Linguistic Research 40(3): 587-606
DOI: 10.17250/khisli.40.3.202312.009

Exploring lexical stress processing in L2 English: A comparative eye-tracking study of native English listeners and Japanese listeners*

Jeonghwa Shin
(Korea Military Academy)

Shin, Jeonghwa. 2023. Exploring lexical stress processing in L2 English: A comparative eye-tracking study of native English listeners and Japanese listeners. *Linguistic Research* 40(3): 587-606. This study aims to explore how individuals with a native language characterized by a lexical pitch accent approach lexical stress in a stress-timed L2 during spoken word recognition. To this end, native English listeners and Japanese listeners of English participated in two phases of experiment: a three-day training and a subsequent eye-tracking experiment. The eye-tracking results revealed distinct processing patterns. Native English listeners predominantly recognized trochaic words by relying on the initial stressed syllable. In contrast, for iambic words, they utilized both the initial unstressed and the second stressed syllables for recognition. Japanese listeners of English demonstrated a different pattern of processing. They initiated lexical access within the first syllable of trochaic stress patterns and slightly later, still relying on first-syllable information, for iambic words. This finding implies that a single initial syllable is enough for Japanese listeners of English to utilize word stress information during L2 English spoken word recognition unlike native English listeners. The equal efficiency in employing two lexical stress patterns in L2 English suggests that lexical processing strategies transferred from the L2 listeners' native language could facilitate word recognition in the target language. While this study underscores the advantages of L1 prosodic structures in L2 English word recognition, it does not imply that Japanese listeners of English process English word stress in the same manner as native English listeners do during overall English word recognition. **(Korea Military Academy)**

Keywords eye-tracking, lexical stress, L2 word recognition, Japanese listeners of English

* This study was supported by 2023 research fund of Korea Military Academy (Hwarangdae Research Institute) for publication and was also funded by a Targeted Investment grant from the Ohio State University. I also would like to appreciate the valuable comments that reviewers provided for this paper. An earlier version of this paper was published in the proceedings of the 10th Tokyo Conference on psycholinguistics. The usual disclaimers apply.

© 2023 Jeonghwa Shin, published by *Linguistic Research* (KHU ISLI). This work is licensed under the Creative Commons Attribution 4.0 International License.

1. Introduction

Research into second language (L2) acquisition has consistently revealed the profound influence of learners' first language (L1) experience on their perception and production of L2 sounds (Kuhl and Iverson 1995; Yamada 1995; Dupoux et al. 1999; Strange et al. 2001; Munro and Bohn 2007; Major 2008; Kusumoto 2012). Central theories in L2 phonology and phonetic acquisition, such as the Perceptual Assimilation Model (Best 1995) and the Speech Learning Model (Flege 1987; Flege 1995), while focusing on varying aspects of L2 sound acquisition, share a common assertion that the ease and proficiency of L2 sound acquisition depends on the extent of similarity between the phonological and phonetic systems of the learner's L1 and L2. Essentially, the greater the similarity between a non-native sound and a native sound, the more effortless and accurate the perception and production of the target sound by L2 learners.

While it is well-established that a learner's L1 segmental phonology significantly affects the acquisition of L2 phonemic contrasts, there remains a notable gap in our understanding of how L1 autosegmental phonology (including elements like tones and metrical structures) affects the processing of prosodic information in L2. If, akin to variations observed in L2 phonemic acquisition, there are language-specific differences in the acquisition of prosody, it stands to reason that the structures and mechanisms of prosody in one's L1 will have implications for L2 sound acquisition and contribute to our comprehension of spoken language processing in L2.

The primary focus of this study is to explore the role of a specific prosodic element, lexical stress also known as word stress, in the context of spoken word recognition of English as an L2. We investigate the utilization of English lexical stress in spoken word recognition by Japanese-speaking learners of English, hereafter referred to as Japanese listeners of English. Specifically, we examine the temporal dynamics of their English lexical stress processing by tracking eye movements during a auditory word recognition task and comparing their performance to that of monolingual English-speaking controls.

The following sections will begin with an overview of prosodic structure at the lexical level in Japanese and English, shedding light on how native listeners from each language utilize prosodic cues for spoken word recognition. This foundation will inform our predictions regarding the behavior of English L1 listeners and Japanese listeners of English in the context of spoken English word recognition. Subsequently, we will provide a description of our research methodology, involving training listeners to associate

English nonwords with varying stress assignments with images of extraterrestrial "aliens." We then monitor their eye movements as they identify the correct alien based on its name in a spoken sentence. Lastly, we will present and discuss the findings, comparing how Japanese listeners of English and monolingual English controls process English lexical stress information throughout the course of spoken word recognition.

2. Word-level prosodic structure in English and Japanese

English is known as a "stress-timed language," where speakers adjust the timing between stressed syllables to maintain a relatively constant rhythm, which may result in the compression or expansion of unstressed syllables. To maintain such a temporal organization of speech rhythm, English exhibits significant variations in stress at both the individual word level (word stress) and the phrasal level where pitch accents emphasize stressed syllables. Stressed syllables in English differ from their unstressed counterparts in terms of acoustic characteristics such as fundamental frequency (F0), duration, and intensity (Fry 1955, 1958; Lieberman 1960; Lehiste 1976). Typically, stressed syllables in isolated words display higher F0, longer duration, and greater intensity compared to unstressed syllables. These acoustic distinctions contribute to perceptual prominence differences, creating trochaic (SW) and iambic (WS) stress patterns in many English word pairs. Moreover, most word pairs contrasting in stressed syllable location also differ in vowel quality, with unstressed vowels reduced to forms like schwa (Delattre 1969). This full vs. reduced vowel information serves as a reliable cue for determining syllable stress.

