Prosodic Effects on Spoken Word Recognition in Second Language: Processing of Lexical Stress by Korean-speaking Learners of English

DISSERTATION

Presented in Partial Fulfillment of the Requirements for
the Degree Doctor of Philosophy in the
Graduate School of The Ohio State University

By

Jeonghwa Shin, M.A.

Graduate Program in Linguistics
The Ohio State University

2012

Dissertation Committee:
Prof. Shari R. Speer, Advisor
Prof. Mary E. Beckman
Prof. Cynthia G. Clopper
ABSTRACT

Prosody, the melody and rhythm of speech, is known to influence how listeners interpret the sequence of sounds, syllables, and higher order organizational units and thus how lexical access proceeds during spoken word recognition. The present study explores how language-specific variation in prosodic structure affects L2 learners’ processing of prosodic categories during spoken word recognition. Specifically, the study examines how Korean-speaking learners of English use English lexical stress during spoken word recognition.

English employs stress reliably at the lexical level. Utterances often put pitch accents on lexically stressed syllables and most lexically unstressed syllables contain reduced vowels. In contrast, Korean as a first language (L1) has neither lexical-level stress, nor reduced vowels, but instead has phrase-level prosodic structure ((T)HLH %) (Jun, 1993). Utterances often align the melody of the accentual phrase - a sequence of phrase-initial tone before a high tone - with the beginning of each content word. The phrase initial tone, T, is realized as either high (H) or low (L) tone depending on the phonation type of the initial consonant (tense/aspirated vs. lax). H is realized with a relatively high fundamental frequency (f0) and L with a relatively low f0. The Perceptual Assimilation Model (PAM; Best, 1995) predicts that if the relevant L2 sounds have no corresponding categories in L1, as is the case with the English lexical stress contrast for Korean learners of English, discrimination of the two sounds should be good because the discrimination is less strongly affected by native phonological equivalence classes. However, this prediction was not borne out in
a preliminary eyetracking study (Shin, 2008). In Shin (2008), native English speakers, but not Korean-speaking learners of English were able to use lexical stress information during English spoken word recognition. Post-hoc analyses showed that Korean learners of English were good at identifying H-pitch accented trochaic words, but were poor at identifying iambic words that began with what they heard as aspirated sounds, suggesting that Korean phrase level prosodic structure might have constrained processing of English lexical stress in Korean learners of English.

In the present study, an eyetracking experiment examined the role of both word-initial phonation type and vowel reduction in unstressed syllables in the processing of English lexical stress during spoken word recognition. Results showed that both native English speakers and Korean learners of English used the reduced vowel cue to identify target words, initiating lexical activation during the initial syllable of words. Still, Korean L2 learners had relatively delayed processing for iambic word recognition. The effect of aspirated vs. lax word-initial segment on the processing of lexical stress during word recognition was not supported statistically. However, we could observe that aspirated initial sounds particularly facilitated H-pitch accented trochaic word recognition in Korean L2 learners of English. A cross-modal gating experiment further investigated this processing difficulty by focusing on the effects of phonation type on the perception of English lexical stress without the reduced vowel cue. This gating experiment showed very similar results to the preliminary eyetracking experiment, suggesting word-initial phonation type influences the perception of English lexical stress in Korean L2 learners of English, but that iambic stress poses general processing difficulty to Korean speakers. These findings suggest that L1 experience with both lexical processing strategy and relevant acoustic phonetic cues influences the perception and processing of L2 prosodic units during spoken word recognition.
To my parents, who are always praying for me.
ACKNOWLEDGMENTS

During my doctoral years at OSU Linguistics, I had many people around who helped me to complete the Ph.D dissertation from various sides. I first want to express my gratitude to my advisor, Dr. Shari Speer. She has supported and guided me with endless advice, read numerous drafts, and helped me through to the end. I also want to send my special thanks to my committee members, Dr. Mary Beckman, for sharing her insights, knowledge, and enthusiasm in linguistics which gave me a big inspiration in the study of linguistics and Dr. Cynthia Clopper, for giving invaluable comments and stimulating discussions, and for believing in me and my work in the rough times. I’m grateful to Dr. Dan Mirman, Dr. Florian Jaeger, and Alex Fine who helped me with their time and expertise in statistical approaches to eyetracking data analysis. Thanks to Dr. Mark Pitt for providing helpful instructions about nonword construction. Dr. Ann Bradlow and Dr. Ewa Jacewicz were also of great help with clear and fast responses to my questions about vowel space measurement. Dr. Kiwako Ito introduced me to eyetracking methodology and gave helpful comments on the eyetracking data analysis. Ping Bai wrote many perl scripts to convert raw gaze data into analyzable data set. I do also thank Dr. JJ Nakayama who was pleasantly on the reading committee for my doctoral candidacy exam. Liz McCullough, Sarah Bibyk, and Laurie Maynell gave generously of their time for my stimulus recordings. Ross Metusalem, Sharon Rose, Stephen Boxwell helped me with ToBI transcription of the stimuli. Joeseph Marulli helped me with Praat scripts and English transcription. Ronnie Rigdon
and Ashley Coolman finished English production accuracy judgment and transcription on time. DJ Hovermale and Jon Dehdari provided LATEX help and troubleshooting. Thanks to Kodi Weatherholtz for proofreading and also for getting me off from heavy work and mind with good energy. Jeff Holliday also proofread three chapters of my dissertation. I also wish to express my appreciation to the wonderful people in Linguistics at the Ohio State University for providing me with a friendly and stimulating work environment.

Lastly, I’d like to thank Dr. Kyung-Ja Park and Dr. Ranghyeyoon Kim at Korea University for their constant support and encouragement with belief in me. Most importantly, I thank with my whole heart to my parents, sisters, and our human-like dog, Bokdol, in Korea for their unquestioning support and love.

Remaining errors, inconsistencies, and flaws in the dissertation are my own.
VITA

2000 .......................... B.A., English Language and Literature,
Korea University, South Korea
2003 .......................... M.A., English Language and Literature,
Korea University, South Korea

PUBLICATIONS

Conference Papers and Books


FIELDS OF STUDY

Major Field: Linguistics
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>ii</td>
</tr>
<tr>
<td>Dedication</td>
<td>iv</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>v</td>
</tr>
<tr>
<td>Vita</td>
<td>vii</td>
</tr>
<tr>
<td>List of Tables</td>
<td>xi</td>
</tr>
<tr>
<td>List of Figures</td>
<td>xv</td>
</tr>
<tr>
<td>Chapters</td>
<td></td>
</tr>
<tr>
<td>1. Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Prosody in L1 Spoken Word Recognition</td>
<td>6</td>
</tr>
<tr>
<td>1.2 Prosody in L2 Spoken Word Recognition</td>
<td>11</td>
</tr>
<tr>
<td>1.3 Overview of Chapters</td>
<td>18</td>
</tr>
<tr>
<td>2. Theoretical Background</td>
<td>20</td>
</tr>
<tr>
<td>2.1 L2 Speech Model</td>
<td>20</td>
</tr>
<tr>
<td>2.1.1 Perceptual Assimilation Model (PAM; Best, 1995)</td>
<td>21</td>
</tr>
<tr>
<td>2.1.2 Speech Learning Model (SLM; Flege, 1995)</td>
<td>24</td>
</tr>
<tr>
<td>2.2 Intonational Structure</td>
<td>28</td>
</tr>
<tr>
<td>2.2.1 English Metrical and Intonation Structure</td>
<td>29</td>
</tr>
<tr>
<td>2.2.2 The Role of Lexical Stress in English Spoken Word Recognition</td>
<td>32</td>
</tr>
<tr>
<td>2.2.3 Korean Intonation Structure</td>
<td>42</td>
</tr>
<tr>
<td>2.2.4 Korean Speaker’s Use of English Lexical Stress</td>
<td>46</td>
</tr>
<tr>
<td>2.3 Research Question and Prediction</td>
<td>49</td>
</tr>
</tbody>
</table>

viii
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Assimilation type, perceptual assimilation pattern, and degrees of goodness-of-fit by Perceptual Assimilation Model (PAM; Best, 1993, 1994, 1995; Best et al., 2001).</td>
<td>22</td>
</tr>
<tr>
<td>2.2</td>
<td>Postulates and hypotheses of Speech Learning Model for second language sound acquisition (SLM; Flege, 1995, p239).</td>
<td>26</td>
</tr>
<tr>
<td>3.1</td>
<td>Eight phonemes used in the trisyllabic nonwords and mean Nlog frequency.</td>
<td>61</td>
</tr>
<tr>
<td>3.2</td>
<td>Mean percent accuracy (%) and mean RT (ms) by L1 and Training day. Values in parentheses represent standard errors.</td>
<td>90</td>
</tr>
<tr>
<td>3.3</td>
<td>Summary of L1 and Training day effects on accuracy and RTs in mixed effects models.</td>
<td>91</td>
</tr>
<tr>
<td>3.4</td>
<td>Mean percent accuracy (%) and mean RT (ms) by L1, Stress pattern, and Phonation type. Values in parentheses represent standard errors.</td>
<td>92</td>
</tr>
<tr>
<td>3.5</td>
<td>Summary of L1, Stress pattern, and Phonation type effects on accuracy and RTs in the mixed effects model.</td>
<td>93</td>
</tr>
<tr>
<td>3.6</td>
<td>Summary of L1, Stress pattern, Phonation type, and Time effects on gaze to the target object (empirical logit scale) in Quasi-MLR (N = 11040; logLik = -17634).</td>
<td>99</td>
</tr>
<tr>
<td>3.7</td>
<td>Summary of L1, Stress pattern, and Phonation type effects on gaze to the target object (empirical logit scale) in Quasi-MLR for segmented windows.</td>
<td>102</td>
</tr>
<tr>
<td>3.8</td>
<td>Goodness of fit of English lexical stress patterns to Korean AP-initial tonal patterns.</td>
<td>104</td>
</tr>
</tbody>
</table>
4.1 Eight phonemes used in the trisyllabic nonword and mean Nlog frequency. . 119

4.2 Mean percent accuracy (%) and mean RT (ms) by L1 and Training day. Values in parentheses represent standard errors. . . . . . . . . . . . . . . . 127

4.3 Summary of L1 and Training day effects on accuracy and RTs in mixed effects models. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 127

4.4 Mean percent accuracy (%) and mean RT (ms) by L1, Stress pattern, and Phonation type. Values in parentheses represent standard errors. . . . . . 128

4.5 Summary of L1, Stress pattern, and Phonation type effects on accuracy and RTs in the mixed effects model. . . . . . . . . . . . . . . . . . . . . . . 129

4.6 Summary of L1, Stress pattern, Phonation type, and Time effects on gaze in Quasi-MLR (N = 8930; logLik = -14393). . . . . . . . . . . . . . . . . . 136

4.7 Summary of L1, Stress pattern, and Phonation type effects on fixation like-lihood in Quasi-MLR for segmented windows. . . . . . . . . . . . . . . . . 138

4.8 Proportion of trials with target object at the top right object position for each condition. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 141

5.1 Mean percent accuracy (%) and mean RT (ms) by L1 and Training day. Values in parentheses represent standard errors. . . . . . . . . . . . . . . . 163

5.2 Summary of L1 and Training day effects on accuracy and RTs in mixed effects models. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 164

5.3 Mean percent accuracy (%) and mean RT (ms) by L1, Stress pattern, and Phonation type. Values in parentheses represent standard errors. . . . . 166

5.4 Summary of L1, Stress pattern, and Phonation type effects on accuracy and RTs in the mixed effects model. . . . . . . . . . . . . . . . . . . . . . . 166

5.5 Mean percent accuracy (%) and mean RT (ms) by L1, Stress pattern, Phonation type, and Fragment size. Values in parentheses represent standard errors. 168

5.6 Summary of L1, Stress pattern, Phonation type, and Time effects on accuracy in Quasi-MLR (N = 3120; logLik = -1667). . . . . . . . . . . . . . . . . . 172
F.1 Mean RT by L1, Stress pattern, and Phonation type after on-target trials removed. Values in parentheses represent standard errors. . . . . . . . . . . 205

F.2 Summary of L1, Stress pattern, and Phonation type effects on RT in the mixed effect model (N = 452; logLik = -3496). The RTs for on-target trials were removed. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 205
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Waveform and fundamental frequency contours for an H* pitch accented English lexical stress minimal pair. Left: Trochaic pattern. Right: Iambic pattern.</td>
<td>30</td>
</tr>
<tr>
<td>2.2</td>
<td>Waveform and fundamental frequency contours for a Korean phrasal tone pattern. Left: Aspirated - AP initial (HHL%), ‘pine tree’. Right: Lax - AP initial (LHL%), ‘nut tree’.</td>
<td>43</td>
</tr>
<tr>
<td>3.1</td>
<td>Left: Spectrum of intensity (dB) against frequency (Hz) in a 10 ms time window. Right: Excitation of loudness (phone) against frequency (Bark) in the same 10 ms time window.</td>
<td>64</td>
</tr>
<tr>
<td>3.2</td>
<td>Top: Presentation of visual stimuli for look-and-listen session and picture naming task. Bottom: Presentation of visual stimuli for picture-choice task and eye-tracking experiment. Text labels are expository and did not appear to participants.</td>
<td>70</td>
</tr>
<tr>
<td>3.3</td>
<td>English speakers’(N = 23) mean fixation proportion for target, competitor, and distractor objects during word recognition. Left: Trochaic nonwords. Right: Iambic nonwords.</td>
<td>96</td>
</tr>
<tr>
<td>3.4</td>
<td>Korean L2 learners’(N = 19) mean fixation proportion for target, competitor, and distractor objects during word recognition. Left: Trochaic nonwords. Right: Iambic nonwords.</td>
<td>96</td>
</tr>
<tr>
<td>3.5</td>
<td>Mean fixation likelihood (empirical logits) for target objects during trochaic and iambic nonword recognition. Left: English speakers (N = 23). Right: Korean learners of English (N = 19).</td>
<td>98</td>
</tr>
</tbody>
</table>
3.6 English speakers (N = 23)’ mean fixation likelihood (empirical logits) of targets for aspirated- (voiceless-) and lax- (voiced-) initial nonwords. Left: Trochaic nonwords. Right: Iambic nonwords. 105

3.7 Korean speakers’(N = 19) mean fixation likelihood (empirical logits) of targets for aspirated- and lax-initial nonwords. Left: Trochaic nonwords. Right: Iambic nonwords. 107

4.1 English speakers’(N = 24) mean fixation proportion for target, competitor, and distractor objects during word recognition. Left: Trochaic nonwords. Right: Iambic nonwords. 132

4.2 Korean L2 learners’(N = 21) mean fixation proportion for target, competitor, and distractor objects during word recognition. Left: Trochaic nonwords. Right: Iambic nonwords. 132

4.3 Mean fixation likelihood (empirical logits) for target objects during trochaic and iambic nonword recognition. Left: English speakers (N = 24). Right: Korean learners of English (N = 21). 134

4.4 English speakers’(N = 24) mean fixation likelihood (empirical logits) of targets for aspirated- (voiceless-) and lax- (voiced-) initial nonwords. Left: Trochaic nonwords. Right: Iambic nonwords. 140

4.5 Korean speakers’(N = 21) mean fixation likelihood (empirical logits) of targets for aspirated- and lax-initial nonwords. Left: Trochaic nonwords. Right: Iambic nonwords. 142

E.1 Bias toward target over competitor (logits) during trochaic and iambic nonword recognition. Left: English speakers (N = 23; 151 trochaic and 144 iambic nonwords). Right: Korean learners of English (N = 19; 132 trochaic and 125 iambic nonwords). 202

E.2 English speakers (N = 23)’ bias toward target over competitor (logits) for aspirated and lax initial nonwords. Left: Trochaic nonwords (98 aspirated and 53 lax initial nonwords). Right: Iambic nonwords (97 aspirated and 47 lax initial nonwords). 203
E.3 Korean speakers ($N = 19$)' bias toward target over competitor (logits) for aspirated and lax initial nonwords. Left: Trochaic nonwords (80 aspirated and 52 lax initial nonwords). Right: Iambic nonwords (83 aspirated and 42 lax initial nonwords).
CHAPTER 1: INTRODUCTION

Prosody as a melodic and rhythmic pattern of speech comprises the patterns of stress, rhythm, and intonation in a language (Beckman & Venditti, 2011). Prosody is characterized by suprasegmental acoustic parameters such as fundamental frequency (f0), duration, and intensity that individually or in combination contribute to the perception of pitch, timing, and loudness (Beckman, 1986, 1996; Fry, 1958; Lehiste, 1970; Lieberman, 1960, 1965). Prosody plays an important role in language comprehension and production at various levels of language processing, including phoneme detection, word recognition, sentence processing, and discourse comprehension. Speakers and listeners use this prosodic information to structure the sequence of sounds, syllables, and higher order organizational units in utterance.

The present study explores how bilinguals who are learning a second language (L2) after the mastery of native language process L2 word prosody during spoken word recognition and how the prosodic structure of their native language affects the processing of L2 word prosody during spoken word recognition. The term ‘bilingual’ broadly refers to the individuals that obtain the knowledge and use of more than one language (Butler & Hakuta, 2004). A large number of people learn more than one language from birth or later in life, becoming bilingual or multilingual for different purposes. Even though the number of bilinguals has been growing to outnumber the monolingual speakers in the world population (Tucker, 2003), the underlying mechanisms by which they perceive and process an L2 are still not very well understood.
The major difference between monolinguals and bilinguals is that bilinguals often have language experience with their first language (L1) prior to L2 learning. Previous work on L1 acquisition has suggested that infants are born as “universal listeners” who can discriminate virtually all possible sound contrasts used in all human languages even without relevant language experience, but by the age of 10 to 12 months they selectively discriminate only those consonants and vowels that are phonemic in their native language (Best, McRoberts, LaFleur, & Silver-Isenstadt, 1995; Polka & Werker, 1994; Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Werker & Tees, 1984a, 2002). Such perceptual adjustment is also made to discern phonotactically possible and impossible strings of segments in their native language by age of nine months, being sensitive to low and high frequent phonotactic sequences and also to the dominant stress pattern (Jusczyk, Cutler, & Redanz, 1993; Jusczyk, Houston, & Newsome, 1999; Jusczyk, 2003; Morgan & Saffran, 1995; Soderstrom, Seidl, Kemler Nelson, & Jusczyk, 2003).

Given that knowledge of L1 sound structure, including segmental and suprasegmental features, is established before one year old, unless bilinguals learn different languages simultaneously from birth (i.e. balanced bilinguals), it is more likely that they will not have equal control over both of their languages, and L1 has a great impact on L2 learning and processing. The term, language transfer, has been used to describe the influence that L1 knowledge has on the comprehension and production of a subsequently learned language. According to Lado (1957, p. 11), the scope of transfer can encompass the entire native phonological system, from the phonemes and their variants, stress and rhythmic patterns, to intonational patterns and their interaction with other phonemes. With regard to L2 spoken language processing, a listener’s L1 phonological system may operate as a phonological “filter” in the way that the differences and similarities between the L1 and
L2 sound systems modulate the L2 learners’ comprehension and production of the target language.

Studies about L1 influence on L2 speech processing have been centered around perception and production of non-native contrasts at the segmental level and have revealed two important findings. (Abramson & Lisker, 1970; Polka, Bohn, et al., 1996; Strange, Akahane-Yamada, Kubo, Trent, & Nishi, 2001; Sundara, Polka, & Baum, 2006; Werker & Tees, 1984b). First, discrimination or categorization of different non-native segments is not equally difficult (Best & Strange, 1992; Best, McRoberts, & Goodell, 2001; Best, Traill, Carter, Harrison, & Faber, 2003; Polka, 1991). Some non-native contrasts are discriminated quite poorly, whereas other contrasts are discriminated quite well. For instance, American speaking English listeners discriminated African Zulu /l/ vs. /ɬ/ (voiceless vs. voiced lateral fricatives) quite well, whereas they discriminated Zulu /b/ vs. /ɓ/ (plosive vs. implosive bilabial stops) quite poorly (Best et al., 2001). The English speaking listeners seemed to assimilate the bilabial stops as an English voiced stop, /b/.

Secondly, the relative ease and difficulty in discriminating a non-native contrast varies according to the listener’s native language. For example, English speaking adult listeners discriminated African Zulu click consonant contrasts better than did listeners from other African languages, Isizulu and Sesotho, that have certain click contrasts (Best, McRoberts, & Sithole, 1988; Best et al., 2003). This was because, unlike African language speaking listeners, English speaking listeners’ perception of African click contrasts was not influenced by any sounds in L1 English.

The perceptual variations across non-native sound contrasts and across listeners’ L1s demonstrate that perception of L2 sounds is affected by the perceived similarities/dissimilarities between acoustic phonetic properties of the non-native sounds and those of the listener’s
native phonology. Relative to the advancement in research for non-native consonant and vowel perception, however, there has been much less work on non-native prosodic perception. A few studies have addressed the issue by examining discriminability of tonal contrasts in listeners with different L1 backgrounds (tonal languages and non-tonal languages) and have provided evidence that non-native tonal perception is also affected by L1 experience at both phonological and phonetic levels.

Gandour (1983), for example, examined dissimilarity judgments of Mandarin tonal features in native speakers of Cantonese Chinese, Mandarin Chinese, and English, and found that English-speaking listeners used pitch height to discriminate different tones, whereas Mandarin-speaking listeners and Cantonese-speaking listeners used both pitch height and pitch direction to differentiate tonal patterns. The results indicate that L1 experience with relevant prosodic cues can affect the way listeners attend to cues to discriminate non-native tonal contrasts. Wayland and Guion (2004) compared Mandarin speaking listeners and English speaking listeners, neither of which had L2 experience with Thai level tones (mid vs. low tones). Chinese listeners discriminated Thai level tones better than English speaking listeners did, suggesting that experience with a tonal language facilitates the perception of non-native tonal contrasts. So and Best (2010) also showed that Cantonese speaking listeners and Japanese speaking listeners were better at discriminating Mandarin tone pairs than were English speaking listeners, supporting that linguistic experience with native tones/pitch variation facilitates identification of non-native tones. However, they also found that Cantonese speaking listeners significantly misidentified Tone 4 (high falling tone) as Tone 1 (high level tone) because these two contours are allotones of Tone 1 (high tone) in Cantonese. The listeners also misidentified Tone 2 (mid rising tone) as Tone 3 (low falling) because they are phonetically similar to Cantonese Tone 2 (high rising tone). These results
imply that listeners are sensitive to only the gradient phonetic information that is relevant to their L1 prosodic contrasts.

Though the above studies addressed lexical tone in relation to non-native prosody processing, the observation was limited to the perception of prosodic contrasts at the lower level of processing in the task of discrimination or categorization. In addition, it was confined to tonal perception, the domain of which was restricted to a single syllable. However, as pointed out by Lehiste (1970, p. 2), suprasegmental features are inherently different from segments as it is established by comparing units in a sequence and has a “overlaid” function of inherent features of speech sounds. It can involve multiple levels above phonetic segments, including syllables, words, phrases, and sentences (Ladefoged & Johnson, 2010; Lehiste, 1970). Therefore, in order to have a more complete picture of prosodic processing, we need to explore the issue at a higher level of processing.

Regarding L2 prosodic processing, we should also notice the cross-linguistic variation of prosodic structures. Languages vary not only in the phonological representation of prosodic features (e.g. word-level prosody, phrase-level prosody, and/or sentential intonation), but also in the acoustic-phonetic representation of the prosodic structures (e.g. f0 cue for Japanese lexical pitch accent, and f0, duration, and intensity cues for English lexical stress). Therefore, questions can be raised as to how and to what extent experience with the features of the prosodic system of one’s native language (e.g. lexical tones, pitch accent, contrastive or emphatic stress, and intonation) affects the perception and processing of non-native prosodic contrasts during L2 processing.

In this light, we direct our attention to the perception and processing of word prosody in the context of lexical processing in the present study. Models of spoken word recognition such as the Race model (Cutler & Norris, 1979; Cutler, Mehler, Norris, & Segui,
1987), TRACE model (McClelland & Elman, 1986b, 1986a; McClelland, Mirman, & Holt, 2006), and the Cohort model (Marslen-Wilson & Welsh, 1978; Marslen-Wilson & Tyler, 1980; Marslen-Wilson, 1987; Marslen-Wilson & Zwitserlood, 1989) agree that continuous evaluation of speech input results in simultaneous activation of multiple lexical candidates which at any moment are supported by the input. Eventual word recognition proceeds on the basis of further input information but also via a process of activation/inhibition or selection between the lexical candidates. If prominence differences across syllables (e.g. INsight and inCITE for English lexical stress; upper case henceforth denotes a stressed syllable) is a part of lexical representation, the manipulations of the relevant suprasegmental acoustic cues may serve as the cues to structure the speech signal in terms of such categories as stressed and unstressed syllables and help listeners’ lexical access during spoken word recognition.

1.1 Prosody in L1 Spoken Word Recognition

The use of suprasegmental variation in lexical activation and recognition has been explored in a number of different languages exhibiting different kinds of suprasegmental structure at the lexical level (e.g. lexical tone, pitch accent, lexical stress). Dutch, Spanish, and English are languages that are known to employ a lexical stress system in which placement is not fixed within the word but is largely determined by syllable weight (and so often called as the “free-stress” language) (Booij, 1999; Zonneveld, Trommelen, Jessen, Bruce, & Arnason, 1999). Both English and Dutch exhibit predominance of stress placement on word-initial syllables, but the tendency is stronger in English (Cutler & Carter, 1987; Cutler, 2005). In Spanish, a majority of multisyllabic words has stress on the penultimate syllable. In addition, though both English and Dutch have reduced vowels, English
employs the segmental cue (i.e. reduced vowels in unstressed syllables) more often to differentiate stressed and unstressed syllables than Dutch does (N. Cooper, Cutler, & Wales, 2002). Spanish doesn’t employ reduced vowels in unstressed syllables. Spanish and Dutch indeed have many stress minimal pairs which members are contrastive only by lexical stress cues in the first two syllables (N. Cooper et al., 2002).

Despite some differences in the placement of stress and the use of reduced vowel as a correlate to lexical stress, studies that have examined the role of lexical stress in these languages have shown that lexical stress information contributes to the activation of target words during spoken word recognition (N. Cooper et al., 2002; Cutler & Van Donselaar, 2001; Soto-Faraco, Sebastián-Gallés, & Cutler, 2001; Van Donselaar, Koster, & Cutler, 2005). Soto-Faraco et al. (2001) examined the effects of stress and segmental cues on Spanish word recognition by employing cross-modal fragment priming experiments. In this study, native Castilian Spanish speakers responded significantly faster in a lexical decision task when they heard a segmentally (e.g. sardi- for the target SARDINA ‘sardine’) or prosodically matching fragment (e.g. PRINci- for the target PRINcipe ‘prince’) than when they heard a phonologically unrelated control fragment (e.g. mos- from the word mosQUITO ‘mosquito’ for the target PRINcipe ‘price’). The response was slower when the listeners heard a segmentally (e.g. sarda- from the word sardana ‘Catalan dance’ for the target SARDINA ‘sardine’) or prosodically mismatching fragment of the target word (e.g. prinCI- from the word prinCIpio ‘beginning’ for the target PRINcipe ‘prince’) than when they responded to the unrelated control fragment. These results suggest that prosodic features contribute to the activation and inhibition of lexical candidates as much as segmental features of a word do during Spanish spoken word recognition.
Van Donselaar et al. (2005) replicated Soto-Faraco et al. (2001) in Dutch with bisyllabic and trisyllabic words, and found a very similar processing pattern. Dutch listeners’ lexical decision for visually presented words was faster and more accurate when the words were preceded by the auditory primes of first two syllables with appropriate stress (e.g. okTO- for OKTOBER) than when the words were preceded by phonologically unrelated primes, whereas their responses took longer and were less accurate when the primes carried inappropriate stress (e.g. OKto- for OKTOBER) than when the primes were phonological unrelated ones. The facilitation effect was found even when the prime was only the initial syllable of a word, but the inhibition effect of inappropriate stress was found when the prime was the first two syllables of a word. These results indicate that lexical stress plays a role in constraining lexical activation and recognition in Dutch, but it is more effective when listeners have at least first two syllables of a word.

N. Cooper et al. (2002) also conducted cross-modal fragment priming experiments using English words, and found that English-speaking listeners also make use of suprasegmental stress cues in recognizing spoken words. They examined lexical decision for visually presented target words with English word pairs of ADmiral - admiRAtion type: one member had primary stress on the first syllable with an unstressed second syllable, and the other member had primary stress on the third syllable with secondary stress on the first syllable and an unstressed second syllable. Lexical decision was faster after stress-matching primes than after control fragments (i.e. control auditory primes were segmentally mismatched to the target words) for both initial primary stress words (e.g. We were sure the word was ADmi- for ADMIRAL) and initial secondary stress words (e.g. We were sure the word was admi- for ADMIRATION). Though initial primary stress words was recognized more rapidly than initial secondary stress words, the results indicate that English speaking
listeners can distinguish primary stress from secondary stress in lexical processing. This facilitation was also found when the primes were single syllable fragments taken from English word pairs of *Music - muSEum* type in which one member had primary stress in the first syllable, and the other was unstressed in the first syllable with no vowel quality change. Interestingly, however, there was no inhibition effect from stress-mismatching primes for the bisyllabic primes with no significant difference in response times between stress-mismatched primes (e.g. *We were sure the word was adm* for ADMIRAL) and control primes. This is quite different processing pattern from what’s been found in Spanish-speaking listeners (Soto-Faraco et al., 2001) and Dutch-speaking listeners (Van Donselaar et al., 2005) who showed significant longer response times for stress-mismatched primes than for control primes. These results suggest that English-speaking listeners are less efficient in using stress cues to the resolution of word competition as compared to Spanish-speaking listeners and Dutch-speaking listeners. N. Cooper et al. (2002) ascribed it to the differences between English, Spanish, and Dutch in the regularity of stress placement and the phonetic manifestation of lexical stress. As a majority of English words have stress in the first syllables and a reduced vowel in unstressed syllables, the secondary stress in first syllables that were relatively strong relative to the unstressed second syllables (e.g. *We were sure the word was adm*) might have been heard as stressed syllables, in turn partially facilitating the activation of stress competitor words (e.g. ADMIRAL), instead of inhibiting the activation.

Just as free-stress languages have shown some prosodic effects on lexical activation, languages of other word prosody types (e.g. Chinese lexical tone, Japanese lexical pitch accent) also showed strong prosodic effects on lexical processing. Unlike lexical stress which is realized over the duration of polysyllabic words, the realization of pitch variations
in Mandarin/Cantonese tone is mostly local to a syllable, though the neutral tone in Mandarin Chinese has a phonetic value as a part of continuous tone of the previous full tone.

In standard Putonghua Mandarin, almost every syllable is lexically specified for one of the phonemically contrasting tones - Tone 1 ‘high level tone’, Tone 2 ‘mid rising’, Tone 3 ‘low falling’, and Tone 4 ‘high falling’ (Peng, 2005). For morphemes such as the agreement-participle -ba, the verbal suffix -le, the nominal suffix -zi, and the pragmatic particles -ma, -a that are inherently unspecified for tone, tone 5 ‘neutral tone’ is assigned. The phonetic value of the neutral tone is determined by that of the preceding full-tone syllable. For example, in a citation form of word with Tone 1 or Tone 2 followed by a neutral tone, the f0 reaches the peak at the end of Tone 1 or 2 and gradually falls to a mid level in a neutral tone. Due to the strong locality of tonal patterns, studies on Mandarin tone processing have been focused on the comparison between the access timing to tonal and segmental information during spoken word recognition (C. Lee, 2007; Liu & Samuel, 2007; Malins & Joanisse, 2010; Yip, 2001). A recent eyetracking study, Malins and Joanisse (2010) showed that when Mandarin listeners identified a target word (chuang with tone 2, ‘bed’), they considered the segmental competitor word (chuang with tone 1, ‘window’) and the tonal competitor word (niu with tone 2, ‘cow’) at a similar point in time after the word onset, suggesting that segmental and tonal information are accessed concurrently and play comparable roles in constraining the activation of lexical competitors during spoken word recognition in Mandarin Chinese.

On the other hand, Japanese pitch accent is a lexical phenomenon that changes the pitch contour of the accentual phrase (AP). The lexically specified aspects of the tone pattern in Japanese include the contrast between “accented (HLL)” words (which have a lexically specified pitch fall on a designated mora) and “unaccented (LHH)” words (which have no
such fall) and properties of the initial mora which govern the timing of a phrase-initial rise. The AP, as a prosodic unit above words in Japanese intonational phonology, is typically characterized by a rise to a high tone around the second mora of the phrase and a subsequent fall to a low tone at the right phrase edge (Beckman & Pierrehumbert, 1986; Venditti, 2005). Its high tone is preempted by a pitch accent if the AP-initial syllable is accented. The use of pitch accent in Japanese word recognition was consistently supported by many word identification gating studies and lexical decision studies (Cutler & Otake, 1999; Otake & Cutler, 1999; Minematsu & Hirose, 1995; Sekiguchi & Nakajima, 1999; Sekiguchi, 2006). Japanese listeners’ identification of target words was made when they heard only the initial mora of a target word (e.g. JI- for JIdou ‘children’ HLL, ji- for jiDOU ‘automatic’ LHH), suggesting that pitch accent information of a word is immediately integrated during lexical access and sufficient to constrain the selection of candidates during Japanese spoken word recognition.

The findings from the studies on Spanish, Dutch, English, Chinese, and Japanese spoken word recognition have provided explicit evidence that the input that listeners attend to during spoken word recognition is both segmental and autosegmental cues in these languages. Thus, evidence suggests that within activation-competition frameworks of spoken word recognition, the lexical candidates that match the prosodic and/or segmental features are more activated than any others during lexical access.

1.2 Prosody in L2 Spoken Word Recognition

While the field of psycholinguistics has progressed to acknowledge prosodic processing during L1 word recognition, little is known about the processing of prosody during L2 word recognition. As shown by the strong contribution of prosodic cues to lexical access in
the languages above, prosodic information should be adequately encoded to make native-like word recognition possible. Also, the existence of suprasegmental correlates of stress variation seems to be important as it affects the nature of the information to which listeners attend if stress is to be used to constrain lexical activation on line.

Regarding L2 speech perception and processing, a particularly vexing problem comes from the need to perceive and produce an aspect of the L2 that is not present in the L1. There are indeed as many fixed-stress languages as free-stress languages, such as Finnish, French, Hungarian, Czech (Goedemans, 2003, unpublished), and also languages that do not employ word-level stress. The findings from L2 sound perception studies allow the prediction that the perceptual discrimination of an L2 prosodic contrast that does not exist in L1 may be good such as the discrimination of Zulu clicks in English listeners (Best et al., 1988, 2003), but the question still remains on whether it may or may not be used during lexical processing.

Some previous studies explored this issue by examining French monolinguals and French L2 learners of Spanish/English in a series of perception and word recognition experiments. French language does not have word-level stress, but has a fixed phrasal ‘accent’, which is realized as final syllable lengthening at the right edge of the word or the phrase (Vaissière, 1991).

Dupoux, Pallier, Sebastian, and Mehler (1997) first examined perceptual distinction of Spanish lexical stress with native French speakers who were naive to Spanish and native Spanish speakers in the ABX (nonwords produced by three different Dutch speakers) and AX (nonwords produced by a Dutch speaker) discrimination tasks. Unlike their prediction that French-speaking listeners would be “deaf” to contrastive stress, French-speaking listeners showed significantly poorer performance than Spanish-speaking listeners only in
the ABX task where high phonetic variability was manipulated for the stimuli by speaker variations. French-speaking listeners’ discrimination of stress contrasts was as good as Spanish-speaking listeners in the AX task, which contained less phonetic variability and thus was cognitively less demanding than the ABX task. As it’s been known that noncontrastive or allophonic information is used to parse speech into phonological constituents at an abstract level (Church, 1987), French speaking listeners’ inability to use allophonic variations of contrastive stress yielded the conclusion that French speakers’ “stress deafness” might not be the problem at the phonetic level of processing, but at the phonological level of processing.

To explore further whether French-speaking listeners’ good discrimination of stress contrasts in the AX task was driven by the simplicity of the task, Dupoux, Peperkamp, and Sebastián-Gallés (2001) conducted a series of short-term memory sequence repetition tasks to increase memory load in the perception tasks. Prior to the task, participants learned the correct association between the words, either a segmental contrast (e.g. /tuku/ vs. /tupu/) or a stress contrast (e.g. /piki/ vs. /piˈki/) and the number keys (e.g. /tuku/ was associated to the key [1], while /tupu/ was associated to the key [2]). During the task, participants listened to repetitions of two words, either a segmental contrast or a stress contrast and were asked to recall each sequence by typing the associated keys in the correct order. The phonetic variability of stimuli was manipulated not only by speaker variation, but also by pitch variation for a single speaker. Similar to the results from Dupoux et al. (1997), the “stress deafness” emerged in French speakers only when the stimuli involved high phonetic variability, suggesting that French speakers have difficulty distinguishing stress contrasts only when they were given in the input with high phonetic variation and high memory load.
Dupoux, Sebastián-Gallés, Navarrete, and Peperkamp (2008) questioned whether French speakers’ poor performance in the task with high phonetic variability was due to the lack of a proper phonological representation of lexical stress or the difficulty with metalinguistic access to the contrast required in the repetition task (which asked participants to associate two stimuli to two response keys). To address this question, they compared performance of French monolinguals, French speaking L2 learners of Spanish, and Spanish monolinguals in the same sequence recall task with high phonetic variability. As it was assumed that French monolinguals did not establish a metalinguistic representation of Spanish stress pattern, whereas French L2 learners of Spanish developed the metalinguistic knowledge of Spanish stress because they took explicit L2 training about Spanish stress contrast, French L2 learners of Spanish were expected to perform better than French monolinguals. The results, however, showed no significant difference between the French monolinguals and the L2 learners. Both French L1 groups showed lower accuracy for the stress contrast than for the phoneme contrast with poorer performance compared to the Spanish speakers, suggesting that French speakers’ poor performance with higher phonetic variability is not due to metalinguistic limitations, but due to a processing deficit. Dupoux et al. (1997, 2001, 2008) claimed that acoustic-phonetic representation of lexical stress does not pose a problem to French speakers in the short-term memory task, but in the memory task that requires abstract phonological representation of word stress.

Tremblay (2008) supported this claim by examining Canadian French speaking L2 learners of English with different English proficiencies in a cross-modal word identification task. Native English speakers and French speaking learners of English were asked to identify a target word after listening a truncated auditory stimulus (e.g. MYS- taken from MYSTery or mis- taken from misTAKE). The French learners of English were not good at
using English stress information for word identification across all L2 fluency levels (inter-
mediate L2 learners 58%, low-advanced L2 learners 59%, high advanced L2 learners 59%),
with accuracy rates slightly above chance level but significantly lower than that of native
English speakers (73%). Tremblay (2009) reported how the Canadian French speaking L2
learners of English that participated in Tremblay (2008) behaved in an AXB perception task
with and without phonetic variability of English lexical stress. In the discrimination task,
they showed very similar results to those reported in Dupoux et al. (1997, 2001). The L2
learners were significantly less accurate and slower in response than native English speak-
ers only when the stimuli had a high phonetic variability, with no significant differences
between L2 fluency levels.

In contrast to the experimenters’ prediction that French L2 learners are less likely to
learn to use L2 lexical stress information that does not play a significant role in the ac-
tivation of L1 words, the previous studies on French-speaking listeners of Spanish and
French-speaking learners of English provide evidence that French speakers are not quite
“stress deaf”. The results rather support that they are able to detect the acoustic correlates
of L2 prosodic contrasts that are not used for L1 prosodic contrasts, to be consistent with
the finding that non-native sounds with no L1 correspondences are well discriminated (i.e.
Zulu clicks for English listeners).

Dupoux et al. (1997, 2001, 2008) and Tremblay (2008, 2009) explained the process-
ing difficulties that French-speaking listeners have shown in the word identification task
and in the ABX task with high phonetic variability by appealing to different memory loads
required for lexical stress processing at the phonetic vs. phonological levels. The L2 lis-
teners who do not have lexical stress in their L1 can sense the acoustic phonetic difference
between syllables induced by lexical stress, but they may have trouble representing and
storing the stress pattern in working memory, leading to poor performance at a higher level of L2 processing. This seems to be especially true for L2 learners whose L1 has a fixed-stress pattern with no contrastive tone, pitch accent, or stress variations.

Here a question is raised as to how L2 learners who do not have word stress in L1 as in French, but have a phrase-level prosody structure with some tonal pattern in L1 unlike French, process L2 word stress during spoken word recognition. This question is related to the general research question earlier in this chapter: how and to what extent language experience with the features of prosodic system of native language (e.g. lexical tones, pitch accent, contrastive or emphatic stress, and intonation) affects perception and processing of non-native prosodic contrasts during L2 processing.

To address this question, we examine Korean speaking learners of English trying to master the use of lexical stress in English during spoken word recognition. Korean has a prosodic system that is typologically quite different from English. It does not employ contrastive stress at the word level, but it has tonal patterns that are associated with the domain of the accentual phrase (Jun, 2005, 2007). Since the accentual phrase may consist of several lexical items, tonal pattern is less likely to be a part of lexical representation for Korean speakers as is the phrase-final accent for French speakers. Thus, Korean L2 learners of English may be good at perceiving English lexical stress but poor at using the information during spoken word recognition, like French-speaking listeners of Spanish or English. Nevertheless, since most accentual phrases comprise one phonological word, and 90% of content words appear in the accentual phrase initial position in Korean (Jun, Fougeron, et al., 2000; Schafer & Jun, 2002), the association of a tone with a word-initial segment may be perceptually salient to Korean speakers (Cho, 1996; M. Kim, Beddor, & Horrocks, 2002) and affect processing of lexical stress during word recognition.
Early attempts to explain and model this process were based on contrastive analyses (e.g. Lado, 1957), predicting that when L1 and L2 rules conflict, interference between L1 and L2 increases the likelihood of errors, while when the rules are similar for both languages, processing of L2 would be facilitated. A contrastive analysis predicts that perception and processing of L2 sounds should be impeded if the target language sound category does not exist in the L1. Thus, the contrastive view predicts uniform difficulty when Korean-speaking learners of English must learn to distinguish words that contrast only in the location of the lexically stressed syllable, e.g. *INsight* and *inCITE*, because their L1 doesn’t have such prosodic categories at the lexical level.

Although this view identifies general areas of processing difficulty for particular L1 - L2 combinations, it provides less insight into the subtle distinctions among non-native sounds that listeners and speakers of a second language must use to perceive and produce such ‘unavailable’ sounds. Indeed, the contrastive analysis approach cannot account for French speakers’ good discrimination of Spanish/English lexical stress in Dupoux et al. (1997, 2001, 2008) and Tremblay (2009). A full understanding of learners’ L2 processing difficulties would require detailed phonetic descriptions of speech sounds at issue rather than presence or absence of phones.

To address this kind of problem in L2 speech model, two cross-language speech perception models, the Perceptual Assimilation Model (PAM; Best, 1995) and the Speech Learning Model (SLM; Flege, 1995) provided more explanatory and testable insights into the variability in the perceptual differentiation of L2 sounds. The two models were designed to explain the difficulties listeners have with phonemic contrasts from a language other than L1. However, as the predictions were established based on the acoustic-phonetic approach to the similarities and differences between L1 and L2 categories, we can refer to
the two constructs as we venture into the area of the use of suprasegmental phenomena during L2 spoken word recognition and production. In addition, because the realization and phonological status of word prosody differs across languages, we take into consideration the cross-linguistic variation of prosodic structures in English and Korean to account for Korean L2 learners’ prosodic processing during English spoken word recognition.

