_5 | 0.7560 | 4.9221 | 0.6918 | 2.3700 | -0.6434 | 4.0214 |
| HS | HS_6 | -0.7640 | 9.0672 | 0.9963 | 14.0000 | 1.8649 | 9.5404 |
| HS | HS_7 | 0.1830 | 6.3701 | -0.4698 | 0.0000 | -0.5413 | 4.0710 |
| HS | HS_8 | -0.2700 | 4.1949 | 1.4871 | 0.5920 | 2.0421 | 0.2676 |
| HS | HS_9 | -2.0200 | 6.6758 | 2.4990 | 0.7680 | 0.2498 | 1.7551 |
| HS | HS_10 | 0.2030 | 7.0746 | 1.0007 | 6.9800 | 4.0081 | 5.8449 |
| HS | HS_11 | -0.6030 | 7.1023 | -1.8743 | -0.5120 | 1.9215 | 1.3452 |
| HS | HS_12 | -1.8200 | 5.1904 | 0.6476 | 1.5000 | 3.0375 | 1.3427 |

Table 78. Duration and spectral cue weights of participants in experiment 4 with old talker stimuli.

| Group | Participants | i_duration weight | i_spec weight | e_dur weight | e_spec weight | u_duration weight | u_spec weight |
|---|---|---|---|---|---|---|---|
| LS 2 | LS 2_1 | 0.3660 | 1.8776 | -0.5620 | -0.2412 | 0.8820 | 1.0381 |
| LS 2 | LS 2_2 | 0.0000 | 12.8568 | -0.2740 | 2.2315 | 1.4700 | 1.7292 |
| LS 2 | LS 2_3 | -0.8610 | 12.7769 | 0.0745 | 0.8840 | 0.1080 | 2.9412 |
| LS 2 | LS 2_4 | 1.1100 | 3.1409 | 0.7260 | -0.2925 | 1.7400 | 1.5002 |
| LS 2 | LS 2_5 | -0.2690 | 12.2850 | -0.3350 | 1.7662 | 0.0000 | 5.6105 |
| LS 2 | LS 2_6 | 0.0000 | 13.4931 | -0.1760 | 2.0136 | 1.1600 | 3.4196 |
| LS 2 | LS 2_7 | 0.2980 | -0.8844 | -0.0725 | 0.4336 | -0.7360 | 1.3679 |
| LS 2 | LS 2_8 | 0.0000 | 12.9885 | -0.9410 | -1.0172 | 1.4300 | 2.6861 |
| LS 2 | LS 2_9 | 0.2230 | 7.9818 | -0.7600 | 1.2553 | 0.6220 | 3.5854 |
| LS 2 | LS 2_10 | 0.8040 | 7.0244 | 0.0000 | 0.0708 | 1.0800 | 0.8560 |
| LS 2 | LS 2_11 | -1.3500 | 14.0709 | 0.6510 | 4.1579 | -0.1760 | 6.1015 |
| LS 2 | LS 2_12 | 1.4400 | 2.5524 | 1.2100 | 0.3080 | 1.5400 | 0.8304 |
| LS 2 | LS 2_13 | -0.5800 | 18.0433 | -0.7360 | 2.2542 | 0.0000 | 8.7799 |
| LS 2 | LS 2_14 | -0.2230 | 8.4944 | 4.6700 | 2.2559 | 3.4400 | -0.1135 |
| LS 2 | LS 2_15 | 0.2780 | 4.5609 | -1.1200 | 2.1431 | 0.0000 | 12.2543 |
| LS 2 | LS 2_16 | 0.4190 | 7.4534 | -0.4150 | 1.6811 | 0.9120 | 3.3697 |
| LS 2 | LS 2_17 | -0.2830 | 10.3257 | -0.3740 | 0.8169 | 1.4800 | 0.4001 |
| LS 2 | LS 2_18 | 0.2890 | 12.8976 | 1.1900 | 1.4146 | 2.1600 | -0.7497 |
| LS 2 | LS 2_19 | -0.4120 | 4.3216 | -0.5580 | 0.0929 | -0.1180 | 3.3665 |
| LS 2 | LS 2_20 | 0.0000 | 2.4238 | -0.3620 | -0.4344 | -0.1520 | 1.0515 |
| LS 2 | LS 2_21 | -1.0300 | 5.8638 | -1.6600 | -0.0928 | 1.1300 | 1.2858 |
| LS 2 | LS 2_22 | 0.1840 | 6.4529 | -3.6700 | 5.2265 | -2.7300 | 1.1179 |
| LS 2 | LS 2_23 | 0.5190 | 9.6061 | 1.7800 | -0.2534 | 2.8400 | 1.4761 |
| LS 2 | LS 2_24 | -0.8610 | 12.7769 | 0.1470 | 0.8033 | -0.8190 | 5.5320 |
| LS 2 | LS 2_25 | 1.1100 | 3.1409 | 1.3000 | -0.2349 | 1.8600 | 1.9464 |