However, research exploring the acoustic cues used by native English listeners to distinguish stressed from unstressed syllables has yielded mixed results. For instance, Cutler (1986) found in a cross-modal priming experiment that minimal stress pairs, such as "FORbear" vs. "forBEAR," activated and facilitated both members, suggesting that lexical stress might not play a central role in lexical access. On the other hand, other studies have shown that listeners are sensitive to lexical stress patterns during word processing. Misplaced stress (e.g., "CHEmist" pronounced as "cheMIST") consistently hinders target word recognition (Cutler and Clifton 1985; Small, Simon, and Goldberg 1988; Slowiaczek 1990; Mattys and Samuel 2000). In a recent eye-tracking investigation (Creel et al. 2006), the impact of lexical stress on the capacity of native English listeners to acquire an artificial lexicon with English-like characteristics was examined. Notably,

while disparities in lexical stress did not lead to increased confusion between stress-matched and stress-mismatched cohort items when pronounced in isolation (e.g., “KAdazu” vs. “kaDAzei”), these stress distinctions became valuable when the words were embedded within sentences. This observation implies that lexical stress plays a crucial role in the process of segmenting lexical items from the continuous speech stream, particularly in the context of sentence-level comprehension.

In contrast to English, Japanese is a pitch-accent language where pitch, or the fundamental frequency of the voice, is used to convey lexical or grammatical meaning. Japanese has two distinct prosodic patterns at the word level: accented and unaccented. Accented words feature an F0 fall from high to low on one “mora[1]” of the word, while unaccented words lack this fall and begin with relatively low F0 on the first mora, followed by relatively higher F0 on subsequent morae (Sekiguchi 2006). Japanese intonational phonology organizes words into “accentual phrases,” where accent location and height are largely determined by lexical concatenation rules (Venditti 2005).

The majority of studies delving into the role of Japanese pitch accent have consistently demonstrated that Japanese native speakers (L1) employ lexical pitch accent when identifying words. Minematsu and Hirose (1995) probed the significance of lexical pitch accent by comparing the recognition of words with misaccented and correctly accented patterns, both in isolation and in context. Their findings indicated that isolated misaccented words posed a greater recognition challenge compared to correctly accented words. However, when presented within a contextual framework, the accent's impact on recognition was reduced. Cutler and Otake (1999) employed a combination of identification and gating tasks to investigate how Japanese listeners process the HL (High -Low) and LH (Low-High) accent contrast in bimoraic Japanese words. Their results revealed that listeners could predict the target word even when only the initial sequence of a consonant and a vowel was presented. Notably, Japanese listeners showed a good perceptual sensitivity to the F0 characteristics of the mora, prioritizing it over duration or intensity cues in the perception of accentual information.

In summary, while both English and Japanese use prosodic features to distinguish lexical items, these features exhibit different acoustic and perceptual characteristics. Native speakers tune their word recognition processes to their native language's acoustic

1 In Japanese phonetics, a “mora” is a unit of sound that carries linguistic significance. It is not equivalent to a syllable in English. A mora can be a single vowel, a consonant followed by a vowel, or the “n” sound. In terms of timing, each mora is roughly equal in duration.

cues. Japanese listeners may rely primarily on F0 cues from the first mora, while English listeners use a more complex set of cues, including syllable duration, F0, intensity, and vowel quality.

This study investigates how Japanese listeners of English process English lexical stress during online spoken word recognition in L2 English. We compared their word identification with that of English L1 listeners by monitoring eye movements in the picture-choice task. As the eye-tracking methodology can reveal the time course of processing while participants listen to words with different prosodic features, the findings are expected to elucidate whether and how English lexical stress influences the early stages of lexical access in native and non-native listeners. Additionally, the study aims to determine whether pitch accent patterns of L1 Japanese facilitate or hinder spoken word recognition in English as the L2. To achieve this, all participants underwent three days of training to learn the nonword-picture associations, followed by an eye-tracking experiment with a picture-choice task. Section 3 outlines the training phase, and Section 4 presents the eye-tracking experiment.

3. Training phrase

3.1 Method

3.1.1 Materials

To account for potential influences stemming from word frequency, word familiarity, and vocabulary size, we devised 48 trisyllabic nonword stimuli by randomly combining the eight most frequently occurring phonemes at each position within English trisyllabic words (C1V1C2V2C3V3; C = Consonant, V = Vowel). The selection of these phonemes was determined based on their Nlog frequency, as documented in the LDC American English Spoken Lexicon (AESL, available at http://www.ldc.upenn.edu/cgi-bin/aesl/aesl).

To ensure that these nonwords were designed to examine the processing of English lexical stress during spoken word recognition, we excluded the reduced vowel and its variant in the first two vowels. For experimental consistency, the first two syllables of the test nonwords featured the same full vowels. Table 1 provides an overview of the eight phonemes employed in each position within the trisyllabic nonword stimuli.