1.3 Overview of Chapters

To address the research question raised in this chapter, the remainder of the dissertation begins with the introduction of the two theoretical constructs of L2 speech model and the description of intonation structures of English and Korean. Then, it is devoted to the description of the preliminary eyetracking study and the demonstration of analyses and results of the dissertation eyetracking and gating experiments. The chapters are organized as follows:

Chapter 2 introduces the two L2 speech models, PAM and SLM, in Section 2.1 and the intonation structures of English and Korean in Section 2.2. An autosegmental approach is taken to specify the metrical and intonational structures of the two languages. Realization of reduced vowel schwa is also described in its relation to unstressed syllables for English lexical stress pattern. Section 2.3 presents this dissertation’s research questions and predictions.

Chapter 3 reviews the preliminary eyetracking experiment (Shin, 2008) which explored the effect of lexical stress on spoken word recognition in native English speakers and Korean speaking learners of English. It also reports the design and methodology in detail as they are also adopted by a dissertation eyetracking study.
Chapter 4 reports the eyetracking experiment of the dissertation. The experiment explores the possible L1 phrasal level prosodic effect on processing of English lexical stress during word recognition with the reduced vowel, schwa, in the unstressed syllable.

Chapter 5 presents the cross-modal fragment gating experiment of the dissertation. The experiment examines the effect of L1 phrasal level prosody on processing of English lexical stress without involving any segmental cues to lexical stress, such as reduced vowels. Word identification performance was observed over three blocks of successively incremented word fragments. This experiment employs very similar materials to those used for the preliminary study.

Chapter 6 provides general discussion and conclusions, and implications of the study.
CHAPTER 2: THEORETICAL BACKGROUND

This chapter first introduces the two L2 speech models, the Perceptual Assimilation Model (PAM; Best, 1995) and the Speech Learning Model (SLM; Flege, 1995). As the two models were proposed to account for L2 sound perception, it would be hard to make a direct application of the models to the predictions and explanation of the L2 lexical stress processing. However, they provide reasonable insights into L2 speech perception, based on which we can infer how suprasegmental aspects of L2 speech are perceived and processed during L2 spoken recognition. Secondly, we describe English metrical and intonation structure and Korean intonation structure to examine the differences and similarities between the two prosodic structures that may or may not facilitate processing of English lexical stress in Korean speaking L2 learners of English. Previous studies that explored the processing of English lexical stress during spoken word recognition are discussed for native English speakers and Korean L2 learners of English.

2.1 L2 Speech Model

The PAM and SLM have focused on somewhat different aspects of L2 speech: the PAM is primarily concerned with naive perception of unfamiliar perceptual distinctions, whereas the SLM aims to account for L2 phonetic acquisition over time. However, they agree that the relative acoustic-phonetic (SLM) or articulatory-phonetic (PAM) similarity between L1 and L2 sounds determines correct categorization and discrimination of L2 sounds.
2.1.1 Perceptual Assimilation Model (PAM; Best, 1995)

The PAM explains the cross-linguistic effect of L1 on L2 phoneme discrimination by an assimilation process and provides specific predictions about sound identification and discrimination. Assimilation here refers to the cross-linguistic influence which occurs in the receptive phonological process by which the phonetics of an L2 speech sound is perceived to be more like that of a (different) sound in L1. In the PAM, the assimilation process is determined by perceived degree of similarity of L2 segments to native (L1) segments based on their articulatory gestural constellations in the vocal tract and the degree of proximity of L2 contrasts to corresponding L1 contrasts in acoustic-phonetic space.

Based on the assimilation process, the PAM provides three perceptual assimilation patterns of a non-native sound against a native sound, and six assimilation patterns of non-native contrasts against native contrasts. It predicts that the discrimination of non-native contrasts can be aided, hindered, or unaffected by native phonology depending on how the non-native sound is related to native phonemes and contrasts (Best, 1993, 1994, 1995; Best et al., 2001). Table 2.1 summarizes the assimilation type, perceptual assimilation pattern, and corresponding degrees of perceptual goodness-of-fit between L1 and L2 sounds.

The three assimilation patterns include Categorized (C), Uncategorized (U), and Nonassimilable (N) assimilation. Categorized assimilation occurs for the non-native phone that assimilates to a native phoneme. Uncategorized assimilation is the case that a phone falls within native phonological space but somewhere in between native phonemes (i.e., similar to two or more phonemes). Nonassimilable assimilation occurs when a phone has no similarity to any native phonemes, and thus is heard as a non-speech sound.
<table>
<thead>
<tr>
<th>Assimilation type</th>
<th>Perceptual assimilation pattern</th>
<th>Goodness-of-fit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single assimilation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Categorized (C)</td>
<td>A non-native phone assimilates to an exemplar of some native phoneme.</td>
<td>Excellent to poor</td>
</tr>
<tr>
<td>Uncategorized (U)</td>
<td>A non-native phone assimilates to an uncategorized consonant or vowel that falls somewhere in between native phonemes.</td>
<td>Poor</td>
</tr>
<tr>
<td>Nonassimilable (N)</td>
<td>A nonspeech sound bears no detectable similarity to any native phonemes.</td>
<td>Excellent</td>
</tr>
<tr>
<td>Pairwise assimilation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Two-Category (TC)</td>
<td>Two non-native phones assimilate to two separate native phones.</td>
<td>Excellent</td>
</tr>
<tr>
<td>Category-Goodness (CG)</td>
<td>Both non-native phones assimilate to a single native phone, but one fits better than the other.</td>
<td>Moderate to very good</td>
</tr>
<tr>
<td>Single-Category (SC)</td>
<td>Both assimilate equally well or poorly to a single native phoneme.</td>
<td>Poor</td>
</tr>
<tr>
<td>Both Uncategorizable (UU)</td>
<td>Both non-native phones are uncategorized speech segments.</td>
<td>Fair to good</td>
</tr>
<tr>
<td>Uncategorized vs. Categorized (UC)</td>
<td>One non-native phone is uncategorized, while the other is categorized to a native phone.</td>
<td>Very good</td>
</tr>
<tr>
<td>Nonassimilable (NA)</td>
<td>Both non-native phones’ articulatory properties are quite discrepant from any native phonemes.</td>
<td>Very good</td>
</tr>
</tbody>
</table>

Table 2.1: Assimilation type, perceptual assimilation pattern, and degrees of goodness-of-fit by Perceptual Assimilation Model (PAM; Best, 1993, 1994, 1995; Best et al., 2001).

For discrimination of non-native contrasts, PAM provides six pairwise assimilation patterns - Two-Category (TC), Category-Goodness (CG), and Single-Category (SC) assimilation for L2 sound categories that are phonetically similar to one or more L1 categories, and Both Uncategorizable (UU), Uncategorized vs. Categorized (UC), and Nonassimilable (NA) assimilation for L2 segments, one or both of which are not phonetically similar to any L1 category.
Native language phonology aids discrimination of L2 contrasts if the non-native phones are clearly separable in the native phonology, as indicated by Two Category assimilation (TC Type), but hinders discrimination if two non-native phones assimilate to the same native sound, lacking phonological contrast, as defined by Single Category assimilation (SC Type). The discrimination of Category Goodness difference (CG Type) is expected to be moderate to very good, but discrimination between the two sounds will be inferior to that for TC contrasts, because the two non-native sounds are assimilated to a single native category. Identification will be very good for the non-native sound that deviates least from the native category. Both Uncategorized (UU Type) assimilation refers to two sounds that fall within the native language phonetic space but outside of any particular native category. As discrimination is less strongly affected by native phonological equivalence classes, it should range between poor and very good, dependent on perceived similarity of the non-native phones to each other and to the set of nearby native phonemes. Uncategorized-Categorized (UC Type) assimilation refers to non-native contrasts, one of which is assimilated to a native category while the other isn’t, and so the discrimination is expected to be very good. Nonassimilable (NA Type) refers to the non-native contrasts that do not sound like speech sounds and so the discrimination is expected to be good to very good.

Many studies have provided evidence for this model by showing that the degree of acoustic-phonetic similarity between L1 and L2 sounds indeed affects L2 learners’ sound discrimination for both consonant and vowel contrasts (Best & Strange, 1992; Best, 1993; Best et al., 2001; Flege, 1995; Polka, 1991, 1992; Werker & Tees, 1984b). For example, Best et al. (2001) examined English speakers’ perceptual discrimination of three different types of non-native consonant contrasts taken from Zulu - voiceless vs. voiced lateral fricatives (/ɬ/-/ɮ/, TC assimilation), voiceless aspirated vs. ejective velar stops (/kʰ/-/kʰ/,
CG assimilation), and plosive vs. implosive bilabial stops (/b/-/ɓ/, SC assimilation). Discrimination of the fricative voicing distinctions, which are signaled by a laryngeal gesture of glottal abduction vs. vibration in both English and Zulu, was near ceiling (TC assimilation), and discrimination of the velar stops, of which the aspirated stop is a better fit to English /k/ than the ejective one is, was between moderate and good (CG assimilation), while discrimination of the bilabial contrast, where the plosive and implosive bilabial stops are nearly equal exemplars of American English /b/ (SC assimilation), was bad. Also, Best, Faber, and Levitt (1996) showed American English speakers’ discrimination of a Norwegian unrounded vs. rounded high front vowel contrast (/i/-/y/) was poor as a SC assimilation, whereas their discrimination for French front unrounded high vs. mid vowels (/i/-/ɛ/) was very good to fit the description of TC assimilation.

To summarize, the PAM incorporates both contrastive and noncontrastive phonetic influence of native language in its prediction about discrimination of diverse types of non-native contrasts (Best et al., 2001). The goodness of fit would be TC (>NA >UC >UU) >CG >SG (Best, 1994, 1995; Best et al., 2001).

2.1.2 Speech Learning Model (SLM; Flege, 1995)

The SLM has the basic assumption that accuracy of L2 sound production has a perceptual basis and thus perception of L2 sounds guides sensorimotor learning of L2 sounds. However, it also acknowledges that incorrect production sometimes can be driven by motoric output constraints based on syllable type in an L1. Furthermore, the model accounts for age-related limits on the ability to produce L2 vowels and consonants.
The model proposed four postulates (P1-P4) and seven hypotheses (H1-H7) to explain the L2 speech learning mechanism (Flege, 1995, p. 239). Table 2.2 summarizes the postulates and hypotheses.

The four postulates shown in Table 2.2 indicate that L2 learners phonetic space established in L1 is maintained over the life span, affecting L2 sound perception. But the space can accommodate new sound contrasts as L2 learners become exposed to other new sounds. As L2 learning proceeds, learners adjust the boundaries of their phonetic categories to some extent to accommodate new L2 sounds. This means that L2 learners can apparently learn to make use of new cues for identifying categories and reallocate the mental space for separating them.

Based on the four postulates, the SLM proposed seven hypotheses of L2 speech learning. Similar to the PAM’s predictions, it hypothesizes that learning of an L2 sound varies depending on the degree of similarity between L1 and L2 sounds. The greater the perceived distance between an L2 and the closest L1 sound, the easier it will be for L2 learners to establish a phonetic category for the L2 sound (H3), and the L2 sound is easily learned if it is similar enough to be merged as a diaphone for an L1 counterpart (H5).1 However, this model allows the possibility that L2 learners can learn a new L2 sound that is similar to a diaphone of the corresponding L1, but not similar enough to be merged with this category; such learning proceeds once the learner perceives the subtle acoustic phonetic differences between the L2 sound and the closest L1 sound (H2).

1The SLM proposed a hypothesis (H1) that L2 learners perceptually relate positional allophones in the L2 to the closest positionally defined allophone in L1 (Flege, 1995, p. 239). The L1 and L2 sounds that are perceptually related in this way were termed as “diaphones” by Weinreich (1957). Flege (1995) adopted this term to refer to the L1 and L2 sounds that are perceptually similar to one another at the position-sensitive allophonic level.
Postulates

P1 The mechanisms and processes used in learning the L1 sound system, including category formation, remain intact over the life span, and can be applied to L2 learning.

P2 Language-specific aspects of speech sounds are specified in long-term memory representations called phonetic categories.

P3 Phonetic categories established in childhood for L1 sounds evolve over the life span reflect the properties of all L1 and L2 phones identified as a realization of each category.

P4 Bilinguals strive to maintain contrast between L1 and L2 phonetic categories, which exist in a common phonological space.

Hypotheses

H1 Sounds in the L1 and L2 are related perceptually to one another at a position-sensitive allophonic level, rather than at a more abstract phonemic level.

H2 A new phonetic category can be established for an L2 that differs phonetically from the closest L1 sound if bilinguals discern at least some of the phonetic differences between the L1 and L2 sounds.

H3 The greater the phonetic dissimilarity between an L2 sound and the closest L1 sound, the more likely it is that phonetic differences between the sounds will be discerned.

H4 The likelihood of phonetic differences between L1 and L2 sounds, and between L2 sounds that are not contrastive in the L1, being discerned decreases as AOL (Age of Learning) increases.

H5 Category formation for an L2 sound may be blocked by the mechanism of equivalence classification. When this happens, a single phonetic category will be used to process perceptually linked L1 and L2 sounds (diaphones). Eventually, the diaphones will resemble one another in production.

H6 The phonetic category established for L2 sounds by a bilingual may differ from a monolingual’s if: 1) the bilinguals’ category is “deflected” away from an L1 category to maintain phonetic contrast between categories in a common L1-L2 phonological space; or 2) the bilingual’s representation is based on different features, or feature weights, than a monolingual’s.

H7 The production of a sound eventually corresponds to properties represented in the phonetic category representation.

Table 2.2: Postulates and hypotheses of Speech Learning Model for second language sound acquisition (SLM; Flege, 1995, p239).
When such an L2 sound comes into the learner’s phonetic space, the closest L1 sound and the L2 sound tend to be dispersed to keep the contrast within the same phonetic space (H6). In addition, this model predicts the factors such as the position in a word, speech rate, degree of stress, speaking style (H1), and learners’ age of learning (AOL) (H4) that can potentially affect the phonetic analysis of an L2 sound. It further hypothesizes that L2 production reflects phonetic category established in the perceptual space (H7). That is, if L2 learners have difficulty establishing new perceptual categories, their production is more likely to result in foreign accents.

Supporting these hypotheses are many empirical studies. For instance, Strange (1992) found that Japanese L2 learners of English perception and production of the English /l/-/ɾ/ was more accurate in word-final position than in word-initial position, supporting that the perceptual distinction of the same non-native sounds is influenced by the position-in-word (H1). Takagi (1995) further showed that Japanese L2 learners better discerned phonetic differences between Japanese /ɾ/, which is perceptually closer to English /l/ than /ɾ/, and English /ɾ/ than between Japanese /ɾ/ and English /l/. This result provides evidence that L2 learners assess perceptual distance of an ambiguous L2 contrast to an L1 category by referring to the perceived degree of similarity to the corresponding L1 sound, supporting H2. In addition, Mack (1990) showed that French-English bilinguals produced /b, d, g/ with short-lag VOT in both languages, while /p, t, k/ with average VOT of 66 ms in French, which is much longer than the one for the French counterparts, and with average VOT of 108 ms in English which is also much longer than the VOT for English counterparts. The production pattern shows that when the VOTs of voiceless stops, /p, t, k/, were closer to each other in L1 and L2, the bilinguals try to make the phonetic contrast between three
categories in L1 and L2, at the expense of accuracy in both languages, and such perceptual distinctions are reflected in the production (H6)

To summarize, with regard to L2 speech processing, the SLM approaches L2 perception in association with production and age effects. The postulates and hypotheses provide predictions on how a non-native phone is categorized in relation to L1 sounds. Though the model is mainly concerned with perception and production at the phoneme level, it opens the possibility that a failure to discern cross language phonetic differences may arise at one or more processing levels (Flege, 1995, p. 241). That is, though L2 listeners can access the sensory properties which distinguish L2 contrasts or L1 and L2 sounds, they may not be able to so during the on-line processing of speech.

2.2 Intonational Structure

The two L2 speech learning models and subsequent studies have provided solid evidence that the degree of acoustic phonetic similarity between L1 and L2 sound systems has great influence on L2 speech perception. The existence of segmental and suprasegmental correlates of stress variation in L1 is potentially important to our investigation of L2 word prosody processing as it affects the nature of the information to which L2 listeners attend when processing lexical stress during L2 spoken word recognition.

To understand how Korean L2 learners of English whose L1 does not employ lexical stress perceive and process L2 English lexical stress during spoken word recognition, we need to examine the metrical and intonational structures of English and Korean. In addition to that, previous studies on native English speakers’ processing of lexical stress during spoken word recognition and Korean speaking L2 learners’ use in English lexical stress during spoken word recognition are discussed.
2.2.1 **English Metrical and Intonation Structure**

English as a target language has meaningful differences in stress at the level of individual words (“lexical stress”, e.g. the minimal pair INsight vs. inCITE where capitals indicate stressed syllables) and at the phrasal level “pitch accents”, e.g. I want CHOcolate pie (not vanilla) vs. I want chocolate PIE (not cake) where capitals indicate pitch accented stressed syllables) (Pierrehumbert, 1990; Beckman, Hirschberg, & Shattuck-Hufnagel, 2005).

English stress at the level of words has its basis in the structure of the foot, defined as a combination of heavy (strong) and light (weak) syllables. A syllable is heavy, if its head is a tense vowel, a diphthong, or an unreduced lax vowel as in the first and third syllables of the word *automatic* (a vowel-based definition, Beckman, 1996; Cutler & Norris, 1988; Cutler & Butterfield, 1992). The heaviness of a syllable is also determined by a perceived prominence (stress) difference relative to the light syllable (a stress-based definition, Lehiste, 1970; Lieberman, 1965; Slowiaczek, 1990). Stressed syllables of words produced in isolation with an H* pitch accent are generally characterized by a higher fundamental frequency ($f_0$) or larger $f_0$ excursion, longer duration, and greater intensity than any other syllables as shown in Figure 2.1.

These suprasegmental acoustic phonetic cues generally give rise to perceived prominence differences across syllables in a word and they can be categorized either as trochaic (SW: Strong associated with the syllable carrying a primary stress, Weak associated with the syllable carrying a secondary or no stress) or iambic (WS) stress pattern at the lexical level.

In addition to the suprasegmental variation, English unstressed syllables are often accompanied by a reduced vowel such as a schwa /ə/ (e.g. the initial vowel differs for OBject...
vs. object) or a very short form of a full vowel (e.g. the /o/ in shallow) (Cutler & Norris, 1988). The acoustic correlates of vowel quality reduction are determined by formant change. The completely reduced vowel, schwa /ə/, is a centralized vowel in a vowel space with mid-range frequency values for the first formant (F1) and the second formant (F2) (Bolinger, 1989; Flemming & Johnson, 2007; Gay, 1978; B. Lee, Guion, & Harada, 2006).

Although every English content word carries its own lexical stress pattern based on its foot structure, transcription conventions of English intonation structure (Beckman & Ayers, 1997; Beckman & Pierrehumbert, 1986) don’t include foot structure. However, feet are important for the discussion of English lexical stress because it is the head of a foot (i.e., a strong stressed syllable) that serves to anchor a pitch accent at the phrasal level.

In the English ToBI (Tones and Break Indices) annotation system (Beckman & Ayers, 1997), a widely accepted standard transcription convention of English autosegmental intonation structure, there are five different types of pitch accents (H*, L*, L+H*, L*+H, H+!H*) assigned at a phrasal level. Each utterance is composed of at least one intonation phrase (IP), with a boundary tone (H%, L%) at its right edge. A boundary tone is realized

Figure 2.1: Waveform and fundamental frequency contours for an H* pitch accented English lexical stress minimal pair. Left: Trochaic pattern. Right: Iambic pattern.
in the phrase final syllable and characterized by phrase final lengthening and/or a follow-
ing pause. Intonation phrases are exhaustively subdivided into intermediate phrases (ip) ending in a phrase accent (H-, L-). An intermediate phrase must contain at least one pitch accent, and the phrase accent is realized over the region between the final pitch accent and the IP boundary tone. A pitch accent aligns with at least one lexically stressed syllable in each intermediate phrase. Each of the pitch accent types has its own pitch shape to serve a specific pragmatic function in the discourse (Beckman et al., 2005; Pierrehumbert, 1990; Pitrelli, Beckman, & Hirschberg, 1994). H*, a simple high tone, is usually realized in a canonical declarative sentence, while L*, a simple low tone, is often realized in the answer to yes-no questions. The bitonal pitch accent, L+H*, carries a pitch tone rising from low to high and serves to make a contrastive focus, while L*+H, a pitch tone with scooped late rising, is used to mark an item in the utterance that conveys pragmatic uncertainty. H+!H* is a pitch pattern that steps down from a higher pitch at the end of the preceding phrase or from a preceding H pitch accent in the same phrase onto the stressed syllable to indicate pragmatic inference. H+!H* is used when the material preceding the accent is clearly high pitched but unaccented. Otherwise, the pitch accent is marked with a simple !H*.

When there is more than one pitch accent in a phrase, the last and most prominent pitch accent is referred to as a nuclear pitch accent. f0 values on pre-nuclear words without pitch accent are determined by interpolating between tonal targets of adjacent pitch accents in continuous speech (Pierrehumbert, 1990). If words are unaccented in post-nuclear positions, the f0, duration and intensity differences that would distinguish a stressed syllable from an unstressed one tend to be neutralized (Huss, 1978). Therefore, whether a primary stressed syllable gets a higher f0 than other syllables in a word is dependent on its status of being (nuclear) pitch accented in the utterance and the type of pitch accent placed in the
syllable. That is, if the melody were L*+H H% for the one-word utterances, *INsight* and *inCITE*, in Figure 2.1, the second syllable of the word, *insight*, would have a higher mean \( f_0 \) than the second syllable of the word, *incite*.

In sum, though English lexical stress pattern is characterized by \( f_0 \), duration, and intensity cues and vowel quality, the way that lexical stress is realized is also influenced by the prosodic structure at the phrasal level. Pitch accented words are produced with expanded \( f_0 \), duration, and intensity ranges as compared to unaccented words. Since not every word receives a pitch accent in the English intonation system, the words that are not pitch accented lose or reduce their prominence differences between stressed and unstressed syllables in a word. This may explain why native English speakers are reported to less reliably use lexical stress information for word recognition than did native speakers of Dutch/Spanish (Soto-Faraco et al., 2001; Van Donselaar et al., 2005). Also, it may also motivate native English speakers to weight segmental correlates (i.e., reduced vowels) as much as or more than suprasegmental cues for their perception of lexical stress.

### 2.2.2 The Role of Lexical Stress in English Spoken Word Recognition

Many previous studies have explored the role of English lexical stress in spoken word recognition, concerning whether lexical stress as a part of lexical representation facilitates target word recognition. They explored the issue with and without reduced vowels in unstressed syllables and have revealed somewhat mixed results.

Early studies showed that when English listeners heard a member of a stress minimal pair (e.g. *FORbear* or *forBEAR*) in cross-modal priming experiments, the associates of both meanings (e.g. ‘ancestor’, ‘tolerate’) were activated (Cutler, 1986). The result suggests that the prominence differences between syllables are not used in the same way as segmental
information to constrain lexical access in English. Small, Simon, and Goldberg (1988) supported this claim by showing that the mis-stressing word (e.g. $INsert$) didn’t inhibit the activation of its stress competitor word (e.g. $inSERT$) in English listeners. However, the inhibition effect of mismatching stress was found for words which do not have minimal stress pair partners (e.g. $chemist$ pronounced as $cheMIST$).

In connection to the mixed findings in Small et al. (1988), Cutler (2005, p. 279) pointed out that the priming methodology with homophones may not provide the best window on the exploitation of lexical stress information, first because such pure minimal stress pairs are very rare in English, and second because Dutch speaking listeners who are well known to use lexical stress information for spoken word recognition (Cutler & Van Donselaar, 2001; Koster & Cutler, 1997; Leyden, Heuven, et al., 1996; Van Donselaar et al., 2005; Heuven, 1985) also showed similar semantic priming effects for stress minimal pairs. Jongenburger and Heuven (1995) and Jongenburger (1996) who replicated Cutler (1986) to examine the role of lexical stress in Dutch word recognition showed that when Dutch speaking listeners were given initially stressed member of Dutch stress minimal pair words (e.g. $VOORnaam$), they activated both associated meanings (“first name” and “respectable”) of the stress minimal pair words at similar rates.

Given these task-related concerns, other studies tested larger sets of materials with various methodologies and provided evidence that lexical stress plays an important role in activating words during English spoken word recognition (N. Cooper et al., 2002; Lindfield, Wingfield, & Goodglass, 1999; Mattys, 1997, 2000; Slowiaczek, 1990; Small et al., 1988). For example, Connine, Clifton, and Cutler (1987) explored the role of English lexical stress in phoneme perception. When listeners were asked to categorize an ambiguous consonant along a continuum between $/t/-/d/$ in either $TIgress-DIgress$ or $diGRESS-tiGRESS$, more
/d/ responses were obtained from the diGRESS-tiGRESS continuum than from the TIgress-DIgress continuum. That is, phoneme identification was facilitated when the phoneme was presented with appropriate stress information in the source word. This suggests that English lexical representation includes not only phonemic but also prosodic representation.

Cutler and Clifton (1984) and Slowiaczek (1990) that employed lexical decision tasks and shadowing tasks respectively showed that native English listeners responded faster when they were given correctly stressed words (e.g. DIplomat) than when they were given incorrectly stressed words (e.g. diPLOmat).

The facilitation effect of appropriate lexical stress on word recognition was also found in a word-onset gating experiment. Lindfield et al. (1999) examined word identification with gated word onset stimuli that were give in three different conditions. In the first condition, participants were given the first 50 ms of the word onset, and then 100 ms of the word onset and 150 ms of the word onset without any other information in the rest of the word. In the second condition, participants heard the word onset with 50-ms increments as in the first condition but the remaining duration of the word that was filled with white noise. In the third condition, the subjects heard the word onsets with increase in 50-ms increments as in the other two conditions, but the remaining duration of the word was filled with information signaling the number of syllables and the place of syllabic stress. These stimuli were created by using a low bandpass cutoff at 325 Hz, which allowed the detection of the prosodic pattern without revealing the word itself. Lindfield et al. (1999) reported that gated fragments were identified faster when stress information was included than when it wasn’t, especially when the given gate onset was the shortest, 50 ms of word onset. The result suggests that English listeners register lexical stress information in their lexical
representation of words and use the stress information in the speech signal for spoken word recognition.

N. Cooper et al. (2002), as briefly introduced earlier in this chapter, showed that the stress-matching prime facilitated lexical decision not only when it entailed two syllables (e.g. *We were sure the word was ADMIRAL*), but also when it had the initial syllable only (e.g. *MU- for MUSIC*). As the monosyllabic stimuli didn’t contain a reduced vowel in the unstressed syllable, unlike the bisyllabic stimuli that had reduced vowels in the unstressed second syllables, the results showed English listeners can exploit suprasegmental stress cues to activate target words. However, it was also found that the performance was significantly improved with two-syllable primes relative to single-syllable primes, and that there was not a significant inhibition effect of stress-mismatching prime. In addition to the results from native English speakers, they conducted a forced-choice identification experiment with native Dutch listeners. The results showed that the Dutch listeners as non-native listeners (72%) outperformed native English listeners (59%) at correctly assigning monosyllabic fragments to the target words. They insisted that English speaking listeners’ less inefficient use of English lexical stress information compared to Dutch speaking listeners might be because English does not supply its listeners with as much opportunity for suprasegmental processing in word recognition as some other languages do. Words of more than two syllables commonly have full vowels in both first and second syllables in Dutch, whereas such sequences are very rare in English.

With this regard, a recent eyetracking study (Creel, Tanenhaus, & Aslin, 2006) examined the use of lexical stress during lexical access for newly learned nonword minimal stress pairs. They tested near minimal stress pairs which were segmentally the same but were contrastive in stress pattern in the first two syllables (e.g. /kadazu/ vs. /ka'dazen/). The
use of nonwords allowed a wider selection of test items that were prosodically contrastive in the first two syllables than were available in English, while also avoiding frequency and familiarity effects. The stress pattern was manipulated not only by the accent cues but also by vowel quality. The results showed that activation of initial-stress words were more easily and more rapidly activated with less confusion than were word-medial stress words, regardless of vowel quality in the initial syllables. This finding supports the claim that lexical stress information modulates lexical learning in adult English speakers, and provides evidence that the non-initial stress with and without accompanying vowel reduction does not facilitate lexical learning and lexical access in English-speaking listeners.

The studies of English lexical stress were also conducted in relation to its role in word segmentation (Cutler & Norris, 1988; Cutler, 1989; Cutler & Butterfield, 1992). Cutler and Norris (1988), in the proposal of Metrical Segmentation Strategy (MSS), defined “metrical stress” as a sequence of strong (full vowel) and weak syllables (reduced vowel) and hypothesized that English listeners initiate lexical access only when they encounter a strong syllable. The vowel-based definition of strong stress allows the strong syllable to be any syllable that contains a full vowel, whether the syllable is primary stressed (e.g. the first syllable in “fascinating”), secondary stressed (e.g. the first syllable in “hesitation”), or unstressed unreduced (e.g. the first syllable in “fulfill”). The strategy to postulate strong syllables as word onsets was claimed to be a good account for how English speakers segment word boundaries because more than 90% of English words were found to begin with a strong syllable based on the vowel-based distinction of stress (single syllable content words were included as beginning with a strong syllable, (Cutler & Carter, 1987)).
Supporting the claim, Cutler and Norris (1988) showed that English listeners detected the word ‘mint’ faster in ‘mintef’ /mɪntəf/ than in ‘mintayf’ /mɪntɛɪf/ (the latter was constructed to have two stressed syllables). The researchers interpreted the result to indicate that ‘mintayf’ consisted of two stressed syllables and thus was segmented into two words sharing the middle consonant, t, causing longer detection time.

Word spotting experiments in McQueen, Norris, and Cutler (1994) and Norris, McQueen, and Cutler (1995) also provided supportive evidence for the MSS. In these studies, participants were asked to find a word (e.g. mass or sack) in a nonword that has Weak-Strong (e.g. /dəmaes/ for the word mass) or Strong-Weak metrical stress patterns (e.g. /sækraʃ/ for the word sack). Word spotting was faster in Weak-Strong nonwords than Strong-Weak nonwords, suggesting segmentation occurred at the strong syllable of the Weak-Strong nonwords to reveal the target word, mass, but not at the weak syllable of the Strong-Weak nonword, making it more difficult to spot the word, sack.

A cross-splicing word acceptability rating test by Fear, Cutler, and Butterfield (1995) lends further support to claim that English listeners pay more attention to the distinction between full vs. reduced vowels than to the distinction between primary vs. secondary vs. unreduced unstressed vowels. In the study, listeners heard words such as audience, auditorium, audition, addition where initial vowels were cross-spliced and exchanged between words, and rated the acceptability of the cross-spliced words and the unspliced original words. Cross-spliced words were rated significantly less acceptable than unspliced words when reduced vowels were exchanged with other full vowels. Though unstressed unreduced vowels were perceived to be different from both reduced and stressed vowels, they were perceived as more like stressed vowels.
With regard to the role of metrical stress in spoken word recognition, the vowel-based distinction of strong stress allows the prediction that, all else being equal, processing of Weak-Strong words should be more time consuming than processing of Strong-Weak words regardless of the degree of stress in strong syllables. The results from the above studies that showed equal activation of homophone words (Cutler, 1986; Small et al., 1988) are consistent with what the MSS predicts about lexical processing. Word-initial unstressed syllables that do not have reduced vowel information (i.e., secondary stress or unreduced unstressed syllable) were more likely to be perceived as strong syllables and activated the lexical competitors with primary stress in the first syllables.

However, extant methodological issues, especially for the priming paradigm with minimal stress pairs, calls into question this evidence (Cutler, 2005). In addition, even though the studies that explored the processing of lexical stress with and without schwa have shown that words with unstressed initial syllable induced more processing difficulty than the words with initial primary stress (Creel et al., 2006) as predicted by the MSS, the MSS didn’t provide any direct account for how such non-initial primary stress syllables (e.g. “appear”, “fantastic”) are processed during word recognition (Mattys & Samuel, 1997; Mattys, 2000; Mattys & Samuel, 2000). Also, the vowel-based conceptualization of stress doesn’t explain the segmentation process in many English words because many English words contain several full vowels that can potentially lead to detection of several word boundaries within a single word (Mattys & Samuel, 1997; Mattys, 2000; Mattys & Samuel, 2000). Even though it’s been reported that over 60% of English words are embedded in longer words (e.g. dip in diplomat) (P. A. Luce, 1986) and 84% of polysyllabic English words contained at least one word within them (McQueen, Cutler, Briscoe, & Norris, 1995), MSS doesn’t provide any
good accounts for how such polysyllabic English words are processed smoothly (Mattys & Samuel, 1997; Mattys, 2000; Mattys & Samuel, 2000).

Mattys (2000) addressed these problems by exploring whether/how primary vs. secondary stress initial words are different in lexical segmentation and lexical access. In the study, participants guessed the origin of word-initial fragments (e.g. /PRA-/ for prosecutor vs. /pra-/ for prosecution) which are contrastive only by primary vs. secondary initial stress. English listeners’ guessing performance was above chance level whether the word-initial fragments were primary stressed (52%) or secondary stressed (56%), which indicates that English listeners can make fine distinctions of the two types of stress and therefore would be able to use the cues for lexical segmentation and lexical access. Performance was significantly improved when the fragments were composed of the first two syllables of words (59% for initial-primary stress fragments and 65% for initial-secondary stress fragments). This further suggests that having access to a reduced vowel as a reference point increased English listeners’ acuity in stress perception.

Vroomen (1997) also showed that word spotting was delayed when the word onset was secondary stressed than when it was primary stressed, suggesting primary stress is a more potent lexical activator than secondary stress is. Mattys and Samuel (1997) adopted a migration paradigm for phoneme detection tasks that showed that the probability of vowel misperception in a secondary stressed syllable was significantly lower in a word context than in a nonword context, but the lexical effect was not found when the vowel was in a primary stressed syllable. The result provides evidence that primary stress is more reliable cue to lexical access than is secondary stress. In phoneme monitoring tasks by Mattys and Samuel (2000), English listeners’ phoneme monitoring was more delayed when the target phoneme was given in non-initial stress words (e.g. /k/ in campAIGN) than when it
was presented in initial stress words (e.g. /k/ in CAMPus), suggesting that processing of non-initial primary stress words needs extra processing time.

To account for the processing difficulty of non-initial primary stress words, the retroactive processing hypothesis was proposed for spoken word recognition (Cluff & Luce, 1990; P. Luce & Cluff, 1998; Mattys & Samuel, 1997). The hypothesis assumes that processing of late-stress words places a substantial demand on memory to maintain early information until delayed integration is accomplished for word recognition. The integration process requires a speech system that repairs faulty prosody-driven segmentation and lexical access by backtracking to information presented earlier in time, that is, retroactivity and/or delayed commitment to the earlier information.

Mattys and Samuel (2000) tested this hypothesis by increasing memory load during initial vs. non-initial stress lexical processing tasks. If non-initial stress words require more temporary storage than initial stress words, the disruptive effect of memory load was expected to be more evident for non-initial stress words than for initial stress words. In the study, participants were asked to do a dual task designed to increase memory load. On each trial, participants memorized two visual monosyllabic pseudo words (e.g. VONG and TREANE), while performing a syllable detection task and then a visual word memory task. In the syllable detection task, listeners listened to a syllable twice and then a word that contained the syllable (e.g. /n@/ followed by either panoRAma or GEnerator) and were asked to answer “Yes” or “No” to the question of whether they heard the syllable (e.g. /n@/) in the word. The two words used in the phoneme detection tasks shared a syllable but one of them had primary stress in the first or the other had primary stress on the third syllable. In the following memory task, participants were asked to determine whether a word (e.g. FLEEN) rhymed with either one of the pseudowords presented at the beginning.
of the trial (e.g. VONG and TREANE). Syllable detection produced longer response times and lower accuracy for non-initial stress words than for initial stress words, suggesting non-initial stress words cost more processing time and accuracy relative to initial stress words. Rhyme judgment was also less accurate after the non-initial stress words than after the initial stress words, supporting the prediction that the processing of non-initial stress words requires more temporary memory storage than for initial-stress words.

A large number of studies on English lexical stress so far demonstrate that native English listeners use lexical stress information for spoken word segmentation and recognition. Yet, the results were not as consistent across studies as compared to the studies on other languages such as Dutch and Spanish (see the earlier section of Prosody in L1 Spoken Word Recognition), suggesting that native English speakers do not make maximal use of stress information compared to other comparable languages that have lexical stress. This lesser use may be due to several structural aspects of English: the acoustic phonetic realization of English stressed syllables is greatly influenced by pitch accent type and location of the word in an utterance, which are determined at the higher-order level of prosodic structure in the English intonation system, and English speakers often reduce unstressed syllables, providing an additional segmental cue to lexical stress.

Though Mattys (2000) and Mattys and Samuel (2000) provided more specified processing mechanism for primary vs. secondary initial stress words than did the MSS (Cutler & Norris, 1988), both stress-based and vowel-based models agree that initial primary stress facilitates lexical access and initial reduced unstressed syllables hamper lexical access. However, relatively few studies directly examined how native English speakers process non-initial stress words with and without reduced vowels during online spoken word recognition in comparison to initial primary stress words. The current dissertation study addresses this
issue in Chapter 4. If syllables with reduced vowels are perceived as unstressed syllables and so lexical segmentation is delayed as was predicted by the MSS (Cutler & Norris, 1988) and the stress-based model (Mattys, 2000; Mattys & Samuel, 2000), target word activation will be delayed until lexical access based on the late stressed syllable is accomplished.

2.2.3 Korean Intonation Structure

Whether Korean has lexical stress or not has been an issue of considerable controversy. Some argue that there is word-level stress, while others disagree. Among those who claim the existence of lexical stress in Korean, there has been controversy about the location of stress - the first or second syllable (H. Lee, 1973), the second syllable (Huh, 1985), or the final syllable (Choi, 1935). Some further claim that Korean lexical stress is sensitive to syllable weight in such a way that a word-initial heavy syllable is stressed, and if the initial syllable is not heavy, the following second syllable is stressed (H. Lee, 1999; Lim, 2001). According to this heavy syllable view, the stressed syllable is a closed syllable (CVC) or a syllable that is acoustically realized with a longer duration than other syllables.

Since lexical stress is produced with $f_0$, duration, and intensity cues (Lehiste, 1970; Lieberman, 1965; Beckman & Pierrehumbert, 1986), however, Jun (1995b) examined the acoustic cues of stress on the word-initial syllable, /ta-/ or /na-/, at different prosodic positions, and found that the location of so-called ‘stressed’ syllable is dependent on its position in an accentual phrase. The so-called ‘stressed’ syllable was realized with the $f_0$ peak when the word was produced in isolation. However, when the same word was used in an utterance-medial position, high $f_0$ occurred only when the word was placed at the accentual phrase initial position, suggesting Korean stress is phrase-level prosody, not word-level prosody.
A Korean intonation phonological model by Jun (1993, 1995a, 1996, 2000, 2005) posits two hierarchical levels of prosodic phrases above phonological words, that is, accentual phrase (AP) and Intonation phrase (IP). An IP consists of more than one AP, which in turn comprises one or more phonological words. The AP has no phrase-final lengthening or optional pause, and instead is defined by a fixed prosodic pattern, (T) HLHa (LHLHa or HHLHa; -a marks the AP boundary), which demarcates the beginning and the end of a phrase. T is realized as a high tone (H) with a relatively high f0 if the AP-initial consonant is aspirated or tense [+ stiff vocal cords], /h/ or /s/, and as a low tone (L) otherwise, as shown in Figure 2.2.

![Waveform and fundamental frequency contours for a Korean phrasal tone pattern](image)

Figure 2.2: Waveform and fundamental frequency contours for a Korean phrasal tone pattern. Left: Aspirated - AP initial (HHL%), ‘pine tree’. Right: Lax - AP initial (LHL%), ‘nut tree’.

Whether a word-initial syllable is pronounced with a particular tone is dependent on the position of the word in the AP and on the initial segment. Thus, the initial syllable of a word in AP-initial position will be produced and perceived with a L tone if it begins with a lax consonant. However, when the same syllable is pronounced with the word in AP-medial position, it may be produced with an H or interpolated tone (Jun, 2005).
An AP can contain more than one phonological word (i.e., a lexical item followed by prepositions or case markers), but the most common AP contains one word on average and most likely contains 3-4 syllables (Jun et al., 2000; Schafer & Jun, 2002; S. Kim, 2004). If an AP contains three or fewer syllables, the medial L or H or both is likely to be undershot, resulting in a simple rise (LHa), an early rise (LHHa), or a late rise (LLHa) for the L-initial APs, and a high plateau (HHa) or a fall-rise (HLHa) pattern for the H-initial APs. Production data showed that the choice of tones undershot varies across speakers and discourse contexts (Jun, 1996, 2005). An AP final tone is generally Ha, but it is sometimes realized as La before another H-initial AP or before an IP-final AP with a low boundary tone, in which case the AP tonal pattern becomes a falling (HLa, HLLa, or HHLa), a low plateau (LLa), or rise fall pattern (LHLa). The low AP-boundary tone can be realized whether the AP is short or long.

If an AP has five syllables, the first two syllables determine the first two AP tones and the last two syllables mark the last two tones of an AP. The $f_0$ value of the syllable between the second and penult of the AP interpolates the $f_0$ target of the second and penult syllable. H. Lee (1999) showed the first two H tones in HHLHa carry much higher $f_0$ than the H tone in the middle of LHLHa. As for the initial tone, there are more studies which suggest that the second AP tone is loosely linked to the second syllable of an AP. De Jong (2000) showed that the second H tone is realized at the end of the first syllable when the first syllable is a closed syllable. J. Kim et al. (1997) reported that when speakers have a tendency to undershoot L in a five-syllable word, the second H tone can be realized on the third syllable (27 - 40%). If a potential AP has more than six syllables, it is more likely to form two different APs. The different realizations of AP tonal pattern do not have to do with contrastive meaning.
IP boundary tones vary in nine different types in the phonology model of Korean intonation - L%, H%, LH%, HL%, HLH%, LHLH%, LHLHL% (% symbol indicates the right edge of the IP boundary). An IP is usually marked by phrase final lengthening and following optional pause. When an AP is final to an IP, the IP-final syllable is realized with the IP boundary tone, which preempts the AP-final tone. For example, the IP boundary tone L% preempted an AP-final tone,-Ha, in the single-word utterance (’nut tree’) shown in the right panel of Figure 2.2. LH% and H% is different in their timing of rising f0 in the way that LH% rises later than H% by showing a f0 valley at the beginning of the IP-final syllable. The same time-related f0 contour differences are reflected in the transcription labels for HL% vs. LHL%, HLH% vs. LHLH%. It’s been reported that the association between boundary tones and functions quite vary in Korean (H. Lee, 1999; Park, 2003). Typically, the high boundary tone, H%, conveys questions and realized in an interrogative sentence type, while the low boundary tone, L%, is realized for a statement in a declarative form. However, the association between boundary tones and functions are quite vary in Korean. Even a question sentence can carry L% or HL% boundary tone as well as H% or LH%. Also, the realization of IP boundary tones is contingent on sentence endings. Park (2003) reported that the sentence ending in -guna takes HL% and ne takes LH% even though both of them deliver the same meaning of discovery and confirmation.