Table 79. Duration and spectral cue weights of participants in experiment 4 with new talker stimuli.

| Group | Participants | i_duration weight | i_spec weight | e_dur weight | e_spec weight | u_duration weight | u_spec weight |
|---|---|---|---|---|---|---|---|
| LS 2 | LS 2_1 | 0.5260 | 0.4514 | 0.2560 | 0.5130 | 0.4310 | -0.3590 |
| LS 2 | LS 2_2 | 1.5500 | 7.7377 | 0.7850 | 2.4300 | 1.5900 | 1.5900 |
| LS 2 | LS 2_3 | 0.0000 | 4.5664 | -0.2310 | 1.2100 | -1.3300 | 1.3300 |
| LS 2 | LS 2_4 | 0.8260 | 1.4583 | -0.3350 | 1.7700 | 2.4300 | 2.1600 |
| LS 2 | LS 2_5 | -0.2010 | 7.2815 | 0.9570 | 1.2600 | 1.1300 | 4.4800 |
| LS 2 | LS 2_6 | 0.1450 | 4.8293 | 0.4010 | 1.4800 | 1.0800 | 5.0700 |
| LS 2 | LS 2_7 | 0.0742 | -0.5910 | 0.6580 | 0.3670 | -0.2950 | -0.5890 |
| LS 2 | LS 2_8 | -0.4550 | 4.9072 | -0.4550 | -0.9740 | 0.0764 | 0.5330 |
| LS 2 | LS 2_9 | 0.2730 | 4.3030 | -0.5580 | 1.9500 | 1.6300 | 5.0100 |
| LS 2 | LS 2_10 | -0.2890 | 2.5176 | 2.0900 | 1.0900 | 0.5580 | 2.9200 |
| LS 2 | LS 2_11 | -0.4600 | 8.5126 | 1.4000 | 3.9700 | 0.4750 | 2.4700 |
| LS 2 | LS 2_12 | 3.0600 | 0.4379 | 2.3300 | 0.9470 | 4.0000 | 1.5700 |
| LS 2 | LS 2_13 | 1.3200 | 4.1272 | 0.7670 | 0.9910 | 0.0000 | 5.8400 |
| LS 2 | LS 2_14 | 1.0300 | 11.8263 | 2.4400 | 0.9630 | 2.1300 | 0.5560 |
| LS 2 | LS 2_15 | 0.6290 | 13.1342 | 0.2990 | -0.5230 | 0.6090 | 4.8100 |
| LS 2 | LS 2_16 | 0.5000 | 3.7413 | -1.7000 | 1.7000 | 0.7290 | 0.0735 |
| LS 2 | LS 2_17 | 0.1960 | 2.0322 | 0.5980 | -0.4490 | 1.0600 | -0.5730 |
| LS 2 | LS 2_18 | 0.0000 | 6.9356 | 1.4700 | 0.3100 | 2.7500 | 0.6860 |
| LS 2 | LS 2_19 | 0.2150 | 7.2880 | -1.0400 | -2.8300 | 0.0000 | 8.6900 |
| LS 2 | LS 2_20 | 0.6230 | 1.2256 | 0.7250 | -0.1460 | -0.2130 | -0.1420 |
| LS 2 | LS 2_21 | 0.4750 | 4.8004 | -0.5160 | 0.1480 | 1.1000 | 0.8470 |
| LS 2 | LS 2_22 | 1.5600 | 14.7623 | -2.9200 | 3.5100 | -1.4900 | 2.8000 |
| LS 2 | LS 2_23 | -2.1700 | 12.2897 | 8.2200 | 0.0000 | 8.7300 | 2.3100 |
| LS 2 | LS 2_24 | 0.2830 | 4.6670 | 0.7290 | 0.2440 | -0.2280 | 3.3500 |
| LS 2 | LS 2_25 | 0.0000 | 3.0143 | 0.5390 | 1.0800 | 2.0200 | 0.6320 |