Table 1. Eight phonemes used in the trisyllabic target nonwords (C1V1C2V2C3V3)

C1 = / p, t, k, g, s, f, m, dʒ /	V1 = / ɑ, æ, ɛ, eɪ, i, ɪ, o, u / *
C2 = / p, t, k, g, v, s, m, n /	V2 = / ɑ, æ, ɛ, eɪ, i, ɪ, o, u / *
C3 = / k, t, d, v, s, n, ʒ, dʒ /	V3 = / ɑ, ɑɪ, eɪ, i, o, u, ə, ɚ /

*Nlog frequency ≥ -4.56

Out of the 48 stimuli, 32 were comprised of 16 minimal stress pairs, characterized by having identical cohorts in the first two syllables while being phonetically distinct in the final syllable (e.g., /ˈdʒɑkunaɪ/ vs. /dʒɑˈkunɚ/). The remaining 16 nonwords were included as filler items. All nonwords underwent a thorough screening process by native speakers of both English and Japanese to ensure that they bore no resemblance to actual words in either language. The full list of target nonwords is presented in Appendix.

The nonword stimuli designated for training purposes were recorded in isolation by a female native English speaker in a sound-attenuated booth. Within each minimal stress pair, the members shared the same pitch accent associated with the lexically stressed syllable, either H* or L+H*. The auditory stimuli were recorded at a sampling rate of 44.1 kHz using Praat.

Paired *t*-tests conducted on the nonwords in isolation revealed that trochaic words exhibited longer first syllables (0.13 s vs. 0.10 s), higher mean F0 (309 Hz vs. 262 Hz), and greater intensity (78 dB vs. 73 dB) compared to the first syllables of iambic words (all *ts* (15) > 2, all *ps* < 0.001). Conversely, iambic words displayed longer second syllables (0.16 s vs. 0.10 s), higher mean F0 (298 Hz vs. 211 Hz), and greater intensity (78 dB vs. 72 dB) in comparison to the second syllables of trochaic words (all *ts* (15) > 2, all *ps* < 0.001).

The visual stimuli used as referents of the 48 nonwords were line-drawings of space aliens that were constructed specifically for use in a word-learning paradigm (drawings taken from Gupta et al. 2004).

3.1.2 Participants

Twenty three native English listeners aged 18-34 years (*M* = 21.5, *SD* = 2.88) participated in the experiment as a control group. They were born in the U.S. and learned English as L1. Some of them had history of learning foreign language in high school or college, but none of them had lived abroad and no other language was spoken at home. Additionally, fifteen standard Japanese speakers who were primarily undergraduate

or graduate students at the time of participation were recruited. None of the Japanese speakers had any English-immersion experience before entering the U.S. They entered the U.S. at an average age of 20.2 years (range: 19 – 24 years; *SD* = 1.23) and lived in the U.S. less than 1 month (*SD* = 0.83) at the time of recruitment. The mean beginning age of learning L2 English in Japan was 12 years old (range: 9 – 17 years; *SD* = 3.56). In the language background questionnaire, Japanese 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 (range: 2.2 – 3.8; *SD* = .45).

The participants reported here were the individuals who demonstrated a minimum of 90% accuracy in the final training session and achieved an average of 80% accuracy in the eye-tracking word recognition task. None of the participants had received formal phonetic training, nor did they report any history of speech or hearing impairments.

3.1.3 Procedure

Participants took part in a three-day training, with each day comprising 12 learning blocks. Each learning block was divided into two sessions. The first session involved participants learning 8 nonword-picture associations, while the second session assessed their progress through a naming task (first six blocks) or a picture-choice task (second six blocks). The trials were managed using E-prime (version 1.2, Psychology Software Tools Inc.).

Within each trial of the learning sessions, a picture of a space alien was displayed in one of three positions on the screen, accompanied by the corresponding nonword sound played twice in a sequence. In the naming sessions, participants encountered the alien picture once again, where they were instructed to vocally identify it by saying "*This is the (target word)*" into the microphone. These productions were digitally recorded using Audacity (version 1.2.6). Subsequently, participants could confirm the accuracy of their response by pressing a button for auditory feedback. In the picture-choice sessions, participants selected the named alien from a set of three pictures after hearing the nonword. Immediate feedback was provided.

Over the course of the 12 training blocks, participants were exposed to the 48 nonwords, each paired with a corresponding alien image reference, a total of 10 times

in each of the three one-hour training sessions. Therefore, throughout the three consecutive training days, all participants were equally exposed to each nonword-picture pairing a total of 30 times. Once the last training session was completed, participants took a break for 5 minutes before the main eye-tracking experiment.

3.2 Results

Accuracy and response times were obtained from the picture-choice sessions of all three training days. Table 2 shows the average accuracy and response time for both groups across the three days of training.