As for vowel quality in relation to the tonal variation, just as Korean doesn’t employ word prosody, it doesn’t have a vowel reduction process in unstressed syllables, nor the vowel variety distinction of /ɔ/ and /i/. The 7 or 8 canonical Korean vowels, /i, i, e, e, a, o, u/, are all full vowels in Korean phonology (Yang, 1996) and each maintains its vowel quality regardless of tonal variations at the phrasal level (Cho & Jun, 2000; Cho, Jun, & Ladefoged, 2000, 2002; Cho & Keating, 2001).
In sum, Korean intonational phonology differs markedly from English intonational phonology. Korean has an AP between the prosodic word and the IP. The intonational structure of Korean is independent of word-level prosody, and in fact, Korean does not have prosody such as lexical stress, lexical tone, and lexical pitch accent. The tonal shape of a Korean word is contingent on its location in an AP. If a word is in AP-initial position, the phonation type of word-initial segment determines the AP-initial tone. The realization of phrase-medial in an AP is often determined by the syllable weight and the number of syllables in an AP.

### 2.2.4 Korean Speaker’s Use of English Lexical Stress

There have been few studies that explored Korean speaking L2 learners’ perception and processing of English lexical stress during spoken word recognition. The most relevant study was Guion (2005). In this study, she presupposed that late bilinguals from language backgrounds with word level prosody can learn the distribution of lexical stress pattern as a function of lexical class (Guion, Harada, & Clark, 2004) and examined whether Korean L2 learners’ experience with L1 phrase-level prosodic structure adversely affects their ability to associate L2 lexical stress pattern to the proper syllabic structure (i.e., stressed syllables associated with heavy syllables) and lexical class (i.e., nouns with initial primary stress, whereas verb with non-initial primary stress). In the study, native English speakers, and early and late Korean-English bilinguals heard bisyllabic nonwords with four different syllabic structures (e.g., CVV CVCC, CV CVCC, CV CVC, or CV CVVC) in two different contexts, “I’d like a ( )” followed by “I’d like to ( )”, which required noun and verb lexical class, respectively. The participants listened to two versions of a nonword in the same
context (e.g. “I’d like a [ˈbeɪtɪst]” followed by “I’d like a [ˈbeɪtɪst]”) and responded which one sounded the most like a real English sentence.

Results showed that the word with initial primary stress was preferred for the noun context more than for the verb context across all four syllable structures only by native English speakers. Late bilinguals were significantly worse than early bilinguals at mapping lexical stress to the lexical class. However, the preference of stress pattern for the four syllabic structures was similar in both native and non-native English speakers - initial stress was preferred for the nonword with a long vowel in the initial syllable (CVV CVCC), whereas final stress was preferred for the nonword with a long vowel in the final syllable (CV CVVC), and final stress was preferred for the final consonant cluster (CV CVCC) than for the final singletone consonant (CV CVC).

Overall, Korean late bilinguals had less solid knowledge about the distribution of English lexical stress across lexical classes than early bilinguals and, in turn, than native English speakers. Guion (2005) ascribed Korean-English bilinguals’ lack of ability to detect statistical properties of stress patterns across the lexicon to their lack of experience with word prosody in L1. This implies that linguistic experience with L1 prosody affects L2 learners’ use of word prosody in L2 learning and processing. However, we still do not have a good picture of how Korean speakers’ experience with phrase-level prosodic structure influences their processing of lexical stress during spoken word recognition in L2 English.

Previous studies on French speaking learners of Spanish/English suggest the prediction that Korean L2 learners of English would be able to perceive prominence differences between syllables based on the relevant prosodic features but would have difficulty in integrating lexical stress information into lexical access (Dupoux et al., 1997, 2001, 2008; Tremblay, 2008, 2009).
Shin (2008), a preliminary study to the dissertation study, examined the processing of American English lexical stress with native English speakers, Korean L2 speakers, and Japanese L2 speakers. The study employed an eyetracking methodology to observe the perception and processing of lexical stress through the lexical competition over the time course of spoken word recognition. The results showed that, compared to native English speakers and Japanese speaking learners of English, Korean speaking learners of English were less efficient in using lexical stress information in an English word recognition task. The finding was consistent with the PAM’s prediction because Japanese has word-level pitch accents (HL- accented vs. LH- unaccented) corresponding to English lexical stress categories (SW vs. WS) and so the Japanese speakers’ discrimination of English lexical stress categories during word recognition should fit the TC type of assimilation, yielding excellent performance in the word recognition task.

However, Korean speakers’ inability to use English lexical stress cannot be accounted for by the PAM. PAM predicts that discrimination of uncategorizable non-native elements (UU) is less strongly affected by native phonological equivalence classes, and so discrimination should be good. Korean L2 learners’ poor word recognition might be because they couldn’t process the lexical stress information at a phonological level as claimed for French L2 learners of English in Tremblay (2008). However, an alternative prediction is that there is an (as yet unknown) L1 phonological factor that constrains the use of English lexical stress in Korean-speaking L2 learners of English, leading to poorer performance in comparison to Japanese speaking L2 learners of English and native English speakers. This prediction is reasonable because, unlike in French, Korean phrase-level prosody involves some tonal patterns associated with f0 height differences of the phrase initial segments. In
fact, Shin (2008) found that Korean phonological structure, both segmental and autosegmental, had an impact on what Koreans perceived in English as an L2. A post-hoc analysis revealed significantly more recognition of a H-pitch accented trochaic word when the word began with what Koreans heard as aspirated sounds than when it began with what Koreans heard as lax sounds (perception was evaluated in a transcription task). More details are discussed in Chapter 3.

However, since Shin (2008) didn’t aim to examine the phonation type effect, the number of test nonwords was not controlled for the aspirated vs. lax initial contrast in the experimental design, and thus any conclusions remained tentative. To verify how Korean L1 phrase-level prosody affects the processing of lexical stress in L2 English, we explore the issue in the dissertation study.

### 2.3 Research Question and Prediction

The dissertation study expands the preliminary eyetracking study by Shin (2008) to further explore how language-specific prosodic variation influences L2 spoken word recognition. The study particularly concerns whether and how Korean L2 learners’ perception and processing of English lexical stress is influenced by the phrasal prosodic structure of their L1, with experimental designs that directly manipulate variables that are common to the acoustic form of both Korean and English (e.g. High f0 and aspirated/tense initial consonant in word-initial position).

We predict that L1 learned phonetic categories in long term memory constrain the learning and recognition processes available for L2. For example, when Korean learners of English hear a H pitch accented trochaic word that begins with what Koreans hear as a
tense/aspirated sound in an utterance, it might sound familiar, and be available for assimilation to a word with phrase-initial high tone in Korean prosodic structure. This assimilation could serve as a recognition strategy for the trochaic English word, allowing the stressed first syllable to be immediately recognized as a prominent syllable.

The study also investigates how Korean L2 learners of English use reduced vowel schwa to process English lexical stress during spoken word recognition. Regarding Korean L2 learners’ use of English lexical stress and reduced vowels, Yun and Jackson (2006) showed that Korean L2 learners produce English schwa with higher F1 and lower F2 values than native English speakers do, but produce it to be distinctive from other canonical vowels in English and Korean (Yun & Jackson, 2006). Though production doesn’t necessarily reflect perceptual categorization, if Korean speakers can distinguish English reduced vowel, schwa, from most of English vowels as demonstrated in Yun and Jackson (2006), it is predicted that they use the reduced vowel in unstressed syllables as a segmental cue to more easily distinguish the (near) minimal stress pair members such as *OBJ*ect and *ob*JECT during the first syllable than they would do for the stress minimal pair, *IN*sight and *in*CITE. In this case, Korean L2 learners of English can show very similar performance to the one of native English speakers in a word recognition task by distinguishing a target word from its stress competitor early in the word recognition.

To explore these issues, an eyetracking experiment (Chapter 4) and a cross-modal gating experiment (Chapter 5) were conducted. Since the preliminary eyetracking study, Shin (2008) (Chapter 3) examined the processing of lexical stress without involving reduced vowels in unstressed syllables, we employ the same eyetracking methodology to be comparable in our investigation into the role of reduced vowel in processing lexical stress during
spoken word recognition. The fragment gating methodology was chosen for the dissertation because it measures successive responses as the spoken word unfolds, allowing examination of the processing of the initial syllable and subsequent syllables separately. The method is especially useful to examine how the phonation contrast in the initial syllable is processed in the stressed/unstressed initial syllable during word recognition.
CHAPTER 3: PRELIMINARY STUDY

This chapter describes the preliminary eyetracking study by Shin (2008), which investigates how language-specific prosodic variation influences L2 English lexical stress processing during spoken word recognition. This study specifically explored how Korean speaking learners of English process English lexical stress during L2 English word recognition in comparison to a control group of native English speakers. To this end, we addressed the following research questions: 1) how do native English speakers as a control group process lexical stress information for spoken word recognition when reduced vowels are not available in unstressed syllables? Are there any differences between trochaic and iambic stress processing during word recognition? 2) how do Korean L2 learners of English use English lexical stress during L2 English word recognition when reduced vowels are not available in unstressed syllables? Do Korean L2 learners of English show a processing pattern similar to the that of native English, or do they rely primarily on disambiguating segmental information for word recognition?

Previous studies revealed that English speakers can make use of suprasegmental stress cues in word recognition if they are given the opportunity, though not as efficiently as native speakers of other languages that employ lexical stress (N. Cooper et al., 2002; Cutler & Norris, 1988; Mattys, 2000). Native English listeners use lexical stress information more efficiently when they hear the first two syllables in a word than when they hear just an initial syllable (N. Cooper et al., 2002). In addition, Mattys (1997) showed that English listeners are faster to recognize trochaic words than iambic words.. Therefore, it was expected
that English listeners use lexical stress for spoken word recognition, with a processing advantage for trochaic words over iambic words. On the other hand, since Korean speakers do not have any incentive to pay attention to stress in their L1, they were expected to have more difficulty in exploiting suprasegmental stress cues for English word recognition than native English speakers do, relying more on any available disambiguating segmental cues.

The remainder of the chapter is organized as follows: Section 3.1 introduces the eye-tracking methodology with visual word paradigm. Sections 3.2 through 3.4 describe the experimental design and testing materials, participants, and procedures. Section 3.5 describes the statistical methodology used for data analysis (data have been re-analyzed for the dissertation presentation). Sections 3.6 and 3.7 present the results of the preliminary study and discussion. In section 3.8, conclusions and implications of the preliminary study will be discussed in relation to the L1 prosodic effects on processing of L2 suprasegmental stress cues in Korean L2 learners of English.

### 3.1 Methodology

The preliminary study and the dissertation study employed eyetracking methodology with visual world paradigm to investigate how individuals process lexical stress over the time course of spoken word recognition in English as an L1 and an L2. This section presents the rationale for using eyetracking methodology in the visual world paradigm by introducing the underlying assumptions of the visual world paradigm, the working mechanisms of the eye tracker, and the operations of eye movements.
3.1.1 Visual World Paradigm

The visual world paradigm is the most widely used eyetracking paradigm for testing the integration of language and visual attention. In the visual world paradigm, a small set of objects are presented in a well defined scene display, and the participants are typically instructed to look at, click on, or move one of the objects. Participants eye movements are monitored as they listen to spoken utterances containing target items. The timing and pattern of fixations to potential referents in the visual display are used to draw inferences about language comprehension.

The visual world paradigm is based on three basic assumptions about the role of eye movements in scene perception (Findlay, 2004). First, human vision dynamically seeks out visual input as it is needed for cognitive and behavioral activity, as in other animals that have developed a visual system. Secondly, because attention plays an important role in visual processing, and because eye movement is a behavioral manifestation of the allocation of attention in a scene, eye movements to objects in visual world serve as a window into the operation of attention. Thirdly, the most important aspect of eye movement in relation to its application to linguistic research is that eye movements provide an unobtrusive, sensitive, and real time behavioral index of ongoing cognitive processing.

These assumptions have been widely accepted in various disciplines including human-computer interaction, marketing research, vehicle simulation, and even medical research as well as the brain and cognitive sciences (Liversedge & Findlay, 2000). The first visual world eyetracking study was conducted by R. Cooper (1974). In this study, participants heard a story and saw a set of nine pictures. Participants who were given semantically relevant objects in the picture set looked to an object that was semantically relevant to the word revealed in the auditory story. Their look to the relevant picture was made toward the
end of or right after the end of the word in the story. Whereas, the participants who were
given pictures that were not relevant to the story did not show preferential looking to any
of the available pictures.

Since R. Cooper (1974) showed such a close time-related link between visual search
and spoken word recognition, many psycholinguistic studies have adopted the visual paradigm
eyetracking methodology to study language processing at various levels, including phonetic
perception (Dahan, Magnuson, & Tanenhaus, 2001; McMurray, Tanenhaus, & Aslin, 2002;
Salverda, Dahan, & McQueen, 2003; Salverda et al., 2007), word recognition (Alloppenna,
Magnuson, & Tanenhaus, 1998; Magnuson, Tanenhaus, Aslin, & Dahan, 2003), syntactic
processing (Altmann & Steedman, 1988; Kamide, Altmann, & Haywood, 2003; Snedeker
& Trueswell, 2004), and discourse comprehension and production (Ito & Speer, 2008;
Speer & Ito, 2009).

Interest in the visual world studies has been particularly growing for spoken word
recognition studies for several reasons. First of all, existing models and theories of word
recognition can be linked to eye movement behavior over the time course of lexical ac-


55

...
are detected over time between lexical candidates and the continuing speech input. So when listeners hear the word candy, words such as can or candle are also activated but their activation begins to decline at the third consonant of the word candy. Given that the interpretation of all behavior measures depends on a theory or the linking hypothesis that maps the response measures onto the theoretical constructs, the existence of explicit word recognition models and linking hypotheses is an advantage of the visual world eyetracking paradigm for the study on spoken word recognition.

Secondly, the advantage of eyetracking technique lies in the continuity of the measurement. Eye movements provide a continuous measure of spoken language processing in which the response is closely time locked to the speech input of testing phrase without interruptions. The time course measure of eye movement behavior can show where the activation, facilitation, or inhibition of lexical candidate occurs during lexical access. The exact locus of any effects on word recognition such as frequency, rhyme, cohort, neighborhood density effect can be also detected by tracking eye movements in visual word paradigm (Tanenhaus, Magnuson, Dahan, & Chambers, 2000; ?, ?). Given that the existing experimental paradigm such as cross modal priming can test only a part of word at a time, providing a snapshot of language processing, visual word eyetracking methodology has clear advantage over the other methodologies in observing the the dynamics of language processing over time.

Thirdly, sensitivity is another advantage of the eyetracking methodology. Dahan, Magnuson, Tanenhaus, and Hogan (2001) showed that eyetracking methodology is sensitive enough to capture the processing of subtle differences in phonetic cues. In addition, the methodology is sensitive to lexical knowledge outside of the closed set of objects displayed. In Magnuson (2001) and Magnuson et al. (2003), cohort and neighborhood density
effects were found on spoken word recognition even though cohorts and non-cohort neighbors were never pictured or mentioned throughout the course of the experiment, suggesting that eyetracking methodology is sensitive enough to reflect lexical effects coming from the full lexicon, despite the closed set of visually-available items on a given trial.

A fourth advantage is the naturalness of the tasks used for studying word recognition. In natural environments, people usually fixate on an object before they reach for it (?). That is, attention precedes eye movements, and eye movements precede action. Since eyetracking experiments asked participants to move or to grasp an object before performing an action with it, attention and then fixation are directed to an object naturally without requiring any metalinguistic awareness.

Despite the ample advantage of visual world eyetracking paradigm, there is a critical concern about the use of visual world paradigm for spoken word recognition studies. Most notably, the language to be studied has to be picturable, and the number of pictured objects and possible actions is limited (?). However, the picturability problem affects experiment design and choice of stimuli rather than the validity of the paradigm.

3.1.2 Eyetracking Methodology

Eyetracking methodology in the visual world paradigm involves a process of monitoring eye movements around a visual scene in response to concurrent linguistic input.

The measurement device used for measuring eye movements is commonly known as an eye tracker. The eye positions and motions are monitored by an eye tracker. In general, there are two types of eye movement monitoring techniques: those that measure the position of the eye relative to the head, and those that measure the orientation of the eye.
in space, or the “point of regard” (Duchowski, 2007). The latter measurement is typically used when the concern is the identification of elements in a visual scene.

Possibly the most widely applied eye tracker for the visual world paradigm is the video-based corneal reflection eye tracker. The video-based eye tracker comes in two types, head-mounted and table-mounted remote optical eye trackers. The preliminary and dissertation experiments used an ASL E6000 head-mounted eye tracker. The head-mounted tracker utilizes cameras and image processing hardware to compute the fixation point in real-time. A camera focuses on the left eye and records movements as the viewer looks at some kind of visual stimulus. Like other modern eye-trackers, it uses contrast to locate the center of the pupil and infrared and near-infrared non-collimated light to create a corneal reflection (CR). The eye tracker measures the rotation of the eye based on the relative positions of the center of the pupil and the corneal reflection. A head-mounted magnet provides information about the participant’s head location and orientation. The vector between these two features is used to compute gaze intersection with a surface after a simple calibration for an individual participant.

3.1.3 Eye Movements

To understand eye movement behavior in the visual world paradigm, we should know the nature and timing of eye movements as well as eye movement control. Eye movement consists of two temporal phases, fixations and saccades that occur in a succession. The pattern and timing of saccades and the resulting fixations are the most widely used response measures in the cognitive sciences (Liversedge & Findlay, 2000; Rayner, 1998).

Fixations are periods when the eye is held relatively still. During fixation, the eye monitors about 200° visual field but receives highly acute visual information only from the
central area of fovea, which is about 2° of visual angle at the fixation region (Hayhoe, 2000; Henderson & Ferreira, 2004). Fixations last about 200 ms or more depending on the task and the complexity of visual display (Richardson & Spivey, 2004; ?, ?).

The rapid ballistic eye movements between fixation locations are called saccades. On average, eyes move to a new fixation location about 3-5 times per sec during scene viewing though it varies across individuals and across scenes (Henderson & Ferreira, 2004). Since eye movements are motor movements that require time to plan and execute, there is a latency period associated with making saccades. The minimal latency of a saccadic eye movement was estimated to be about 150 ms for a simple visual display and about 200 - 300 ms for multiple target locations or a complicated task (Matin, Shao, & Boff, 1993; ?, ?). During a saccade, the eye is in motion for 20 to 60 ms (?, ?) and the saccadic eye movements are fast, moving 700° of visual angle or better per sec (Carpenter, 1988).

The sensitivity to visual information is dramatically reduced just before, during, and immediately after a saccade (Findlay & Liversedge, 2002). This phenomenon is known as saccadic suppression (Thiele, Henning, Kubischik, & Hoffmann, 2002; Irwin, 2004; Dodge, 1900; ?, ?). This saccadic suppression implies that uptake of visual information occurs only during fixations and no eyetracking data can be collected during saccades. However, given the saccadic latency, the decision to move the eye to a certain location must be made 150 - 200 ms prior to the onset of the saccade. This means that saccades can be considered as part of the following fixation, and eye movement responses in the experiments occur at least 150 - 200 ms after the auditory information that triggered the eye movements. Therefore, when eye movements are the dependent measure in spoken language experiments, it is assumed that saccades are planned and executed on the basis
of speech available before the saccade begins. For the preliminary and dissertation eyetracking study, the delay needs to be taken into account when interpreting eye movement responses.

### 3.2 Design and Materials

As the preliminary eyetracking study aimed to examine how native and non-native English speakers use suprasegmental stress cues during spoken word recognition, the experiment was designed with a 2 * 2 factorial design with Stress pattern (trochaic vs. iambic) as a within-subjects factor and L1 (English vs. Korean) as a between-subjects factor.

#### 3.2.1 Auditory Stimulus Materials

The preliminary eyetracking study used trisyllabic nonword stimuli (CVCVCV). Artificial linguistic materials have been successfully used to provide precise control over distributional information in several domains of language acquisition and processing (Creel et al., 2006; Gomez & Gerken, 1999; Magnuson et al., 2003; Saffran, Newport, & Aslin, 1996). In spoken word recognition, studies with artificial lexicons complement traditional studies by allowing researchers to create stimuli to test precise hypotheses while greatly reducing the set of uncontrolled variables. In the preliminary study, nonwords were used to provide a good number of test stimuli that are near stress minimal pairs and also to control the potential effects of frequency, familiarity, and vocabulary size associated with real words, especially because these variables differ for different L1 speakers (e.g. they may be dependent on curriculum, years speaking the language, etc.).

Forty eight trisyllabic nonwords were constructed. Test nonwords were 16 near stress-minimal pairs, e.g. /dʒəkunaI/ - /dʒaˈkunə/: the first two syllables differed only in stress,
either trochaic (SW) or iambic (WS), and the last syllable was segmentally disambiguating. The last syllable was designed to be disambiguating for the listeners who may not use suprasegmental stress cues for word recognition. The phonemes used in the nonwords were the 8 most frequent phonemes in each position of English trisyllabic words and the Nlog frequency was counted based on the LDC American English Spoken Lexicon (AESL, http://www.ldc.upenn.edu/cgi-bin/aesl/aesl). Table 3.1 shows the eight consonants and eight vowels used to create the test nonwords and the mean Nlog frequency.

<table>
<thead>
<tr>
<th>Syllable</th>
<th>Consonants</th>
<th>Mean Freq. (Nlog)</th>
<th>Vowels</th>
<th>Mean Freq. (Nlog)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>/p, t, k, g, f, s, m, dʒ/</td>
<td>- 2.49</td>
<td>/i, ɪ, e, ə, a, æ, ə, u/</td>
<td>- 3.02</td>
</tr>
<tr>
<td>2</td>
<td>/b, t, k, g, v, s, m, n/</td>
<td>- 2.55</td>
<td>/i, ɪ, ɛ, ə, a, æ, ə, u/</td>
<td>- 3.04</td>
</tr>
<tr>
<td>3</td>
<td>/k, t, d, v, s, n, ʒ, dʒ/</td>
<td>- 2.66</td>
<td>/i, ɪ, e, aɪ, o, ə, ə, ɛ/</td>
<td>- 2.91</td>
</tr>
</tbody>
</table>

Table 3.1: Eight phonemes used in the trisyllabic nonwords and mean Nlog frequency.

The first two syllables in the test nonwords consisted of the same vowels. The reduced vowel /ə/ was not used in the first two syllables of the test nonwords for the experimental purpose despite the high frequency, but they were included in the last syllable. The consonants /r, l, j, w, h/ were excluded as they convey their own formant values and are more likely to contain coarticulatory cues to the perception of the preceding and following segments. The remaining 16 nonwords were fillers that were created to be phonologically unrelated to any of the target members. They were also trisyllabic nonwords, but contained schwas in the first and second unstressed syllables.

The 48 nonwords were recorded in citation form for use in training sessions and also recorded in a carrier sentence, “Click on the (target word) now”, for use in the subsequent eyetracking word recognition task. A phonetically trained female native English speaker
produced the nonwords in a sound attenuated booth. The recordings were made with a QE Headset microphone (Samson SWA3CE). The auditory stimuli were digitally recorded at a sampling rate of 44.1 kHz and 16 bit using Praat (version 5.0.09; Boersma & Weenink, 2008). Members of the same minimal stress pair were produced with the same pitch accent, L+H*. A full list of nonword stimuli is presented in Appendix A.1.

A ToBI trained English native speaker was employed to transcribe the pitch accent type of auditory stimuli, and confirmed that the 48 nonword recordings in isolation and in the carrier sentence were all produced with L+H* pitch accent. Nonwords were screened by native speakers of English and Korean separately to ensure that the nonwords did not sound similar to any real words in English and Korean.

### 3.2.2 Acoustic Phonetic Measurements

Acoustic measures were made on the auditory stimuli for both suprasegmental and segmental lexical stress cues to make sure that the stress minimal pair members were contrasted only in suprasegmental correlates of primary vs. secondary stress, and not in vowel quality.

**Duration, f0, and Intensity**

The suprasegmental acoustic parameters that were traditionally measured and manipulated in perception experiments were duration, average f0, and intensity (Beckman, 1986; Fry, 1958; Lehiste, 1970; Lieberman, 1960, 1965) These measures were therefore made in the first two syllables of the 16 pairs of test nonword tokens as the main acoustic parameters to stress.

**Duration** Vowel duration (ms) was measured from the onset to the offset of the steady-state voiced formant structure in the vowel. The steady-state vowel portion was identified
to exclude portions of the formant transitions from the preceding consonant and to the
following consonant.

*Fundamental Frequency* \( f_0 \) (Hz) was measured as the average value over the vowel
duration using the pitch tracking function in Praat. As the speaker of the stimuli was female,
the pitch range was set to 100-500 Hz as recommended by the Praat manual. Mean \( f_0 \) was
remeasured by changing the pitch range in the Pitch settings when the pitch contour was
absent or displayed incompletely and when the \( f_0 \) value was high or low compared to the
rest of the speaker’s recordings.

*Intensity/Loudness* Tones of the same intensity, but of different frequency, are perceived
as being of different loudness (Fletcher, 1933; Robinson & Dadson, 1956). For example,
in order for a pure tone at 100 Hz to sound as loud as a 20 dB pure tone at 1000 Hz, the
intensity needs to be nearly 50 dB according to the equal loudness contours (ISO, 2003).
To take this frequency dependency into account for loudness measures, we develop the
following algorithm: we extracted a discrete Fourier transform (DFT) spectrum over a
10 ms Hanning window at every 10 ms in each vowel, and convert this spectrum into an
estimation of the excitation pattern. That is, the intensity (dB) against frequency (Hz) is
converted into loudness (phone) against frequency (Bark). Figure 3.1 shows how intensity
in spectrum (left panel) is different from perceptual loudness in excitation pattern (right
panel).

The conversion of the dB per Hz spectrum into the phones per Bark excitation pattern
takes the frequency dependency of the relationship between amplitude and loudness into
account. However, the conversion from dB to phones does not provide a linear scale for
loudness, particularly for amplitude levels that are well above threshold at frequencies that
are far from the reference frequency of 1000 Hz. Therefore, the vector of loudness values
in phones was converted to sones, a unit that estimates loudness relative to a 40 phons reference level. These sone values were then summed to return a single overall loudness value. The Praat manual gives the following formula (see Equation 3.1) for this two-step conversion from the vector of phones per Bark into a single overall loudness estimate in sones.

\[
\int df 2^{(e(F) - 40\text{phon})/10}
\]

(3.1)

Finally, the sequence of loudness values over all of the 10-ms windows across the vowel were averaged, to return the mean loudness in sones.

Paired t-tests showed that trochaic words in isolation had first syllables that on average were longer (130 ms vs. 100 ms), \( t(15) = 6.3, p < .001; \) higher in mean \( f_0 \) (310 Hz vs. 258 Hz), \( t(15) = 14.4, p < .001; \) and greater in loudness (42 sone vs. 30 sone), \( t(15) = \)
6.7, \( p < .001 \), than first syllables of iambic words in isolation. Conversely, iambic words in isolation had second syllables that were longer (170 ms vs. 110 ms), \( t(15) = 7.59, p < .001 \); higher in mean f0 (300 Hz vs. 212 Hz), \( p(15) = 15.2, p < .001 \); and greater in loudness (41 sone vs. 23 sone), \( t(15) = 8, p < .001 \), than second syllables of trochaic words in isolation.

In addition, first syllables were different from second syllables within a word for the two stress conditions: first syllables of trochaic words in isolation were longer (130 ms vs. 110 ms), \( t(15) = 2.3, p < .05 \); higher in mean f0 (310 Hz vs. 212 Hz), \( t(15) = 16.3, p < .001 \); and greater in loudness (42 sone vs. 23 sone), \( t(15) = 4.7, p < .001 \), than the second syllables of trochaic words in isolation. Words with iambic stress patterns had second syllables that were longer (170 ms vs. 100 ms), \( t(15) = 4.1, p = .001 \); higher in mean f0 (300 Hz vs. 258 Hz), \( t(15) = 5.6, p < .001 \); and greater in loudness (41 sone vs. 30 sone), \( t(15) = 2, p < .05 \), than the first syllables.

The same measurements were taken for the test nonwords recorded in the carrier sentence. Paired t-tests showed that trochaic words had first syllables that were longer (110 ms vs. 70 ms), \( t(15) = 11.7, p < .001 \); higher in mean f0 (318 Hz vs. 249 Hz), \( t(15) = 16.2, p < .001 \); and greater in loudness (47 dB vs. 30 dB), \( t(15) = 7.1, p < .001 \), than the first syllables of iambic words. Conversely, iambic words had second syllables that were longer (130 ms vs. 90 ms), \( t(15) = 9.9, p < .001 \); higher in mean f0 (307 Hz vs. 233 Hz), \( t(15) = 12.3, p < .001 \); and greater in loudness (43 dB vs. 29 dB), \( t(15) = 7.8, p < .001 \), than the second syllables of trochaic words.

First syllables were significantly different from second syllables within a word for the two stress conditions: first syllables in trochaic words were longer (110 ms vs. 90 ms), \( t(15) = 2, p = .05 \); higher in mean f0 (318 Hz vs. 233 Hz), \( t(15) = 18.5, p < .001 \); and greater in loudness (47 dB vs. 29 dB), \( t(15) = 3.4, p < .05 \), than the second syllables. Words with
iambic stress patterns had second syllables that were longer (130 ms vs. 70 ms), \( t(15) = 4.3, p < .001 \); higher in mean f0 (307 Hz vs. 249 Hz), \( t(15) = 7, p < .001 \); and greater in loudness (43 dB vs. 30 dB), \( t(15) = 3.3, p < .05 \), than the first syllables.

In addition to the acoustic measures of the nonwords, mean f0 of the word the used in the carrier sentence, “Click on the (target word) now”, was measured in the vowel to ensure that there was no significant early f0 rise before a trochaic target word. Paired t-test revealed no significant difference between the mean f0s of the vowel in the word, the, produced before the trochaic word and the ones produced before the iambic word (232 Hz vs. 227 Hz, \( t(15) = 1.2, p > .05 \)).

Here, as the word-initial phonation type effect (aspirated vs. lax) was examined as a post-hoc analysis for its interaction with lexical stress, mean f0s of the initial syllables were compared to see whether the word-initial syllables were produced with higher f0 in what Koreans heard as aspirated-initial words than in what Koreans heard as lax-initial words for both trochaic and iambic words. The phonation type of word-initial consonants was determined based on Korean participants’ transcription of the nonwords in a listen-and-orthographic transcription task (see Section 3.4 for more details about the task). According to the post-eyetracking orthographic transcription for which Korean L2 learners used Hangul, a Korean phonetic alphabet that can mark the distinction of aspirated, lax, and tense contrast, a large number of targets (10 / 16 target pairs) was transcribed as aspirated-initial words by the Korean L2 learners (98.7% transcription agreement for the phonation type (aspirated vs. lax) of word initial consonants (See Table C.1 in Appendix C). In the same task, English speakers transcribed what Korean speakers heard as aspirated and lax-initial consonants as voiceless and voiced initial consonants, respectively (100% transcription agreement for the voicing contrast (voiceless vs. voiced) of word-initial consonants.
English speakers’ transcription is also consistent with the intended voicing of the English speaker who produced the materials.

T-tests showed that first syllables of aspirated-initial trochaic words produced in isolation were higher in mean f0 (318 Hz vs. 295 Hz), \( t(14) = 3.0, p < .05 \) than first syllables of lax-initial trochaic words. Iambic words with an aspirated-initial sound also had first syllables that were higher in mean f0 (266 Hz vs. 245 Hz), \( t(14) = 3.4, p < .05 \) than the first syllables of lax-initial words. aspirated-initial trochaic words in the carrier sentence also had first syllables that were higher in mean f0 (328 Hz vs. 302 Hz), \( t(13.6) = 3.5, p < .05 \) than first syllables of lax-initial trochaic words. Iambic words with an aspirated-initial sound had first syllables that were higher in mean f0 (262 Hz vs. 228 Hz), \( t(9.4) = 4.2, p < .05 \) than the first syllables of lax-initial words.

The acoustic measurements indicate that suprasegmental cues might conceivably differentiate cohort words that are phonemically identical early on. If stress location is utilized in lexical learning, the cues can be used by listeners to discern the pairs that differ in stress.

**Vowel Quality**

To confirm that the vowels in unstressed syllables were not produced in reduced form, Euclidean distances of the vowels in the first two syllables from the central point of the speaker’s F1 × F2 (Hz) space were measured. F1 and F2 frequencies were taken from the midpoint of the first two vowels in the test nonwords. The measures for the diphthongs were made at the formant transition point between the two vowels in the syllable. Values that differed by more than 200 Hz from the mean for the vowel category were hand checked and corrected if necessary. The central point of F1 × F2 (Hz) vowel space was calculated separately for the vowels in words produced in isolation and for those produced in the carrier sentence with the Equation 3.2. The F1 and F2 frequencies of the vowel space
mid point were obtained separately based on the the minimum and maximum formant frequencies (Hz) that were from the eight vowels produced in stressed syllables.

\[
F_{1 \times 2 \text{ central point}} = \min F{1(0r F 2)} + \frac{(\max F 1(0r F 2) - \min F 1(0r F 2))}{2}
\]

The Euclidean distance of each vowel from the central point was calculated with the Equation 3.3, following the method in Bradlow, Torretta, and Pisoni (1996).

\[
\text{Euclidean distance} = \sqrt{(F 1 - \text{central point } F 1)^2 + (F 2 - \text{central point } F 2)^2}
\]

Paired t-tests showed that first vowels of trochaic and iambic words in isolation were comparable in the distance from the midpoint of the speaker’s vowel space (594 Hz vs. 625 Hz), \(t(15) = 0.5, p > 0.05\). Second vowels of trochaic and iambic words in isolation were also comparable in the vowel dispersion from the midpoint of the speaker’s vowel space (623 Hz vs. 663 Hz), \(t(15) = 0.3, p > 0.05\).

For the test nonwords recorded in the carrier sentence, paired t-tests showed that first vowels of trochaic and iambic words had a comparable vowel dispersion from the midpoint of the speaker’s vowel space (657 Hz vs. 562 Hz), \(t(15) = 1.8, p > 0.05\). Second vowels of trochaic and iambic words were also comparable in vowel space dispersion (583 Hz vs. 606 Hz), \(t(15) = 0.2, p > 0.05\).

The measures of vowel distance from the midpoint of speaker’ vowel space indicate that the prosodic contrast in the first two syllables of the test nonwords didn’t involve a
segmental cue to lexical stress. Thus, if participants are shown to use lexical stress information for lexical learning in the preliminary experiment, they would necessarily rely on suprasegmental cues to lexical stress.

### 3.2.3 Visual Stimulus Materials

Visual stimuli were forty eight line-drawings of space aliens like those shown in Figure 3.2. The source of the alien pictures was Gupta et al. (2004). Aliens were chosen because they have little resemblance to known objects and thus were less likely to evoke memory strategies based on pre-existing images. And they are more interpretable, and thus easier to remember than the abstract shape configurations used in other artificial lexicon studies (Creel et al., 2006; Magnuson et al., 2003).

Alien pictures (120×140 pixel bitmaps) were associated with nonwords in such a way as to control distinctiveness of the aliens trial-by-trial throughout the experiment. Specifically, no two triplet members should have: the same marked features in head (e.g. horns, eyes on ears, tentacles, etc.), body size, body features (e.g. a tail, multiple arms, etc.), clothing type (e.g. long coats, long boots, a cape, soldier looks, etc.), so that pictures presented in the same test trial had an equal distinctiveness. A full list of visual stimuli is presented in Appendix B.

### 3.3 Participants

Twenty-three native English speakers and nineteen Korean-speaking learners of English participated in the experiment. The participants were those who scored accuracy above 90% in the picture choice task on the last day of training and above 75% accuracy in the following eyetracking word recognition task (78.6%, 44 out of 56 participants who
completed all training and eyetracking tasks). This accuracy indicates that participants had acquired the nonwords at a level minimally adequate to assess their processing of prosody during word recognition. Data from two Korean-speaking learners of English were additionally excluded due to poor eye calibration.
Native English speakers were the speakers of midwestern American English aged 18-27 ($M = 21.3$, $SD = 2.08$). They were born and grew up in the U.S., learning English as their L1. Some had history of foreign language study in high school or college, but none of them had lived abroad and no other language was spoken in the home during childhood. They received partial course credit in a linguistic course and a small fee for their participation.

Korean speaking learners of English were undergraduate or graduate students of the Ohio State University, aged 21-34 ($M = 26.9$, $SD = 3.63$) at the time of recruitment. They reported to speak Seoul Korean as an L1 with no residence history in the area of Kyungsang dialect. Unlike the standard Seoul Korean dialect, Kyungsang dialect is known as a tone language, preserving tonal contrasts at the lexical level (Jun, 1996, 2005) and so the participants who had residency history were excluded. None of the Korean speakers had any English-immersion experience before entering the U.S. They were immersed into the U.S. environment at average 24.5 years old ($SD = 3.60$, range from 16 to 30 years of age), and had lived in the U.S. for three months to seven years ($M = 2.5$, $SD = 2.48$) at the time of recruitment. The mean beginning age of learning L2 English in Korea was 13.2 years old ($SD = 1.75$, range from 10 to 17 years old). In the language background questionnaire (see Appendix D), Korean speakers were asked to evaluate their comprehensive English proficiency on a scale from 1 to 4 ($1 = \text{barely}; 2 = \text{poorly}; 3 = \text{passably}; 4 = \text{fluently}$). The average self-evaluation of English fluency was 3 ($SD = .74$, range from 2.2 to 4). They were all recruited voluntarily and paid for their participation.

None of the participants reported speech or hearing disorders and vision impairment at the time of testing.
3.4 Procedures

The experiment consisted of three parts - training (which included a production component), an eyetracking word recognition experiment, and a listening-with-orthographic transcription task. Participants were asked to fill out a language background questionnaire (see Appendix D) prior to the training, and completed three consecutive days of training to learn spoken nonword “names” for the 48 alien drawings. On the third day they also completed a nonword recognition test, during which their eye movements were monitored as they heard the nonword name and chose the matching alien from a set of three drawings (see the bottom panel in Figure 3.2 for an illustration of the display).

3.4.1 Training Session

All training sessions took place in a sound-attenuated booth where participants were seated in front of a Dell desktop computer and a button box, wearing a QE Headset microphone (Samson SWA3CE). Visual stimulus presentation and response recording were implemented by E-Prime software, Version 1.0 (Psychology Software Tools, Inc.). Auditory stimulus materials were presented through a computer-mounted speaker.

Each day participants were presented with alien names in two look-and-listen sessions, the first followed by a picture naming task and the second by a picture choice task to evaluate their learning progress. The look-and-listen session comprised six learning blocks, each of which presented eight alien pictures and the auditory names four times in random order. Participants learned the associations between nonwords and aliens with an auditory name repeated twice for each line drawing of an alien (see the top panel in Figure 3.2). The position of alien pictures was counterbalanced across the trials in a block and only one member of a minimal stress pair was given in a learning block. Individual participants
proceeded to the next trial at their own learning paces by pressing a designated button on the button box.

After completing a learning block in the first look-and-listen task, participants were asked to name a nonword for an alien picture in the carrier phrase, “This is the (target word)”, and then they heard the nonword again as an auditory feedback to their production. They were instructed to speak naturally at a typical rate and loudness level, and allowed to say “I don’t know” if they couldn’t recall the corresponding name of the alien. Production of the 48 nonwords was recorded using Audacity (version 1.3.4, http://audacity.sourceforge.net/) and Project Series Tube MP preamplifier (ART Pro Audio). The speech tokens were sampled at a rate of 22.05 kHz with a quantization of 16 bits and saved as a sound file.

The second look-and-listen task followed the first look-and-listen task. After each of the six learning blocks in the second look-and-listen task, participants completed a picture-choice task, choosing one of three aliens on a screen by pressing a corresponding button on the button box upon hearing a nonword name. Again they were given feedback on their accuracy. The number of appearances of a target alien in the three picture positions was counterbalanced across the trials. Only one member of a minimal stress pair was given in a block.

Across the training sessions and tasks, participants were exposed to the 48 nonwords and alien image referents 10 times per each of the three 1-hour training sessions. Different combinations of eight aliens were given for each block in the second and the third day of training. Thus, for the three consecutive days of training, all the participants were equally exposed to each nonword-picture pair 30 times.
3.4.2 Eyetracking Word Recognition Experiment

The subsequent word recognition experiment was preceded by a 10 minute break after the last picture-choice task. In a separate room, participants were seated at a comfortable distance from a computer monitor with two speakers placed on the left and right side of the computer. They wore a plastic headband with a magnetic receiver and a 60 Hz camera, and their eye position was calibrated before the experiment began. Another 60 Hz camera was mounted to the wall behind the participants and provided a view of the computer screen. Separate screen and eye displays were available for monitoring by the experimenter throughout the experiment. The ASL E6000 head-mounted eye tracker recorded gaze position and duration from the onset to the offset of each trial. A trial was terminated when the participant clicked on a picture object on the computer screen. Stimulus presentation and response recording were made by the E-Prime software, Version 1.0, as in the training session.

In the eyetracking word recognition test, participants were asked to mouse-click on one of the three alien pictures on the computer screen upon hearing the English instruction, “Click on the (target word) now”. The three pictures were the target, with either a trochaic or iambic name (e.g. /dʒakuna/), a stress competitor (e.g. /dʒa'kuna/), and a distractor (e.g. /gæsita/) that was phonologically unrelated to the target and competitor, as shown in the bottom panel of Figure 3.2. Participants were instructed to respond at their own pace. Feedback was not provided. To ensure that each trial began with the participant fixating the neutral central location, participants were asked to click on a central fixation cross to begin the next trial.

Two stimulus lists were used: a participant was given a spoken nonword list which included 8 trochaic nonwords and 8 iambic nonwords with only one member of the stress
minimal pairs. In addition to the 16 test trials, 13 filler trials were placed in pseudo-randomized positions throughout the list. A set of three filler trials was presented at the beginning of the experiment to familiarize participants to the procedure. The positions of target objects were counterbalanced across trials in such a way that the target objects did not appear in the same position in a series of trials.

3.4.3 Listening-with-Orthographic Transcription Task

After completing the eyetracking word recognition test, participants were asked to listen to the instructions again, twice for each word, and write down the target word in their L1 orthography. This task was designed to examine how the participants heard the test stimuli. Korean orthography is “shallow”, that is, it provides a very fine-grained phonetic representation for each corresponding symbol, and thus the phonation type contrast of aspirated, lax, and tense for the word-initial consonants is interpretably transcribed.

3.5 Data Analysis

The preliminary study produced five sets of data. The training session provided production data in the picture naming tasks and accuracy and response time (RT) data in the picture choice tasks. The eyetracking word recognition experiment generated accuracy and RTs, and also gaze data to the three Areas of Interest (AOI), target, competitor, and distractor object positions, over time. Listening-with-orthographic transcription task provided transcription data for test nonwords.

To understand how English lexical stress is processed over the time course of spoken word recognition, the eyetracking data from the word recognition experiment is of the most
interest. However, as reliable learning of the 48 nonwords is prerequisite to the task performance required in the following eyetracking word recognition experiment, the accuracy and RT data from the three-day training sessions are analyzed as a good index of participants’ learning progress. The accuracy rates and RTs for the 32 test nonwords from the eyetracking word recognition were analyzed to examine the general behavior pattern for the test nonwords used in the eyetracking word recognition. The transcription data was used to confirm the phonation type (voicing contrast) of word-initial consonants and used in the post-doc analysis. The production in the picture naming is not be examined for the purpose of the dissertation study.