Table 2. Accuracy and response time (RT) in the training phase

L1	Day1		Day2		Day3	
	Accuracy (%)	RT (ms)	Accuracy (%)	RT (ms)	Accuracy (%)	RT (ms)
English (n = 23)	85.6	1974.5	97.3	1534.3	98.5	1312.7
Japanese (n = 15)	85.0	2116.3	94.0	1559.9	95.8	1446.6

Repeated measures ANOVAs were conducted with L1 as a between-subjects factor and training day as a within-subjects factor, while considering both subjects and items as random effects, for the 48 nonwords. It was observed that Japanese listeners of English exhibited lower learning accuracy compared to English L1 speakers, leading to a main effect of L1 in the item analysis ($F1(1, 36) < 1$, $p < .001$; $F2(1, 47) = 26.1$, $p < .001$). As training progressed, performance consistently improved, resulting in a main effect of training day ($F1(2, 72) = 7.71$, $p < .001$; $F2(2, 94) = 13.93$, $p < .001$). Notably, there was no significant interaction between these two factors (both $Fs < 1$, $ps > .05$). On the final training day, both groups achieved accuracy levels exceeding 95%, with English L1 listeners at 98.5% and Japanese listeners of English at 95.8%.

Response times for both groups decreased as the training days advanced, establishing a main effect of training day in both subject and item analyses ($F1(2, 72) = 89.7$, $p < .001$; $F2(2, 94) = 74.73$, $p < .001$). English L1 speakers exhibited slightly shorter response times than Japanese listeners of English, although the main effect of L1 was only evident in the item analysis ($F1(1, 36) < 1$, $p > .05$; $F2(1, 47) = 6.86$, $p < .001$). The interaction between these two factors was not significant in either the subject or item analysis (both $Fs < 1$, $ps > .05$).

3.3 Discussion

These training outcomes reveal that both listener groups demonstrated enhancements in accuracy and reductions in response times as they took training over the three consecutive days. Despite Japanese listeners of English displaying numerically lower accuracy than English L1 listeners, their average accuracy exceeded 90% on the final training day, and their response times were comparable to those of the English L1 controls.

4. Eye-tracking experiment

4.1 Method

4.1.1 Materials

The 48 nonwords used in the training phase were recorded within the carrier sentence, "*Click on the (target word) now*". The speaker responsible for recording the nonwords in the training phase were asked to consistently employ an L+H* tone on the target nonword in the carrier sentence.

The 32 instructional sentences that encompassed both members of 16 stress minimal pairs (e.g., /ˈdʒɑkunɑɪ/ vs. /dʒɑˈkunə/) served as the target sentences for the final word recognition task. An additional 16 items served as fillers, and they featured words that were phonologically unrelated to the target pair members (e.g., /ˈgæsitə/).

To confirm the acoustic characteristics, we submitted the duration, mean F0, and intensity in the initial two syllables of the target words within the carrier sentence to paired *t*-tests. It was observed that words with trochaic stress exhibited longer first syllables (.11 s vs. .07 s), higher mean F0 (304 Hz vs. 244 Hz), and greater intensity (81 dB vs. 76 dB) compared to the first syllables of iambic words (All *ts* (15) > 2, All *ps* < .001). Conversely, iambic words featured second syllables that were longer in duration (.13 s vs. .09 s), higher mean F0 (307 Hz vs. 232 Hz), and greater intensity (79 dB vs. 76 dB) than the second syllables of trochaic words (All *ts* (15) > 2, All *ps* < .001).

Additionally, we measured the mean F0 on the word "*the*" within the carrier sentence

to investigate any early F0 rise noted before trochaic target words. A paired *t*-test yielded no significant difference in F0 between trochaic and iambic target words (232 Hz vs. 227 Hz; t (15) = 1.17, p = .26).

Collectively, these acoustic measurements suggest that if lexical stress is utilized by listeners to identify target words, stress minimal pairs can be discerned through these acoustic cues.

Regarding the visual stimuli of the eye-tracking experiment, we concurrently presented three objects while providing the instruction, "*Click on the (target word) now*", as shown in Figure 1. These visual stimuli consisted of images corresponding to two members of a minimal stress pair, one designated as the target and the other as the competitor, alongside a distractor. Each of the images was placed at an equal distance from the central fixation cross (+).

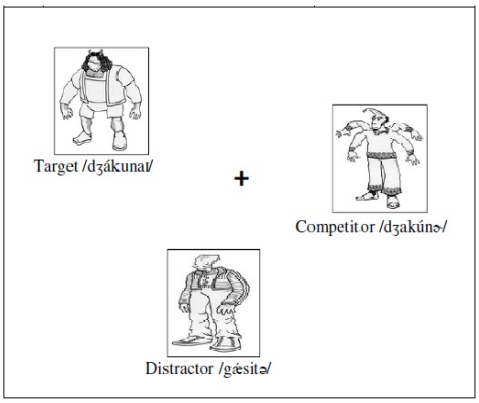

Figure 1. Presentation of visual stimuli in the eye-tracking experiment
Note: Text labels are expository and did not appear on the screen.

4.1.2 Participants

The same group of 23 native English listeners and 15 Japanese listeners of English who took part in the training phase participated in the subsequent eye-tracking experiment.

4.1.3 Procedures

The eye-tracking experiment employed a spoken word recognition task with visual world paradigm. Sixteen target sentences and 16 filler sentences were presented to the participants, each accompanied by a set of three alien images as shown in Figure 1. Half of the target trials featured trochaic stress, while the other half featured iambic stress. We balanced two sets of experimental stimuli lists to ensure that the same stress minimal pair members did not appear together in the same list. Across all test trials in a list, each target image appeared an equal number of times in each position on the screen.

During a trial, participants were instructed to select one of the three alien images in response to the instruction, "*Click on the (target word) now*," at their own pace. Clicking on the target image automatically led to a blank screen with a central fixation cross. Participants were required to click on the central cross when they were ready to proceed to the next trial. No feedback was provided during this phase. Participants' eye movements toward the target images during the task were tracked using an ASL Eye-Trac 6000 head-mounted eye tracker equipped with a 60 Hz camera and head-movement correction.

4.2 Analysis of eye-tracking data

The eye-tracking data analysis was performed on the "correct trials" only (See Table 3 in Section 4.2.1). Eye-tracking data from the "correct trials" underwent arcsine transformation to calculate fixation proportions for the three picture objects: the target nonword, the stress competitor, and the distracter. The data were synchronized to the onset of the target nonword for each trial (i.e., 0 ms), and subsequent syllable durations were marked by dotted vertical lines in Figures of mean fixation proportions. For trochaic nonwords, the average durations of the first and second syllables were 238 ms and 164 ms, respectively. In the case of iambic words, the durations for the initial two syllables were 180 ms and 230 ms.

To pinpoint the moment during the instructional timeline when the fixation proportions for the target and competitor became significantly distinct, *t*-tests were employed to compare the fixation proportion difference between the target and competitor to zero. These analyses scrutinized gaze patterns during the initial syllable of the target

word, employing a 238 ms window for trochaic words and 180 ms for iambic words. Additionally, consecutive 200 ms windows were assessed, taking into consideration the time required to plan and execute a saccadic eye movement during the eye-tracking experiment (Matin, Shao, and Boff 1993).

4.3 Results

4.3.1 Accuracy and response time

Table 3 provides a summary of the mean accuracy rate and mean response time for each language group during the spoken word recognition task.

Table 3. Mean accuracy and mean response time (RT) in the word recognition task

L1	Accuracy (%)	RT (ms)
English L1 Controls (n = 23)	88.6	2375.4
Japanese listeners of English (n = 15)	87.3	2302.8

To assess the differences in performance, we conducted a one-way repeated measures analysis of variance (ANOVA) with L1 as a between-subjects factor. This analysis revealed no significant accuracy differences between English L1 controls and Japanese listeners of English, as indicated by both subject and item analyses (both $Fs < 1$). While the response times for Japanese listeners were numerically shorter than those for English L1 listeners, this distinction only reached statistical significance in the item analysis ($F1(1, 36) < 1$, $p > .05$; $F2(1, 47) = 3.56$, $p < .05$). That is, word recognition accuracy and latencies did not significantly differ between native and non-native listeners. However, we need to track the activation of target words over the time course of word recognition by referring to the eye tracking data to address the research questions of the present study.

4.3.2 English L1 listeners

English controls' fixation proportions to the three pictures over the time course of the instruction, "*Click on the (target) now*" are shown in Figures 2a and 2b for trochaic and iambic target nonwords, respectively.

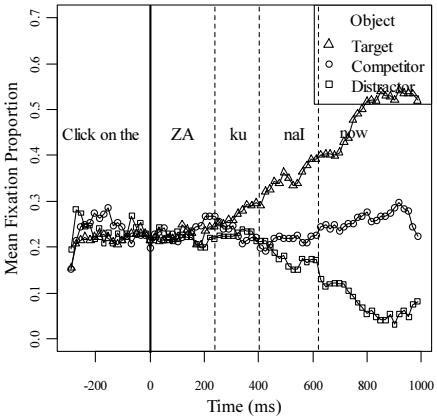

Figure 2a. Mean fixation proportions for trochaic condition, English L1 controls

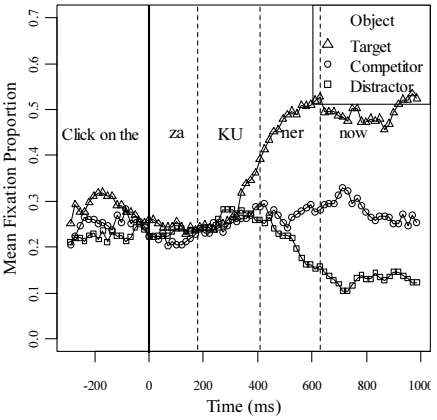

Figure 2b. Mean fixation proportions for iambic condition, English L1 controls

Visual examination of Figures 2a and 2b makes it apparent that English L1 listeners initiated the recognition of the correct target during the second syllable for both trochaic and iambic words. They did not entirely rule out the stress competitor in either condition when they entered into the second syllable.

English L1 controls displayed differences in the processing timeline for trochaic words compared to iambic words. In the case of trochaic words, gaze patterns towards the target started to differentiate from the competitor and distracter early in the second syllable, indicating sensitivity to stress information in the first syllable. A *t*-test conducted within the first 200 ms following the end of the first syllable approached statistical significance (t (1, 22) = 1.78, p = .08), and all subsequent 200 ms intervals showed significances (All *ts* (1, 22) > 2, All *ps* < .05).

For the iambic condition, gaze patterns towards the target began to diverge from the competitor and distracter later in the second syllable, suggesting a delayed response to variations in stress information between the first and second syllables. Fixation differences between the target and competitor were not statistically significant in the 200 ms window following the first syllable (t (1, 22) < 1, p > .05) but were significant in all subsequent 200 ms time intervals (All *ts* (1, 22) > 2, All *ps* < .05). These outcomes indicate that English L1 listeners utilized stress information from the first syllable in the case of trochaic words and from both syllables (or the second syllable) in the case of

iambic words to initiate their looks to the target object.