This section describes the statistical approaches for the analyses of accuracy and RT data and for the analysis of eyetracking data. Section 3.5.1 explains the mixed effects model used for the accuracy and RT behavioral data analyses. Section 3.5.2 describes the reduction and coding of eyetracking data, and a Quasi-Multilevel logistic regression used for eyetracking data analysis.

3.5.1 Accuracy and RT Data Analysis

The data submitted to analysis came from participants whose accuracy on the training picture choice task was above 90% and whose accuracy on the eyetracking word recognition task was above 75%. The accuracy (correct/incorrect responses) and RT data from the three-day training sessions and the ones from the eyetracking word recognition task are analyzed by using a mixed effects model.

The mixed effects model, a type of generalized linear mixed model, is used as it is considered to be more powerful than canonical methods such as repeated measures analysis of variance (ANOVA) (Jaeger, 2008; Baayen, Davidson, & Bates, 2008). Unlike ANOVAs
which accommodates only a continuous dependent variable, a mixed effects model can accommodate continuous dependent variables such as RTs, and categorical dependent variables such as correct/incorrect binary responses, without prior averaging by items or subjects. Also, while ANOVAs can test only categorical predictors, a mixed effect model can simultaneously evaluate categorical and continuous predictor variables. That is, it can test effects that unfold over the time course of an experiment while testing other potentially relevant covariates simultaneously (Baayen et al., 2008). This is an advantage over repeated measures ANOVAs for the analysis of training behavioral data as we are interested in the training effect on learning.

For the training data analyses, we include the data for all 48 nonwords (32 test nonwords and 16 filler nonwords) because the training session aimed to train participants to learn the 48 nonword-picture associations. A mixed effects model fitted in software program R (R Development Core Team) is conducted on the 48 binary correct/incorrect responses (coded as ‘1’/‘0’ respectively)\(^2\) and RTs for the correct responses\(^3\) separately, using the \texttt{lmer()} function from the package \texttt{lme4} (Bates & Sarkar, 2007). The RTs were measured from the onset of each trial, which coincided with the onset of the nonword stimulus. The RTs that were two standard deviations greater or smaller than the mean RT of an L1 group for each day and RTs shorter than 100 ms were removed as outliers.

The effects of the two predictors, \textit{L1} and \textit{Training day}, and the interaction are estimated with coefficients, \textit{Z} scores (accuracy) or \textit{t} values (RT), and \textit{p} values for the intercept and slope terms. Interpretation of the parameter estimates for the categorical predictor depends on the numerical coding scheme used for the predictor in the model. For the current model,

\(^2\text{\texttt{lmer(accuracy \sim lg * day + (1 + lg * day|subj) + (1|item), data, family = “binomial”)}}\)

\(^3\text{\texttt{lmer(rt \sim lg * day + (1 + lg * day|subj) + (1|item), data)}}\)
we employ a treatment coding scheme which is used to compare each level of a categorical predictor (i.e., \( L1 \)) against a reference level - Korean L1 was coded as a “1” and compared against English L1 that was coded as a “0” to be taken as the baseline. \( Training \ day \) is treated as a continuous predictor. The intercept term corresponds to the effect of \( L1 \) on the first day, while the slope term is associated with the effect of the continuous predictor, \( Training \ day \). The corresponding coefficient tells the direction (+/-) and magnitude of an effect or an interaction. The coefficient for the continuous predictor, \( Training \ day \), indicates the effect change as one unit of the continuous predictor increases (e.g. training day for \( Training \ day \)). The Z score, the coefficient divided by its standard error, describes how distant the coefficient estimate is from zero in terms of its standard error. If the Z score is large enough, the test returns significance in the \( p \) value. The \( t \) values are estimated for RTs by taking the parameter estimates and dividing them by their standard errors. The \( p \)-values are calculated by treating the \( t \) values as if they are drawn from a normal distribution and using the \( \text{pnorm} \) function in R.\(^4\)

The mixed effects model uses the link function of logit in Equation 3.4 when treating the binary correct/incorrect response data (coded as ‘1’/‘0’), and thus the estimates of any effects on the binary measures are in the log odds (the natural logarithmic of odds or logit) scale.

\[
\text{Log odds} = \log \left( \frac{Y}{N} \right)
\]  

(3.4)

In Equation 3.4, \( N \) indicates the proportion of trials for which a participant provided incorrect answers for a condition and \( Y \) indicates the proportion of trials for which a participant

\(^4p\) values are calculated by using a R function, \( 2^{*(1-\text{pnorm}(abs(ts)))} \)
provided correct answers for the condition. The logit is centered around 0 corresponding to a probability of .5, and ranges from negative to positive infinity. If the logit is positive, target event (i.e., correct response) is more likely to occur than not \( (p > .5) \), whereas if the logit is negative, the target event is less likely to occur than not \( (p < .5) \).

The analyses of the accuracy and RT data for the eyetracking word recognition task are conducted in the same manner as the ones for the training session, except for the following two modifications. First, the responses submitted to the model are the ones for the 32 test nonwords to be consistent with the ones submitted to the eyetracking data analysis. The RTs were measured from the beginning of the instruction, “Click on the (target word) now”, up to the time point when participants chose one of the three picture referents on the scene display. The RTs that are two standard deviations greater or smaller than the mean RT for an L1 group and RTs shorter than 100 ms are removed as outliers. Secondly, the predictors do not include training day, but instead include some other categorical variables such as Lexical stress and Phonation type. To examine the interaction between the predictors, a sum-coding scheme was used: stress condition (trochaic pattern = -.5 and iambic pattern = .5), phonation type (aspirated (voiceless) initial word = -.5 and lax (voiced) initial word = .5), L1 (English = -.5, Korean = .5). The interaction reflects a contrast between levels of predictors, and the way to test significance is equivalent to the standard ANOVA test for an interaction.

### 3.5.2 Eyetracking Data Analysis

This section describes how the eyetracking data were coded and illustrated, and what/how statistical method was applied to the analysis.
Collection and Coding of Eyetracking Data

Since the eyetracking experiment presented a target, a competitor, and a distractor picture in each trial, there were three AOIs in the top left, mid right, and bottom mid positions in a scene display, as illustrated in Figure 3.2. ASL E6000 software was used to create AOIs for relevant areas of the computer screen, and to calculate the gaze location at a sampling rate of 60 Hz from the beginning to the end of a trial. The sampling rate of 60 Hz samples the eye position once every 17 ms. The beginning of a trial was the onset of the instruction “Click on the (target word) now”, and the end of the trial was the time point when a participant clicked on an object.

As the positions of target, competitor, and distractor picture referents were randomized across trials, a perl script was used to associate the gaze recordings to the corresponding target, competitor, and distractor AOIs specified to each trial. Data were coded as a “1” if there was a look to a position, and as a “0” if there was no look to a position. For instance, if there was a gaze to a competitor up to 150 ms from the onset of the trial, a “1” was coded for the competitor of the trial and a “0” was coded for the target and distractor of the trial at each time frame between 0 ms and 150 ms after the onset of the auditory instruction. Thus, there should be eight gaze recordings for each AOI during the time window. If the participants looked at an area other than the three AOIs or blinked at a given time point, all three AOIs was coded as a “0”.

Visual Presentation of Eyetracking Data

The eyetracking data are presented in two ways for the trials with correct responses. One way is to present mean fixation proportions to the target, competitor, and distractor
over time from the onset of the auditory instruction to the end of analysis window (x-axis). The mean fixation proportions to the three AOIs (y-axis) were computed every 17 ms by dividing the number of trials where a gaze was made to a corresponding AOI by the total number of trials (or items) for a given condition. The grand mean fixation proportion for each AOI was an aggregate of the mean fixation proportions over participants for the condition of interest.

The visual presentation of the mean fixation proportions for the target, competitor, distractor for each condition allows visual inspection of Anticipatory Baseline Effects (ABEs; Barr, Gann, & Pierce, 2011). The ABEs are biases in visual attention existing prior to the presentation of a critical speech stimulus (Barr et al., 2011). If there was already such a gaze preference for the target object to the competitor or distractor object before and at the onset of the target word, the gaze data after the target word onset cannot be as informative as it should be because one cannot make a saccade to what one is already looking at (Brown-Schmidt, 2009, p. 896). According to Tanenhaus, Frank, Jaeger, Salverda, and Masharov (2008), whenever ABEs are present, the estimation of the slope (gaze rate over time) will be biased, such that a higher-baseline condition (a condition that has more such “on-target” trials) shows less of a change in slope than a lower-baseline condition (a condition that has fewer “on-target” trials).

The lack of informativeness of the on-target trial with respect to further processing of input is a critical issue to the current study, because we examine how lexical stress information that is embedded in the first syllable of the target word is processed over time, and ABEs on the target can blur the effect of lexical stress on word recognition. As a solution to the issue, Tanenhaus et al. (2008) proposed eliminating all on-target trials from the data set and performing the analyses on the remaining off-target trials. Though the
elimination of on-target trials may lead into potentially extreme data loss (Barr et al., 2011), if there are any ABEs even after our effort to control such effects through the experimental control such as counterbalancing the position of target across trials and having a neutral fixation point before a trial, we can control the effect by eliminating the on-target trials. The on-target trials are those which gaze to the target AOI is coded as a “1” at the onset of the target word. For the trials which gaze to the target object is coded as a “0” due to eye blink or on-going eye movements toward the target object at the word onset, if there are gaze to the target object at 17 ms before and after the word onset, the trials are also removed for analysis.

Once the ABEs are inspected and controlled, the mean gaze to the target object for a condition of interest is plotted in an empirical log odds scale (x-axis) over time (y-axis). Empirical log odds of gaze make the data presentation be consistent with the Multilevel Logistic Regression (MLR), a statistical model adopted for the eyetracking data analysis (see Section 3.5.2 for details). An empirical logit is computed every 17 ms from the onset of the target word to the end of analysis window using the Equation 3.5.

\[
\text{Empirical logit} = \log \left( \frac{Y + .5}{N + .5} \right) \quad (3.5)
\]

In the equation, \(Y\) is the number of trials for which gaze was on the critical object (i.e., target AOI), while \(N\) is the number of trials where gaze was not on the critical object. The grand mean empirical log odds to the target is computed by aggregation of the empirical log odds over participants for a condition of interest and presented in a plot.

One important issue to the graphic presentation of eye gaze data over time, either in a proportion or an empirical logit scale, is where to align the series of gaze data for different
items and participants. The alignment point is useful to grasp a more precise time point where an expected effect occurs. The alignment is especially important when prosodic cues are manipulated in the same part of the target utterance. For instance, the first syllables of trochaic and iambic words are different in duration - stressed syllables are generally longer than unstressed syllables. To better observe processing differences between the trochaic and iambic first syllables during word recognition, a series of gaze data was aligned at the time point between the first and second syllables of the target word for each item and participant. Aligning between the two critical parts of the target word allows the best possible alignment for both first and second syllables: first syllables are perfectly aligned at their offsets, second syllables at their onsets. The first syllable onset and second syllable offset were estimated by averaging the durations of first and second syllables of trochaic and iambic test words, separately.

**Quasi-Multilevel Logistic Regression**

There have been several different approaches in the field that were employed for eye-tracking data analyses. The most prevalent and canonical eyetracking data analysis method has been the repeated measures analysis of variances (ANOVAs) by subject ($F_1$) and item ($F_2$) analyses. This method treats the continuous time variable as a categorical variable by binning time into a series of consecutive analysis windows and performing a separate analysis for each window.

The binary gaze data are made continuous by calculating proportions or percentages, i.e., the ratios of the number of observed gazes over the number of binary trials for a certain time window. Though many published studies have taken this approach for the eyetracking data analyses, several issues have been pointed out for a long time regarding
the appropriateness and reliability of the method (Baayen et al., 2008; Barr, 2008; Jaeger, 2008; Mirman, Dixon, & Magnuson, 2008).

First of all, though the strength of eyetracking methodology lies in its capability of time-locked observations of human language processing, ANOVAs cannot evaluate time as a continuous measure because ANOVAs can only accommodate categorical predictor variables. This not only goes against the intrinsic nature of the time variable in the eyetracking data, but also treating time as discrete can lead the analyst to overlook important patterns in the data. Secondly, the transformation of binomially distributed gaze data from a categorical to continuous variable is more likely to lead to spurious null effects (Type II error). Jaeger (2008) showed it is the case even with arcsine square root transformed proportions if a substantial number of the proportions is equal to 0 or 1, or if the number of trials is small. 5

To remedy these problems, alternative approaches including mixed effects analyses (Baayen et al., 2008; Jaeger, 2008), growth curve analyses (GCA; Mirman et al., 2008), and multilevel logistic regression analyses (MLR, also known as hierarchical or “mixed effect” approach; Barr, 2008) have been suggested recently for eyetracking data analyses. All these approaches are claimed to be more powerful and legitimate than the canonical method for testing time-related data, accommodating both categorical and continuous predictors and dependent variables.

5When proportions or percentages are used as the dependent variable in an ANOVA or regression, one important assumption is that the population proportion as the percent of outcomes expected over many replications of the experiment should have homogeneous variance. If the computed proportions lie between 0.3 and 0.7, this or any other transformation should have only a small difference in the analytical results. However, when a sizable number of the observed proportions are either relatively small (i.e., 0 <p <0.2) or large (i.e., 0.8 <p <1.0), the variances are more likely to be heterogeneous. The arcsine transformation is known to assist a data analyst when working with proportions and percentages because a proportion can be made nearly “normal” if the square root of proportion is used with the arcsine transformation.
The preliminary and dissertation studies employ the MLR for the eyetracking data analyses. The method was chosen as the other approaches have some weaknesses to the analysis of eyetracking data for the preliminary and dissertation studies. First of all, a mixed effects analysis which includes the binary data observed from all time points in participants’ trials will necessarily be statistically too powerful, producing Type I errors (i.e., detecting an effect when none exists). The reason lies in the eyetracker’s data collection mechanism, which records eye location every 17 ms, that is, many times during a single trial. The nature of eye movements makes the eyetracking observations be more likely to be correlated than other response data (eye movements are ballistic and hence the position of the eye at a given moment in time is related to its previous). Therefore, if the observations are treated as independent, it can underestimate standard errors, causing the Type I errors. The standard solution to this problem is to use proportional data aggregated over participants or items, but but the cost of this approach is a substantial loss of statistical power in the mixed effects model (Barr, 2008).

Secondly, the GCA which is a standard regression model that captures the time effect using fourth-order orthogonal polynomials - the linear, quadratic, cubic, and quartic term respectively reflects a monotonic change in fixation proportion over time, an increase followed by a decrease in fixation proportion, and a steepness of the curve (cubic and quartic terms) - and thus appeals for the precise evaluation of dynamics of processing, cannot capture ABEs at the y-intercept. This is because its intercept term is estimated by mean fixation proportion over the entire analysis time window under the orthogonal polynomial scheme. If it cannot take the gaze advantage at the onset of the analysis window into account when evaluating effects of predictors over time, the interpretation of the slope effect is subject to be confounded especially when the data entails any ABEs. Mirman et al. (2008) says that
GCA can model the y-intercept by using natural polynomials, but this has the disadvantage of losing the independence of the intercept term from the higher order polynomial terms, restricting interpretation of the parameter estimates. In addition to the incompatibility between the intercept term and the higher-order terms, the GCA evaluates predictor effects only on the basis of subject variances, ignoring item variances.

In this light, the MLR method, specifically Quasi-MLR (weighted empirical logit regression in a linear mixed effects model), is used for the preliminary and dissertation eye-tracking data analyses. The MLR (first proposed by Haldane, 1955) follows a logit regression framework - the regression framework enables treatment of time as a continuous variable, while the logistic transformation accommodates the categorical dependent variable of gaze. The Quasi-MLR can solve the nonindependency issue of observations in visual world experiments by allowing aggregation of observations into a series of time bins or into a condition over a certain time bins through the quasi-logit transformations (see the Equation 3.5). Therefore, the inflation of Type I errors noticed in the use of the mixed effects analysis can be avoided in this model.

The Quasi-MLR solves the problems of spurious significances and heterogeneous variances by combining the transformations with an appropriate weighting of observations (Jaeger, 2008). The equation for weighting is in Equation 3.6.

\[ Empirical \, Weights = \frac{1}{Y + .5} + \frac{1}{N + .5} \]  

(3.6)

In addition, unlike the GCA, the MLR fits the slope curve using natural polynomials, and thus the ABEs can be captured in the intercept term that corresponds to the likelihood
of looking at the target region at the very start of the analysis window. The effects related to integration of information over time are given in slope terms that evaluate the gaze changes away from the baseline levels (Barr, 2008; Barr et al., 2011).

The implementation followed the recommendations by Agresti (2002), Barr (2008), and McCullagh and Nelder (1989). The analysis window was first defined by assessing the lowest and highest peak of looks to the target in the grand mean logits from target nonword onset. A grand mean of log odds across all experimental conditions was used to provide unbiased evaluation of the relevant predictors. For the preliminary eyetracking data, the grand mean log odds indicated -153 ms was the earliest time point after word onset where looks to the target begin to rise and 867 ms was the highest peak. The binary gaze data that fell within the analysis window were transformed into the empirical log odds using the empirical logit function in Equation 3.5. The gaze data were aggregated into a series of 50 ms bins (three eyemovement observations at a rate of 60 Hz in each bin because 60 Hz is really measuring every 16.666667 ms, which does come out to be 50 ms when multiplied by 3) up to the level of individual trials for each participant so that we can fit a model with crossed random effects of both subjects and items. The empirical logit transformation was performed for each bin by adding a small constant (i.e., .5) to the numerator and denominator in the fraction as in Equation 3.5. This is because we may encounter zero observations for which the log of zero cannot be specified. To deal with heterogeneous variances, a weighted linear regression with weights (1/empirical weights) was calculated for each bin with the Equation 3.6 (McCullagh & Nelder, 1989).

The time covariates were coded in seconds rather than in milliseconds from the onset of the analysis window, otherwise the parameter estimates would be too small since little happens in the course of a millisecond (Barr, 2008). Since the time variable was aligned
at the offset of the first syllable of target words, the onset of analysis window was coded as a negative time stamp. The negative onset time stamp and subsequent time bins of the analysis window were incremented to make the analysis window begin with 0.00 sec. This makes the data frame prepared for the natural polynomials where the y-intercept is estimated at 0 time point.

The Quasi-MLR analysis was conducted with R code in (1) using the software program R (R Development Core Team) with package lme4 (Bates & Sarkar, 2007).

\[(1) \text{lmer(empirical logit} \sim \text{stress pattern} \times \text{phonation type} \times L1 \times \text{time} + (1|\text{subject}) + (1|\text{item}), \text{data, weights} = 1/\text{empirical weights})\]

The outcomes include main effects and interactions of predictor variables at the intercept and on the slope. The intercept terms show the effects of predictors at the onset of the analysis window (where time equals 0 in the data frame and -153 ms on the graphic representations), and thus represent anticipatory baseline effects if there are any, whereas the slope terms represent the interaction between categorical predictors and the continuous time variable, representing the predictor effects over time. The intercept and slope terms show the effect of predictors in terms of coefficients in empirical logits, \(t\)-value, and \(p\)-value. The coefficient estimates are the index of the direction and size of effects, while \(t\)-value and \(p\)-value indicate the significance level. To examine the interaction between predictors, two-level predictors such as Stress pattern, word-initial Phonation type, and L1 were sum-coded as in the mixed effects model. Time was treated as a continuous predictor.

**Segmented Logistic Regression**

One pitfall of the MLR to the analysis of eyetracking data is that it cannot capture the curvature of slope even if there are any inflection points on the curve. This is because the
MLR is basically a linear regression model. Even the GCA, which is known to be more precise in capturing such curve changes than any other methods, cannot tell the locus of time where effects occur.

In this light, the present study conducted a Quasi-MLR with some time breakpoints to quantify abrupt changes of gaze response by function of any influential predictors. Therefore, the time breakpoint should be estimated at a critical, safe, or threshold time value beyond or below which (un)desired effects occur. As the gaze data were time-locked to the offset of the first syllable (= 0 ms) to better capture the stressed/unstressed initial syllable effect, the first breakpoint was made to coincide with the offset of the first syllable (i.e., -153 - 0 ms). The subsequent breakpoints were made at every 200 ms from the offset of the first syllable as saccade programming latency is estimated to take about 150 - 200 ms (Matin et al., 1993).

For each breakpoint time window, a Quasi-MLR model was conducted without the time variable, using the R code in (2).

\[ \text{subset(data, time < .153), weights = 1/empirical weights} \]

Since the practice of collapsing time into discrete categories can obscure theoretically relevant data patterns by confounding anticipatory effects with rate effects (Barr, 2008), if there are any anticipatory effects in the first time window (-153 - 0 ms), the interpretation of corresponding effects or interactions in the subsequent 200 ms windows should be made with caution.
3.6 Results

3.6.1 Training Result

Participants in the experiment are those who reached accuracy rates above 90% for the 48 nonwords given in the picture-choice task on the last day of training and above 75% for the 32 nonwords tested in the eyetracking word recognition task. Table 3.2 shows the mean percent accuracy for the 48 nonwords and the mean RT for the correct responses tested on each training day for each L1 group. Outlier RTs were discarded (4.2%, 243 / 5775 correct responses). Both native English speakers and Korean L2 learners of English showed increasing accuracy and decreasing RTs as they took more training sessions over the three days.

<table>
<thead>
<tr>
<th>Training day</th>
<th>English speakers</th>
<th></th>
<th>Korean speakers</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Accuracy</td>
<td>RT</td>
<td>Accuracy</td>
<td>RT</td>
</tr>
<tr>
<td>Day 1</td>
<td>89.6 (0.92)</td>
<td>1931.7 (25.69)</td>
<td>90.1 (0.99)</td>
<td>2216.3 (32.28)</td>
</tr>
<tr>
<td>Day 2</td>
<td>97.3 (0.49)</td>
<td>1499.7 (19.12)</td>
<td>98.7 (0.38)</td>
<td>1652.4 (19.93)</td>
</tr>
<tr>
<td>Day 3</td>
<td>98.5 (0.37)</td>
<td>1299.7 (14.96)</td>
<td>99.0 (0.33)</td>
<td>1413.6 (16.71)</td>
</tr>
</tbody>
</table>

Table 3.2: Mean percent accuracy (%) and mean RT (ms) by L1 and Training day. Values in parentheses represent standard errors.

To examine whether native English speakers and L2 learners made comparable learning progress for the nonword and picture associations, a mixed effects model was separately conducted on the binary correct (coded as a “1”)/incorrect (coded as a “0”) responses and the RTs for the correct responses presented in the picture choice task in all three training days, with L1 and Training day as two predictors and Subject and Item as crossed random effects. The results are presented in Table 3.3.
The results on accuracy revealed a significant main effect of Training day with no effect of L1 and no interaction between L1 and Training day. The positive coefficient for the Training day effect on accuracy ($\beta = 1.157$, $p < .0001$) indicates that participants were more accurate in identifying nonword-picture associations as they had more training sessions over the three days. The accuracy rate was not significantly different between English and Korean speakers from the first day, being improved at an equal rate throughout the three training sessions.

The results of RTs show main effects of L1 and Training day, and also a significant interaction between L1 and Training day. The negative coefficient for Training day effect on RTs ($\beta = -318.56$, $p < .0001$) suggests that the RTs decreased as participants had more training over the three training days. The positive coefficient L1 ($\beta = 262.74$, $p < .05$) indicates that Korean L2 learners took longer to make the picture choices on the first day of training than did native English, but the negative estimate parameter for the interaction with Training day ($\beta = -90.91$, $p < .05$) demonstrates that the L1 difference on the first day of training was significantly decreased toward the end of the three-day training session.
Overall, the results showed that both native English speakers and Korean L2 learners made comparable learning progress for the 48 nonword and picture associations, with similar similar accuracy increases and response time decreases by the end of the three-day training.

### 3.6.2 Eyetracking Result

#### Accuracy and RT

The mean percent accuracy for the twenty nine trials including sixteen test trials and thirteen filler trials in the eyetracking word recognition task was 88.16% \( (SE = 1.25) \) for English speakers and 90.74% \( (SE = 1.24) \) for Korean speakers.

To examine how native English speakers and Korean L2 learners were different in their accuracy rates and RTs for the trochaic and iambic words during the word recognition task, the mean percent accuracy for the sixteen test trials and the mean RT of the correct responses are presented in Table 3.4. The accuracy and RTs for the two stress condition are also presented by the function of phonation type that was a predictor in the post-hoc analysis. Outlier RTs were discarded (3.8%, 21 out of 552 correct responses).

<table>
<thead>
<tr>
<th>Stress</th>
<th>Phonation</th>
<th>English speakers</th>
<th>Korean speakers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Accuracy</td>
<td>RT</td>
</tr>
<tr>
<td>Trochaic</td>
<td>Aspirated</td>
<td>85.2 (3.32)</td>
<td>2507.7 (79.37)</td>
</tr>
<tr>
<td></td>
<td>Lax</td>
<td>76.8 (5.12)</td>
<td>2269.9 (100.94)</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>82.1 (2.84)</td>
<td>2428.4 (63.19)</td>
</tr>
<tr>
<td>Iambic</td>
<td>Aspirated</td>
<td>84.4 (3.40)</td>
<td>2310.6 (70.92)</td>
</tr>
<tr>
<td></td>
<td>Lax</td>
<td>68.1 (5.65)</td>
<td>2607.8 (121.55)</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>78.3 (3.05)</td>
<td>2411.8 (63.35)</td>
</tr>
</tbody>
</table>

Table 3.4: Mean percent accuracy (%) and mean RT (ms) by \( L1, \) Stress pattern, and Phonation type. Values in parentheses represent standard errors.
The mean percent accuracy was higher for trochaic words than for iambic words in both native English speakers and Korean L2 learners of English. However, the processing times for the two stress conditions were similar to each other in both native English speakers and L2 learners.

A mixed effects model was separately conducted on the binary correct (coded as a “1”)/incorrect (coded as a “0”) responses and the RTs of the correct responses with L1, Stress pattern, Phonation as three predictors and Subject and Item as crossed random factors. Table 3.5 shows the results.

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Est.</th>
<th>SE</th>
<th>Wald Z</th>
<th>p</th>
<th>Est.</th>
<th>SE</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1.640</td>
<td>0.141</td>
<td>11.661</td>
<td>.0001</td>
<td>2534.82</td>
<td>79.32</td>
<td>31.96</td>
<td>.0001</td>
</tr>
<tr>
<td>L1 = Korean L1</td>
<td>0.444</td>
<td>0.281</td>
<td>1.579</td>
<td>n.s.</td>
<td>208.21</td>
<td>124.88</td>
<td>1.67</td>
<td>n.s.</td>
</tr>
<tr>
<td>Stress1 = Iambic</td>
<td>-0.408</td>
<td>0.224</td>
<td>-1.822</td>
<td>n.s.</td>
<td>68.62</td>
<td>69.05</td>
<td>0.99</td>
<td>n.s.</td>
</tr>
<tr>
<td>Phonation1 = Lax</td>
<td>-0.449</td>
<td>0.224</td>
<td>-2.007</td>
<td>.05</td>
<td>44.76</td>
<td>119.75</td>
<td>0.37</td>
<td>n.s.</td>
</tr>
<tr>
<td>L1:Stress1</td>
<td>-0.287</td>
<td>0.450</td>
<td>-0.640</td>
<td>n.s.</td>
<td>-16.16</td>
<td>137.87</td>
<td>-0.12</td>
<td>n.s.</td>
</tr>
<tr>
<td>L1:Phonation1</td>
<td>0.639</td>
<td>0.450</td>
<td>1.428</td>
<td>n.s.</td>
<td>-45.88</td>
<td>138.21</td>
<td>0.33</td>
<td>n.s.</td>
</tr>
<tr>
<td>Stress1:Phonation1</td>
<td>-1.017</td>
<td>0.450</td>
<td>-2.272</td>
<td>.05</td>
<td>313.87</td>
<td>137.55</td>
<td>2.28</td>
<td>.05</td>
</tr>
<tr>
<td>L1:Stress1:Phonation1</td>
<td>-1.254</td>
<td>0.895</td>
<td>-1.401</td>
<td>n.s.</td>
<td>-334.33</td>
<td>274.71</td>
<td>-1.22</td>
<td>n.s.</td>
</tr>
</tbody>
</table>

Table 3.5: Summary of L1, Stress pattern, and Phonation type effects on accuracy and RTs in the mixed effects model.

The results for accuracy show no significant effects of L1 and Stress pattern. An interaction between L1 and Stress pattern was not significant. However, there was a significant effect of Phonation type. The negative coefficient for the Phonation type effect ($\beta = -0.449$, $p < .05$) with no significant interaction with L1 indicates that both native speakers and L2 learners were more accurate for aspirated-initial word identification than for lax-initial word identification in the eyetracking word recognition task. The interaction
between Stress pattern and Phonation type was also significant with no interaction with L1. The negative coefficient ($\beta = -1.017$, $p < .05$) indicates that the Phonation type effect was stronger for trochaic words than for iambic words in both native English speakers and Korean L2 learners.

The results for RTs show no main effects and no interaction between L1 and Stress pattern. This suggests that the response times for the trochaic word recognition and the iambic word recognition were not significantly different, and native speakers and L2 learners were similar in the response latencies for the two stress conditions. There was not an effect of Phonation type, but the interaction between Stress pattern and Phonation type was significant. As there was not a significant interaction between Stress pattern, Phonation type, and L1, the positive coefficient for the interaction between Stress pattern and Phonation type ($\beta = 313.87$, $p < .05$) indicates that both native English speakers and Korean L2 learners took longer to recognize lax-initial words than aspirated-initial words for iambic words than for trochaic words.

In sum, the accuracy and RT behavioral data analyses of the eyetracking experiment show that both native speakers and L2 learners were comparable in their accuracy rates and response times for both trochaic and iambic word recognition. However, word-initial Phonation type had some effects on the Stress pattern in both accuracy and response times. Both native English speakers and Korean L2 learners were more accurate and faster in recognizing aspirated-initial words than lax-initial words for trochaic word recognition than for iambic word recognition.
Eyetracking Data

Mean fixation proportions for the target, competitor, distractor AOIs of the trials with correct responses were first plotted over time from the beginning of the auditory instruction, “Click on the (target word) now” for visual inspection to detect any ABEs. English speakers provided a total of 151 correct responses for trochaic word recognition and 144 correct responses for iambic word recognition. Korean speakers provided 132 correct responses for trochaic word recognition and 125 correct responses for iambic word recognition.

Figure 3.3 presents native English speakers’ mean fixation proportions for the three AOIs during trochaic and iambic word recognition and Figure 3.4 presents Korean L2 learners’ fixation proportions. The data were time-aligned trial-by-trial, with the first syllable offset at 0 ms (shown with a solid vertical line in the figures). Dotted vertical lines indicate mean syllable durations. The mean durations of the first two syllables are 238 ms (SD = .02) and 164 ms (SD = .03) for trochaic words and 180 ms (SD = .02) and 230 ms (SD = .04) for iambic words. Thereby, the first and second vertical dotted lines respectively correspond to the mean onset of first syllable and the mean offset of second syllable.

Looking at Figure 3.3, we can find no noticeable difference in fixation proportion between the target and competitor objects at the beginning of the target word for both trochaic (-238 ms in the left panel) and iambic stress conditions (-180 ms in the right panel). One-way ANOVAs conducted on the fixation proportions by subjects and by items at the word onset also showed no significant difference between targets and competitors in fixation proportion for both trochaic ($F1(1,22) = 0.015, p > .05; F2(1, 15) = 0.017, p > .05$) and iambic word conditions ($F1(1,22) = 1.393, p > .05; F2(1, 15) = 0.136, p > .05$). In Figure 3.4, Korean L2 learners also show almost equal distribution of fixation proportions for the target
Figure 3.3: English speakers’ \((N = 23)\) mean fixation proportion for target, competitor, and distractor objects during word recognition. Left: Trochaic nonwords. Right: Iambic nonwords.

Figure 3.4: Korean L2 learners’ \((N = 19)\) mean fixation proportion for target, competitor, and distractor objects during word recognition. Left: Trochaic nonwords. Right: Iambic nonwords.
and competitor objects at the word onset for both trochaic and iambic stress conditions. ANOVA tests also showed no significant difference in fixation proportion between the target and competitor objects at the onset of the words for both trochaic \((F1(1,18) = 0.173, p > .05; F2(1, 15) = 0.111, p > .05)\) and iambic conditions \((F1(1,18) = 0.406, p > .05; F2(1, 15) = 0.018, p > .05)\).

As there was not a substantial constant advantage in gaze probability to the target over the competitor for English and Korean participants, all gaze data to the target object of correct responses was included in the analyses. Figure 3.5 shows the mean empirical logits of gaze to the target object for the two stress conditions at every 17 ms for native English speakers (left panel) and Korean L2 learners of English (right panel). In each plot, the gaze data were time-aligned with the boundary between the first and second syllables at 0 ms. The dotted lines before and after the alignment time point (i.e., 0 ms) correspond to the mean onsets of first syllables and the mean offsets of second syllables for trochaic words (black lines) and iambic words (gray lines).

The left panel of Figure 3.5 shows that English speakers’ mean fixation likelihood for the target object was higher for the trochaic condition than for the iambic condition at around the offset of the first syllable (0 ms). The relatively higher gaze to trochaic over iambic target objects started off toward the end of the first syllable, while the gaze rate of iambic words was relatively static until the end of the second syllable. This suggests that trochaic stress pattern provided a processing advantage for the target word activation at an early stage of lexical processing. However, the fixation likelihood of iambic words increased to be temporarily higher than that of trochaic words at around the offset of the second syllable (230 ms). Given that it takes 150 - 200 ms to plan and execute a saccadic eye movement (Matin et al., 1993), this finding suggests that English speakers used stressed
syllable information to initiate the activation of target word - the first syllable of the trochaic word and the second syllable of the iambic word, leading to the processing advantage for trochaic word recognition over iambic word recognition.

Korean speakers’ fixation likelihood in the right panel of Figure 3.5 shows that the fixation likelihoods of the two stress conditions remained relatively static up to the end of the second syllable (at around 180 ms) and increased almost in parallel shortly after that time point. This suggests that Korean L2 learners needed at least the first two syllables of nonwords to activate the target words and they could use prominence differences between syllables to initiate target word recognition. The parallel increase of fixation rates was then followed by higher fixation likelihood of trochaic words over iambic words at around 350 ms. The processing advantage was made after they heard the disambiguating last syllable, but it’s not reasonable to ascribe the advantage solely to the last syllable as the last
syllable was equally disambiguating in both stress conditions. Therefore, the right panel in Figure 3.5 suggests that Korean speakers could use lexical stress information for target word recognition, and they also seemed to show a processing advantage for trochaic over iambic words though it was much delayed compared to that of native English speakers.

To test the effects of lexical stress pattern and L1 on word recognition, the gaze data to the target AOI were submitted to a Quasi-MLR within the analysis window from -153 ms to 867 ms. The model was conducted with $L1$, $Stress$ pattern, and $Time$ as predictor variables and $Subject$ and $Item$ as crossed random effects. In addition to the three predictors, the predictor, $Phonation type$, was also included in the same model to estimate the $Phonation type$ effect together with other predictors, but will be discussed later in the section as a post-hoc analysis. The three categorical predictors, $L1$, $Stress$ pattern, and $Phonation type$ were sum coded as described in 3.5.2. Table 3.6 shows the results.

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Est.</th>
<th>SE</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-0.904</td>
<td>0.127</td>
<td>-7.133</td>
<td>.0001</td>
</tr>
<tr>
<td>$L1 = Korean L1$</td>
<td>0.049</td>
<td>0.199</td>
<td>0.246</td>
<td>n.s.</td>
</tr>
<tr>
<td>Stress1 = Iambic</td>
<td>-0.214</td>
<td>0.100</td>
<td>-2.136</td>
<td>.05</td>
</tr>
<tr>
<td>Phonation1 = Lax</td>
<td>0.052</td>
<td>0.187</td>
<td>0.279</td>
<td>n.s.</td>
</tr>
<tr>
<td>L1:Stress1</td>
<td>-0.059</td>
<td>0.200</td>
<td>-0.290</td>
<td>n.s.</td>
</tr>
<tr>
<td>L1:Phonation1</td>
<td>-0.650</td>
<td>0.200</td>
<td>-3.246</td>
<td>.001</td>
</tr>
<tr>
<td>Stress1:Phonation1</td>
<td>-0.705</td>
<td>0.200</td>
<td>-3.522</td>
<td>.0001</td>
</tr>
<tr>
<td>L1:Stress1:Phonation1</td>
<td>0.638</td>
<td>0.400</td>
<td>1.593</td>
<td>n.s.</td>
</tr>
<tr>
<td>Time</td>
<td>0.952</td>
<td>0.089</td>
<td>10.656</td>
<td>.0001</td>
</tr>
<tr>
<td>L1:Time</td>
<td>-0.438</td>
<td>0.179</td>
<td>-2.453</td>
<td>.05</td>
</tr>
<tr>
<td>Stress1:Time</td>
<td>0.185</td>
<td>0.179</td>
<td>1.035</td>
<td>n.s.</td>
</tr>
<tr>
<td>Phonation1:Time</td>
<td>-0.247</td>
<td>0.179</td>
<td>-1.382</td>
<td>n.s.</td>
</tr>
<tr>
<td>L1:Stress1:Time</td>
<td>-0.338</td>
<td>0.357</td>
<td>-0.947</td>
<td>n.s.</td>
</tr>
<tr>
<td>L1:Phonation1:Time</td>
<td>0.596</td>
<td>0.357</td>
<td>1.668</td>
<td>n.s.</td>
</tr>
<tr>
<td>Stress1:Phonation1:Time</td>
<td>1.371</td>
<td>0.357</td>
<td>3.839</td>
<td>.0001</td>
</tr>
<tr>
<td>L1:Stress1:Phonation1:Time</td>
<td>-1.848</td>
<td>0.714</td>
<td>-2.587</td>
<td>.05</td>
</tr>
</tbody>
</table>

Table 3.6: Summary of $L1$, $Stress$ pattern, $Phonation type$, and $Time$ effects on gaze to the target object (empirical logit scale) in Quasi-MLR ($N = 11040$; logLik = -17634).
At the intercept, there was only a main effect of *Stress pattern*. Because the variable was coded as trochaic = -.5 and iambic = .5, the negative parameter estimate ($\beta = -0.214, p < .05$) implies that the chance of looking at the target object at the intercept was higher for trochaic words than for iambic words. An effect of *L1* and an interaction between *L1* and *Stress pattern* were not significant at the intercept, suggesting that native English speakers and L2 learners were comparable in the fixation likelihoods for the two stress conditions at the intercept.

For the slope estimate for the continuous predictor, *Time*, there was a significant effect. The positive coefficient ($\beta = 0.952, p < .0001$) implies that the overall fixation likelihood was significantly increased as one unit of the time variable was increased (i.e., slope change). An interaction between *Time* and *L1* was found. The negative coefficient ($\beta = -0.438, p < .05$) indicates that the increase rate of fixation likelihood for the target object over time was higher for English speakers than for Korean speakers, consistent with their faster performance overall. An effect of *Stress pattern* on the slope was not significant, showing that the fixation likelihood to the target object for iambic words increased over time at a similar rate to the one for trochaic words. The interaction between *L1* and *Stress pattern* on the slope was not significant, either. This indicates the native English speakers and Korean L2 learners were not significantly different in the increase of fixation rates for the two stress conditions.

The effects of *L1* and *Stress pattern* and their interaction on the slope allow observation of how the two lexical stress patterns were processed by native speakers and L2 learners over the time course of spoken word recognition. So far, the effect on the slope was significant only for *L1*, with higher fixation likelihood increase for native English speakers than for Korean L2 learners.
However, looking at the left panel of Figure 3.5 and Figure 3.4, Korean listeners’ shallower slope seems to be primarily driven by the low increase of fixation likelihood for iambic word recognition. Though a significant interaction between \( L1 \) and \textit{Stress pattern} on the slope was not detected, its negative coefficient (\( \beta = -0.338, p > .05 \)) implies that the difference between native English speakers and L2 learners in the slope change was stronger for iambic word recognition. In addition, the slope changes over time for the two stress patterns do not look quite linear, especially for Korean L2 learners, with some fluctuations in the slope change over time.

To explore how the processing of two stress patterns changed over the time course of word recognition, a Quasi-MLR was conducted for a series of segmented time windows -153 - 0 ms corresponding to the first syllable and every 200 ms window after the offset of first syllable (0 - 200 ms, 200 - 400 ms, 400 - 600 ms, and 600 - 867 ms). \( L1 \), \textit{Stress pattern}, and \textit{Phonation type} were the predictors and \textit{Subject} and \textit{Item} were the crossed random effects. Table 3.7 summaries the results.