To investigate potential variations in the sensitivity of English listeners to the three relevant acoustic cues to English stress, a multiple linear regression was employed. The analysis focused on the differences in fixation proportion within the initial 200 ms following the end of the target word. Independent variables included mean F0, mean amplitude, and duration of the first syllable. Here the amplitude was obtained by converting intensity (dB) over F0 (Hz) in the 1st and 2nd vowel portions. The conversion was made because the tones of the same intensity, but of different frequency are perceived as being of different loudness (Fletcher and Munson 1933; Robinson and Dadson 1956). Notably, the mean amplitude of the first syllable emerged as the most consistent predictor of the fixation proportion difference between the target and competing objects among English L1 listeners (r^2 = .15, p < .05), suggesting the combinatorial interplay of F0 and intensity is the most reliable cue for native English listeners to use stress information during spoken word recognition.

4.3.3 Japanese listeners of English

Figure 3a and 3b present the mean fixation proportions over time for Japanese listeners of English.

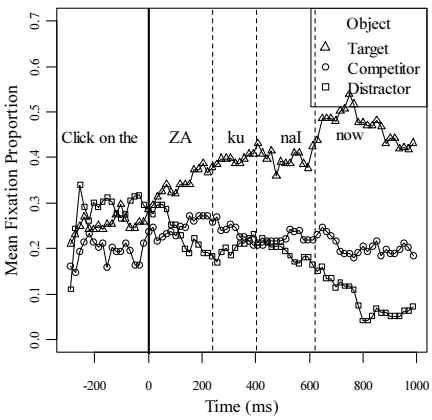

Figure 3a. Mean fixation proportions for trochaic condition by Japanese Listeners

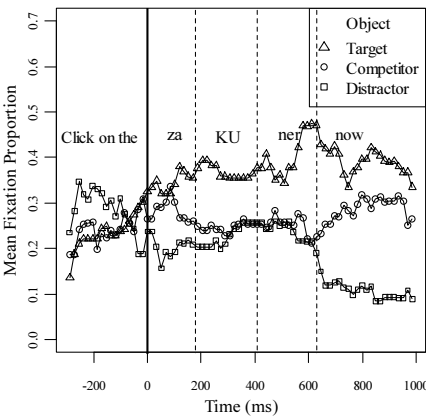

Figure 3b. Mean fixation proportions for iambic condition by Japanese Listeners

A visual examination of the data presented in Figures 3a and 3b reveals a noteworthy distinction in the response of Japanese listeners of English compared to the English control counterparts. Japanese listeners exhibited a tendency to identify the correct target as early as the first syllable, regardless of whether the word was trochaic or iambic. Fixation proportions directed at the target began to rise and diverge from the competitor and distracter almost immediately after the onset of the initial syllable of the target word. However, akin to English listeners, they did not entirely rule out the stress competitor, continuing to direct some of their attention toward the stress competitor throughout the entire measurement period in both conditions.

T-tests conducted on the fixation proportion difference between the target and competitor indicated that this difference was statistically significant in the first syllable of the trochaic condition (All *ts* (1, 14) > 2, All *ps* < .05). In the case of the iambic condition, significance emerged from the first 200ms window following the end of the first syllable (All *ts* (1, 14) > 2, All *ps* < .05). The results suggest that Japanese listeners were capable of utilizing either a stressed or unstressed initial syllable to identify the target.

Sensitivity to the acoustic cues was explored again using a multiple linear regression on the Japanese listeners' fixation proportion differences during the first 200 ms window after the offset of the target words with mean F0, mean amplitude and duration of first syllable as independent variables. Only the mean F0 of the first syllable approached significance (r^2 = .11, p = .05), suggesting F0 was the most reliable cues for Japanese learners of English to process English lexical stress during word recognition.

This outcome aligns with findings from Japanese L1 studies (Cutler and Otake 1999), suggesting that Japanese L1 listeners can match incoming speech input against the target word using just a single initial syllable and the prosodic cues within it. This processing strategy shares similarities with English L1 listeners in terms of their ability to incorporate lexical stress information into their lexical access. However, in contrast to English L1 listeners, Japanese listeners of English exhibited equal efficiency in using both lexical stress patterns, implying a strong sensitivity to absolute pitch height.

4.4 Discussion

Eye-tracking allows for millisecond-level temporal precision, enabling researchers to

capture the exact timing of visual attention shifts and cognitive processes during a task. The eye-tracking results of the present study revealed distinct processing patterns in the processing of trochaic and iambic stress patterns and also in the native vs. non-native listeners. English L1 controls recognized trochaic target nonwords by relying on the initial stressed syllable. For trochaic stress patterns, their lexical access was initiated upon the full realization of the first syllable. However, for iambic words, they utilized both the initial unstressed and the second stressed syllables for target word recognition. Iambic word recognition was delayed until the end of the second syllable in native English listeners. This delay in processing iambic words suggests that English L1 listeners needed extra time to integrate later information into the encoding of earlier segments to access the words. This outcome aligns with the findings of Mattys and Samuel (2000) and can be explained by their retroactive processing hypothesis, which posits that stress in non-initial positions activates irrelevant lexical candidates that must be subsequently deactivated. Consequently, we can say that processing non-initial stress words demands more phonetic memory capacity than processing initial stress words in native English listeners.