In the first two time windows, -153 - 0 ms and 0 - 200 ms, there were not significant effects of \( L1 \) and \textit{Stress pattern}, nor the interaction between \( L1 \) and \textit{Stress pattern}. Given that saccadic eye movements take approximately 150 - 200 ms to plan and execute (Matin et al., 1993), the results indicate the stress information in the first syllable did not have a robust effect on the activation of target words in either English or Korean speakers. In the window from 200 to 400 ms which approximates the time after the second syllable and in the following window 400 - 600 ms, interactions of \( L1 \) and \textit{Stress pattern} were significant (\( p < .05 \)), with no other effects and interactions approaching significance. The negative coefficients for the interaction between \( L1 \) and \textit{Stress pattern} (\( \beta = -0.462 \) for 200 - 400 ms; \( \beta = -0.502 \) for 400 - 600 ms) indicate that \( L1 \) effect was stronger for iambic words than for
<table>
<thead>
<tr>
<th>Predictor</th>
<th>Est.</th>
<th>SE</th>
<th>t</th>
<th>p&lt;</th>
<th>Est.</th>
<th>SE</th>
<th>t</th>
<th>p&lt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analysis window</td>
<td>-153 ms - 0 ms</td>
<td>(N = 1656; logLik = -2482)</td>
<td>0 ms - 200 ms</td>
<td>(N = 2208; logLik = -3407)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>-0.863</td>
<td>0.184</td>
<td>-4.684</td>
<td>.0001</td>
<td>-0.750</td>
<td>0.160</td>
<td>-4.687</td>
<td>.0001</td>
</tr>
<tr>
<td>L1 = Korean L1</td>
<td>0.082</td>
<td>0.261</td>
<td>0.313</td>
<td>n.s.</td>
<td>-0.085</td>
<td>0.251</td>
<td>-0.340</td>
<td>n.s.</td>
</tr>
<tr>
<td>Stress1 = Iambic</td>
<td>-0.222</td>
<td>0.121</td>
<td>-1.832</td>
<td>n.s.</td>
<td>-0.158</td>
<td>0.110</td>
<td>-1.434</td>
<td>n.s.</td>
</tr>
<tr>
<td>Phonation1 = Lax</td>
<td>0.016</td>
<td>0.287</td>
<td>0.056</td>
<td>n.s.</td>
<td>-0.010</td>
<td>0.227</td>
<td>-0.044</td>
<td>n.s.</td>
</tr>
<tr>
<td>L1:Stress1</td>
<td>-0.030</td>
<td>0.242</td>
<td>-0.119</td>
<td>n.s.</td>
<td>0.071</td>
<td>0.221</td>
<td>0.323</td>
<td>n.s.</td>
</tr>
<tr>
<td>L1:Phonation1</td>
<td>-0.496</td>
<td>0.242</td>
<td>-2.054</td>
<td>.05</td>
<td>-0.622</td>
<td>0.221</td>
<td>-2.816</td>
<td>.05</td>
</tr>
<tr>
<td>Stress1:Phonation1</td>
<td>-0.487</td>
<td>0.241</td>
<td>-2.024</td>
<td>.05</td>
<td>-0.359</td>
<td>0.220</td>
<td>-1.631</td>
<td>n.s.</td>
</tr>
<tr>
<td>L1:Stress1:Phonation</td>
<td>0.854</td>
<td>0.481</td>
<td>1.774</td>
<td>n.s.</td>
<td>0.245</td>
<td>0.440</td>
<td>0.557</td>
<td>n.s.</td>
</tr>
</tbody>
</table>

| Analysis window | 200 ms - 400 ms | (N = 2208; logLik = -3541) | 400 ms - 600 ms | (N = 2208; logLik = -3584) |
| Intercept | -0.446 | 0.160 | -2.796 | .05 | -0.203 | 0.170 | -1.192 | n.s. |
| L1 = Korean L1 | -0.174 | 0.218 | -0.797 | n.s. | -0.165 | 0.245 | -0.673 | n.s. |
| Stress1 = Iambic | -0.083 | 0.116 | -0.714 | n.s. | -0.143 | 0.118 | -1.207 | n.s. |
| Phonation1 = Lax | -0.035 | 0.260 | -0.134 | n.s. | -0.026 | 0.264 | -0.097 | n.s. |
| L1:Stress1 | -0.462 | 0.231 | -2.000 | .05 | -0.502 | 0.236 | -2.126 | .05 |
| L1:Phonation1 | -0.375 | 0.232 | -1.619 | n.s. | -0.280 | 0.236 | -1.185 | n.s. |
| Stress1:Phonation1 | -0.363 | 0.231 | -1.571 | n.s. | 0.118 | 0.235 | 0.500 | n.s. |
| L1:Stress1:Phonation | -0.406 | 0.461 | -0.880 | n.s. | -0.766 | 0.470 | -1.628 | n.s. |

| Analysis window | 600 ms - 867 ms | (N = 2760; logLik = -4570) |
| Intercept | -0.173 | 0.126 | -1.371 | n.s. |
| L1 = Korean L1 | -0.342 | 0.230 | -1.485 | n.s. |
| Stress1 = Iambic | -0.065 | 0.108 | -0.604 | n.s. |
| Phonation1 = Lax | -0.215 | 0.150 | -1.439 | n.s. |
| L1:Stress1 | -0.155 | 0.215 | -0.722 | n.s. |
| L1:Phonation1 | -0.160 | 0.215 | -0.744 | n.s. |
| Stress1:Phonation1 | 0.551 | 0.215 | 2.570 | .05 |
| L1:Stress1:Phonation | -0.722 | 0.429 | -1.684 | n.s. |

Table 3.7: Summary of \( L1, \) Stress pattern, and Phonation type effects on gaze to the target object (empirical logit scale) in Quasi-MLR for segmented windows.

trochaic words. That is, the chance of looking at the target was significantly higher in native English speakers than in L2 learners for iambic words than it was for trochaic words. In the last window (600- 867 ms), the effects and interaction of \( L1 \) and Stress pattern were not
significant. This indicates that Korean speakers’ fixation likelihoods for the trochaic and iambic targets became comparable to those of native English speakers toward the end of word recognition. It also indicates that L2 learners’ iambic word recognition was relatively delayed after the last syllable of words.6

Korean L2 learners’ relatively less reliable processing of iambic words might be because they had general processing difficulty for iambic stress pattern. However, it might also be because many iambic nonwords began with what Korean heard as aspirated sounds, for which relatively high f0s in the following vowel may not serve as a good fit with the iambic stress pattern, causing processing difficulty for iambic words. According to the post-eyetracking orthographic transcription, a large number of targets (10 / 16 target pairs) was transcribed as aspirated-initial words by the Korean L2 learners (98.7% transcription agreement for the phonation type (aspirated vs. lax) of word-initial consonants (See Table C.1 in Appendix C).

If there is any correspondence between English lexical stress pattern and Korean AP-initial tonal pattern in terms of f0 height, the goodness of fit can be summarized as in Table 3.8.

An additional t-test was conducted, following the log-ratio analysis in Arai, Gompel, and Scheepers (2007). In the analysis, the log gaze probability ratios of the target AOI over the competitor AOI was compared against 0. Time variable was collapsed into the same discrete time windows as for the Quasi-MLR analyses with segmented time windows, each represented with the mean log ratios by subjects. This analysis is useful to estimate the gaze probability of targets directly against its competitors. The measure of log ratio yields a score of zero if both target and competitor objects were inspected equally frequently; a positive score means that there were more looks to the target object than to the competitor object, and a negative score implies that there were more looks to the competitor object than to the target object. Its absolute ratio value reflects the magnitude (i.e., the strength of bias), while the sign (see +/- of values) expresses the direction of visual bias. The results presented in Appendix E. The results provide the evidence that English speakers’ preference for trochaic target words over competitors was significantly made at around the offset of the first syllable (0 - 200 ms, t(22) = 2.06, p <.05; see Table E.1 and the left panel in Figure E.1), and it was after the offset of the second syllable for iambic words (200 - 400 ms, t(22) = 2.36, p <.05). On the other hand, Korean speakers showed delayed preference for both trochaic and iambic target words over the competitors compared to English speakers. Korean L2 learners’ preference for the trochaic target over the competitor was significant after the third syllable (400 - 600 ms, t(18) = 2.98, p <.05; see Table E.1 and the right panel in Figure E.1) and it was not significant in any windows for the iambic words.
<table>
<thead>
<tr>
<th>Initial Consonant, AP type</th>
<th>Trochaic</th>
<th>Iambic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aspirated, HHLH</td>
<td>Good-fit (e.g. TEmido)</td>
<td>Poor-fit (e.g. teMIdo)</td>
</tr>
<tr>
<td>Lax, LHLH</td>
<td>Poor-fit (e.g. ZAkuńa)</td>
<td>Good-fit (e.g. zaKUnα)</td>
</tr>
</tbody>
</table>

Table 3.8: Goodness of fit of English lexical stress patterns to Korean AP-initial tonal patterns.

As aspirated sounds go with H tones in the initial syllable of Korean words, what Korean L2 learners heard as aspirated-initial words can be a good fit to the H* pitch accented trochaic words that had relatively higher f0 in the first syllable as compared to the f0 of the second syllable (see the acoustic measures in Section 3.2.2). On the other hand, as lax sounds go with L tones in the initial syllable of Korean words, what Korean L2 learners heard as lax-initial words can be a good fit to iambic words that had relatively lower f0 in the first syllable as compared to the f0 of the H* pitch accented second syllable.

Therefore, if Korean L2 learners had general processing difficulty for iambic words, the word-initial phonation type should not affect the processing of English stress pattern during word recognition. If there was Korean L1 prosodic influence, we can expect the interaction between stress pattern and word-initial phonation type in Korean speakers in such a way trochaic word recognition is facilitated by what Korean heard as aspirated-initial words, whereas iambic word recognition is facilitated by what Korean heard as lax-initial words.

In this light, we further investigate the processing of lexical stress pattern in its interaction with word-initial phonation type as a post-hoc analysis. For the analysis, the test nonwords were classified into what Korean L2 learners heard as an ‘aspirated’ vs. ‘lax’ initial words in the transcription task. Of the correct responses for trochaic (n = 132) and iambic word recognition (n = 125) in Korean speakers, there were 80 aspirated and 52 lax-initial trochaic words and 83 aspirated and 42 lax-initial iambic words. English speakers
provided correct responses for 98 aspirated (voiceless) and 53 lax (voiced) initial trochaic words \((n = 151)\) and 97 aspirated (voiceless) and 47 lax (voiced) initial iambic words \((n = 144)\).

Figure 3.6 shows native English speakers’ mean empirical logits of gaze to the target object over time for the two stress conditions as a function of the word-initial phonation type.

![Figure 3.6: English speakers \((N = 23)\)' mean fixation likelihood (empirical logits) of targets for aspirated- (voiceless-) and lax- (voiced-) initial nonwords. Left: Trochaic nonwords. Right: Iambic nonwords.](image)

The left panel of Figure 3.6 shows a steeper slope for aspirated (voiceless) initial trochaic words than for lax (voiced) initial trochaic words, whereas the right panel shows a steeper slope for lax (voiced) initial iambic words than for aspirated (voiceless) initial iambic words, suggesting that the word-initial voicing contrast affected English speakers processing of English lexical stress during word recognition. As the aspiration of voiceless
stops tends to be longer in stressed syllables than unstressed syllables and in word-initial position than word medial position (Klatt, 1975; Ladefoged & Johnson, 2010; Lavoie, 2001), the relatively long aspiration of voiceless sounds in word-initial position might have provided a supplementary cue to the processing of trochaic words, rather than to iambic words. Also, as it’s been repeatedly demonstrated that f0 at the onset of vowels is significantly higher after voiceless onsets than after voiced onsets (Ohde, 1984; Whalen, Abramson, Lisker, & Mody, 1990; Whalen, Abramson, Lisker, & Mody, 1993), the f0 height difference in the voicing contrast of word-initial consonants could potentially affect the perception and processing of initial stress/unstressed syllables in native English speakers. However, as the fixation likelihood for lax (voiced) initial trochaic words started off very high even at the onset of the analysis window (i.e., -153 ms) and stayed as high after the word onset as compared to the one for aspirated (voiceless) initial trochaic words, the steeper slope change for aspirated (voiceless) over lax (voiced) initial trochaic word is not purely attributable to the longer aspiration and higher f0 in the aspirated (voiceless) initial syllable.

Figure 3.7 shows Korean L2 learners’ mean empirical logits of gaze to the target object over time for the two stress conditions as a function of the word-initial phonation type contrast.

Unlike English speakers, Korean L2 learners had very similar fixation likelihoods for aspirated and lax-initial words at the onset of analysis window (i.e., -153 ms) for both trochaic (left panel) and iambic conditions (right panel). The left panel of Figure 3.7 shows that the slope change was steeper for aspirated-initial trochaic words than for lax-initial trochaic words. Also, there was a slight early advantage of gaze likelihood for aspirated-initial words temporarily after the first syllable (i.e., 68 - 187 ms) and later in the word
for the trochaic words, suggesting the Korean L2 learners had a processing advantage for aspirated-initial words over lax-initial words during trochaic word recognition. The right panel of Figure 3.7 shows that overall slope steepness was similar for both aspirated and lax-initial iambic words. However, there was a clear advantage in fixation likelihood for the aspirated-initial words over the lax-initial words right after the first syllable until the end of the third syllable (at around 500 ms). This goes contrary to the prediction that lax-initial words would facilitate iambic word recognition.

The Quasi-MLR result in Table 3.6 is revisited here to examine the interaction between Stress pattern and Phonation type. At the intercept, it revealed no significant effect of Phonation type, but significant interactions between L1 and Phonation type and between Stress pattern and Phonation type. The interaction between L1 and Phonation type indicates that native English speakers and L2 learners were significantly different in the likelihood of fixating the target object for aspirated (voiceless)-initial and lax (voiced)-initial
words at the onset of the analysis window (i.e., -153 ms). The negative parameter estimate ($\beta = -0.650, p < 0.05$) tells that there was significantly higher chance of looking to the lax (voiced) initial word than to the aspirated (voiceless) initial word in English speakers than it was in Korean speakers at the onset of the analysis window. The interaction between Stress pattern and Phonation type suggests that the two stress patterns were significantly different in the fixation likelihood for the aspirated vs. lax-initial words at the onset of the analysis window. The negative coefficient for the interaction ($\beta = -0.705, p < 0.05$) tells that the chance of looking at the target was higher for lax (voiced) initial words than for aspirated (voiceless) initial words in the trochaic condition than it was in the iambic condition. The interaction between L1, Stress pattern, and Phonation type was not significant, suggesting the interaction between Stress pattern and Phonation type at the intercept was not significantly different in native English speakers and Korean L2 learners.

For the slope, there was not an effect of Phonation type, nor an interaction of Phonation type and L1. This means the significant interaction between Phonation type and L1 at the intercept was no longer significant. Rather, the use of phonation type contrast over time became similar for both native speakers and L2 learners. However, there were significant interactions between Stress pattern and Phonation type and between L1, Stress pattern, and Phonation type. The positive parameter estimate for the interaction between Stress pattern and Phonation type on the slope ($\beta = 1.371, p < .001$) indicates that the increase in fixation likelihood was higher for the aspirated-initial word than for the lax-initial word in the trochaic condition, whereas it was the opposite in the iambic condition. The negative estimate for the interaction between L1, Stress pattern, and Phonation type on the slope ($\beta$
= -1.848, \( p > .05 \)) indicates that the interaction of Stress pattern and Phonation type on the slope was stronger for English speakers than for Korean L2 learners.\(^7\)

The result of the segmented windows in Table 3.7 also revealed an significant interaction between L1 and Phonation type and of Stress pattern and Phonation type in the first window, -153 - 0ms, to be consistent with the results at the intercept in the Quasi-MLR analysis with the entire analysis window. The interaction between L1 and Phonation type was also significant in window 0 - 200 ms (\( \beta = -0.622, \ p < .05 \)), suggesting the chance of looking was higher for the lax (voiced) initial words than for the aspirated (voiceless) initial words in English speakers than it was in Korean L2 learners. The effect of Phonation type and its interaction with other predictors were not found until the last window, 600 - 867 ms. In the last window, the interaction between Stress pattern and Phonation type was significant with higher fixation likelihood for the aspirated (voiceless) initial words than for the lax (voiced) initial words in the trochaic condition and higher fixation likelihood for the lax (voiced) initial words than for the aspirated (voiceless) initial words in the iambic condition (\( \beta = 0.551, \ p < .05 \)). Though this interaction result supports the prediction that aspirated-initial words facilitate trochaic word recognition and lax-initial words facilitate

\(^{7}\)The analyses on the log gaze probability ratios of the target over the competitor also support the results with regard to the interaction of Stress pattern and Phonation type. Table E.2 in Appendix shows that English speakers preferred the aspirated initial trochaic target over the competitor after the first syllable (0 - 200 ms, \( t(22) = 2.06, \ p < .05 \)), whereas they preferred the lax-initial trochaic target over the competitor before 400 ms (\( p < .05 \)) due to the ABE for lax-initial trochaic target words but not in the rest of the windows (see Figure E.2). For the iambic word, their preference for the aspirated-initial target word was significant after 600 ms (\( t(22) = 3.14, \ p < .05 \)), whereas the preference for the lax-initial target word was significant after 400 ms, \( t(22) = 2.74, \ p < .05 \). The results suggest that native English speakers had overall processing advantage for trochaic words, and the voicing contrast of word-initial obstruents affected processing of stress patterns during lexical access. On the other hand, Table E.3 shows that Korean L2 learners’ preference for the aspirated-initial target word over the competitor was significant after the last syllable for the trochaic words (400 - 600 ms, \( t(18) = 2.24, \ p < .05 \)), whereas their preference for the lax-initial trochaic word was not significant in any windows, suggesting that Korean L2 learners had a processing advantage for the aspirated initial trochaic word over the lax-initial trochaic word. For the iambic word, significance was not found in any windows for both aspirated and lax-initial words, supporting that iambic stress pattern posed great processing difficulty to Korean L2 learners.
iambic word recognition, the effect appeared too late in time to have the processing result have to do with phonation type in the first syllable. Contrary to the result from the Quasi-MLR analysis with the entire analysis window, the interaction between Stress pattern, Phonation type, and L1 was not significant, suggesting the interaction between Stress pattern and Phonation type was not significantly different in English and Korean speakers toward the end of the word recognition.

To sum up, the Quasi-MLR analyses with and without segmentation of the analysis window between -153 ms and 867 ms, show that native English speakers were better at integrating lexical stress information during word recognition than were Korean L2 learners of English, with steeper slope change. The significant interaction between L1 and Stress pattern during the segmented windows from 200 ms to 600 ms suggests that Korean L2 learners had a processing advantage for trochaic over iambic words right after the second syllable as compared to native English speakers. The significant interaction between Stress pattern and Phonation type on the slope suggests that the phonation type (voicing contrast) of word-initial sounds affected perception and processing of English lexical stress pattern in such a way that what Koreans heard as aspirated and lax-initial words respectively facilitated trochaic word recognition and iambic word recognition. However, as there were the interaction between Stress pattern and Phonation type at the intercept, we cannot use the interaction between Stress pattern and Phonation type on the slope to test the prediction that word-initial phonation type affects the processing of English lexical stress in Korean L2 learners of English.
3.7 Discussion

For the eyetracking experiment, it was expected that English speakers would recognize words using stressed syllable information in a word, and thus would have earlier activation of trochaic words, to be consistent with the literature (N. Cooper et al., 2002; Cutler & Clifton, 1984; Cutler & Norris, 1988; Cutler, 1989; Mattys & Samuel, 2000; Mattys, 2000). On the other hand, Korean L2 learners of English were expected not to be able to use the lexical stress information as efficiently as native speakers do, but instead use the last segmentally disambiguating syllable for word recognition, if they are “stress deaf” like French speaking learners of Spanish/English (Dupoux et al., 1997, 2001, 2008; Tremblay, 2008, 2009).

The training results showed that native English speakers and Korean L2 learners of English learned the 48 nonwords at a comparable level throughout the three-day training session. The accuracy and RT data of the eyetracking word recognition experiment showed that the two L1 groups were overall not significantly different from each other in their accuracy rates and response times for both trochaic and iambic word recognition.

However, the gaze data revealed more detailed processing patterns that were not captured in the accuracy and RT data analyses. The eyetracking data analyses showed that native English speakers were better at integrating lexical stress information during word recognition than Korean L2 learners of English, with steeper increase of fixation likelihood over time. Though the interaction between Lexical stress and L1 was not significant on the slope, the analysis on the segmented windows showed that the activation rates of trochaic and iambic target words were more comparable in English speakers than in Korean L2 learners after the second syllable (200 - 600 ms). The t-test on the log fixation probability ratios of the target over the competitor further provides evidence that English
speakers looked to targets for trochaic words right after the first syllable (0 ms) and to those for iambic words after the second syllable (200 ms), whereas Korean L2 learners looked to trochaic targets after 400 ms and iambic targets toward the end of the word. This accounts for the finding that fixation likelihoods for the trochaic and iambic words were more comparable after the second syllable (200 ms) in native English speakers than they were in Korean L2 learners.

The results further provide evidence for the MSS model (Cutler & Norris, 1988; Cutler, 1989; Cutler & Butterfield, 1992). The MSS allows all strong syllables with full vowels to activate words. If the delayed activation of iambic target words in native English speakers was because they were looking at the trochaic competitor in the iambic word recognition, the result support what the MSS predicts about lexical stress and lexical access - a full vowel initiates lexical access. This seems to be the case for the stimuli used in this experiment, where unstressed syllables in first and second syllable position contained no reduced vowels. Therefore, the delayed activation of the iambic words might be because English speaking listeners needed information from both stressed and unstressed syllables in order to establish the relative prominence difference between the first two syllables of the iambic word, thus determining that an unstressed initial syllable with a full vowel was indeed unstressed. The results suggest that unreduced vowels can serve as cues to unstressed syllables for native English listeners. However, the preliminary study cannot address whether vowel reduction is more readily integrated into the process of target word activation than relative prominence would do because the stimuli didn’t include reduced vowels in unstressed syllables for comparison.

Korean speakers also showed an explicit processing advantage for trochaic over iambic stress pattern, though their activation of both trochaic and iambic words were much delayed
as compared to the native English speakers. To explore whether such a processing advantage for trochaic over iambic words was due to the influence of Korean L1 phrase-level tonal pattern, the word-initial Phonation type was included in the Quasi-MLR model. Interestingly, the significant interaction between L1, Stress pattern, and Phonation type on the slope indicates that English speakers were more responsive to the word-initial phonation type contrast (voicing contrast for English speakers) for their processing of stress contrast in a word than were Korean speakers. However, as English speakers had a strong tendency to look at the target object at the word onset for lax (voiced) initial trochaic words than for aspirated (voiceless) initial trochaic words, the estimation of the slope could be biased, such that a lax (voiced) initial trochaic condition shows less of a change in slope than a aspirated (voiceless) initial trochaic condition. Therefore, the slope estimates for the interaction between L1, Stress pattern and Phonation type cannot be used to the answer for the question of whether word-initial phonation type contrast influences the processing of English lexical stress in Korean L2 learners of English. It is also unreasonable to refer to the intercept estimate to examine the the interaction between Stress pattern and Phonation type because the intercept was estimated much earlier in time before a critical auditory stimulus was sufficiently given and thus cannot be interpreted as a reflection of processing result.

Although we couldn’t find statistical evidence to prove the potential influence of L1 phrase-level prosodic structure on L2 lexical stress processing in Korean L2 learners, as the strong baseline effect was primarily found in English speakers’ trochaic word recognition and Korean L2 learners who showed relatively comparable fixation likelihoods for the aspirated and lax initial trochaic words at the onset of the analysis window had an early activation advantage for aspirated-initial words right after the first syllable during
trochaic word recognition, we still hold the prediction and explore the issue in the dissertation experiments. The relatively weak phonation effect on iambic word processing in Korean, however, is still attributable to Korean L1 prosodic effect. In the Korean intonation phonology, the phrase initial tones, H and L, are not characterized by the manipulation of duration and amplitude cues, but by the manipulation of f0 height. Therefore, even though the iambic stress pattern is potentially a better fit to the lax-initial words than to the aspirated-initial words in terms of f0 height, the information in the English unstressed initial syllable, which entailed reduction of all acoustic cues, would be psychoacoustically difficult to detect, yielding more processing difficulty to Korean L2 learners regardless of the word-initial phonation type.

Taken the results together, we could observe that the L1 experience with lexical processing strategy can affect L2 lexical stress processing during spoken word recognition in Korean L2 learners of English. As Korean speakers do not have incentive to pay attention to the prominence difference between syllables for spoken word recognition, processing of English lexical stress for word recognition may have created additional memory load and processing burden to Korean L2 learners of English. This was evidenced by the strong competition between the target and the stress competitor in the middle of word recognition and the delayed target word activation in Korean L2 learners of English as compared to in native English speakers. However, we also noticed that Korean L2 learners’ processing of English lexical stress was affected by their perceptual experience with L1 phrase-level prosodic structure when they integrated lexical stress information during word recognition. The tonal shape of trochaic words with H pitch accents may be perceptually similar to the shape of Korean HHLH accentual phrases that begin with aspirated consonants. This may have caused perceptual assimilation, even though the relevant cue was employed at
different levels of linguistic representation, phrasal vs. lexical prosodic structures. Their processing of iambic words seems to be also constrained by L1 prosodic structure where acoustic cues such as duration and intensity do not distinguish phrase-initial tonal contrasts and thus are less likely to motivate any facilitating assimilation.

3.8 Implications

In the preliminary study, we explored how Korean speaking learners of English process English lexical stress during L2 spoken word recognition and observed the possibility that L1 phrase-level prosodic structure can affect processing of L2 word-level prosody in Korean L2 learners of English. However, because of the unexpected baseline effect for aspirated (voiceless) initial trochaic condition, we couldn’t verify the interaction between word-initial phonation type and lexical stress processing in Korean L2 learners of English in the study. In addition, as the number of aspirated vs. lax word-initial consonants was not controlled in the study, any conclusions drawn about prosodic influence on word recognition must remain tentative. Furthermore, we cannot say that the stimulus set of the preliminary study fully represented the properties of English lexical stress because the initial syllables of the stimuli did not include reduced vowels, a segmental correlate to lexical stress in English. Therefore, the dissertation study explores the issue by including reduced vowels in unstressed syllables.
CHAPTER 4: EXPERIMENT 1

The first dissertation experiment employs eyetracking methodology to explore how Korean L2 learners of English process lexical stress with reduced vowels in unstressed syllables. The specific research questions are as follows: 1) How do Korean L2 learners of English use English lexical stress during L2 English word recognition when reduced vowel information is available in the unstressed syllable? 2) How do native English speakers as a control group use lexical stress information for spoken word recognition when a reduced vowel cue is available in unstressed syllables? Does the processing of trochaic stress and iambic stress differ during word recognition? 3) Do L1 and L2 speakers show similar processing patterns in the use of reduced vowel information for word recognition? 4) Does Korean L1 phrase-level prosodic structure influence processing of English lexical stress in Korean L2 learners?

Relatively few studies have directly examined how reduced vowels are used during spoken word recognition in English as an L1 and an L2. If Korean L2 learners of English use the English reduced vowel, schwa, as a distinctive vowel from most of English canonical vowels as reported in Yun and Jackson (2006), they would be able to use the reduced vowel as a segmental cue to distinguish a target word from its stress competitor during the first syllable of the target word. And if they use the segmental cue in the initial syllables of words, rather than lexical stress information, for word recognition, the impact of L1 phrase-level prosodic structure on the processing of English lexical stress is more likely to be diminished than it was in the preliminary study.
As for English listeners’ processing of lexical stress with reduced vowels, the MSS (Cutler & Norris, 1988) and the stressed-based lexical processing models (Mattys, 2000; Mattys & Samuel, 2000) allow the prediction that native English speakers would not be able to use reduced unstressed syllables to initiate lexical access. If this is the case, we can expect delayed processing of iambic words as compared to trochaic words. However, speech perception research examining reduced and unreduced unstressed vowels also hints that native English speakers may use reduced vowels to identify spoken words. For example, Fear et al. (1995) showed that native English listeners perceived unstressed unreduced vowels more like stressed vowels, but not reduced vowels. Though the study didn’t test the role of full and reduced vowel syllables in lexical access, if native English speakers can discriminate full and reduced vowels when the vowels were given with lexical stress, they might be able to use the vowel cue to promote similar processing of stress patterns. That is, the rate of target word activation would be comparable for the trochaic and iambic word recognition in English speakers. With regard to the use of phonation type contrast of word-initial consonants, if native English speakers hear the word-initial phonation contrast as a voicing contrast and if they are sensitive to suprasegmental cues as were the native English speakers in the preliminary study, English speakers’ processing of lexical stress would be affected by the voicing contrast of word-initial consonants in a way similar to how Korean L2 learners’ processing of lexical stress would be influenced by word-initial phonation type contrast.

The remainder of this chapter is organized as follows. Sections 4.1 through 4.3 describe the experimental design and materials, participants, and procedures. Section 4.4 presents the statistical analysis methods employed for the data analyses. Sections 4.5 and 4.6 present the results of the eyetracking study and discussion in relation to the reduced vowel
effect and the L1 phrase-level prosody effect on the processing of lexical stress in Korean L2 learners of English as compared to native English speakers.

4.1 Design and Materials

The current eyetracking experiment examines how native English speakers and Korean speaking learners of English use a segmental correlate of English lexical stress, schwa, during spoken word recognition. In addition, we explore whether Korean L1 phrase-level prosodic structure affects the use of English lexical stress. The experiment uses 2 * 2 * 2 factorial design with Stress pattern (trochaic vs. iambic) and word-initial Phonation type (aspirated vs. lax for Korean L2 learners, voiceless vs. voiced for native English speakers; see the summary of transcription in Table C.2 in Appendix A.2) as within-subject factors and L1 (English vs. Korean) as a between-subjects factor.

4.1.1 Auditory Stimulus Materials

The auditory stimuli were constructed in the same manner as in the preliminary eye-tracking study, with two exceptions. First, the sixteen paired nonwords include the reduced vowel, schwa, in unstressed syllables (e.g. /tI'ma(do/ vs. /tI'mI'da/). The vowels used for stressed syllables were /i, I, a, æ, e, eI, o, ə/ for the first and second syllables. The vowels for final syllables did not include schwa. Second, the word-initial consonants were counterbalanced according to perceptual judgments made by two native Korean speakers in a pre-listening test. Half of the 16 pairs began with obstruents that were heard and transcribed as aspirated (/p, t, k, s/), and half began with those heard as lax (/b, d, g, m/). The consonants for the second and third syllables were the same as in the preliminary study. Table 4.1 shows the eight consonants and eight vowels used to create the test nonwords and
the Nlog frequency counted for each syllable position based on the LDC American English Spoken Lexicon.

<table>
<thead>
<tr>
<th>Syllable</th>
<th>Consonants</th>
<th>Mean Freq. (Nlog)</th>
<th>Vowels</th>
<th>Mean Freq. (Nlog)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>/p, t, k, s, b, d, g, m/</td>
<td>- 2.76</td>
<td>/i, i, e, ei, a, æ, o, æ/</td>
<td>- 2.97</td>
</tr>
<tr>
<td>2</td>
<td>/b, t, k, g, v, s, m, n/</td>
<td>- 2.55</td>
<td>/i, i, e, ei, a, æ, o, æ/</td>
<td>- 1.35</td>
</tr>
<tr>
<td>3</td>
<td>/k, t, d, v, s, n, ʒ, dʒ/</td>
<td>- 2.66</td>
<td>/i, ei, a, ai, o, u, au, æ/</td>
<td>- 3.73</td>
</tr>
</tbody>
</table>

Table 4.1: Eight phonemes used in the trisyllabic nonword and mean Nlog frequency.

Sixteen additional nonword fillers were created to be phonologically unrelated to any of the corresponding target pair members. A full list of nonword stimuli is presented in Appendix A.2.

The 48 nonwords were recorded and transcribed in the same manner as in the preliminary eyetracking experiment. The speaker and transcriber were different from the ones in the preliminary study, but the speaker was still a female native English speaker. Nonword recordings were screened by both a native English speaker and a Korean native speaker to ensure that the nonwords were not heard as real words of either English or Korean.

4.1.2 Acoustic Phonetic Measurements

Acoustic measures were made on the auditory stimuli to confirm that the stress minimal pair members carried both suprasegmental and segmental acoustic cues to lexical stress in the first two syllables and that the word-initial phonation type contrast involved an f0 height difference in the vowels of word-initial syllables.
Duration, f0, and Intensity

The suprasegmental acoustic measures include duration, mean f0, and mean intensity in the first two vowels of the 16 pairs of test nonwords. The measurements of the relevant acoustic cues were made in the same manner as in the preliminary study (see Section 3.2.2). Paired t-tests showed that trochaic words in isolation had first syllables that were longer (130 ms vs. 30 ms), $t(15) = 11.1, p < .001$; higher in mean f0 (258 Hz vs. 206 Hz), $t(15) = 12.5, p < .001$; and greater in loudness (32 sone vs. 15 sone), $t(15) = 10.5, p < .001$, than the first syllables of iambic words in isolation. Conversely, iambic words in isolation had second syllables that were longer (150 ms vs. 50 ms), $t(15) = 9.5, p < .001$; higher in mean f0 (254 Hz vs. 225 Hz), $t(15) = 4.2, p < .001$; and greater in loudness (31 sone vs. 18 sone), $t(15) = 6.6, p < .001$, than the second syllables of trochaic words in isolation.

In addition, first syllables were different from second syllables within a word for the two stress conditions: first syllables of trochaic words were longer (130 ms vs. 50 ms), $t(15) = 7.3, p < .001$; higher in mean f0 (258 Hz vs. 225 Hz), $t(15) = 6.5, p < .001$; and greater in loudness (32 sone vs. 18 sone), $t(15) = 12.1, p < .001$, than second syllables of trochaic words. Words with iambic stress patterns had second syllables that were longer (150 ms vs. 30 ms), $t(15) = 13.4, p = .001$; higher in mean f0 (254 Hz vs. 206 Hz), $t(15) = 13.8, p < .001$; and greater in loudness (31 sone vs. 15 sone), $t(15) = 20.9, p < .001$, than first syllables of iambic words.

The same measurements were conducted on the test nonwords recorded in the carrier sentence. Paired t-tests showed that trochaic words had first syllables that were longer (130 ms vs. 40 ms), $t(15) = 13.1, p < .001$; higher in mean f0 (273 Hz vs. 199 Hz), $t(15) = 20.8, p < .001$; and greater in loudness (43 sone vs. 19 sone), $t(15) = 13.2, p < .001$, than the first syllables of iambic words. Conversely, iambic words had second syllables that were longer
(150 ms vs. 50 ms), \( t(15) = 10.2, p < .001 \); higher in mean f0 (267 Hz vs. 213 Hz), \( p(15) = 10.5, p < .001 \); and greater in loudness (39 sone vs. 23 sone), \( t(15) = 6.7, p < .001 \), than the second syllables of trochaic words.

Also, first syllables were significantly different from second syllables within a word for the two stress conditions: first syllables in trochaic words were longer (130 ms vs. 50 ms), \( t(15) = 8.6, p < .001 \); higher in mean f0 (273 Hz vs. 213 Hz), \( t(15) = 10, p < .001 \); and greater in loudness (43 sone vs. 23 sone), \( t(15) = 19.8, p < .001 \), than the second syllables of trochaic words. Words with iambic stress patterns had second syllables that were longer (150 ms vs. 40 ms), \( t(15) = 13.8, p = .001 \); higher in mean f0 (267 Hz vs. 199 Hz), \( t(15) = 14.4, p < .001 \); and greater in loudness (39 sone vs. 23 sone), \( t(15) = 18.7, p < .001 \), than the first syllables of iambic words.

In addition to the acoustic measures of lexical stress, mean f0 of the initial syllables was compared to make sure aspirated-initial words began with higher f0s than lax-initial words for both trochaic and iambic words. Paired t-tests showed that first syllables of aspirated-initial trochaic words produced in isolation were higher in mean f0 (271 Hz vs. 246 Hz), \( t(7) = 6.6, p < .001 \) than first syllables in lax-initial trochaic words. Iambic words with aspirated-initial sounds had first syllables that were higher in mean f0 (290 Hz vs. 256 Hz), \( t(7) = 2.5, p < .05 \) than the first syllables of lax-initial words. Aspirated initial trochaic words produced in the carrier sentence also had first syllables that were higher in mean f0 (290 Hz vs. 256 Hz), \( t(7) = 9.1, p < .001 \) than first syllables of lax-initial trochaic words. Iambic words with aspirated-initial sound had first syllables that were higher in mean f0 (209 Hz vs. 189 Hz), \( t(7) = 4.7, p < .05 \) than the first syllables of lax-initial words.

Lastly, mean f0 of the word *the* used in the carrier sentence was measured to ensure that there was no significant f0 rise before the trochaic target word that might serve as a
premature cue to upcoming lexical stress. Paired t-test revealed no significant difference in the mean f0s between the vowels in the word, *the*, produced before trochaic words and the one produced before iambic words (194 Hz vs. 198 Hz, \( t(15) = 1.7, p > .05 \)).

The acoustic measurements indicate that suprasegmental cues might conceivably differentiate the prominence in the first two syllables of stress minimal pair words. If stress location is utilized in lexical learning, listeners can use the cues to discern the pairs that differ in stress. And if listeners are sensitive to the f0 difference between aspirated- and lax- (or voiceless and voiced) initial syllables and the relationship of this f0 difference to lexical stress (aspirated/voiceless initial syllables with relatively higher f0s can be a good fit to trochaic stress pattern, whereas lax/voiced initial syllables with relatively lower f0s can be a good fit to iambic stress pattern), the phonation type (or voicing contrast) in the word-initial syllable is expected to affect the perception and processing of English lexical stress during spoken word recognition.

**Vowel Quality**

To confirm that the unstressed syllables carried reduced vowels (i.e., /ə/), F1 and F2 frequencies were taken from the first two vowels in the test nonword tokens to measure vowel quality, in the same manner as for the preliminary study. The Euclidean distance of each vowel token from the central point in the speaker’s F1 × F2 (Hz) space was computed, separately for the vowels produced in words in isolation and for those produced in the carrier sentence.

Paired t-tests showed that trochaic words in isolation had first vowels that were farther from the midpoint of the speaker’s vowel space than first vowels of iambic words in isolation (501 Hz vs. 364 Hz), \( t(15) = 2.3, p < .05 \). Conversely, second vowels of iambic words
in isolation had larger vowel space dispersion than second vowels of trochaic words (516 Hz vs. 308 Hz), \( t(15) = 3.1, p < .05 \).

For the vowels produced in the carrier sentence, paired t-tests showed that trochaic words had first vowels that were greater in vowel space dispersion than first vowels of iambic words (519 Hz vs. 400 Hz), \( t(15) = 2.1, p = .05 \). Second vowels of iambic words had greater vowel space dispersion than second vowels of trochaic words (515 Hz vs. 291 Hz), \( t(15) = 3.5, p < .05 \).

The measures of vowel distance from the midpoint of speaker’s vowel space indicate that the prosodic contrast in the first two syllables of nonwords involved vowel quality difference. The vowels in unstressed syllables were produced in a reduced form and thus the distance from the vowel space midpoint was significantly smaller than the vowels produced in stressed syllables. Therefore, if listeners are sensitive to vowel quality difference between the stressed and unstressed syllables during lexical learning, the segmental correlate to lexical stress could be used by listeners to discern the word pair members early in lexical processing.

4.1.3 Visual Stimulus Materials

The visual stimuli were the same line-drawings of space aliens as in the preliminary study. A full list of visual stimuli is presented in Appendix B.

4.2 Participants

Twenty-four native English speakers and twenty-one Korean-speaking learners of English participated in the experiment. The participants were those who scored above 90% accuracy in the picture choice task on the last day of training and above 75% accuracy in
the eyetracking word recognition task (54 out of 63 participants, 85.7%, who completed all
training and eyetracking tasks). Data from five native English speakers and four Korean-
 speaking learners of English were additionally excluded due to poor eyetracking calibra-
tion.

Native English speakers were speakers of midwestern American English aged 17-27
\( (M = 20.2, \ SD = 2.11) \). They were born and grew up in the U.S., learning English as their
L1. Some had history of foreign language study in high school or college, but none of
them had lived abroad and no other language was spoken in the home during childhood.
They received partial course credit in a linguistics course and a small payment for their
participation.

Korean speaking learners of English were undergraduate or graduate students of the
Ohio State University, aged 21-34 \( (M = 25.4, \ SD = 4) \) at the time of recruitment. They re-
ported speaking Seoul Korean as an L1 and no residence history in the area of Kyungsang
dialect. None of the Korean speakers had any English-immersion experience before enter-
ing the U.S. These Korean speakers moved to the U.S. at an average age of 23.5 years \( (SD
= 4.76, \ range \ from \ 14 \ to \ 33 \ years \ of \ age) \), and had lived in the U.S. for three months to
eight years \( (M = 1.9, \ SD = 2.33) \) at the time of recruitment.

The average beginning age of learning L2 English in Korea was 11 years of age \( (SD
= 2.23, \ range \ from \ 7 \ to \ 14 \ years \ old) \). In the language background questionnaire, Korean
speakers were asked to evaluate their comprehensive English proficiency on a scale from
1 to 4 (1 = barely; 2 = poorly; 3 = passably; 4 = fluently). The average self-evaluation of
English fluency was 3 \( (SD = .52, \ range \ from \ 2 \ to \ 4) \). They all voluntarily participated and
were paid for their participation.
None of the participants had participated in the preliminary study. None of them reported speech or hearing disorders or vision impairment at the time of testing.

4.3 Procedures

As in the preliminary eyetracking study, the current experiment consisted of three days of training sessions, a word recognition test with eye movement monitoring, and a listening-with-orthographic transcription task. Training consisted of the same learning tasks used in the preliminary study. All participants were trained with all 16 test nonword pairs and 16 fillers. The 48 nonwords were presented ten times in every 1-hour training session. The word recognition with eye-tracking test used the same picture-choice task as in the preliminary study. All participants were asked to listen to and then transcribe the target words in their L1 orthography after finishing the eyetracking experiment.

Unlike the preliminary study, two experimental lists of 48 nonwords were used in training sessions to randomize the picture-nonword associations for test pair members. The 16 picture referents used as targets in one list were used as competitors in the other list. Therefore, in the following word recognition test, four experimental lists were used, each of which contained 8 trochaic (4 aspirated and 4 lax initials) and 8 iambic nonwords (4 aspirated and 4 lax initials) so that members of the same pair did not occur in the same list. In addition, the same 16 filler nonwords were included in each list.

4.4 Data Analysis

As in the preliminary study, the current experiment includes the behavioral data from the training sessions and the eyetracking word recognition task, and the gaze data from
the eyetracking word recognition task. The accuracy and response time (RT) data from the picture choice task in the training session is used to confirm that both native English speakers and Korean L2 learners of English made comparable progress in learning the 48 nonword-picture associations. The accuracy and RT data from the eyetracking word recognition task are used to examine the general behavioral pattern for the 32 test nonwords used in the eyetracking word recognition. Our primary interest lies in the gaze data collected over time in the eyetracking word recognition task. We use the gaze data to examine how lexical stress information with a reduced vowel correlate and word-initial phonation type are incorporated into the process of spoken word recognition.

The statistical approaches to the analyses of the accuracy and RT data and the eyetracking data are identical to the ones employed in the preliminary study. The behavioral data are analyzed with a mixed effects model (see Section 3.5.1), and the gaze data to the target AOI are analyzed by using a Quasi-Multilevel logistic regression model (see Section 3.5.2).

4.5 Results

4.5.1 Training Result

Table 4.2 shows the mean percent accuracy for the 48 nonwords and the mean RT for the correct responses tested in the picture-choice task on each training day. Outlier RTs were discarded (4.1%, 248 / 6112 correct responses). Both native English speakers and Korean L2 learners of English showed increasing accuracy rates with decreasing RTs as they underwent more training sessions over three days.

To examine whether native English speakers and Korean L2 learners made comparable learning progress in learning the 48 nonwords, responses for the 48 nonwords (correct coded as “1”, incorrect coded as “0”) and RTs for the correct responses in all three training
days were separately submitted to a mixed effects model with L1 (English vs. Korean) and Training day as two predictors and Subject and Item as crossed random factors. The results are presented in Table 4.3.

Table 4.3: Summary of L1 and Training day effects on accuracy and RTs in mixed effects models.

The results revealed significant main effects of Training day for both accuracy and RTs, with no other effects or interactions approaching significance. The positive parameter estimate for the Training day effect on accuracy (\( \beta = 1.101, p < .0001 \)) indicates that both native English speakers and Korean L2 learners were more accurate in identification of the 48 nonword-picture associations as training increased over the three days of training. The negative parameter estimate for Training day on RTs (\( \beta = -393.72, p < .0001 \)) indicates that participants were faster to identify correct nonword-picture pairings as training increased.
For both accuracy and RTs, there was no main effect of L1 and no interaction between L1 and Training day, indicating that native English speakers and Korean L2 learners were not significantly different in their baseline accuracies and response latencies or in their learning trajectories over the three-day training sessions.

In sum, the training results demonstrate that both native English speakers and Korean L2 learners of English made comparable progress in learning the 48 nonword-picture associations over the three days of training, with a similar increase in accuracy rate and decrease in RT.

4.5.2 Eyetracking Result
Accuracy and RT

The mean percent accuracy for the sixteen test trials and thirteen filler trials in the eye-tracking word recognition experiment was 90.23% (SE = 1.13) for native English speakers and 90.97% (SE = 1.16) for Korean L2 learners of English.

The mean percent accuracy of the sixteen test trials and the mean RT for the correct responses by functions of stress pattern, phonation type, and L1 are given in Table 4.4. Outlier RTs were discarded (4.5%, 27 out of 606 correct responses).