In contrast, Japanese listeners of English showed a different processing pattern. For trochaic stress patterns, their lexical access commenced within the first syllable of the target word, and slightly later, but still based on first syllable information, for iambic patterns. Notably, their responses to iambic patterns showed a significant difference between the target and competitor immediately after the initial unstressed syllable. These findings suggest that, for Japanese listeners of English, a single initial syllable provided sufficient information to initiate word recognition in English as the L2. This indicates that Japanese listeners of English may be more adept at using lexical stress information than English L1 listeners, as both stress patterns were equally efficient for them in identifying target words. This result is consistent with previous studies on the role of pitch accent in Japanese spoken word recognition. In Japanese, the native Japanese listeners make use of pitch height in the first syllable to identify accented/unaccented mora during spoken word recognition. Therefore, the equal efficiency in employing two lexical stress patterns in English can be attributed to the positive transfer of accent processing from their L1.

5. Conclusion

The present study investigated the processing of English lexical stress in both native English listeners and Japanese listeners of English. By employing eye-tracking methodology, we could make precise temporal measurements of processing patterns for English word stress over the time course of spoken word recognition. English L1 controls displayed distinct recognition strategies for trochaic and iambic words, relying on the initial stressed syllable for trochaic patterns but incorporating both initial unstressed and second stressed syllables for iambic words, causing a delayed iambic word recognition. This aligns with Mattys and Samuel's (2000) retroactive processing hypothesis, claiming extra processing time needed for non-initial stress words in English. Conversely, Japanese listeners exhibited faster recognition, initiating lexical access within the first syllable for both stress patterns in L2 English. Surprisingly, they showed equal efficiency in identifying target words for both patterns, possibly indicating enhanced proficiency in using lexical stress information. This is potentially attributed to positive transfer from their native language's pitch accent processing during spoken word recognition.

While this study emphasizes the advantages of L1 prosodic structures for L2 spoken word recognition, it does not suggest that Japanese listeners of English outperform English L1 listeners in overall English word recognition. Instead, it highlights the particular benefits that Japanese listeners of English may derive from their familiarity with L1 prosodic structures when identifying words in L2 English. Furthermore, the study's focus on trochaic and iambic stress patterns in English does not encompass the extensive variety of accent patterns present in the English, where a significant number of words feature reduced vowels in unstressed syllables. Future studies are expected to address the role of diversity of English stress patterns in L2 spoken word recognition.

Reference

Best, Catherine T. 1995. A direct realist perspective on cross-language speech perception. In Winifred Strange (ed.), *Speech perception and linguistic experience: Issues in cross-language research,* 171-204. Timonium, MD: York Press.

Bond, Zinny S. and Larry H. Small. 1983. Voicing, vowel, and stress mispronunciations in continuous speech. *Perception and Psychophysics* 34(5): 470-474.

Cooper, Nicole, Anne Cutler, and Roger Wales. 2002. Constraints of lexical stress on lexical access in English: Evidence from native and non-native listeners. *Language and Speech* 45(3): 207-228.

Creel, Sarah C., Michael K. Tanenhaus, and Richard N. Aslin. 2006. Consequences of lexical stress on learning an artificial lexicon. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 32(1): 15-32.

Cutler, Anne. 1986. Forbear is a homophone: Lexical prosody does not constrain lexical access. *Language and Speech* 29(3): 201-220.

Cutler, Anne and Charles Clifton. 1984. The use of prosodic information in word recognition. In Herman Bouma and Don G. Bouwhuis (eds.), *Attention and performance X*, 183-196. Hillsdale, NJ: Earlbuam.

Cutler, Anne and Takashi Otake. 1999. Pitch accent in spoken-word recognition in Japanese. *Journal of the Acoustical Society of America* 105(3): 1877-1888.

Delattre, Pierre. 1969. An acoustic and articulatory study of vowel reduction in four languages. *International Review of Applied Linguistics in Language Teaching* 7(4): 294-325.

Dupoux, Emmanuel, Kazuhiko Kakehi, Yuki Hirose, Christophe Pallier, and Jacques Mehler. 1999. Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance* 25(6): 1568-1578.

Flege, James Emil. 1987. The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics* 15(1): 47-65.

Flege, James Emil. 1995. Second-language speech learning: Theory, findings, and problems. In Winifred Strange (ed.), *Speech perception and linguistic experience*, 233-277. Timonium, MD: York Press.

Fletcher, Harvey and Wilden A. Munson. 1933. Loudness, its definition, measurement and calculation. *Bell System Technical Journal* 12(4): 377-430.

Fry, Dennis B. 1955. Duration and intensity as physical correlates of linguistic stress. *Journal of Acoustical Society of America* 27(4): 765-768.

Fry, Dennis B. 1958. Experiments in the perception of stress. *Language and Speech* 1(2): 126-152.

Gupta, Prahlad, John Lipinski, Brandon Abbs, Po-Han Lin, Emrah Aktunc, David Ludden, Nadine Martin, and Rochelle Newman. 2004. Space aliens and nonwords: Stimuli for investigating the learning of novel word-meaning pairs. *Behavior Research Methods, Instruments, and Computers* 36(4): 599-603.