<table>
<thead>
<tr>
<th>Stress</th>
<th>Phonation</th>
<th>English speakers</th>
<th>Korean speakers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Accuracy</td>
<td>RT</td>
</tr>
<tr>
<td></td>
<td>Trochaic</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Aspirated</td>
<td>81.3 (4.00)</td>
<td>2250.2 (84.27)</td>
</tr>
<tr>
<td></td>
<td>Lax</td>
<td>87.5 (3.39)</td>
<td>2059.6 (79.25)</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>84.4 (2.73)</td>
<td>2152.4 (58.09)</td>
</tr>
<tr>
<td></td>
<td>Iambic</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Aspirated</td>
<td>78.1 (4.24)</td>
<td>2001.5 (70.50)</td>
</tr>
<tr>
<td></td>
<td>Lax</td>
<td>87.5 (3.39)</td>
<td>2267.5 (76.81)</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>82.8 (2.63)</td>
<td>2138.9 (53.25)</td>
</tr>
</tbody>
</table>

Table 4.4: Mean percent accuracy (%) and mean RT (ms) by L1, Stress pattern, and Phonation type. Values in parentheses represent standard errors.
The binary correct/incorrect responses (coded as “1” and “0”, respectively) and the RTs of correct responses were separately submitted to a mixed effects model with \( L1 \), Stress pattern, and Phonation type as three predictors and Subject and Item as crossed random factors. Table 4.5 shows the results.

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Est.</th>
<th>SE</th>
<th>Wald Z</th>
<th>p</th>
<th>Est.</th>
<th>SE</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Accuracy</strong> (( N = 720; \log \text{Lik} = -305.3 ))</td>
<td>1.835</td>
<td>0.149</td>
<td>12.335</td>
<td>.0001</td>
<td>2145.90</td>
<td>52.35</td>
<td>40.99</td>
<td>.0001</td>
</tr>
<tr>
<td><strong>RT</strong> (( N = 579; \log \text{Lik} = -4503 ))</td>
<td>4.84</td>
<td>98.54</td>
<td>0.05</td>
<td>n.s.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L1 = Korean L1</td>
<td>0.170</td>
<td>0.263</td>
<td>0.646</td>
<td>n.s.</td>
<td>4.84</td>
<td>98.54</td>
<td>0.05</td>
<td>n.s.</td>
</tr>
<tr>
<td>Stress1 = Iambic</td>
<td>-0.518</td>
<td>0.225</td>
<td>-2.30</td>
<td>.05</td>
<td>108.10</td>
<td>49.66</td>
<td>2.18</td>
<td>.05</td>
</tr>
<tr>
<td>Phonation1 = Lax</td>
<td>0.546</td>
<td>0.265</td>
<td>2.062</td>
<td>.05</td>
<td>41.16</td>
<td>60.98</td>
<td>0.67</td>
<td>n.s.</td>
</tr>
<tr>
<td>L1:Stress1</td>
<td>-0.828</td>
<td>0.450</td>
<td>-1.838</td>
<td>n.s.</td>
<td>207.17</td>
<td>99.15</td>
<td>2.09</td>
<td>.05</td>
</tr>
<tr>
<td>L1:Phonation1</td>
<td>-0.102</td>
<td>0.450</td>
<td>-0.227</td>
<td>n.s.</td>
<td>-38.42</td>
<td>99.33</td>
<td>-0.39</td>
<td>n.s.</td>
</tr>
<tr>
<td>Stress1:Phonation1</td>
<td>-0.395</td>
<td>0.450</td>
<td>-0.872</td>
<td>n.s.</td>
<td>293.32</td>
<td>99.30</td>
<td>2.95</td>
<td>.05</td>
</tr>
<tr>
<td>L1:Stress1:Phonation1</td>
<td>-1.202</td>
<td>0.900</td>
<td>-1.335</td>
<td>n.s.</td>
<td>-383.24</td>
<td>198.26</td>
<td>-1.93</td>
<td>n.s.</td>
</tr>
</tbody>
</table>

Table 4.5: Summary of \( L1 \), Stress pattern, and Phonation type effects on accuracy and RTs in the mixed effects model.

The results for accuracy show main effects of Stress pattern and Phonation type, with no other effects or interactions approaching significance. The negative coefficient (\( \beta = -0.518, p < .05 \)) of Stress pattern indicates that participants were overall more accurate for trochaic word recognition than for iambic word recognition (86.9% and 81.4%). The positive coefficient of Phonation type (\( \beta = 0.546, p < .05 \)) suggests that participants provided more correct answers for the words that began with what Koreans heard as a lax-initial sound than for the words that began with what Koreans heard as an aspirated-initial sound.

The results for RTs revealed a main effect of Stress pattern, and significant interactions of \( L1 \) and Stress pattern and of Stress pattern and Phonation type. Other effects and interactions were not significant. The positive coefficient for the Stress pattern effect (\( \beta \)
indicates that RTs were longer for iambic word recognition than for trochaic word recognition. The positive coefficient for the interaction of L1 and *Stress pattern* ($\beta = 207.17, p < .05$) indicates that Korean L2 learners of English took more time for iambic word recognition than for trochaic word recognition (2026.4 ms for trochaic words and 2234.3 ms for iambic words) than did English speakers (2152.4 ms for trochaic words and 2138.9 ms for iambic words). The positive coefficient for the interaction of *Stress pattern* and *Phonation type* ($\beta = 293.32, p < .05$) with no interaction with L1 indicates that the trochaic and iambic words elicited significantly different RTs for the aspirated and lax-initial words for both native English speakers and Korean L2 learners of English. Trochaic words had longer RTs when they began with aspirated sounds than when they began with lax sounds, while iambic words had longer RTs when they began with lax sounds than when they began with aspirated sounds. Though Table 4.4 shows this pattern explicitly only for English participants, the interaction between *Stress pattern*, *Phonation type*, and L1 was not significant.  

To sum up, the accuracy and RT data of the eyetracking experiment shows that the overall accuracy and RT were not significantly different between native English speakers and Korean L2 learners of English in the word recognition task. However, Korean speakers were less accurate and had more processing difficulty for iambic word recognition than for trochaic word recognition as compared to native English speakers. Unlike the prediction that aspirated (voiceless) initial trochaic words and lax (voiced) initial iambic words would

---

7For eyetracking data analysis, the trials in which eye gaze was made on the target AOI before the onset of the target word in the instruction, “Click on the (target word) now”, were removed (22.4%, 136 out of 606 correct responses). See the following Section 4.5.2 for more details. The result on RTs after removing the on-target trials is presented in Appendix F. Outlier RTs were removed (3.8%, 18 out of 470 trials). The results were the same as in the analysis result with all trials: there was a main effect of *Stress pattern* ($\beta = 109.51, p < .05$) and significant interactions between *Stress pattern* and *Phonation type* ($\beta = 270.65, p < .05$), and between *L1* and *Stress pattern* ($\beta = 218.04, p < .05$).
be easier to process compared to their word-initial phonation (voicing) correspondents, the interaction between Stress pattern and Phonation type for RTs suggests that both English and Korean speakers had processing difficulty for aspirated (voiceless) initial trochaic words and lax (voiced) initial iambic words.

Eyetracking Data

Visual inspection of ABEs of targets over competitors was made for correct responses over the entire time course of the auditory instruction, the same as in the preliminary study. There were 162 correct responses for trochaic word recognition and 159 correct responses for iambic word recognition in English speakers. Korean L2 learners provided 151 correct responses for trochaic word recognition and 134 correct responses for iambic word recognition.

Figure 4.1 shows the mean fixation proportions of the target, competitor, and distractor AOIs during the trochaic and iambic word recognition for native English speakers and Figure 4.2 shows the proportions for Korean L2 learners of English. The data were time-aligned with the boundary between the first and second syllables at 0 ms as in the preliminary study. The mean durations of the first two syllables are 243 ms (SD = .02) and 120 ms (SD = .03) for trochaic words and 126 ms (SD = .02) and 240 ms (SD = .05) for iambic words. In the s, the first vertical dotted line is associated with the mean onset of the first syllable, while the second dotted line corresponds to the mean offset of the second syllable.

Figure 4.1 shows that English speakers looked at the target object more than the competitor object from the beginning of the instruction and continued to look at it at the onset of the target word for both trochaic (-243 ms in the left panel) and iambic stress conditions (-126 ms in the right panel). One-way ANOVAs were conducted on the fixation proportions at the word onset, separately for trochaic and iambic conditions. A significant difference
Figure 4.1: English speakers’ ($N = 24$) mean fixation proportion for target, competitor, and distractor objects during word recognition. Left: Trochaic nonwords. Right: Iambic nonwords.

Figure 4.2: Korean L2 learners’ ($N = 21$) mean fixation proportion for target, competitor, and distractor objects during word recognition. Left: Trochaic nonwords. Right: Iambic nonwords.
in fixation proportions between targets and competitors was found for the iambic condition by subject analysis \((F1(1,23) = 5.83, p < .05; F2(1, 15) = 0.33, p > .05)\), but no significant difference was found for the trochaic condition \((F1(1,23) = 1.07, p > .05; F2(1, 15) = 0.65, p > .05)\). Looking at Figure 4.2, we could observe a very similar gaze behavior in Korean L2 learners. A significant baseline effect for the target object was found for the iambic condition by subject analysis \((F1(1,20) = 5.30, p < .05; F2(1, 15) = 1.99, p > .05)\), but not for the trochaic condition \((F1(1,20) = 0.65, p > .05; F2(1, 15) = 0.28, p > .05)\).

Despite design and counterbalancing efforts, including a centralized fixation cross and counterbalanced screen position of target alien image, exploration of data with item-by-item plots of the fixation proportions to the three visual objects revealed that more on-target fixations at the onset of the critical words were made in the trials where target objects were placed at the top right position on the scene display. It seems that participants searched picture referents starting from the top right visual object position and continued to fixate on this object if it was named by the target word in the auditory instruction. As such an ABE could bias the estimation of the intercept and slope in the regression model (see previous Section 3.5.2 for discussion), the trials where gaze location was the target AOI at 17 ms before the onset of the target word were excluded for subsequent analyses (22.4% of correct responses, 136 out of 606 correct responses, were eliminated by this criterion). The remaining correct responses were for 118 trochaic words and 118 iambic words for English speakers, and for 125 trochaic words and 109 iambic words for Korean L2 learners of English.

The gaze data (coded either as a “1” or a “0”) to the target object of the remaining trials were plotted on the empirical logits scale in Figure 4.3. For each plot, the gaze data were time-aligned with the boundary between the first and second syllables at 0 ms. The dotted
lines before and after the alignment time point (i.e., 0 ms) correspond to the mean onsets of first syllables and the means offsets of second syllables of trochaic words (black lines) and iambic words (gray lines).

Figure 4.3: Mean fixation likelihood (empirical logits) for target objects during trochaic and iambic nonword recognition. Left: English speakers (N = 24). Right: Korean learners of English (N = 21).

The left panel of Figure 4.3 shows that English speakers’ fixation likelihood for the target object was rapidly increased toward the end of the first syllable for trochaic words (at -51 ms) and upon the offset of the first syllable for iambic words (at 0 ms). The fixation likelihood for iambic words increased to catch up with the fixation rate of trochaic words right after the first syllable (at 78 ms). Given the time taken to plan and execute saccadic eye movements 150 - 200 ms (Matin et al., 1993), the first syllable information contributed target word activation for both trochaic and iambic word recognition in native
English speakers. This is quite different from the processing pattern found in the preliminary study, where iambic word activation was relatively delayed until the second syllable offset. Therefore, the qualitative differences between stressed and unstressed initial syllables of the nonwords must have provided an important cue for native English speakers to distinguish the target word from its stress competitor at an early stage of lexical processing.

The right panel of Figure 4.3 shows that Korean speakers also had a rapid increase in fixation likelihood to the target object from the first syllable of both trochaic and iambic words. The fixation rates increased in parallel up to the end of the second syllable (240 ms) for both trochaic and iambic words, suggesting that the vowel quality difference in the first syllables of stress minimal pair words was also a critical cue for Korean speakers to distinguish target words from competitors at an early stage of word recognition. However, while trochaic target words were reliably recognized, with fixations increasing until the end of the target phrase, fixations to iambic word targets were much lower, reaching a comparable level to the trochaic words only after the completion of the entire instruction. The activation of iambic words was greatly reduced temporarily after the second syllable up to the time point, 646 ms.

To test the effects of lexical stress pattern and L1 on word recognition, the gaze data to the target AOI were submitted to a Quasi-MLR with the analysis window from -119 ms to 867 ms. The analysis window was defined as the lowest and highest peak in the grand mean of log odds of gaze to target objects across conditions at every 17 ms from the onset of target nonwords. The regression model included L1, Stress pattern, Phonation type, and Time as predictors and Subject and Item as crossed random factors. Table 4.6 shows the results. The results about L1 and Stress pattern are discussed here first.
Table 4.6: Summary of $L1$, Stress pattern, Phonation type, and Time effects on gaze in Quasi-MLR ($N = 8930; \log \text{Lik} = -14393$).

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Est.</th>
<th>SE</th>
<th>$t$</th>
<th>$p &lt;$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-0.906</td>
<td>0.114</td>
<td>-7.932</td>
<td>.001</td>
</tr>
<tr>
<td>$L1 = \text{Korean} L1$</td>
<td>0.196</td>
<td>0.164</td>
<td>1.195</td>
<td>n.s.</td>
</tr>
<tr>
<td>Stress1 = Iambic</td>
<td>0.030</td>
<td>0.107</td>
<td>0.284</td>
<td>n.s.</td>
</tr>
<tr>
<td>Phonation1 = Lax</td>
<td>0.112</td>
<td>0.191</td>
<td>0.586</td>
<td>n.s.</td>
</tr>
<tr>
<td>$L1:\text{Stress1}$</td>
<td>0.187</td>
<td>0.215</td>
<td>0.869</td>
<td>n.s.</td>
</tr>
<tr>
<td>$L1:\text{Phonation1}$</td>
<td>0.155</td>
<td>0.214</td>
<td>0.723</td>
<td>n.s.</td>
</tr>
<tr>
<td>Stress1:Phonation1</td>
<td>-1.086</td>
<td>0.215</td>
<td>-5.059</td>
<td>.0001</td>
</tr>
<tr>
<td>$L1:\text{Stress1}:\text{Phonation}$</td>
<td>0.147</td>
<td>0.429</td>
<td>0.342</td>
<td>n.s.</td>
</tr>
<tr>
<td>Time</td>
<td>1.229</td>
<td>0.101</td>
<td>12.204</td>
<td>.0001</td>
</tr>
<tr>
<td>$L1:\text{Time}$</td>
<td>-0.546</td>
<td>0.201</td>
<td>-2.713</td>
<td>.05</td>
</tr>
<tr>
<td>Stress1:Time</td>
<td>-0.323</td>
<td>0.201</td>
<td>-1.603</td>
<td>n.s.</td>
</tr>
<tr>
<td>Phonation1:Time</td>
<td>-0.133</td>
<td>0.201</td>
<td>-0.662</td>
<td>n.s.</td>
</tr>
<tr>
<td>$L1:\text{Stress1}:\text{Time}$</td>
<td>-0.965</td>
<td>0.403</td>
<td>-2.395</td>
<td>.05</td>
</tr>
<tr>
<td>$L1:\text{Phonation1}:\text{Time}$</td>
<td>-0.159</td>
<td>0.403</td>
<td>-0.395</td>
<td>n.s.</td>
</tr>
<tr>
<td>Stress1:Phonation1:Time</td>
<td>1.729</td>
<td>0.403</td>
<td>4.293</td>
<td>.0001</td>
</tr>
<tr>
<td>$L1:\text{Stress1}:\text{Phonation1}:\text{Time}$</td>
<td>-0.979</td>
<td>0.806</td>
<td>-1.216</td>
<td>n.s.</td>
</tr>
</tbody>
</table>

At the intercept, there were no effects of $L1$ and Stress pattern, and no interaction between $L1$ and Stress pattern, which indicates that the fixation likelihoods for the trochaic and iambic target objects were not significantly different at the beginning of analysis window for native English speakers and Korean L2 learners.

For the slope, an effect of Time was significant, suggesting that overall fixation likelihood was significantly increased over time. An effect of $L1$ on the slope was significant, but an effect of Lexical stress was not significant. The negative coefficient for the $L1$ effect on the slope ($\beta = -0.546, p < .05$) indicates that the increase of fixation likelihood for the target object over time was higher for English speakers than for Korean speakers. No significant effect of Lexical stress on the slope means that overall fixation increase rate was not significantly different for trochaic and iambic word recognition. However, an interaction
between L1 and Stress pattern on the slope was significant. The negative coefficient of the interaction ($\beta = -0.965, p < .05$) indicates that gaze increase was significantly higher for trochaic words than for iambic words for Korean speakers than it was for English speakers. The results show that Korean L2 learners had more processing difficulty for iambic word recognition as compared to trochaic words than did native English speakers.

To examine whether the relatively shallower slope for iambic words in Korean L2 learners was due to the decreased fixation likelihood for the iambic target object in the middle of word recognition, a Quasi-MLR was conducted for a series of segmented time windows - -119 ms - 0 ms for the first syllable portion and every 200 ms window after the offset of first syllable (0 - 200 ms, 200 - 400 ms, 400 - 600 ms 600 - 867 ms). L1, Stress pattern, and Phonation type were the predictors and Subject and Item were the crossed random effects. Table 4.7 summaries the results.

The result revealed no effects of L1 and Stress pattern, nor the interaction between L1 and Stress pattern in the first three windows (-119 - 0 ms, 0 - 200 ms, and 200 - 400 ms). This indicates that Korean L2 learners were not significantly different from native English speakers in the chance of looking at the target of both trochaic and iambic words at an early stage of word recognition. It seems that L2 learners exploited the segmental correlate to lexical stress to distinguish the target from the competitor as native English speakers did.

However, the interaction between L1 and Stress pattern turned out to be significant in the window, 400 - 600 ms and 600 - 867 ms which approximates the time after the last disambiguating syllable, with no other effects approaching significance except the main effect of Stress pattern in window 400 - 600 ms. A main effect of Stress pattern in window 400 - 600 ms ($\beta = -0.323, p < .05$) indicates that the chance of looking at the target object was higher for trochaic words than for iambic words even after the third syllable was heard.
<table>
<thead>
<tr>
<th>Predictor</th>
<th>Est.</th>
<th>SE</th>
<th>t</th>
<th>p&lt;</th>
<th>Est.</th>
<th>SE</th>
<th>t</th>
<th>p&lt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analysis window</td>
<td>-119 ms - 0 ms</td>
<td>(N = 940; logLik = -1438)</td>
<td></td>
<td></td>
<td>0 ms - 200 ms</td>
<td>(N = 1880; logLik = -2957)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>1.080</td>
<td>0.104</td>
<td>-10.392</td>
<td>.0001</td>
<td>-0.645</td>
<td>0.111</td>
<td>-5.792</td>
<td>.0001</td>
</tr>
<tr>
<td>L1 = Korean L1</td>
<td>0.133</td>
<td>0.203</td>
<td>0.655</td>
<td>n.s.</td>
<td>0.070</td>
<td>0.201</td>
<td>0.352</td>
<td>0 n.s.</td>
</tr>
<tr>
<td>Stress1 = Iambic</td>
<td>-0.088</td>
<td>0.154</td>
<td>-0.572</td>
<td>n.s.</td>
<td>0.088</td>
<td>0.119</td>
<td>0.743</td>
<td>n.s.</td>
</tr>
<tr>
<td>Phonation1 = Lax</td>
<td>-0.089</td>
<td>0.161</td>
<td>-0.556</td>
<td>n.s.</td>
<td>0.107</td>
<td>0.153</td>
<td>0.699</td>
<td>n.s.</td>
</tr>
<tr>
<td>L1:Stress1</td>
<td>0.068</td>
<td>0.309</td>
<td>0.221</td>
<td>n.s.</td>
<td>0.171</td>
<td>0.238</td>
<td>0.721</td>
<td>n.s.</td>
</tr>
<tr>
<td>L1:Phonation1</td>
<td>-0.082</td>
<td>0.309</td>
<td>-0.264</td>
<td>n.s.</td>
<td>0.115</td>
<td>0.238</td>
<td>0.484</td>
<td>n.s.</td>
</tr>
<tr>
<td>Stress1:Phonation1</td>
<td>-0.893</td>
<td>0.309</td>
<td>-2.891</td>
<td>.05</td>
<td>-0.767</td>
<td>0.238</td>
<td>-3.144</td>
<td>.05</td>
</tr>
<tr>
<td>L1:Stress1:Phonation1</td>
<td>-0.166</td>
<td>0.617</td>
<td>-0.269</td>
<td>n.s.</td>
<td>0.106</td>
<td>0.475</td>
<td>0.223</td>
<td>n.s.</td>
</tr>
<tr>
<td>Analysis window</td>
<td>200 ms - 400 ms</td>
<td>(N = 1880; logLik = -3019)</td>
<td></td>
<td></td>
<td>400 ms - 600 ms</td>
<td>(N = 1880; logLik = -3065)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>-0.350</td>
<td>0.158</td>
<td>-2.219</td>
<td>.05</td>
<td>-0.123</td>
<td>0.143</td>
<td>-0.859</td>
<td>n.s.</td>
</tr>
<tr>
<td>L1 = Korean L1</td>
<td>-0.101</td>
<td>0.229</td>
<td>0.442</td>
<td>n.s.</td>
<td>-0.134</td>
<td>0.193</td>
<td>-0.696</td>
<td>n.s.</td>
</tr>
<tr>
<td>Stress1 = Iambic</td>
<td>-0.085</td>
<td>0.122</td>
<td>-0.701</td>
<td>n.s.</td>
<td>-0.323</td>
<td>0.125</td>
<td>-2.575</td>
<td>.05</td>
</tr>
<tr>
<td>Phonation1 = Lax</td>
<td>0.103</td>
<td>0.249</td>
<td>0.413</td>
<td>n.s.</td>
<td>0.184</td>
<td>0.245</td>
<td>0.752</td>
<td>n.s.</td>
</tr>
<tr>
<td>L1:Stress1</td>
<td>-0.144</td>
<td>0.242</td>
<td>-0.596</td>
<td>n.s.</td>
<td>-0.516</td>
<td>0.250</td>
<td>-2.064</td>
<td>0.5</td>
</tr>
<tr>
<td>L1:Phonation1</td>
<td>0.313</td>
<td>0.243</td>
<td>1.293</td>
<td>n.s.</td>
<td>0.544</td>
<td>0.250</td>
<td>0.218</td>
<td>n.s.</td>
</tr>
<tr>
<td>Stress1:Phonation1</td>
<td>-0.425</td>
<td>0.243</td>
<td>-1.750</td>
<td>n.s.</td>
<td>-0.354</td>
<td>0.250</td>
<td>-1.413</td>
<td>n.s.</td>
</tr>
<tr>
<td>L1:Stress1:Phonation1</td>
<td>-0.273</td>
<td>0.485</td>
<td>-0.563</td>
<td>n.s.</td>
<td>-0.707</td>
<td>0.450</td>
<td>-1.416</td>
<td>n.s.</td>
</tr>
<tr>
<td>Analysis window</td>
<td>600 ms - 867 ms</td>
<td>(N = 2350; logLik = -3845)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>-0.016</td>
<td>0.186</td>
<td>-0.086</td>
<td>n.s.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L1 = Korean L1</td>
<td>-0.238</td>
<td>0.255</td>
<td>-0.932</td>
<td>n.s.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stress1 = Iambic</td>
<td>-0.166</td>
<td>0.111</td>
<td>-1.496</td>
<td>n.s.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phonation1 = Lax</td>
<td>-0.070</td>
<td>0.292</td>
<td>-0.239</td>
<td>n.s.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L1:Stress1</td>
<td>-0.608</td>
<td>0.221</td>
<td>-2.755</td>
<td>.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L1:Phonation1</td>
<td>-0.023</td>
<td>0.220</td>
<td>-0.105</td>
<td>n.s.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stress1:Phonation1</td>
<td>0.417</td>
<td>0.222</td>
<td>1.881</td>
<td>n.s.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L1:Stress1:Phonation1</td>
<td>-0.432</td>
<td>0.441</td>
<td>-0.980</td>
<td>n.s.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.7: Summary of L1, Stress pattern, and Phonation type effects on fixation likelihood in Quasi-MLR for segmented windows.

This provides evidence that the iambic stress pattern posed more processing difficulty than did the trochaic stress pattern even after a disambiguating syllable was heard. The negative coefficients for the interaction between L1 and Stress pattern ($\beta = -0.516, p < .05$ for
400 - 600 ms; $\beta = -0.608, p < .05$ for 600 - 867 ms) indicate that the difference between trochaic and iambic words was larger for Korean L2 learners than for native English speakers. The results provide evidence that Korean L2 learners had chronic processing difficulty for iambic words even after the target words were activated by an additional segmental cue to lexical stress and the last disambiguating syllable was heard.

Another central question we address in Experiment 1 is whether word-initial phonation type modulates the processing of lexical stress during word recognition. We examine the effect of Phonation type and its interaction with other predictors, Stress pattern and L1 over time. For the analysis, the test nonwords were classified into those that Korean L2 learners heard as ‘aspirated’ vs. ‘lax’ initial words, based on the transcription data collected from Korean participants in the post-eyetracking orthographic transcription task (98.8% transcription agreement for the phonation type contrast of word-initial consonants among Korean L2 learners, 99% transcription agreement among the native English speakers for the voicing contrast of what Korean listeners heard as a phonation type contrast (see Table C.2 in Appendix C). Of the correct responses for trochaic ($n = 118$) and iambic word recognition ($n = 118$) in English speakers, there were 63 aspirated (voiceless) and 55 lax (voiced) initial trochaic words, and for 44 aspirated (voiceless) and 74 lax (voiced) initial iambic words. Korean speakers provided correct answers for 61 aspirated and 64 lax-initial trochaic words ($n = 125$), and for 53 aspirated and 56 lax-initial iambic words ($n = 109$).

Figure 4.4 shows native English speakers’ gaze to the target object over time for the two stress conditions as the function of word-initial phonation type contrast.

In the left panel of Figure 4.4, English speakers showed a steeper slope for aspirated (voiceless) initial trochaic words than for lax (voiced) initial trochaic words in the analysis window between -119 ms and 867 ms and it was the opposite for iambic words as shown in
the right panel of Figure 4.4. As English speakers transcribed what the Korean L2 learners heard as aspirated- vs. lax-initial consonants as voiceless- vs. voiced- initial consonants, respectively, the slope change suggests that voiceless word-initial syllable with relatively high f0 and long aspiration facilitated trochaic word recognition and voiced word-initial syllable with relatively low f0 and short aspiration facilitated iambic word recognition. However, Figure 4.4 shows some ABEs for voiced trochaic words over voiceless trochaic words and for voiceless iambic words over voiced iambic words at the onset of the analysis window (i.e. -119 ms). Looking at the radical decrease of fixation likelihood for voiced trochaic words at around 500 ms, the early fixation to the target object led to early recognition of target words for voiced trochaic words. This is consistent with the shorter RT for the lax-initial trochaic word recognition (2059.6 ms) as compared to the one for aspirated-initial trochaic word recognition (2250.2 ms, Table 4.4). As an ABE is more likely to lead
to a shallow slope change as in the case of voiceless initial iambic words, the gaze data after
the onset of the analysis window cannot be as informative as it should be (Brown-Schmidt,
2009, p. 896). Therefore, if we have conditions with baseline effects, it would be hard to
make a direct comparison between the conditions (i.e., aspirated- (voiceless-) vs. lax-
(voiced-) initial words in each stress condition).

Then, what causes the ABE in this study? Since the ABE seemed to be initiated before
the word onset, it’s not reasonable to ascribe the early on-target fixation to the test nonword
stimuli that came after the effect. Looking at the position distribution of the target object
in the trials, the target object appeared in the top left object position more often for lax-
(voiced-) initial trochaic words than for aspirated- (voiceless-) initial trochaic words, and
it appeared at the top right object position more often for aspirated- (voiceless-) initial
iambic words than for lax- (voiced-) initial iambic words, as shown in Table 4.8. Even
though on-target trials were eliminated in the analysis to avoid ABE effects for targets over
competitors, when the positions of target objects were not proportionally balanced between
conditions, we may still have the ABEs.

<table>
<thead>
<tr>
<th>Initial Consonant</th>
<th>Trochaic</th>
<th>Iambic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aspirated (Voiceless)</td>
<td>23.8% (15 out of 63 trials)</td>
<td>43.8% (19 out of 44 trials)</td>
</tr>
<tr>
<td>Lax (Voiced)</td>
<td>40% (22 out of 55 trials)</td>
<td>16.2% (12 out of 74 trials)</td>
</tr>
</tbody>
</table>

Table 4.8: Proportion of trials with target object at the top right object position for each
condition.

Figure 4.5 shows Korean L2 learners’ mean gaze to the target object for the stress
conditions as a function of word-initial phonation type contrast.
Figure 4.5 shows that Korean L2 learners also had strong ABEs at the beginning of the analysis window (i.e., -119 ms) in both trochaic and iambic conditions. Lax-initial words attracted earlier on-target fixation at the onset of the analysis window than did aspirated-initial words for trochaic condition, and it was the opposite for iambic condition. The proportion of target object position in these trials also revealed that the condition with the larger number of trials that had target object in the top left object position was more likely to generate a high fixation likelihood for the target before and at the onset of the target word. Target object appeared in the top left object position more often for lax (voiced) initial trochaic words (29 out of 64 trials, 45.3%) than for aspirated-initial trochaic words (12 out of 61 trials, 19.7%), and it appeared in the top left AOI more often for aspirated (voiceless) initial iambic words (22 out of 53 trials, 41.5%) than for lax (voiced) initial iambic words (12 out of 56 trials, 21.4%). Even though the ABEs are more likely to lead to shallow slopes, the slopes for the aspirated and lax-initial words increased almost in
parallel in the trochaic condition, suggesting that Korean L2 learners were not influenced by L1 accentual phrase form when processing English trochaic stress pattern for word recognition. However, the slope increase of lax-initial trochaic words did not last long toward the end of the analysis window. The slope for aspirated-initial iambic words was not as steep as the one for lax-initial iambic words.

To test the effect of Phonation type and the interactions with other predictors at the intercept and on the slope, we refer back to the Quasi-MLR results in Table 4.6. This analysis revealed no significant effect of Phonation type and no interaction between L1 and Phonation type at the intercept. However, the interaction between Stress pattern and Phonation type was significant at the intercept. The negative estimate for the interaction between Stress pattern and Phonation type ($\beta = -1.086, p > .05$) indicates that the chance of looking at the target started off significantly higher for lax-initial words than for aspirated-initial words when the words were trochaic and it was the opposite when the words were iambic. The interaction between L1, Stress pattern, and Phonation type was not significant at the intercept, suggesting the interaction between Stress pattern and Phonation type was similar for native English speakers and L2 learners.

For the slope term, there was not an effect of Phonation type, nor an interaction of Phonation type and L1, indicating that the use of phonation type contrast over time was similar for native English speakers and L2 learners. An interaction between Stress pattern and Phonation type on the slope was significant to be consistent with the slope changes in Figure 4.4 and Figures 4.5. The positive parameter estimate for the interaction between Stress pattern and Phonation type on the slope ($\beta = 1.729, p > .001$) provides evidence that the increase of fixation likelihood to the trochaic target was significantly higher when the word began with what Korean listeners heard as aspirated sounds than when it began
with what they heard as lax sounds, and the increase rate of fixation likelihood to the
iambic target was significantly higher when the word began with what Korean listeners
heard as lax sounds than when it began with what they heard as aspirated sounds. The
interaction between \( L1 \), \( Stress \ pattern \), and \( Phonation \ type \) on the slope was not significant,
supporting that the effect of phonation type contrast (voicing contrast) on processing of
lexical stress patterns over time was similar for native English speakers and Korean L2
learners of English.

The results from the segmented windows (see Table 4.7) also showed no significant
effect of \( Phonation \ type \) and no interaction between \( L1 \) and \( Phonation \ type \) in any windows
to be consistent with results at the intercept and on the slope. The interaction between
\( Stress \ pattern \) and \( Phonation \ type \) was significant in the first two windows. The negative
parameter estimate (\( \beta = -0.893, p < .05 \)) for the interaction in the first window (-119 - 0
ms) is consistent with the significant interaction at the intercept in the analysis with the
entire window. That is, the chance of looking at the target was significantly higher for
lax-initial words than for aspirated-initial words when the words were trochaic words and
it was the opposite when the words were iambic words. The interaction in the second
window 0 - 200 ms (\( \beta = -0.767, p < .05 \)) showed a similar interaction pattern for word-
initial phonation type and lexical stress. The lack of a significant interaction between \( Stress \pattern \) and \( Phonation \ type \) in the subsequent windows suggests that an interaction between
\( Stress \ pattern \) and word-initial \( Phonation \ type \) at an early stage of word recognition is no
longer available. The interaction between \( Stress \ pattern \), \( Phonation \ type \), and \( L1 \) was not
significant in any of the time windows. This demonstrates that the interaction between
\( Phonation \ type \) and \( Stress \ pattern \) was not significantly different for the two \( L1 \) groups
from the beginning to the end of word recognition. The result supports the conclusion that
both native speakers and Korean L2 learners had similar processing patterns for lexical stress in relation to the word-initial phonation type contrast.

To sum up, the Quasi-MLR analyses with and without segmentation of analysis window between -119 - 867 ms show that native English speakers were overall faster at word recognition than Korean L2 learners of English, with steeper slope change for English speakers than for Korean L2 learners. The significant interaction between $L1$ and Stress pattern on the slope suggests that Korean L2 learners had more processing difficulty for iambic words than for trochaic words as compared to native English speakers. The segmented window analysis also reveals the interaction during the time window of 400 - 867 ms. Given the time point is after the disambiguating third syllable, the processing difficulty that Korean L2 learners had with iambic words must be strong enough to suppress the effects of vowel quality difference in the first syllable and disambiguating segmental cues in the last syllable. With regard to Korean L1 prosodic influence on English lexical stress processing, the interaction between Stress pattern and Phonation type was significant on the slope with no interaction with $L1$, suggesting that the trochaic words were more facilitated when the words began with what Korean heard as aspirated sounds than when the words began with what Korean heard as lax sounds, and vice versa for the iambic words. However, since there were strong ABEs at the intercept that went opposite to the slope, we cannot use the slope results to support the prediction that the word-initial phonation type (voicing contrast) influences the processing of lexical stress during word recognition.
4.6 Discussion

One of the central questions we addressed in Experiment 1 was how native English speakers and Korean L2 learners of English process English lexical stress for word recognition when the lexical stress involves both suprasegmental variables and vowel quality differences in stressed and unstressed syllables as it does in natural speech. It was expected that native English speakers could use the vowel quality difference in the first syllable to constrain the activation of stress competitor words from an early stage of lexical processing. In this case, the activation of target words can be comparable for trochaic and iambic words. Korean L2 learners were also expected to use the vowel quality difference to discern the stress competitors, in which case their word recognition should be as efficient as the one in native English speakers at an early stage of lexical processing. Yet, if they still do not pay attention to the prominence difference in syllables, the processing difficulty for iambic words is more likely to persist, as in the preliminary study.

The training result of the present study showed that eyetracking word recognition task was based on comparable learning of the nonword-picture associations between native English speakers and Korean L2 learners of English. Though the eyetracking experiment did not show any significant difference between native English speakers and L2 learners in accuracy and response times, the eye movement record showed that L2 learners had more processing difficulty for iambic word recognition than did native English speakers.

A closer look at the processing pattern over time revealed that native English speakers were overall faster at recognizing target words than Korean L2 learners of English, with a steeper slope change over time. The interactions between *Lexical stress* and *L1* on the slope and in the time window between 400 and 867 ms provide evidence that activation of trochaic and iambic words was more comparable in English speakers than in Korean L2
learners over the course of word recognition. This suggests that the vowel quality difference in the stressed/unstressed initial syllable was an important cue to the perception of lexical stress in English speakers, facilitating word recognition consistently from an early stage of word recognition. This result supports the previous claim that vowel quality is an important cue for native English speakers to determine stressed and unstressed syllables (Fear et al., 1995; Lieberman, 1965). Furthermore, it provides evidence that, unlike unstressed syllables with full vowels in the preliminary study, unstressed syllables with reduced vowels can trigger early activation of target words as stressed syllables do during spoken word recognition.

The significant interaction between Lexical stress and L1 in the window 400 - 867 ms which approximates the time after the third disambiguating syllable (the interaction was significant after the second syllable in the preliminary study) also suggests that the full vs. reduced vowel quality difference in the first syllable was perceived by Korean L2 learners, initiating reliable activation of target words at an early stage of word recognition as for native English speakers. However, the processing difficulty for the iambic words after the third disambiguating syllable provides evidence that iambic words posed processing difficulty to Korean L2 learners anyway, despite the fact that the words contained unambiguous vowel information in the first two syllables. This result supports the prediction that native English speakers and Korean L2 learners use the initial vowel quality difference in a different way. That is, as reduced vowels do not occur in Korean, the Korean L2 learners were less likely to use vowel reduction as a correlate to the unstressed syllable, but rather as a distinctive vowel category. Though this would make Korean speakers use the segmental cue during early lexical processing, the processing difficulty for iambic words seemed
to be maintained as the unstressed initial syllables were given with a greatly reduced f0, duration, and intensity cues.

Another central question of Experiment 1 was whether the perceived phonation type of the word-initial segment affects the perception and processing of lexical stress during spoken word recognition. The slope results revealed a significant interaction between *Stress pattern* and *Phonation type* without an interaction with *L1* on the slope, suggesting that the word-initial phonation type (voicing) contrast affected processing of lexical stress during word recognition for both English and Korean speakers. Trochaic word recognition was more facilitated when the word began with what Korean listeners heard as an aspirated sound than when it began with what Koreans listeners heard as a lax sound, while iambic word recognition was more facilitated when the word began with what Koreans listeners heard as a lax sound than when it began with what Koreans listeners heard as an aspirated sound.

However, the significant interaction between *Stress pattern* and *Phonation type* at the intercept, with no interaction with *L1*, shows that there were also significant ABEs (higher fixation likelihood for lax-initial words than for aspirated-initial words at the onset of analysis window in the trochaic condition and higher fixation likelihood for aspirated-initial words than for lax-initial words in the iambic condition) that went contrary to the interaction on the slope. The interaction at the intercept indicates that both English and Korean speakers gazed at the target earlier for lax-initial words than for aspirated-initial words when the words carried trochaic stress pattern, and they looked at the target earlier for aspirated-initial words than for lax-initial words when the words were iambic. These ABEs were generated due to an experimental artifact of the target objects being more often in the top left position in a scene display for lax- (voiced-) initial trochaic words than for
aspirated- (voiceless-) initial trochaic words and for aspirated- (voiceless-) initial iambic words than for lax- (voiced-) initial iambic words.

As such a baseline effect occurs when listeners had already fixated to a target object in the visual scene display, the slope steepness that is estimated based on eye movements toward a target object over time is less likely to be high. Therefore, we cannot use the interaction result on the slope to support the claim that word-initial phonation type (voicing contrast) affected the processing of English lexical stress during spoken word recognition.

Overall, the current eyetracking study shows that experience with both phonological representation and acoustic-phonetic cues of L1 prosodic structure modulates the processing of L2 word prosody in Korean L2 learners of English. Korean L2 learners, whose L1 does not employ lexically specified prosodic units and thus no segmental correlate to it, were overall poor at integrating English lexical stress in the process of spoken word recognition compared to native English speakers. Even with the presence of a segmental cue to lexical stress, iambic words still caused the strong retroactive/corrective processing in the middle of word recognition and delayed target word recognition in Korean speakers, but not in native English speakers. This tells us that the reduced vowel in the unstressed initial syllable was not perceived as a stress correlate but as an independent segmental cue by Korean L2 learners of English. When the segmental cue was used in unstressed syllables, it seemed to pose perceptual difficulty to Korean L2 learners and so the processing of iambic words was not as efficient as the processing of trochaic words.
Nevertheless, the finding that the aspirated-initial trochaic words and the lax-initial words had a rapid increase of fixation likelihood over time, based on slope, provides evidence for the prediction that perceptual experience with the L1 phrase-level prosodic system is maintained in the perception and processing of lexical stress during word recognition. However, since there were also strong ABEs, whether there is any effect of word-initial phonation type on lexical stress processing cannot be confirmed in Experiment 1.
CHAPTER 5: EXPERIMENT 2

Experiment 1 examined the effect of lexical stress, including reduced vowels in un-stressed syllables, on word recognition in L1 and L2 English. With vowel cues in un-stressed syllables, Korean-speaking learners of English activated the target word earlier in time during word recognition as did native English speakers. Still, the L2 learners had processing difficulty for iambic words during the middle of word recognition. The effect of word-initial phonation type on lexical stress processing was not fully confirmed in Experiment 1.

Experiment 2 further investigates this issue without changing the vowel quality in stressed and unstressed syllables to examine whether the effect of word-initial phonation type affects the processing of lexical stress. Experiment 2 uses a cross-modal fragment gating paradigm, which was chosen because it measures successive responses as the spoken word unfolds. Listeners hear the first syllable of the nonword, or the first two syllables of the nonword, allowing examination of the processing of the initial syllable and subsequent syllables separately.

The specific research questions to be explored in the cross-modal gating experiment are: 1) How do Korean L2 learners of English use English lexical stress for word recognition in comparison to native English speakers when the lexical stress doesn’t involve a vowel quality cue? Is there any difference between trochaic and iambic word identification? 2) How do Korean L2 learners of English use the word-initial phonation type of aspirated vs. lax contrast, in comparison to native English speakers with the comparable voiceless/voiced
distinction, when processing English lexical stress that doesn’t involve a vowel quality cue?
3) How do Korean L2 learners of English process English lexical stress and word-initial phonation type as they encounter successive syllables of a word, in comparison to native English speakers?

It is expected that native English speakers will show overall better performance than Korean L2 learners of English with higher accuracy and shorter RTs for word identification. For the use of the lexical stress patterns, as many previous studies and the preliminary eyetracking study have shown, both native English speakers and Korean L2 learners are expected to have a processing advantage for trochaic over iambic words, thereby with higher accuracy and shorter RTs for trochaic word recognition than for iambic word recognition. However, this should be more persistent in Korean L2 learners of English compared to native English speakers, who should be more readily integrating the following stressed syllable information for lexical access.

As for the word-initial phonation type effect on the processing of lexical stress, if English speakers perceive what Korean heard as an aspirated vs. lax contrast as a voiceless and voiced contrast as in the previous eyetracking experiments, they should show higher accuracy and shorter RTs for voiceless initial trochaic word identification than for voiced initial trochaic word identification, and higher accuracy and shorter RTs for voiced iambic word identification than for voiceless iambic word identification. If the word-initial phonation type affects Korean L2 learners’ processing of English lexical stress, higher accuracy and shorter RTs are expected for aspirated-initial trochaic word identification than for lax-initial trochaic word identification, and for lax-initial iambic word identification than for aspirated-initial iambic word identification.
If the word-initial phonation (voicing) type affects the processing of lexical stress differently in native English speakers and Korean L2 learners of English, the interaction of stress pattern, phonation type, and L1 is not expected to appear after listeners hear the initial syllable fragment. However, as the presented fragment includes more syllables with more syllable prominence information than did the initial-syllable fragment, higher accuracies and shorter RTs are expected for the long word fragment than for the short word fragment in both English and Korean speakers. When the fragment contains a word in a full form, the segmentally disambiguating last syllable in addition to the prominence difference information in the first two syllables can better lead English and Korean speakers to the target words with higher accuracy and shorter RTs than in the shorter fragment blocks.

The remainder of Chapter 4 is organized as follows: Sections 5.1 through 5.3 describe the experimental design, participants, and procedures. Sections 5.5 and 5.6 present the results and discussion.

5.1 Design and Materials

The current experiment examines how native English speakers and Korean-speaking learners of English process English lexical stress and word-initial phonation type contrast for word identification and how processing changes as they hear successively longer portions of the word. This experiment uses a 2*2*3*2 factorial design with stress pattern (trochaic vs. iambic), word-initial consonant phonation type (aspirated- vs. lax- for Korean listeners), and fragment size (one-syllable vs. two-syllable vs. whole word fragment) as three within-subjects factors, and L1 (English vs. Korean) as a between-subjects factor.
5.1.1 Auditory Stimulus Materials

The auditory stimuli were constructed in the same manner as in Experiment 1, with the same phoneme constituents as the ones used in Experiment 1. However, the stress minimal pairs did not include reduced vowels in unstressed syllables to make the lexical stress cue in nonwords more like the preliminary experiment. The first two vowels of the test pairs were the full vowels, /i, i, a, æ, e, e, o, o/, that were used in Experiment 1. The sixteen near stress minimal pairs were constructed by counterbalancing word-initial phonation type: thus, 8 pairs began with obstruents that Korean native speakers in a pre-listening task heard as aspirated (/p, t, k, s/) and the other 8 pairs begin with those they heard as lax (/b, d, g, m/).

The vowel combinations in the first two syllables of the test pairs were counterbalanced for word-initial phonation types so that the first two vowel sequences were the same in the 8 aspirated-initial nonword pairs and the corresponding 8 lax-initial nonword pairs. In addition to the 16 test pairs, 16 filler nonwords that were phonologically unrelated to any of the corresponding target pair members were created. The 32 test nonwords and 16 fillers are listed in Appendix A.3.

The 48 nonwords were recorded in isolation and also in the contextually neutral carrier sentence, “Click on the (target word) now”, with L+H* on the target word and transcribed in the same manner as in the previous eyetracking experiments. The speaker was the same as in Experiment 1, but the transcriber was a different phonetically trained native English speaker. Nonword recordings were pretested by a native English speaker and a Korean native speaker to ensure that the nonwords are not heard as any real words in both English and Korean.

The nonword recorded with a carrier sentence was then truncated into three different sizes of fragment for use in the gating word identification task: 1) short fragment, where
the phrase contained only the initial syllable of the critical word (e.g. “Click on the /'tɪ-/'”),
2) long fragment, the phrase contained only the segmentally overlapping but prosodically
contrastive portion (e.g. “Click on the /tɪ'mə-/’), and 3) whole word, where the full target
word remains (e.g. “Click on the /'tɪmə do/ now’). As coarticulatory information at the
boundary between syllables can provide the information about the identity of the subse-
quent syllable in a word, the end of the syllable fragment was segmented slightly before
the boundary so as not to include such a transitional cue. For example, the first syllable of
the word, /'tɪ-/, was cut before the spectral cues to the next segment, /ml/, became salient.

5.1.2 Acoustic Phonetic Measures

Acoustic measures were made on the auditory stimuli for both suprasegmental and
segmental lexical stress cues to confirm that the stress minimal pair members were only
prosodically contrastive in the first two syllables and that the word-initial phonation type
contrast involved an f0 height difference in the vowels of word-initial syllables.

Duration, f0, and Intensity

The suprasegmental acoustic measures include duration, mean f0, and mean intensity
in the first two vowels of the 16 test nonword tokens. Paired t-tests showed that trochaic
words in isolation had first syllables that were longer (160 ms vs. 60 ms), \( t(15) = 21.5, \)
\( p < .001 \); higher in mean f0 (264 Hz vs. 204 Hz), \( t(15) = 18.4, p < .001 \); and greater in
loudness (33 sone vs. 21 sone), \( t(15) = 8.6, p < .001 \), than the first syllables of iambic
words in isolation. Conversely, iambic words in isolation had second syllables that were
longer (170 ms vs. 80 ms), \( t(15) = 14.8, p < .001 \); higher in mean f0 (250 Hz vs. 211 Hz),
\( p(15) = 11.8, p < .001 \); and greater in loudness (31 sone vs. 16 sone), \( t(15) = 13.2, p < .001, \)
than the second syllables of trochaic words in isolation.
In addition, first syllables were different from second syllables within a word for the two stress conditions: first syllables of trochaic words in isolation were longer (160 ms vs. 80 ms), \( t(15) = 11.6, p < .001 \); higher in mean f0 (264 Hz vs. 211 Hz), \( t(15) = 11.2, p < .001 \); and greater in loudness (33 sone vs. 16 sone), \( t(15) = 10.2, p < .001 \), than the second syllables of trochaic words in isolation. Words with iambic stress patterns had second syllables that were longer (170 ms vs. 60 ms), \( t(15) = 13.5, p = .001 \); higher in mean f0 (250 Hz vs. 204 Hz), \( t(15) = 14.3, p < .001 \); and greater in loudness (31 sone vs. 21 sone), \( t(15) = 4.5, p < .001 \), than the first syllables.

The same measurements were conducted on the test nonwords recorded in the carrier sentence. Paired t-tests showed that trochaic words had first syllables that were longer (150 ms vs. 60 ms), \( t(15) = 15.0, p < .001 \); higher in mean f0 (272 Hz vs. 206 Hz), \( t(15) = 20.1, p < .001 \); and greater in loudness (37 sone vs. 19 sone), \( t(15) = 9.7, p < .001 \), than first syllables of iambic words. Conversely, iambic words had second syllables that were longer (105 ms vs. 80 ms), \( t(15) = 12.1, p < .001 \); higher in mean f0 (264 Hz vs. 204 Hz), \( p(15) = 18.8, p < .001 \); and greater in loudness (31 sone vs. 17 sone), \( t(15) = 17.0, p < .001 \), than second syllables of trochaic words.

For the two stress conditions, first syllables of the test nonwords recorded in a carrier sentence were significantly different from the second syllable of these words. First syllables of trochaic words in a carrier sentence were longer (150 ms vs. 80 ms), \( t(15) = 8.7, p < .001 \); higher in mean f0 (272Hz vs. 204 Hz), \( t(15) = 18.8, p < .001 \); and greater in loudness (37 sone vs. 17 sone), \( t(15) = 10.1, p < .001 \), than second syllables of trochaic words in a carrier sentence. Words with iambic stress patterns had second syllables that were longer (150 ms vs. 60 ms), \( t(15) = 12.9, p = .001 \); higher in mean f0 (264 Hz vs. 206 Hz), \( t(15) = 17.3, p \)
and greater in loudness (31 sone vs. 17 sone), $t(15) = 19.0, p < .001$, than the first syllables.

In addition to the acoustic measures of lexical stress, mean f0 of the initial syllables was compared as a function of phonation type contrast to make sure aspirated-initial words began with higher f0s than lax-initial words for both trochaic and iambic words. Paired t-tests showed that first syllables in aspirated-initial trochaic words produced in isolation were higher in mean f0 (274 Hz vs. 255 Hz), $t(7) = 3.3, p < .05$ than first syllables in lax initial trochaic words. Iambic words with aspirated-initial sounds had first syllables that were higher in mean f0 (206 Hz vs. 202 Hz), $t(7) = 2.1, p < .05$ than the first syllables of lax initial words. Aspirated-initial trochaic words produced in the carrier sentence also had first syllables that were higher in mean f0 (285 Hz vs. 259 Hz), $t(7) = 5.6, p < .001$ than first syllables of lax initial trochaic words. Iambic words with aspirated-initial sounds had first syllables that were higher in mean f0 (213 Hz vs. 200 Hz), $t(7) = 3.5, p < .05$ than the first syllables of lax initial words.

Lastly, mean f0 of the word *the* used in the carrier sentence was measured to ensure that there was no significant early f0 rise before a trochaic target word. Paired t-tests revealed no significant difference between the mean f0s of the vowel in the word *the* produced before trochaic words and the one produced before iambic words (200 Hz vs. 203 Hz, $t(15) = 1.4, p > .05$).

The acoustic measurements indicate that suprasegmental cues might conceivably differentiate cohort words that are phonemically identical early on. If stress location is utilized in lexical learning, the cues can be used by listeners to discern the pairs that differ in stress. And if listeners are sensitive to the f0 height difference between word-initial aspirated and
lax syllables, the phonation type in the word-initial syllable is expected to affect the perception and processing of English lexical stress.

Vowel Quality

To confirm that the unstressed syllables didn’t involve vowel reduction, F1 and F2 frequencies were taken from the first two vowels in the test nonword to measure vowel quality in the same manner as in the previous eyetracking experiments. The Euclidean distance of each vowel from the central point in the speaker’s F1 × F2 (Hz) space was computed, separately for the vowels produced in words in isolation and for those produced in the carrier sentence.

Paired t-tests showed that the first vowels of trochaic and iambic words in isolation were not significantly different in the distance from the midpoint of the speaker’s vowel space (479 Hz vs. 372 Hz), $t(15) = 1.9, p > .05$. Second vowels of trochaic and iambic words in isolation were also comparable in the vowel dispersion from the midpoint of the speaker’s vowel space (380 Hz vs. 497 Hz), $t(15) = 1.8, p > .05$.

For the test nonwords recorded in the carrier sentence, paired t-tests showed that the first vowels of trochaic and iambic words had a comparable vowel dispersion from the midpoint of the speaker’s vowel space (457 Hz vs. 362 Hz), $t(15) = 1.3, p > .05$. However, second vowels of trochaic and iambic words were not comparable in the vowel space dispersion (405 Hz vs. 502 Hz), $t(15) = 2.4, p < .05$.

Overall, the measures of vowel distance from the speaker’s vowel space midpoint indicate that the stress contrast in the test nonwords did not involve a significant vowel quality difference. Therefore, if listeners use lexical stress information for lexical learning, they would primarily rely on suprasegmental lexical stress cues. Though the dispersion was significantly different in the second syllables of trochaic and iambic words recorded in the
carrier sentence, this significant difference is not considered to be critical to the research questions in the present study, because what we concern with here is how the phonation type of word-initial consonant influences word recognition when the vowel quality of word-initial syllables is not disambiguating for stress minimal pair words.

5.1.3 Visual Stimulus Materials

The visual stimuli were the same line drawings of space aliens as in the previous eye-tracking experiments.

5.2 Participants

Thirty-three native English speakers and thirty-two Korean-speaking learners of English participated in the experiment. The participants were those who scored above 90% accuracy in the picture choice task on the last day of training and above 75% accuracy in the third block of whole words in the gating experiment (65 out of 78 participants, 83%, who completed all training and gating tasks).

Native English speakers spoke midwestern American English and were aged 18-34 ($M = 21.5, SD = 2.88$). They were born in the U.S., learning English as their L1. Some had history of foreign language study in high school or college, but none of them had lived abroad and no other language was spoken in the home during childhood. They received partial course credit in a linguistics course and a small payment for their participation.

Korean speaking learners of English were undergraduate or graduate students of the Ohio State University, aged 18-33 ($M = 24.2, SD = 3.9$) at the time of recruitment. They reported speaking Seoul Korean as an L1 with no residence history in the area of Kyungsang
dialect. None of the Korean speakers had any English-immersion experience before entering the U.S. They entered the U.S. at an average age of 22.3 years ($SD = 4.98$, range from 13 to 32 years of age), and had lived in the U.S. for three months to five years ($M = 1.9$, $SD = 1.83$) at the time of recruitment. The mean beginning age of learning L2 English in Korea was 12 years old ($SD = 2.40$, range from 8 to 18 years old). In the language background questionnaire, Korean speakers were asked to evaluate their comprehensive English proficiency on a scale from 1 to 4 (1 = barely; 2 = poorly; 3 = passably; 4 = fluently). The average self-evaluation of English fluency was 3 ($SD = .60$, range from 2.5 to 4). They were all recruited voluntarily and paid for their participation.

None of the participants had participated in the preliminary eyetracking study or Experiment 1. None of the participants reported speech or hearing disorders or vision impairment at the time of testing.

5.3 Procedures

The cross-modal fragment gating experiment comprised three days of training, a main word identification task, and a listening-with-orthographic transcription task, as in the previous eyetracking experiments. Training tasks modeled on the previous eyetracking experiments were employed to establish the nonword-to-alien drawing associations in participants. All participants were equally exposed to 16 test nonword pairs and 16 fillers with picture referents, 30 times over the three days. Two experimental lists of 48 nonwords were used for training sessions as in Experiment 1 (see Section 4.3 for details).

In the following gating word identification experiment, participants were presented with one of four lists of a picture-choice task. Each experimental list consisted of three blocks - Block 1 included 16 word-initial syllable fragments (short fragment) (e.g. “Click on the
and Block 2 consisted of 16 two-syllable fragments (long fragment) (e.g. “Click on the /ti-ma-/”). In each of the first two blocks, the 16 test nonword fragments were from 8 trochaic (4 aspirated and 4 lax) and 8 iambic (4 aspirated and 4 lax) test nonwords, with the two blocks counterbalanced so that the same pair members did not appear in the same block. For instance, a participant who was presented with a short fragment of the trochaic version of an aspirated-initial word (e.g. “Click on the /ti-/”) in Block 1 heard the long fragment of the corresponding iambic word (e.g. “Click on the /ti-ma-/”) in Block 2. In Block 3, the nonwords tested in Block 1 were given in full forms (whole word fragment) (e.g. “Click on the /tim-de/”) to assess whether participants’ responses to the short word fragments were based on the correct associations between nonwords and corresponding picture referents. The same 16 filler trials were presented in all three blocks with different fragment sizes. The test trials were randomized in each block in such a way that the nonwords that belonged to the same experimental condition did not appear in adjacent trials of the block. There was an interval of 30 seconds between blocks. Each block began with three filler trials.

In all three blocks, participants were asked to identify a source word and mouse-click on the corresponding picture referent among three picture referents as accurately and quickly as possible upon hearing a word fragment in the carrier instruction (e.g. “Click on the (target word fragment)”). The three picture choices were the picture referents of the target word, its stress competitor, and the distractor. The positions of the target picture on the computer screen were pseudorandomized in a block so that the target picture referents did not appear in the same position in the successive trials.

The subsequent orthographic transcription task asked all participants to listen to and transcribe the nonwords in their L1 orthography as in the previous eyetracking experiments.
5.4 Data Analysis

As for the previous eyetracking studies, Experiment 2 concerns the behavioral pattern in the training session and also the one in the gating word recognition task. The accuracy and response time (RT) data from the picture choice task in the training session was used to confirm that both native English speakers and Korean L2 learners of English made comparable learning progress for the 48 nonword-picture associations throughout the three training days. The accuracy rates and RTs for the test nonwords that were given in full form in the third block of the gating experiment are used as an index that participants’ word identification in Block 1 was based on their correct associations between nonwords and corresponding picture referents.

To examine how lexical stress information and word-initial phonation type affect word identification after the first syllable in a word and how the interaction between lexical stress and word-initial phonation type is modulated by the size of the word fragment in the process of word recognition, we are interested in the percentages and RTs of correct identification of the source words across the three blocks. To this purpose, the test words were classified into ‘aspirated’ vs. ‘lax’ initial words based on the transcription by Korean participants in the post-eyetracking orthographic transcription task. Korean L2 learners agreed 99.6% of the time for the transcription of the phonation type contrast of word-initial consonants, and native English speakers agreed 99.6% of the time for the transcription of the voicing contrast to what Korean listeners heard as a phonation type contrast (see Table C.3 in Appendix A.2). The response times were measured from the offset of the word fragment for the comparisons across blocks. RTs that were two standard deviations greater or smaller than the mean RT of an L1 group for each block and RTs shorter than 100 ms were removed as outliers.
The statistical approaches to the analysis of the behavioral data of the training session and to the one of the whole word fragment block in the gating experiment are identical to the ones used in the preliminary study and Experiment 1. The accuracy and RT data are analyzed with a mixed effects model (see Section 3.5.1). The analysis of the behavioral data from the gating experiment is the same as the one used for the analysis of the whole fragment block in the gating experiment, except that the three gating blocks of fragment size increments are treated as a continuous variable, just as Time was treated as a continuous variable in the eyetracking data analysis.

5.5 Results

5.5.1 Training Result

Table 5.1 shows the mean percent accuracy for the 48 nonwords and the mean RT for the correct responses on each training day. Outlier RTs were discarded (4.26%, 379 out of 8895 correct responses). Both native English speakers and Korean L2 learners of English showed increasing accuracy with decreasing RTs as they took more training sessions over the three days.

| Training day | English speakers | | Korean speakers | |
|--------------|------------------|------------------|------------------|
|               | Accuracy | RT       | Accuracy | RT       |
| Day 1        | 90.3 (0.74) | 1966.0 (20.36) | 88.7 (0.81) | 2205.1 (27.12) |
| Day 2        | 96.7 (0.45) | 1488.9 (12.90) | 98.1 (0.35) | 1738.1 (18.79) |
| Day 3        | 97.9 (0.36) | 1349.8 (11.64) | 98.6 (0.30) | 1474.0 (14.40) |

Table 5.1: Mean percent accuracy (%) and mean RT (ms) by L1 and Training day. Values in parentheses represent standard errors.
A mixed effects model was built on correct (coded as a “1”) and incorrect (coded as a “0”) responses for the 48 nonwords and RTs of the correct responses separately with \( L1 \) (English vs. Korean) and Training day as two predictors and Subject and Item as crossed random factors. The results are presented in Table 5.2.

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Est.</th>
<th>SE</th>
<th>Wald Z</th>
<th>( p &lt; )</th>
<th>Est.</th>
<th>SE</th>
<th>t</th>
<th>( p &lt; )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(( N = 9360; ) logLik = -1594)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>2.609</td>
<td>0.194</td>
<td>13.430</td>
<td>.0001</td>
<td>1914.24</td>
<td>56.75</td>
<td>33.73</td>
<td>.0001</td>
</tr>
<tr>
<td>( L1 = ) Korean ( L1 )</td>
<td>0.041</td>
<td>0.307</td>
<td>0.134</td>
<td>n.s.</td>
<td>289.57</td>
<td>86.46</td>
<td>3.35</td>
<td>.001</td>
</tr>
<tr>
<td>Training Day</td>
<td>0.808</td>
<td>0.104</td>
<td>7.796</td>
<td>.0001</td>
<td>-305.76</td>
<td>22.41</td>
<td>-13.65</td>
<td>.0001</td>
</tr>
<tr>
<td>( L1: ) Training Day</td>
<td>0.275</td>
<td>0.205</td>
<td>1.339</td>
<td>n.s.</td>
<td>-71.47</td>
<td>34.72</td>
<td>-2.06</td>
<td>.05</td>
</tr>
</tbody>
</table>

Table 5.2: Summary of \( L1 \) and Training day effects on accuracy and RTs in mixed effects models.

The results for accuracy revealed a significant main effect of Training day with no effect of \( L1 \) and no interaction between \( L1 \) and Training day. The positive coefficient for the Training day effect on accuracy \( (\beta = 0.808, \ p < .0001) \) indicates that all participants significantly improved their accuracy in identifying the 48 nonword-picture associations as they took more training sessions over the three days.

The results for RTs show main effects of \( L1 \) and Training day, and also a significant interaction between \( L1 \) and Training day. The negative coefficient of the Training day effect \( (\beta = -305.76, \ p < .0001) \) suggests that response latency decreased as participants underwent more training over the three days. The positive coefficient of \( L1 \) \( (\beta = 289.57, \ p < .001) \) demonstrates that Korean L2 learners took longer response times in the picture-choice task on the first day of training than did native English speakers. However, the negative estimate parameter of the interaction between \( L1 \) and Training day \( (\beta = -71.47, \ p \)
reveals that the RT difference between native speakers and L2 learners on the first day was significantly decreased as the participants underwent more training over the three days.

In sum, native English speakers and Korean L2 learners of English demonstrated similar learning progress for the 48 nonword-picture associations over the three days of training. Though the L2 learners were slower than native speakers in response to the picture-choice task on the first day of training, the differences were significantly decreased toward the end of the three-day training session.

5.5.2 Gating Result
Accuracy and RT from Whole Word Block

The mean percent accuracy for twenty nine mouse-click responses including the sixteen test trials and thirteen filler trials in the whole word block of the gating experiment was 91.75% (SE = 0.89) for native English speakers and 91.16% (SE = 0.93) for Korean-speaking learners of English.

The mean percent accuracy of the sixteen test trials and the mean RT for the correct responses are given in Table 5.3 by functions of L1, stress pattern, and word-initial phonation type. The RTs were measured from the beginning of the instruction, "Click on the (target word) now", up to the time point when participants clicked on one of the three picture referents on the scene display, to be consistent with the behavioral data analyses in the previous two eyetracking experiments. Outlier RTs were discarded (3.7%, 33 out of 893 correct responses).

The binary correct/incorrect responses (coded as "1" and "0", respectively) and the RTs of the correct responses were separately submitted to a mixed effects model with L1, Stress
Table 5.3: Mean percent accuracy (%) and mean RT (ms) by L1, Stress pattern, and Phonation type. Values in parentheses represent standard errors.

<table>
<thead>
<tr>
<th>Stress</th>
<th>Phonation</th>
<th>English speakers</th>
<th></th>
<th>Korean speakers</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Accuracy</td>
<td>RT</td>
<td>Accuracy</td>
<td>RT</td>
</tr>
<tr>
<td>Trochaic</td>
<td>Aspirated</td>
<td>84.8 (3.13)</td>
<td>2260.7 (90.06)</td>
<td>83.6 (3.29)</td>
<td>2429.4 (104.76)</td>
</tr>
<tr>
<td></td>
<td>Lax</td>
<td>85.6 (3.07)</td>
<td>2346.6 (100.29)</td>
<td>84.4 (3.22)</td>
<td>2607.8 (123.30)</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>85.2 (2.19)</td>
<td>2303.2 (67.24)</td>
<td>84.0 (2.30)</td>
<td>2515.6 (80.54)</td>
</tr>
<tr>
<td>Iambic</td>
<td>Aspirated</td>
<td>87.9 (2.85)</td>
<td>2349.4 (84.44)</td>
<td>89.8 (2.68)</td>
<td>2739.9 (113.94)</td>
</tr>
<tr>
<td></td>
<td>Lax</td>
<td>83.3 (3.26)</td>
<td>2447.4 (93.44)</td>
<td>87.5 (2.93)</td>
<td>2682.7 (123.45)</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>85.6 (2.16)</td>
<td>2398.0 (62.86)</td>
<td>88.7 (1.98)</td>
<td>2712.0 (83.69)</td>
</tr>
</tbody>
</table>

The results for accuracy show no main effects or interactions of L1, Stress pattern, and Phonation type. The accuracy results indicate that English and Korean speakers were comparably accurate for the test nonword identification regardless of the word’s stress pattern.
and initial phonation type when the fragment included the whole word with the segmentally disambiguating last syllable. The results for RTs, however, revealed a main effect of L1, with no other effects and interactions approaching significance. The positive estimate parameter ($\beta = 293.94$, $p < .05$) for the L1 effect suggests that Korean speakers overall took longer to identify the target word than did native English speakers regardless of the stress pattern and the word-initial phonation type even when they were given the word in its full form.

In sum, as predicted, Korean L2 learners had more processing difficulty in word identification than native English speakers though the accuracy was comparable between the native speakers and the L2 learners. The effects and interactions of L1, Stress pattern, and Phonation type were not found for either accuracy or response times.

**Accuracy and RT from Short, Long, and Whole Blocks**

Table 5.5 shows the mean percentages of correct identification of source words for the test nonword fragments and the mean response times for the correct responses calculated for the conditions generated by the design: lexical stress pattern of the source word (trochaic vs. iambic), word-initial phonation type (aspirated vs. lax), and fragment size (first-syllable vs. first-two syllables vs. whole word). Outlier RTs were removed (4.9%, 112 out of 2298).

The accuracy result shows that both native English speakers and Korean L2 learners of English identified the target word at levels above chance for both trochaic and iambic words even when they were given only a single initial syllable of the word. This indicates that both native speakers and L2 learners were able to exploit the prominence information embedded in the first syllable to identify the target word. Yet, native English speakers were overall more accurate than L2 learners, especially for trochaic word identification in the short and
long fragment blocks. Though lower than the accuracy of English speakers, Korean L2 learners also showed a higher accuracy rate for trochaic word identification than for iambic word identification in the short fragment block, suggesting stressed syllables were more useful to the initiation of lexical access than unstressed syllables in both native speakers and L2 learners of English.

With regard to the word-initial phonation type effect on the use of lexical stress information at the initial stage of word recognition (short trials), the percent accuracy results show that English speakers were more accurate in word identification for both trochaic and iambic conditions when the word began with an aspirated (voiceless) initial sound than when the word began with a lax (voiced) initial sound, whereas Korean L2 learners were
only more accurate for trochaic word identification. Korean L2 learners’ iambic word identification was more accurate when the word began with a lax consonant than when it began with an aspirated consonant, providing evidence that Korean L1 phrase-level prosody influenced L2 lexical stress processing during word recognition. However, the numerical difference in accuracy between aspirated and lax initial iambic words was not as great as it was for the aspirated vs. lax initial trochaic words. Similarly, the accuracy difference between aspirated (voiceless) vs. lax (voiced) word identification was also very small for iambic words in native English speakers. This suggests that iambic words are harder to identify than are trochaic words at the initial stage of word identification, regardless of the presence and absence of relevant cues induced by the phonation type (voicing) contrast in the word-initial syllable.

Accuracy differences between English and Korean speakers, between trochaic and iambic word identification, and between aspirated (voiceless) and lax (voiced) initial words (especially for trochaic words) were greatly reduced as listeners received more information (via more syllables) for word identification. In the long and whole fragment block, while the differences are small, Korean L2 learners’ accuracy for iambic words was even higher than the one for trochaic words. The phonation type (voicing) effect on accuracy was different in English and Korean speakers. In the long block, Korean speakers were still more accurate for aspirated-initial trochaic word identification than for lax-initial trochaic word identification, and they were more accurate for lax-initial iambic word identification than for aspirate-initial iambic word identification, suggesting that the phonation type of word-initial consonants had an effect on the accurate identification of the target word based on the lexical stress information. This preference of word-initial phonation type was the opposite
Response time results also show that Korean L2 learners took longer to identify the source word than did English speakers both in the short fragment block and the longer fragment blocks. Overall, times were longer for iambic word identification than for trochaic word identification, with the exception of the long block, where iambic and trochaic times showed a numeric difference in the opposite direction. The disadvantage for iambic identification was particularly pronounced for Korean L2 learners in the short and whole conditions, suggesting that iambic words were particularly hard to process.

As to the word-initial phonation type effect on response times, unlike the accuracy results, both native English speakers and Korean L2 learners of English took longer to identify the source word when the word began with a lax sound than when it began with an aspirated sound for both trochaic and iambic words. That is, even though Koreans showed numerically higher accuracy for lax-initial iambic word identification than for aspirated-initial iambic word identification in the short and long conditions, they had more processing difficulty with lax- than aspirated-initial syllables for both trochaic and iambic word identification. Response time differences between English and Korean speakers, between trochaic and iambic word identification, and between aspirated and lax initial words were greatly reduced and sometimes numerically reversed as listeners heard more syllables in
a word. Both English and Korean speakers were faster in identifying iambic words than in identifying trochaic words when they heard two syllables of words. When they heard the whole word with a disambiguating final syllable, the response times were greatly decreased, though Korean speakers were still slower in response than were native English speakers.

To test the effects and interactions of $L1$, Stress pattern, Phonation type, and Fragment size on accuracy, a mixed effects model was performed on correct (coded as a “1”) and incorrect (coded as a “0”) responses, with $L1$, Stress pattern, Phonation type, and Fragment size as predictors and Subject and Item as crossed random factors. Table 5.6 summarizes the analysis result on accuracy. As we concern with the processing results of word-initial phonation type (voicing contrast) and stress information in the first syllables of words, the intercept terms as the indexes of performance after the initial syllables of words are of as much interest as the slope terms.

At the intercept, the effect of Stress pattern was significant, with no other effects and interactions approaching significance. The negative estimate for the effect of Stress pattern ($\beta = -0.499, p < .0001$) with no interaction with $L1$ or/and Phonation type indicates that overall accuracy was significantly higher for trochaic word identification than for iambic word identification in both native English speakers (72.3% vs. 54.9%) and Korean L2 learners of English (61.3% vs. 55.1%) regardless of the phonation type of the word-initial consonant. The result is consistent with the numerical accuracy results in Table 5.5, supporting the prediction that iambic words pose more processing difficulty than trochaic words for both native speakers and Korean L2 learners. Although the accuracy results in Table 5.5 showed that, unlike native English speakers, Korean L2 learners preferred aspirated-initial words for trochaic word identification and lax initial words for iambic word identification,
<table>
<thead>
<tr>
<th>Predictor</th>
<th>Est.</th>
<th>SE</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.467</td>
<td>0.108</td>
<td>4.311</td>
<td>.0001</td>
</tr>
<tr>
<td>L1 = Korean L1</td>
<td>-0.296</td>
<td>0.165</td>
<td>-1.800</td>
<td>n.s.</td>
</tr>
<tr>
<td>Stress1 = Iambic</td>
<td>-0.499</td>
<td>0.123</td>
<td>-4.066</td>
<td>.0001</td>
</tr>
<tr>
<td>Phonation1 = Lax</td>
<td>-0.226</td>
<td>0.187</td>
<td>-1.212</td>
<td>n.s.</td>
</tr>
<tr>
<td>L1:Stress1</td>
<td>0.471</td>
<td>0.246</td>
<td>1.918</td>
<td>n.s.</td>
</tr>
<tr>
<td>L1:Phonation1</td>
<td>0.078</td>
<td>0.245</td>
<td>0.316</td>
<td>n.s.</td>
</tr>
<tr>
<td>Stress1:Phonation1</td>
<td>0.336</td>
<td>0.245</td>
<td>1.367</td>
<td>n.s.</td>
</tr>
<tr>
<td>L1:Stress1:Phonation</td>
<td>0.609</td>
<td>0.491</td>
<td>1.240</td>
<td>n.s.</td>
</tr>
<tr>
<td>Gate</td>
<td>0.704</td>
<td>0.055</td>
<td>12.825</td>
<td>.0001</td>
</tr>
<tr>
<td>L1:Gate</td>
<td>0.154</td>
<td>0.110</td>
<td>1.402</td>
<td>n.s.</td>
</tr>
<tr>
<td>Stress1:Gate</td>
<td>0.399</td>
<td>0.110</td>
<td>3.642</td>
<td>.001</td>
</tr>
<tr>
<td>Phonation1:Gate</td>
<td>0.056</td>
<td>0.110</td>
<td>0.515</td>
<td>n.s.</td>
</tr>
<tr>
<td>L1:Stress1:Gate</td>
<td>-0.110</td>
<td>0.219</td>
<td>-0.502</td>
<td>n.s.</td>
</tr>
<tr>
<td>L1:Phonation1:Gate</td>
<td>-0.0207</td>
<td>0.219</td>
<td>-0.095</td>
<td>n.s.</td>
</tr>
<tr>
<td>Stress1:Phonation1:Gate</td>
<td>-0.352</td>
<td>0.219</td>
<td>-1.607</td>
<td>n.s.</td>
</tr>
<tr>
<td>L1:Stress1:Phonation1:Gate</td>
<td>-0.083</td>
<td>0.438</td>
<td>-0.190</td>
<td>n.s.</td>
</tr>
</tbody>
</table>

Table 5.6: Summary of $L1$, Stress pattern, Phonation type, and Time effects on accuracy in Quasi-MLR ($N = 3120$; logLik = -1667).

The accuracy difference between aspirated and lax initial iambic words was too small in both native English speakers and Korean L2 learners of English to produce any significant interaction between $L1$, Stress pattern, and Phonation type. The result rather supports that unstressed initial syllables are least likely to contribute to lexical activation at an early stage of word recognition.

For the slope term as the index of fragment size effect, only Stress pattern effect was significant. The positive parameter estimate ($\beta = 0.399$, $p < .001$) shows that the difference in accuracy between trochaic and iambic word recognition was significantly reduced as listeners heard more syllables in a word.
To examine the effects and interactions of $L1$, Stress pattern, Phonation type, and Fragment size on response times, the same mixed effects model was performed on response times for the correct responses. Table 5.7 summarizes the analysis result on response times.

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Est.</th>
<th>SE</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>2947.69</td>
<td>126.58</td>
<td>23.287</td>
<td>.0001</td>
</tr>
<tr>
<td>$L1 = Korean L1$</td>
<td>1544.40</td>
<td>200.23</td>
<td>7.713</td>
<td>.0001</td>
</tr>
<tr>
<td>Stress1 $= Iambic$</td>
<td>471.49</td>
<td>126.74</td>
<td>3.720</td>
<td>.0001</td>
</tr>
<tr>
<td>Phonation1 $= Lax$</td>
<td>331.47</td>
<td>199.53</td>
<td>1.661</td>
<td>n.s.</td>
</tr>
<tr>
<td>$L1$:Stress1</td>
<td>604.05</td>
<td>252.42</td>
<td>2.393</td>
<td>.05</td>
</tr>
<tr>
<td>$L1$:Phonation1</td>
<td>503.10</td>
<td>251.57</td>
<td>2.000</td>
<td>.05</td>
</tr>
<tr>
<td>Stress1:Phonation1</td>
<td>219.31</td>
<td>253.18</td>
<td>0.866</td>
<td>n.s.</td>
</tr>
<tr>
<td>$L1$:Stress1:Phonation1</td>
<td>218.63</td>
<td>504.25</td>
<td>0.434</td>
<td>n.s.</td>
</tr>
<tr>
<td>Gate</td>
<td>-810.50</td>
<td>45.66</td>
<td>-17.751</td>
<td>.0001</td>
</tr>
<tr>
<td>$L1$:Gate</td>
<td>-636.92</td>
<td>91.32</td>
<td>-6.975</td>
<td>.05</td>
</tr>
<tr>
<td>Stress1:Gate</td>
<td>-211.36</td>
<td>91.52</td>
<td>-2.309</td>
<td>.05</td>
</tr>
<tr>
<td>Phonation1:Gate</td>
<td>-147.61</td>
<td>91.25</td>
<td>-1.618</td>
<td>n.s.</td>
</tr>
<tr>
<td>$L1$:Stress1:Gate</td>
<td>-296.70</td>
<td>182.78</td>
<td>-1.623</td>
<td>n.s.</td>
</tr>
<tr>
<td>$L1$:Phonation1:Gate</td>
<td>-175.94</td>
<td>182.50</td>
<td>-0.964</td>
<td>n.s.</td>
</tr>
<tr>
<td>Stress1:Phonation1:Gate</td>
<td>-142.28</td>
<td>182.82</td>
<td>-0.778</td>
<td>n.s.</td>
</tr>
<tr>
<td>$L1$:Stress1:Phonation1:Gate</td>
<td>-332.76</td>
<td>365.13</td>
<td>-0.911</td>
<td>n.s.</td>
</tr>
</tbody>
</table>

Table 5.7: Summary of $L1$, Stress pattern, Phonation type, and Time effects on RTs in Quasi-MLR ($N = 2186$; logLik $= -19339$).

At the intercept, effects of $L1$ and Stress pattern were significant. The positive coefficient of $L1$ ($\beta = 1544.40$, $p < .0001$) indicates that Korean speakers took longer to finish word identification than did English speakers when they were given the initial syllable information for word identification. The positive coefficient of Stress pattern effect ($\beta = 471.49$, $p < .0001$) indicates that it took more time for iambic word identification than for trochaic word identification. However, the significant interaction between $L1$ and Stress pattern ($\beta = 604.05$, $p < .05$) with no interaction with Phonation type suggests that Korean L2 learners had more processing difficulty for iambic word identification than for trochaic
word identification than did English speakers, regardless of the phonation type of the word-initial consonant. There was also a significant interaction between $L1$ and *Phonation type*. The positive coefficient ($\beta = 503.10, p < .05$) means that lax initial words elicited longer response times than did aspirated-initial words in Korean L2 learners than they did in English speakers. As it was shown in Table 5.5, since the longer response time for lax initial words were found not only for trochaic word identification but also for iambic word identification in both native speakers and L2 learners, the interaction of $L1$, *Stress pattern*, and *Phonation type* was not found.

For the slope, only the effects of $L1$ and of *Stress pattern* were significant, with no other effects approaching significance. The negative parameter estimates of $L1$ ($\beta = -636.92, p < .05$) indicate that the significant response time difference between English speakers and Korean L2 learners in the short fragment block was significantly decreased as listeners heard more syllables in a word for word recognition. The negative parameter estimates of the *Stress pattern* effect ($\beta = -211.36, p < .05$) also indicate that the response time difference between trochaic and iambic word recognition in the first block was significantly decreased as listeners heard more syllables in a word. However, there were not significant interactions between $L1$ and *Stress pattern* or between $L1$ and *Phonation type* on the slope, suggesting that the significant response time differences between English speakers and Korean L2 learners for the lexical stress conditions and for the word-initial phonation type contrast were maintained over the three blocks. This means that the processing latencies for iambic words over trochaic words and for lax-initial words over aspirated-initial words were persistent in Korean L2 learners of English even when they identified the target word after hearing the second syllable and the segmentally disambiguating last syllable of a word.
In sum, the mixed effects analyses on the accuracy and response times in the gating word identification revealed some similarities and differences between native English speakers and Korean L2 learners of English. Word identification accuracy was comparable between native English speakers and Korean L2 learners not only in the short fragment block, but also in the subsequent longer word fragment blocks. They both were more accurate for trochaic word identification than for iambic word identification when they were given only the initial syllable in a word for word identification. As the accuracy difference between aspirated and lax initial words were small for iambic words in both native English speakers and Korean L2 learners, the interaction between L1, Stress pattern, and Phonation type was not significant for accuracy.

On the other hand, the results for response times revealed that Korean L2 learners took significantly longer to identify the source word than native English speakers, especially when the target word was given up to the first syllable. This latency difference between native speakers and L2 learners was greater for iambic word identification, even when the size of the word fragment increased. We couldn’t observe a significant interaction between L1, Stress pattern, and Phonation type for response times, because both native English speakers and Korean L2 learners of English had longer response times for lax initial word identification than for aspirated-initial word identification for both stress conditions. The significant L1 and Stress pattern effects were significantly weakened for both accuracy and response times as listeners received more syllable information in the stimulus.

5.6 Discussion

Experiment 2 explored the research questions raised in the previous eyetracking experiments, using an additional task, gating. The preliminary eyetracking study that explored the
processing of L2 lexical stress without alternating vowel quality in stressed and unstressed syllables in words showed that Korean L1 phrase-level prosodic structure could potentially influence the processing of lexical stress during the middle of spoken word recognition in Korean L2 learners of English. The present study examines this issue more closely by observing the processing result after each syllable of words in a cross-modal fragment gating experiment with an experimental control of word-initial phonation type contrast in the test nonwords. To the purpose of Experiment 2, we examined the processing result reflected in accuracy and response times for a short word fragment block and for subsequent longer word fragment blocks separately.

The training results indicate both native English speakers and Korean L2 learners of English learned the nonword-picture associations at comparable levels, scoring above 90% accuracy. However, in the last training session, Korean L2 learners still showed longer response times than did English speakers in choosing a target picture. This behavioral pattern was also found for word identification in the whole word block of the gating experiment. Though the overall accuracy in the whole word condition was comparable between English speakers (85.4%) and Korean speakers (86.4%), response latency was still longer for Korean speakers than for English speakers, suggesting that L2 learners of English overall had more lexical processing difficulty than did native English speakers in identification of target words.

A close look at the identification performance in the gating experiment first allows us to notice that a single initial syllable of a target word was sufficient for both native English speakers and Korean L2 learners to correctly access the target word above chance level. Native English speaker’s good word identification performance with the initial syllable information was reported in Mattys (2000) where English speakers provided 52% accuracy for
initial primary stress words and 56% accuracy for initial secondary stress words. However, given that the test words were the cohorts of the minimal stress pairs that were contrastive only by prosody in the first two syllables, Korean L2 learners’ good guessing performance contradicts with our intuitive prediction that Korean L2 learners of English would be deaf to syllable prominence during L2 spoken word recognition. As this result was found after the exposure to the near minimal stress pair members in the three-day training session, Korean L2 learners’ good identification performance in the short fragment block implies that they can detect and use the syllable prominence information during lexical access if they were given the chance to be exposed to it. Yet, the significantly longer response times in Korean L2 learners as compared to native English speakers in the short fragment block indicates that the L2 learners had more processing difficulty than native English speakers did when identifying the target word just based on the prosodic information in a single syllable.

In addition, we could observe that the good performance in the short fragment block was better for trochaic word identification than for iambic word identification in both native speakers and L2 learners. Response time was also significantly shorter after the stressed initial syllable fragment than after the unstressed initial syllable fragment. That is, both English speakers and Korean L2 learners showed a processing advantage for trochaic over iambic words when they heard just the initial syllable for word recognition. This result is consistent with the preliminary eyetracking study where English speakers were more erroneously looking at the competitor object for iambic words than for trochaic words at an early stage of word recognition, and also consistent with the MSS (Cutler, 1986; ?, ?) and stress-based lexical processing model (Mattys, 2000). The previous studies ascribed this effect to the English speakers’ perceptual and processing tendency to hear an unstressed initial syllable as a stressed syllable when it doesn’t involve reduced vowel information and
to wait until encountering a stressed syllable against which they can evaluate the degree of stress.

The significant interaction between L1 and Lexical stress for response times, not for accuracy, in the short fragment block suggests that Korean L2 learners needed more times to process the initial syllable for iambic word recognition as compared to trochaic word recognition than did native English speakers. The significant effect of Lexical stress on the slope for response times with no interaction with L1 indeed indicates that the processing difficulty for iambic words was greatly reduced in both native English speakers and Korean L2 learners as they heard more syllables of words. This result is consistent with the finding of the preliminary eyetracking experiment where English speakers’ gaze probability for the iambic target object was significantly higher than that for the stress competitor after they heard the stressed second syllables in a word, while it was after the first syllable for trochaic word recognition. On the other hand, Korean L2 learners showed strong competition between the target and the competitor during iambic word recognition even after the last disambiguating syllable. In the present study, Korean L2 learners also had such persistent processing difficulty for iambic words even after the second stressed syllable and the disambiguating last syllable, to be consistent with the initial prediction. This processing pattern is quite different from that of native English speakers. If Korean L2 learners’ lack of efficiency in using the unstressed initial syllable for word recognition is because they need a reference syllable to judge the degree of stress as native English speakers do, they should have shown a similar processing pattern as the native English speakers. To explain the L2 learners’ processing pattern, we argue that as Koreans are less motivated to use syllable prominence information for L1 lexical access, the perceptually less salient initial syllables in L2 words would increase the processing load to retrieve and restore the information
which can continuously restrain the process of lexical access during word recognition, even disturbing the use of the segmentally disambiguating syllable in a word.