Kuhl, Patricia K. and Paul Iverson. 1995. Linguistic experience and the perceptual magnet effect. In Winifred Strange (ed.), *Speech perception and linguistic experience*, 121-154. Timonium, MD: York Press.

Kusumoto, Yoko. 2012. Between perception and production: Is the ability to hear L1-L2 sound differences related to the ability to pronounce the same sounds accurately. *Journal of Polyglossia* 22: 15-33.

Lehiste, Ilse. 1976. Suprasegmental features in speech. In Norman J. Lass (ed.), *Contemporary issues in experimental phonetics*, 225-239. New York: Academic Press.

Lieberman, Philip. 1960. Some acoustic correlates of word stress in American English. *Journal of Acoustical Society of America* 32(4): 451-454.

Major, Roy C. 2008. Transfer in second language phonology. In Jette G. Hansen Edwards and Mary L. Zampini (eds.), *Phonology and second language acquisition*, 63-94. Amsterdam: John Benjamins Publishing Company.

Matin, Ethel, Kuo-Chih Shao, and Kenneth R. Boff. 1993. Saccadic overhead: Information processing time with and without saccades. *Perception and Psychophysics* 53(4): 371-380.

Mattys, Sven and Arthur G. Samuel. 2000. Implications of stress-pattern differences in spoken-word recognition. *Journal of Memory and Language* 42(4): 571-596.

Minematsu, Nobuaki and Keikichi Hirose. 1995. Roles of prosodic features in the human process of perceiving spoken words and sentences in Japanese. *Journal of Acoustical Society of Japan* 16(5): 311-320.

Munro, Murray J. and Ocke-Schwen Bohn. 2007. The study of second language speech learning: A brief overview. In Ocke-Schwen Bohn and Murray J. Munro (eds.), *Language experience in second language Speech Learning*, 3-12. Amsterdam: John Benjamins Publishing Company.

Robinson, Derek W. and R. So Dadson. 1956. A re-determination of the equal-loudness relations for pure tones. *British Journal of Applied Physics* 7(5): 166.

Sekiguchi, Takahiro and Yoshiaki Nakajima. 1999. The use of lexical prosody for lexical access of the Japanese language. *Journal of Psycholinguistic Research* 28(4): 329-454.

Sekiguchi, Takahiro. 2006. Effects of lexical prosody and word familiarity on lexical access of spoken Japanese words. *Journal of Psycholinguist Research* 35: 369-384.

Small, Larry H., Stephen D. Simon, and Jill S. Goldberg. 1988. Lexical stress and lexical access: Homographs versus non-homographs. *Perception and Psychophysics* 44(3): 272-280.

Slowiaczek, Louisa M. 1990. Effects of lexical stress in auditory word recognition. *Language and Speech* 33(1): 47-68.

Strange, Winifred, Reiko Akahane-Yamada, Rieko Kubo, Sonja A. Trent, and Kanae Nishi. 2001. Effects of consonantal context on perceptual assimilation of American English vowels by Japanese listeners. *Journal of the Acoustical Society of America* 109(4): 1691-1704.

Venditti, Jennifer J. 2005. The J_ToBI model of Japanese intonation. In Sun-Ah Jun (ed.), *Prosodic typology: The phonology and intonation of phrasing*, 172-200. Oxford: Oxford University Press.

Yamada, Reiko A. 1995. Age and acquisition of second language speech sounds: Perception of American English /r/ and /l/ by native speakers of Japanese. In Winifred Strange (ed.), *Speech perception and linguistic experience*, 305-320. Timonium, MD: York Press.

Appendix

Target Nonword Stimuli

	Trochaic Nonword	Iambic Nonword
1	/ˈpovɪsu/	/poˈvɪseɪ/
2	/ˈtimʌdo/	/tiˈmʌdɑ/
3	/ˈkʌnædʒɑu/	/kʌˈnædʒu/
4	/ˈsɪbɛtɑ/	/sɪˈbɛti/
5	/ˈbɑkeɪni/	/bɑˈkeɪnə-/
6	/ˈdɛsiʒɑɪ/	/dɛˈsiʒo/
7	/ˈgætɑvə/	/gæˈtɑvɑu/
8	/ˈmeɪgokeɪ/	/meɪˈgokɑ/
9	/ˈpɑneɪku/	/pɑˈneɪkɑ/
10	/ˈtɛbito/	/tɛˈbitu/
11	/ˈkævɑʒeɪ/	/kæˈvɑʒə/
12	/ˈseɪgodə-/	/seɪˈgodi/
13	/ˈbotɪnɑu/	/boˈtɪnei/
14	/ˈdikʌdʒi/	/diˈkʌdʒɑɪ/
15	/ˈgʌmæsɑɪ/	/gʌˈmæsɑu/
16	/ˈmɪsɛvɑ/	/mɪˈsɛvo/

Jeonghwa Shin
Associate Professor
Department of English
Korea Military Academy
574 Hwarang-ro, Nowon-gu,
Seoul 01805, Korea
E-mail: jshin1@kma.ac.kr

Received: 2023. 10. 18.
Revised: 2023. 11. 17.
Accepted: 2023. 11. 17.