The phonation type effect on the processing of English lexical stress for word identification didn’t reveal a significant effect for either accuracy or response times, contrary to the prediction. Though the numerical accuracy results in the short and long fragment blocks showed that, unlike native English speakers, Korean L2 learners were more accurate with aspirated-initial words than with lax-initial words for trochaic word identification, and it was vice versa for iambic word identification, the differences were too small to produce significant effects. Rather, the significant interaction between L1 and Phonation type for response times at the intercept but not on the slope suggests that Korean L2 learners had more difficulty in using lax-initial syllables for both trochaic and iambic word identification than did native English speakers from the early to the end of word recognition. The fact that what Koreans heard as lax sounds did not facilitate trochaic word identification is consistent with the processing results in the previous eyetracking experiments, supporting that L1 perceptual experience with phrase initial tonal distinction influence the perception and processing of English trochaic words which initial syllables are salient enough to exploit the shared cue, f0. The finding that what Korean heard as lax sounds did not facilitate the activation of iambic words particularly in Korean L2 learners is also similar to the processing pattern in the preliminary eyetracking study where Korean L2 learners’ processing of iambic words were much delayed regardless of the phonation type of the word-initial consonants.

Overall, to be consistent with the previous eyetracking studies, Experiment 2 provides supporting evidence that L2 learners’ processing of L2 word prosody for spoken word recognition is largely constrained by their perceptual and processing experience with L1
prosodic structure. Korean L2 learners who do not have word prosody, nor have stress minimal pairs in L1 that are purely contrastive by prosodic cues such as in English *INsight* and *inCIte*, are assumed to be not sensitive to the prominence difference in syllables for word recognition. This processing strategy was maintained during word recognition of the gating word identification task. The L2 learners needed more time than native speakers of English to identify the target word as accurately as native speakers do after listening to the initial syllable in a word. This processing difficulty was particularly strong for iambic word recognition. The perceptually less salient initial syllable seem to pose a processing burden to the L2 learners, causing more delayed target word recognition in Korean L2 learners than in native English speakers.

However, we could still observe that the aspirated-initial word which is a better fit to the trochaic stress pattern facilitated trochaic word identification from early stage of word recognition. That is, Korean L2 learners showed that they can detect the f0 height difference between what they heard as aspirated and lax-initial syllables when the syllable is salient enough to use the relevant information to their perception and processing of L2 lexical stress. Given that the f0 height difference between aspirated- vs. lax-initial syllables is evident only when they are placed at the phrase initial position in the Korean intonational structure, the result demonstrates that the perceptual experience with L1 phrase-level prosodic system can affect the perception and processing of L2 lexical stress during spoken word recognition despite the fact that they belong to different categorical domains, word vs. phrase levels.
CHAPTER 6: GENERAL DISCUSSION AND CONCLUSION

Prosody is known to influence how listeners structure syllables in a word and how lexical access proceeds during spoken word recognition. Even though the importance of prosodic information in spoken word recognition is well known, there have been relatively few studies and theories that have addressed autosegmental processing during L2 spoken word recognition. If prosody is the linguistic information that contributes to native-like language processing, we can raise the question of how the similarities and differences between L1 and L2 in their prosodic structure affect the perception and processing of word prosody during L2 word recognition.

The primary goal of this dissertation was to explore whether L2 learners can process L2 word prosody during spoken word recognition. Given that there are no prosodically-based lexical distinctions in L1, but these distinctions are necessary for word identification in L2, how would L2 learners of this L1 background process prosodic information during L2 spoken word recognition? To address this question, we examined the processing of English lexical stress in Korean L2 learners of English. As Korean language doesn’t have “stress” as part of its word-level prosody, it was intuitively predicted that they would not be able to use English lexical stress information during spoken word recognition. However, Korean has a phrase-level prosodic structure, (T)HLH, that encompasses one word or more than one word in the unit. The initial tone is realized as an H tone when the initial segment is heard as aspirated or tense and as an L tone otherwise. That is, even though the phrase-level prosodic structure is not at the same level as is the word-level prosody such as lexical stress, they
share the acoustic cue of f0 to create prominence difference between syllables. Therefore, by examining the processing of English lexical stress in Korean L2 learners of English, we explored not only whether/how L2 learners use L2 word prosody for word recognition, but also to what extent the L1 prosodic domain is transferable to the L2 prosodic domain during L2 word recognition.

The preliminary eyetracking study examined how native English speakers and Korean L2 learners of English use English lexical stress (trochaic vs. iambic stress pattern) during spoken word recognition when the stressed and unstressed syllables do not involve a segmental cue to lexical stress. Korean L2 learners’ activation of target words was more likely to be delayed until they encountered a disambiguating third syllable in a word, whereas English speakers’ activation of target words were initiated when they encountered a primary stressed syllable in a word. That is, native English speakers initiated lexical access after the stressed first syllable during trochaic word recognition and after the stressed second syllable during iambic word recognition. Though Korean L2 learners also showed relatively earlier activation for trochaic target words than for iambic target words, a strong competition between targets and competitors lasted toward the end of word offset for both trochaic and iambic words.

Processing difficulty for iambic word recognition can be accounted for by the retroactive processing hypothesis that was proposed by the stress-based lexical processing model (Mattys, 2000; Mattys & Samuel, 2000). According to the retroactive processing hypothesis, the initial syllable of non-initial stress words is relatively hard to detect but needs to be maintained until delayed integration is accomplished based on the later stressed syllable. Consequently, processing of late-stress words places a substantial demand on phonetic memory and needs extra processing time to decode the early information. In both Korean
learners of English and native English speakers, decoding of earlier information was found in the reconsideration of a stress competitor that was constrained earlier in word recognition. This right-to-left retroactive processing resulted in delayed recognition of iambic words, but must be a necessary corrective process for correct word recognition.

With regard to the research question of how language-specific prosodic variation influences L2 spoken word recognition, we conducted a post-hoc analysis to examine whether Korean phrase-level prosodic structure influences the processing of English lexical stress in Korean L2 learners. We observed that what Korean heard as aspirated-initial syllable facilitated trochaic word recognition temporarily right after the stressed first syllable in a word. Such a word-initial phonation type effect on lexical stress processing was not found for iambic word recognition. In addition, as there were some unexpected ABEs for lax- (voiced-) initial trochaic word recognition and for aspirated- (voiceless-) initial iambic word recognition in native English speakers, we couldn’t statistically verify the interaction between English lexical stress and Korean L1 phrase-level tonal variation in Korean L2 learners of English.

In addition, because the number of aspirated vs. lax word-initial consonants was not controlled in the preliminary study, any conclusions drawn about prosodic influence on word recognition must remain tentative. Furthermore, we cannot say that the stimulus set of the preliminary study fully represented the properties of English lexical stress because the unstressed syllables of the stimuli did not include reduced vowels, a segmental correlate to lexical stress in English.

Experiment 1 in the dissertation study explored the issues by employing eyetracking methodology. In the experiment, we examined how English speakers and Korean-speaking
learners of English process English lexical stress cues during word recognition when lexical stress involves a reduced vowel, schwa, in unstressed syllables. The lexical processing patterns were quite different from the ones found in the preliminary study. Both English and Korean speakers activated the target word immediately after hearing the first syllable in a word, regardless of whether the word carried trochaic or iambic stress pattern. The results provided empirical evidence that English speakers can use reduced vowel information to initiate lexical access. This is quite different result from what has been predicted by the existing MSS (Cutler & Norris, 1988; Cutler, 1989; Cutler & Butterfield, 1992) and the stressed-based lexical processing model (Mattys, 2000; Mattys & Samuel, 2000) which have predicted delayed processing for non-initial primary stress words. However, given that the participants learned the test nonwords in three days of training prior to the eyetacking word recognition task, native English participants might have detected the segmental difference between stress minimal pair words and used the vowel quality cue as an activator for lexical access.

On the other hand, given that Korean doesn’t have reduced vowels, and a previous study has reported that Korean L2 learners can distinguish English schwa from other vowel categories in English (Yun & Jackson, 2006), they seemed to use schwa in unstressed syllables as a segmental cue to distinguish stress minimal pair words at an early stage of spoken word recognition. However, Korean L2 learners still showed retroactive processing in the middle of iambic word recognition. This suggests that Korean L2 learners didn’t process the reduced vowel as a segmental correlate to lexical stress, but just as a distinctive vowel category, hampering reliable activation of iambic words when it was heard in an unstressed syllable.
As for the word-initial phonation type effect on the processing of lexical stress during word recognition, Experiment 1 also revealed that there were strong ABEs during lax-(voiced-) initial trochaic word recognition and aspirated- (voiceless-) initial iambic word recognition in both English and Korean L2 learners of English. Since the effects can bias the slope estimates between conditions, we couldn’t use the results to test the interaction between word-initial phonation type and lexical stress processing during word recognition.

Experiment 2 further investigated this issue without changing the vowel quality in stressed and unstressed syllables by employing a cross-modal fragment gating paradigm. The study showed very similar results as the preliminary eyetracking experiment. Korean L2 learners showed delayed and less accurate identification of target words as compared to native English speakers. However, the word identification accuracy after the initial syllable of words was comparable between native English speakers and Korean speakers, suggesting that Korean L2 learners could detect the subtle prosodic prominence differences between syllables and used the information if they were given the chance to pay attention to it.

The word-initial phonation type effect on the processing of lexical stress was also similar to the one found in the preliminary study. Korean L2 learners of English preferred trochaic words to iambic words when the word began with what they heard as an aspirated sound and it was opposite when the word began with what they heard as a lax sound. This suggests that L2 word recognition is facilitated when the lexical stress pattern fits L1 phrase-level prosodic pattern in terms of f0 height - a relatively higher f0 of aspirated-initial syllable provides an additional cue to the identification of trochaic words. This influence of word-initial phonation type on the processing of lexical stress was found only in Korean participants when the listeners were presented with two syllables of words. However, the
effect of word-initial phonation type was not strong for iambic word identification to be consistent with the general processing difficulty for iambic words.

Taking the results from the three experiments together, we conclude that L2 learner’s processing of L2 word prosody for spoken word recognition is largely constrained by their perceptual and processing experience with L1 prosodic structure. Korean-speaking learners of English have less incentive to pay attention to prominence difference in syllables during L1 spoken word recognition. The processing strategy that has been developed for L1 word recognition influences the efficiency of using L2 English lexical stress during spoken word recognition. The lack of motivation to use prosodic cues for word recognition was particularly pronounced in the processing of iambic words. The unstressed syllables with much reduced duration, f0, intensity cues seem to make it hard for Korean-speaking learners of English to encode and retrieve relevant information in the syllable, and thus retard proactive lexical processing during iambic word recognition. For this reason, the unstressed initial syllable with a reduced vowel, didn’t lead to reliable lexical activation in Korean-speaking learners of English, though the initial syllable provides a segmentally distinctive cue for Korean participants to suppress stress competitors at an early stage of word recognition. Though Korean-speaking learners of English showed perceptual deficiency for unstressed syllables, we should notice that Korean L2 learners’ processing of trochaic words was facilitated when the words began with what Koreans heard as an aspirated sound, suggesting that L1 phrase-level prosodic structure influences the processing of L2 word-level prosody during spoken word recognition.
6.0.1 Implications

According to Lehiste and Peterson (1959, p. 428), perception of linguistic stress is based upon judgments of the physiological effort involved in producing vowels. In Korean intonational structure, the “effort” is involved in tense/aspirated versus lax contrast of the phrase/word-initial segment and tone is a correlate of these. The study explores the possible influence of non-vowel cues on the perception of linguistic stress during L2 word recognition due to tonal pattern transfer from L1 prosodic structure. This cross-linguistic influence can be good evidence that prosodic transfer doesn’t necessarily operate in a manner that is restricted by linguistic category. The phonological mapping can be made between different categories (e.g., vowels and consonants) if they induce the same features - f0 variations induced by English vowels for stress contrast and by word-initial segments consonants for the AP-initial tonal contrast in Korean.

The finding of prosodic transfer from Korean phrase-level prosodic structure to English word-level prosody also suggests that language transfer in L2 perception isn’t limited to the same phonological domain. Instead, categorical mapping seems to be extendable to a higher (e.g., phrasal) domain if the categories share some features, for example, f0 for lexical-level stress contrasts in English and phrase-level prosodic structure in Korean. Thus, when applying the PAM to accounts for prosodic processing in L2 spoken word processing, we can add evidence for a transferable domain in the model, by not restricting the domain, but paying more attention to acoustic-phonetic realizations.

If a prosodic transfer occurs from Korean phrase-level prosodic structure to English word-level prosody, it also suggests that the elements being transferred do not necessarily share the same functional load. In Korean, phrasal tonal variations do not make lexical contrasts and so its functional load is relatively low for lexical distinction. While, lexical
stress carries a higher functional load in English than does phrase-level tonal structure in
Korean.

Even though a large population of Korea is learning other languages, particularly En-
glish, as a second language, there are relatively few studies that explore the Korean lan-
guage and its cross-linguistic influence on other languages during language processing. In
light of this, the results from the present study are expected to increase our understanding
of the processing mechanism in Korean L2 learners of English, especially their prosodic
transfer during English word recognition and production. This will contribute to the breadth
and depth of our understanding of bilinguals’ language processing.
## APPENDIX A: NONWORD STIMULUS LIST

<table>
<thead>
<tr>
<th>Number</th>
<th>Trochaic Target Word</th>
<th>Iambic Target Word</th>
<th>Filler</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>/mætoker/</td>
<td>/mæˈtoker/</td>
<td>/ˈfəgɪsə/</td>
</tr>
<tr>
<td>2</td>
<td>/ˈtrɪmido/</td>
<td>/ˈtrɪmido/</td>
<td>/ˈdʒəsmu/</td>
</tr>
<tr>
<td>3</td>
<td>/dʒəkunəˈɾi/</td>
<td>/dʒəˈkunəɾi/</td>
<td>/ˈgæsɪtə/</td>
</tr>
<tr>
<td>4</td>
<td>/ˈɡɪsædʒi/</td>
<td>/ɡiˈsædʒi/</td>
<td>/ˈkətɾəvə/</td>
</tr>
<tr>
<td>5</td>
<td>/ˈʃɪbeta/</td>
<td>/ˈʃɪbətə/</td>
<td>/ˈmæəvədʒəɾi/</td>
</tr>
<tr>
<td>6</td>
<td>/ˈkʊnəɾiʒə/</td>
<td>/kʊˈnəɾiʒə/</td>
<td>/ˈtəbədəɾi/</td>
</tr>
<tr>
<td>7</td>
<td>/ˈfeɪɡəbəɾ/</td>
<td>/fɛˈɡəbəɾ/</td>
<td>/ˈsɛtəʒi/</td>
</tr>
<tr>
<td>8</td>
<td>/ˈpəˈvɪsəɾ/</td>
<td>/pəˈvɪsəɾ/</td>
<td>/ˈtəˈməɾəkəɾi/</td>
</tr>
<tr>
<td>9</td>
<td>/ˈməkədʒi/</td>
<td>/məˈkədʒəɾi/</td>
<td>/ˈʃəˈnəɾvə/</td>
</tr>
<tr>
<td>10</td>
<td>/ˈtəmɪsəɾ/</td>
<td>/təˈmɪsəɾ/</td>
<td>/pəˈɡəkəɾi/</td>
</tr>
<tr>
<td>11</td>
<td>/dʒiˈvədəɾ/</td>
<td>/dʒiˈvədəɾ/</td>
<td>/ˈɡəˈkrənu/</td>
</tr>
<tr>
<td>12</td>
<td>/ˈɡætəɾkəɾu/</td>
<td>/ɡæˈtəɾkəɾu/</td>
<td>/ˈpəˈvrədəɾ/</td>
</tr>
<tr>
<td>13</td>
<td>/ˈsɛɾɲuˈɾəɾ/</td>
<td>/sɛɾɲuˈɾəɾ/</td>
<td>/ˈkəˈbəɾəɾi/</td>
</tr>
<tr>
<td>14</td>
<td>/ˈkəsoɾəɾ/</td>
<td>/kəˈsoɾəɾ/</td>
<td>/miˈnəidʒi/</td>
</tr>
<tr>
<td>15</td>
<td>/ˈfəbətəɾ/</td>
<td>/fəˈbətəɾ/</td>
<td>/dʒəˈkəsəɾi/</td>
</tr>
<tr>
<td>16</td>
<td>/ˈpəməɾəɾtəɾ/</td>
<td>/pəˈməɾəɾtəɾ/</td>
<td>/soˈmədʒəɾə/</td>
</tr>
</tbody>
</table>

Table A.1: Nonword stimuli in the preliminary eyetracking experiment
<table>
<thead>
<tr>
<th>Number</th>
<th>Trochaic Target Word</th>
<th>Iambic Target Word</th>
<th>Filler</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>/povəsu/</td>
<td>/pəvəset/</td>
<td>/ɡəmatau/</td>
</tr>
<tr>
<td>2</td>
<td>/tímədo/</td>
<td>/təˈmidaʊ/</td>
<td>/ɡebəsu/</td>
</tr>
<tr>
<td>3</td>
<td>/kənədʒəu/</td>
<td>/kəˈnədʒi/</td>
<td>/dəˈmeta/</td>
</tr>
<tr>
<td>4</td>
<td>/siβəta/</td>
<td>/səˈbita/</td>
<td>/məˈvįʒət/</td>
</tr>
<tr>
<td>5</td>
<td>/bəkanət/</td>
<td>/bəˈkanət/</td>
<td>/ˈtosoDi/</td>
</tr>
<tr>
<td>6</td>
<td>/dəsəʒi/</td>
<td>/dəˈseʒu/</td>
<td>/ˈkɛrtəvo/</td>
</tr>
<tr>
<td>7</td>
<td>/ɡətəvəj/</td>
<td>/ɡoˈtəvə/</td>
<td>/kəˈsema/</td>
</tr>
<tr>
<td>8</td>
<td>/meŋgəkeɪtu/</td>
<td>/məˈɡɛɪka/</td>
<td>/ˈbæsədʒi/</td>
</tr>
<tr>
<td>9</td>
<td>/pməku/</td>
<td>/pəˈmiko/</td>
<td>/diβəsu/</td>
</tr>
<tr>
<td>10</td>
<td>/təˈbətə/</td>
<td>/təˈbətəu/</td>
<td>/səˈkɑnət/</td>
</tr>
<tr>
<td>11</td>
<td>/kəvəʒai/</td>
<td>/kəˈvįʒər/</td>
<td>/ˈsəndʒəu/</td>
</tr>
<tr>
<td>12</td>
<td>/səŋənəvə/</td>
<td>/səˈɡemɨ/</td>
<td>/bəˈtikəu/</td>
</tr>
<tr>
<td>13</td>
<td>/bəməsa/</td>
<td>/bəˈməsaʊ/</td>
<td>/poˈvədəu/</td>
</tr>
<tr>
<td>14</td>
<td>/dəkədʒi/</td>
<td>/dəˈkodʒai/</td>
<td>/mənənəvə/</td>
</tr>
<tr>
<td>15</td>
<td>/ɡəsəduəu/</td>
<td>/ɡəˈsəduə/</td>
<td>/ˈtikəʒəvə/</td>
</tr>
<tr>
<td>16</td>
<td>/matəveɪu/</td>
<td>/məˈtava/</td>
<td>/poˈɡokai/</td>
</tr>
</tbody>
</table>

Table A.2: Nonword stimuli in Experiment 1
<table>
<thead>
<tr>
<th>Number</th>
<th>Trochaic Target Word</th>
<th>Iambic Target Word</th>
<th>Filler</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>/povisu/</td>
<td>/po viseu/</td>
<td>/bo'midzø/</td>
</tr>
<tr>
<td>2</td>
<td>/timato/</td>
<td>/ti mada/</td>
<td>/sa'kmu/</td>
</tr>
<tr>
<td>3</td>
<td>/kanædzau/</td>
<td>/ka'nædzø/</td>
<td>/qabøsi/</td>
</tr>
<tr>
<td>4</td>
<td>/sibeta/</td>
<td>/si beti/</td>
<td>/ke'gono/</td>
</tr>
<tr>
<td>5</td>
<td>/bakemi/</td>
<td>/ba'kemø/</td>
<td>/do'metau/</td>
</tr>
<tr>
<td>6</td>
<td>/desiʒai/</td>
<td>/de'siʒø/</td>
<td>/ta'gøu/</td>
</tr>
<tr>
<td>7</td>
<td>/gætavœ/</td>
<td>/ga'tavau/</td>
<td>/ko'seiko/</td>
</tr>
<tr>
<td>8</td>
<td>/meŋokei/</td>
<td>/me'gokai/</td>
<td>/di'basø/</td>
</tr>
<tr>
<td>9</td>
<td>/paneika/</td>
<td>/pa'nerka/</td>
<td>/ga'sava/</td>
</tr>
<tr>
<td>10</td>
<td>/teбито/</td>
<td>/te'bitu/</td>
<td>/ma'vøga/</td>
</tr>
<tr>
<td>11</td>
<td>/kævaʒet/</td>
<td>/ka'vegaʒø/</td>
<td>/so'naðau/</td>
</tr>
<tr>
<td>12</td>
<td>/serɡoðø/</td>
<td>/ser'godi/</td>
<td>/bo'taku/</td>
</tr>
<tr>
<td>13</td>
<td>/botmæu/</td>
<td>/bo'tmer/</td>
<td>/prævu/</td>
</tr>
<tr>
<td>14</td>
<td>/di'kadʒi/</td>
<td>/di'kædʒæi/</td>
<td>/me'tavei/</td>
</tr>
<tr>
<td>15</td>
<td>/ga'mæsai/</td>
<td>/ga'mæsø/</td>
<td>/po'nødʒøi/</td>
</tr>
<tr>
<td>16</td>
<td>/mi'eva/</td>
<td>/mi'revo/</td>
<td>/ti'kæzi/</td>
</tr>
</tbody>
</table>

Table A.3: Nonword stimuli in Experiment 2
### Table B.1: Nonword visual stimuli

<table>
<thead>
<tr>
<th>Number</th>
<th>Trochaic Target Word</th>
<th>Iambic Target Word</th>
<th>Filler</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><img src="image1" alt="Image" /></td>
<td><img src="image2" alt="Image" /></td>
<td><img src="image3" alt="Image" /></td>
</tr>
<tr>
<td>2</td>
<td><img src="image4" alt="Image" /></td>
<td><img src="image5" alt="Image" /></td>
<td><img src="image6" alt="Image" /></td>
</tr>
<tr>
<td>3</td>
<td><img src="image7" alt="Image" /></td>
<td><img src="image8" alt="Image" /></td>
<td><img src="image9" alt="Image" /></td>
</tr>
<tr>
<td>4</td>
<td><img src="image10" alt="Image" /></td>
<td><img src="image11" alt="Image" /></td>
<td><img src="image12" alt="Image" /></td>
</tr>
</tbody>
</table>

APPENDIX B: VISUAL STIMULUS LIST
<table>
<thead>
<tr>
<th>Number</th>
<th>Trochaic Target Word</th>
<th>Iambic Target Word</th>
<th>Filler</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td><img src="number_5.png" alt="Image" /></td>
<td><img src="trochaic_5.png" alt="Image" /></td>
<td><img src="iambic_5.png" alt="Image" /></td>
</tr>
<tr>
<td>6</td>
<td><img src="number_6.png" alt="Image" /></td>
<td><img src="trochaic_6.png" alt="Image" /></td>
<td><img src="iambic_6.png" alt="Image" /></td>
</tr>
<tr>
<td>7</td>
<td><img src="number_7.png" alt="Image" /></td>
<td><img src="trochaic_7.png" alt="Image" /></td>
<td><img src="iambic_7.png" alt="Image" /></td>
</tr>
<tr>
<td>8</td>
<td><img src="number_8.png" alt="Image" /></td>
<td><img src="trochaic_8.png" alt="Image" /></td>
<td><img src="iambic_8.png" alt="Image" /></td>
</tr>
</tbody>
</table>

Table B.2: Nonword visual stimuli
Table B.3: Nonword visual stimuli
### Table B.4: Nonword visual stimuli

<table>
<thead>
<tr>
<th>Number</th>
<th>Trochaic Target Word</th>
<th>Iambic Target Word</th>
<th>Filler</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td><img src="image1" alt="Image" /></td>
<td><img src="image2" alt="Image" /></td>
<td><img src="image3" alt="Image" /></td>
</tr>
<tr>
<td>14</td>
<td><img src="image4" alt="Image" /></td>
<td><img src="image5" alt="Image" /></td>
<td><img src="image6" alt="Image" /></td>
</tr>
<tr>
<td>15</td>
<td><img src="image7" alt="Image" /></td>
<td><img src="image8" alt="Image" /></td>
<td><img src="image9" alt="Image" /></td>
</tr>
<tr>
<td>16</td>
<td><img src="image10" alt="Image" /></td>
<td><img src="image11" alt="Image" /></td>
<td><img src="image12" alt="Image" /></td>
</tr>
</tbody>
</table>
### APPENDIX C: TRANSCRIPTION RELIABILITY

<table>
<thead>
<tr>
<th>No.</th>
<th>Trochaic Word</th>
<th>Iambic Word</th>
<th>Korean Speakers</th>
<th>English Speakers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Transcript</td>
<td>Agr</td>
<td>Transcript</td>
<td>Agr</td>
</tr>
<tr>
<td>1</td>
<td>/mætoker/</td>
<td>lax</td>
<td>100</td>
<td>voiced</td>
</tr>
<tr>
<td>2</td>
<td>/tɛmido/</td>
<td>aspirated</td>
<td>100</td>
<td>voiceless</td>
</tr>
<tr>
<td>3</td>
<td>/dʒakunä/</td>
<td>lax</td>
<td>100</td>
<td>voiced</td>
</tr>
<tr>
<td>4</td>
<td>/gɪsædʒi/</td>
<td>lax</td>
<td>94.4</td>
<td>voiced</td>
</tr>
<tr>
<td>5</td>
<td>/sɪbɛtä/</td>
<td>aspirated</td>
<td>100</td>
<td>voiceless</td>
</tr>
<tr>
<td>6</td>
<td>/kunëɾzi/</td>
<td>aspirated</td>
<td>100</td>
<td>voiceless</td>
</tr>
<tr>
<td>7</td>
<td>/fɛɾɡabo/</td>
<td>aspirated</td>
<td>100</td>
<td>voiceless</td>
</tr>
<tr>
<td>8</td>
<td>/pɔvisu/</td>
<td>aspirated</td>
<td>100</td>
<td>voiceless</td>
</tr>
<tr>
<td>9</td>
<td>/mɔkrdʒi/</td>
<td>lax</td>
<td>100</td>
<td>voiced</td>
</tr>
<tr>
<td>10</td>
<td>/tuṃsä/</td>
<td>aspirated</td>
<td>100</td>
<td>voiceless</td>
</tr>
<tr>
<td>11</td>
<td>/dʒiʃaɾo/</td>
<td>lax</td>
<td>100</td>
<td>voiced</td>
</tr>
<tr>
<td>12</td>
<td>/gæɾtekə/</td>
<td>lax</td>
<td>94.4</td>
<td>voiced</td>
</tr>
<tr>
<td>13</td>
<td>/sɛɾɡuni/</td>
<td>aspirated</td>
<td>100</td>
<td>voiceless</td>
</tr>
<tr>
<td>14</td>
<td>/keɾoʃaɾo/</td>
<td>aspirated</td>
<td>100</td>
<td>voiceless</td>
</tr>
<tr>
<td>15</td>
<td>/fəbiɾu/</td>
<td>aspirated</td>
<td>100</td>
<td>voiceless</td>
</tr>
<tr>
<td>16</td>
<td>/pəɾævəɾu/</td>
<td>aspirated</td>
<td>100</td>
<td>voiceless</td>
</tr>
</tbody>
</table>

|     | Mean          | 98.7        | 100             |                  |

Table C.1: Transcription of nonword initial segments (aspirated/voiceless vs. lax/voiced) by Korean (n = 19) and English speakers (n = 23) and percent agreement (%) in the preliminary eyetracking experiment.
<table>
<thead>
<tr>
<th>No.</th>
<th>Trochaic Word</th>
<th>Iambic Word</th>
<th>Korean Speakers</th>
<th>English Speakers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Transcript</td>
<td>Agr</td>
</tr>
<tr>
<td>1</td>
<td>/povasu/</td>
<td>/pəˈvosei/</td>
<td>aspirated</td>
<td>100</td>
</tr>
<tr>
<td>2</td>
<td>/timado/</td>
<td>/təˈmidaI/</td>
<td>aspirated</td>
<td>100</td>
</tr>
<tr>
<td>3</td>
<td>/kanodzau/</td>
<td>/kəˈnɒdʒi/</td>
<td>aspirated</td>
<td>100</td>
</tr>
<tr>
<td>4</td>
<td>/sibeta/</td>
<td>/səˈbitə/</td>
<td>aspirated</td>
<td>100</td>
</tr>
<tr>
<td>5</td>
<td>/bakonat/</td>
<td>/bəˈkɒnət/</td>
<td>lax</td>
<td>100</td>
</tr>
<tr>
<td>6</td>
<td>/deszi/</td>
<td>/dəˈsɛzi/</td>
<td>lax</td>
<td>100</td>
</tr>
<tr>
<td>7</td>
<td>/gætəvə/</td>
<td>/ɡəˈtævə/</td>
<td>lax</td>
<td>100</td>
</tr>
<tr>
<td>8</td>
<td>/meŋəkei/</td>
<td>/məˈɡeɪka/</td>
<td>lax</td>
<td>100</td>
</tr>
<tr>
<td>9</td>
<td>/pmaku/</td>
<td>/pəˈmuːko/</td>
<td>aspirated</td>
<td>100</td>
</tr>
<tr>
<td>10</td>
<td>/tæbeto/</td>
<td>/təˈbætə/</td>
<td>aspirated</td>
<td>95.2</td>
</tr>
<tr>
<td>11</td>
<td>/kivəzau/</td>
<td>/kəˈvɪzəu/</td>
<td>aspirated</td>
<td>100</td>
</tr>
<tr>
<td>12</td>
<td>/sərgənə/</td>
<td>/səˈɡɛni/</td>
<td>aspirated</td>
<td>100</td>
</tr>
<tr>
<td>13</td>
<td>/brməsa/</td>
<td>/bəˈməsəu/</td>
<td>lax</td>
<td>100</td>
</tr>
<tr>
<td>14</td>
<td>/dokədzii/</td>
<td>/dəˈkɒdʒi/</td>
<td>lax</td>
<td>100</td>
</tr>
<tr>
<td>15</td>
<td>/gasədau/</td>
<td>/ɡəˈsədei/</td>
<td>lax</td>
<td>85.7</td>
</tr>
<tr>
<td>16</td>
<td>/matovei/</td>
<td>/məˈtəvi/</td>
<td>lax</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><strong>Mean</strong></td>
<td><strong>98.8</strong></td>
</tr>
</tbody>
</table>

Table C.2: Transcription of nonword initial segments (aspirated/voiceless vs. lax/voiced) by Korean (n = 21) and English speakers (n = 24) and Percent Agreement (%) in Experiment 1
<table>
<thead>
<tr>
<th>No.</th>
<th>Trochaic Word</th>
<th>Iambic Word</th>
<th>Korean Speakers</th>
<th>English Speakers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Transcript</td>
<td>Agr</td>
</tr>
<tr>
<td>1</td>
<td>/povisu/</td>
<td>/po'viset/</td>
<td>aspirated</td>
<td>100</td>
</tr>
<tr>
<td>2</td>
<td>/timado/</td>
<td>/ti'mada/</td>
<td>aspirated</td>
<td>100</td>
</tr>
<tr>
<td>3</td>
<td>/kanædʒau/</td>
<td>/kæ'naedʒu/</td>
<td>aspirated</td>
<td>100</td>
</tr>
<tr>
<td>4</td>
<td>/sibeta/</td>
<td>/si'beti/</td>
<td>aspirated</td>
<td>100</td>
</tr>
<tr>
<td>5</td>
<td>/bakemi/</td>
<td>/ba'kemə/</td>
<td>lax</td>
<td>100</td>
</tr>
<tr>
<td>6</td>
<td>/desιtə/</td>
<td>/de'siʃo/</td>
<td>lax</td>
<td>100</td>
</tr>
<tr>
<td>7</td>
<td>/gætavə/</td>
<td>/gæ'tavau/</td>
<td>lax</td>
<td>100</td>
</tr>
<tr>
<td>8</td>
<td>/meɪgokei/</td>
<td>/meɪ'ɡoki/</td>
<td>lax</td>
<td>100</td>
</tr>
<tr>
<td>9</td>
<td>/paneɪku/</td>
<td>/pa'neɪka/</td>
<td>aspirated</td>
<td>96.9</td>
</tr>
<tr>
<td>10</td>
<td>/ˈtebɪto/</td>
<td>/ˈtebiʃu/</td>
<td>aspirated</td>
<td>96.9</td>
</tr>
<tr>
<td>11</td>
<td>/kævaʃet/</td>
<td>/kæ'vaʃə/</td>
<td>aspirated</td>
<td>100</td>
</tr>
<tr>
<td>12</td>
<td>/ˈseɪgoʃə/</td>
<td>/ˈseɪɡodi/</td>
<td>aspirated</td>
<td>100</td>
</tr>
<tr>
<td>13</td>
<td>/ˈbotmaʊ/</td>
<td>/ˈboʃmeɪ/</td>
<td>lax</td>
<td>100</td>
</tr>
<tr>
<td>14</td>
<td>/diˈkædʒai/</td>
<td>/diˈkaʃai/</td>
<td>lax</td>
<td>100</td>
</tr>
<tr>
<td>15</td>
<td>/ɡæmæsə/</td>
<td>/ɡæ'mæsə/</td>
<td>lax</td>
<td>100</td>
</tr>
<tr>
<td>16</td>
<td>/ˌmɪʃəva/</td>
<td>/ˌmɪʃevo/</td>
<td>lax</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mean</td>
<td>99.6</td>
</tr>
</tbody>
</table>

Table C.3: Transcription of nonword initial segments (aspirated/voiceless vs. lax/voiced) by Korean (n = 32) and English speakers (n = 33) and Percent Agreement (%) in Experiment 2
APPENDIX D: LANGUAGE BACKGROUND QUESTIONNAIRE

Name (First, Last) ____________________________  Date ______________

Age ____________  Sex ( M / F / Other )

1. Where have you lived, and what age were you at the time (starting with place of birth)?

Place (City, Area, Country)  Ages (From. To.)

__________________________  __________________________

__________________________  __________________________

__________________________  __________________________

__________________________  __________________________

2. Where were your parents or other caretakers born? Where did they grow up?

Who?  Place of Birth:  Place(s) they grew up:

__________________________  __________________________  __________________________

__________________________  __________________________  __________________________

__________________________  __________________________  __________________________

3. Did your parents or other caretakers speak any language(s) other than English to you at home? If so, which one(s)?

______________________________
4. What other languages do you speak (if any)? At what age did you begin learning these languages? How well can you read, write, speak and understand each one?

<table>
<thead>
<tr>
<th>Language</th>
<th>Age</th>
<th>Reading</th>
<th>Writing</th>
<th>Speaking</th>
<th>Understanding</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1 2 3 4</td>
<td>1 2 3 4</td>
<td>1 2 3 4</td>
<td>1 2 3 4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 2 3 4</td>
<td>1 2 3 4</td>
<td>1 2 3 4</td>
<td>1 2 3 4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 2 3 4</td>
<td>1 2 3 4</td>
<td>1 2 3 4</td>
<td>1 2 3 4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 2 3 4</td>
<td>1 2 3 4</td>
<td>1 2 3 4</td>
<td>1 2 3 4</td>
</tr>
</tbody>
</table>

5. If you have taken a TOEFL test, when did you take it most recently? What score did you get in the test?

<table>
<thead>
<tr>
<th>Year</th>
<th>Reading</th>
<th>Writing</th>
<th>Speaking</th>
<th>Listening</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

6. Circle the highest level of education you have completed so far:

Primary        Junior high        High school        College        Post-graduate

7. What is your profession? (If academic or student, indicate field of study)

8. Are you right-handed or left-handed?

9. Do you have any speech or hearing disorders? Y / N (if yes, please describe)
10. Do you have any vision impairment? Y / N (if yes, please describe).

11. Are you good at playing any musical instruments? Y / N

   If yes, what musical instruments can you play? When did you start learning to play the
   instruments (age)? How long did you play the instruments?

<table>
<thead>
<tr>
<th>Musical Instrument</th>
<th>Age (From-To)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

12. May we contact you for future studies? Y / N

   If you answered yes, please provide your contact phone number and/or e-mail.

   ________________________________
APPENDIX E: PRELIMINARY DATA ANALYSIS ON THE LOG RATIO OF TARGET OVER COMPETITOR

Figure E.1: Bias toward target over competitor (logits) during trochaic and iambic nonword recognition. Left: English speakers ($N = 23$; 151 trochaic and 144 iambic nonwords). Right: Korean learners of English ($N = 19$; 132 trochaic and 125 iambic nonwords).
Table E.1: Summary of stress pattern effect on log gaze probability ratios of target over competitor in t-test (By subject analysis).

<table>
<thead>
<tr>
<th>Analysis window</th>
<th>Trochaic English speakers</th>
<th>Iambic English speakers</th>
<th>Trochaic Korean speakers</th>
<th>Iambic Korean speakers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>t(22) p&lt;</td>
<td>t(22) p&lt;</td>
<td>t(18) p&lt;</td>
<td>t(18) p&lt;</td>
</tr>
<tr>
<td>-153 ms - 0 ms</td>
<td>0.18 n.s.</td>
<td>-0.60 n.s.</td>
<td>0.77 n.s.</td>
<td>0.11 n.s.</td>
</tr>
<tr>
<td>0 ms - 200 ms</td>
<td>2.06 .05</td>
<td>0.47 n.s.</td>
<td>0.32 n.s.</td>
<td>-0.23 n.s.</td>
</tr>
<tr>
<td>200 ms - 400 ms</td>
<td>2.54 .05</td>
<td>2.36 .05</td>
<td>1.30 n.s.</td>
<td>-0.04 n.s.</td>
</tr>
<tr>
<td>400 ms - 600 ms</td>
<td>2.72 .05</td>
<td>3.29 .05</td>
<td>2.98 .05</td>
<td>-0.08 n.s.</td>
</tr>
<tr>
<td>600 ms - 867 ms</td>
<td>5.61 .0001</td>
<td>4.17 .001</td>
<td>5.17 .001</td>
<td>1.48 n.s.</td>
</tr>
</tbody>
</table>

Figure E.2: English speakers (N = 23)’ bias toward target over competitor (logits) for aspirated and lax initial nonwords. Left: Trochaic nonwords (98 aspirated and 53 lax initial nonwords). Right: Iambic nonwords (97 aspirated and 47 lax initial nonwords).

Table E.2: Summary of stress pattern and phonation effect on log gaze probability ratios of target over competitor for English speakers in t-test (By subject analysis).

<table>
<thead>
<tr>
<th>Analysis window</th>
<th>Trochaic Aspirated</th>
<th>Trochaic Lax</th>
<th>Iambic Aspirated</th>
<th>Iambic Lax</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>t(22) p&lt;</td>
<td>t(22) p&lt;</td>
<td>t(22) p&lt;</td>
<td>t(22) p&lt;</td>
</tr>
<tr>
<td>-153 ms - 0 ms</td>
<td>-1.62 n.s.</td>
<td>2.39 .05</td>
<td>0.40 n.s.</td>
<td>-1.49 n.s.</td>
</tr>
<tr>
<td>0 ms - 200 ms</td>
<td>0.48 n.s.</td>
<td>2.89 .05</td>
<td>-0.50 n.s.</td>
<td>0.07 n.s.</td>
</tr>
<tr>
<td>200 ms - 400 ms</td>
<td>0.95 n.s.</td>
<td>2.31 .05</td>
<td>1.46 n.s.</td>
<td>1.53 n.s.</td>
</tr>
<tr>
<td>400 ms - 600 ms</td>
<td>2.06 .05</td>
<td>1.10 n.s.</td>
<td>1.59 n.s.</td>
<td>2.74 .05</td>
</tr>
<tr>
<td>600 ms - 867 ms</td>
<td>5.13 .0001</td>
<td>1.95 n.s.</td>
<td>3.14 .05</td>
<td>2.79 .05</td>
</tr>
</tbody>
</table>
Figure E.3: Korean speakers (N = 19)’ bias toward target over competitor (logits) for aspirated and lax initial nonwords. Left: Trochaic nonwords (80 aspirated and 52 lax initial nonwords). Right: Iambic nonwords (83 aspirated and 42 lax initial nonwords).

Table E.3: Summary of stress pattern and phonation effect on log gaze probability ratios of target over competitor for Korean speakers in t-test (By subject analysis).
APPENDIX F: EXPERIMENT 1- RT BEHAVIORAL DATA ANALYSIS AFTER ON-TARGET TRIALS REMOVED

<table>
<thead>
<tr>
<th>Stress</th>
<th>Phonation</th>
<th>English speakers</th>
<th>Korean speakers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trochaic</td>
<td>Aspirated</td>
<td>2291.1 (85.95)</td>
<td>2048.5 (66.43)</td>
</tr>
<tr>
<td></td>
<td>Lax</td>
<td>2108.9 (98.23)</td>
<td>2049.6 (75.48)</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>2207.2 (65.05)</td>
<td>2049.1 (50.24)</td>
</tr>
<tr>
<td>Iambic</td>
<td>Aspirated</td>
<td>2072.7 (93.49)</td>
<td>2245.5 (109.18)</td>
</tr>
<tr>
<td></td>
<td>Lax</td>
<td>2288.5 (79.10)</td>
<td>2314.8 (78.85)</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>2207.1 (62.12)</td>
<td>2280.8 (66.69)</td>
</tr>
</tbody>
</table>

Table F.1: Mean RT by L1, Stress pattern, and Phonation type after on-target trials removed. Values in parentheses represent standard errors.

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Est.</th>
<th>SE</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>2181.31</td>
<td>51.22</td>
<td>42.59</td>
<td>.0001</td>
</tr>
<tr>
<td>L1 = Korean L1</td>
<td>-17.30</td>
<td>101.56</td>
<td>-0.17</td>
<td>n.s.</td>
</tr>
<tr>
<td>Stress1 = Iambic</td>
<td>109.51</td>
<td>55.72</td>
<td>1.97</td>
<td>.05</td>
</tr>
<tr>
<td>Phonation1 = Lax</td>
<td>31.49</td>
<td>57.56</td>
<td>0.55</td>
<td>n.s.</td>
</tr>
<tr>
<td>L1:Stress1</td>
<td>218.04</td>
<td>111.43</td>
<td>1.96</td>
<td>.05</td>
</tr>
<tr>
<td>L1:Phonation1</td>
<td>-18.20</td>
<td>111.99</td>
<td>-0.16</td>
<td>n.s.</td>
</tr>
<tr>
<td>Stress1:Phonation1</td>
<td>270.65</td>
<td>111.58</td>
<td>2.43</td>
<td>.05</td>
</tr>
<tr>
<td>L1:Stress1:Phonation1</td>
<td>-338.92</td>
<td>223.12</td>
<td>-1.52</td>
<td>n.s.</td>
</tr>
</tbody>
</table>

Table F.2: Summary of L1, Stress pattern, and Phonation type effects on RT in the mixed effect model (N = 452; logLik = -3496). The RTs for on-target trials were removed.
References


Cutler, A. (1986). Forbear is a homophone: Lexical prosody does not constrain lexical access. *Language and Speech, 29*(3), 201. 32, 33, 38, 177


210


Kim, M., Beddor, P., & Horrocks, J. (2002). The contribution of consonantal and vocalic information to the perception of Korean initial stops. *Journal of Phonetics*, 30(1), 77–100. 16


216


