



The role of acoustic cues in lexical stress recognition of real and nonce words by Korean learners of English*

Eunkyung Sung (Cyber Hankuk University of Foreign Studies)



This is an open-access article distributed under the terms of the Creative Commons License, which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received: November 23, 2024

Revised: February 2, 2025

Accepted: February 10, 2025

Sung, Eunkyung
Professor, Department of
English,
Cyber Hankuk University of
Foreign Studies
107 Imun-ro, Dongdaemun-gu,
Seoul, Korea
Tel: 02) 2173-8761
Email: eks@cufs.ac.kr

*This work was supported by the
Cyber Hankuk University of
Foreign Studies Research Fund
of 2024.

ABSTRACT

Sung, Eunkyung. 2025. The role of acoustic cues in lexical stress recognition of real and nonce words by Korean learners of English. *Korean Journal of English Language and Linguistics* 25, 247-263.

This study investigated how Korean learners of English perceived lexical stress, focusing on their reliance on acoustic cues such as duration, f0, and intensity when processing nonce words versus real words. Twenty-one Korean learners of English and 21 native English listeners participated in a perception experiment. The results showed that Korean listeners, whose L1 prosody utilizes f0 for phrasal prominence and segmental distinction in stop categories, tended to transfer this cue when perceiving English lexical stress. In contrast, English listeners primarily relied on duration cues. Furthermore, both listener groups demonstrated a stronger dependence on acoustic cues when processing nonce words compared to real words, indicating that familiarity with real words reduced reliance on detailed acoustic information. These findings indicate the impact of L1 prosody on L2 lexical stress perception and reveal differences in the processing of nonce and real words. This research contributes to the broader understanding of L2 prosodic perception, emphasizing the role of L1 transfer and lexical status.

KEYWORDS

lexical stress, acoustic cues, duration, f0, intensity, nonce words, real words, L2 perception

1. Introduction

Languages vary in their prosodic structure, with elements such as pitch, duration, and intensity being organized differently across languages. Lexical stress emphasizes certain syllables within words, which can change their meanings, as demonstrated in English and Spanish. For example, in English, the noun *record* [ˈrɛkərd] refers to “a stored piece of information”, whereas the verb *record* [rɪˈkɔrd] means “to document something.” Similarly, in Spanish, *papa* [ˈpapa] means “potato”, while *papá* [paˈpa] means “father.” In pitch accent languages, such as Japanese and Serbo-Croatian, pitch variations on specific syllables are used to differentiate meanings, rather than relying on changes in duration or intensity. For instance, in Japanese, the word *hashi* [haɕi] can mean “chopsticks” when its pitch accent is on the first syllable, but it means “bridge” when its pitch accent is on the second syllable. Meanwhile, tone languages like Mandarin Chinese and Thai use tonal distinctions as a primary means of distinguishing words. Mandarin, for example, employs four tones to alter the meaning of syllables like *ma*. Additionally, duration plays a crucial role in languages like Japanese and Finnish, where differences in vowel or consonant lengths can completely change meanings; in Japanese, [biru] refers to “a building,” while [bi:ru] denotes “beer.”

In English, lexical stress is primarily conveyed through a combination of acoustic cues such as pitch, duration, intensity, and vowel quality. Stressed syllables typically have a higher pitch, making them more prominent compared to unstressed syllables (Cutler 2005, Fry 1958, Liberman 1975). Additionally, stressed syllables tend to be longer in duration and produced with greater intensity, enhancing their perceptual salience (Beckman and Edwards 1994, Beckman and Pierrehumbert 1986). Another important cue is vowel reduction. While stressed syllables maintain a clear and full vowel quality, unstressed syllables often undergo vowel reduction, resulting in a more centralized and less distinct vowel sound (Beckman 1986). Furthermore, the relationship between pitch and lexical stress is not straightforward. English words can carry various intonational pitch accents such as H*, L+H*, L*, L*+H, which result in diverse pitch patterns. These pitch accents are influenced by the structure of the discourse and are realized based on the word’s position within phrases and tonal boundaries. For instance, in a statement with falling intonation, the stressed syllable may have a higher pitch than the unstressed one. In contrast, in a yes/no question with rising intonation, the stressed syllable may have a lower pitch than that of the unstressed one (Beckman 1986, Beckman and Pierrehumbert 1986). Research by Sluijter and van Heuven (1996) has shown that when perceiving word stress, pitch is a less reliable cue than other factors like duration, intensity, and vowel quality. This is because pitch is more variable and can change based on the overall intonation of a sentence rather than being consistently tied to lexical stress.

Unlike English, which uses distinct lexical stress patterns, Korean relies on prosodic structures at the phrase level. According to Jun (2005, 2014), Korean pitch patterns are associated with a prosodic unit called the Accentual Phrase (AP). An AP in Korean is a prosodic level lower than an Intonation Phrase (IP) and higher than a Phonological Word (PW), which consists of a lexical item followed by case markers or postpositions. In Korean, especially in the Seoul dialect, an AP is typically characterized by a default pitch contour of Low-High-Low-High (LHLH). That is, a key feature of the AP in Korean is a phrase-final high tone, which is marked by a rise in fundamental frequency (f₀). This increase in pitch at the end of a phrase helps native Korean speakers recognize word boundaries. Thus, pitch variations in Korean function at the phrasal level to signal prosodic boundaries rather than serving as a word-level stress marker.

The role of acoustic cues in lexical stress perception has yet to be explored in a study comparing real and nonce words. This study has two main objectives. First, it examines how listeners’ L1 influences their use of

three acoustic cues (i.e., duration, f_0 , intensity)¹ in perceiving stress. Second, it investigates whether the use of these cues differs between real words and nonce words. Perception experiments on word stress often focus on nonsense words (see Altmann 2006, Guion 2005 for reviews), based on the assumption that real words may fail to reveal phonological processes because their stress patterns might be stored in memory. This idea has been supported by several studies (Port 2007, Vitevitch et al. 1997). Vitevitch et al. (1997) investigated how phonotactic configurations and stress placement in nonsense words affect processing, suggesting that real words might not effectively uncover underlying phonological processes due to existing lexical representations. Similarly, Port (2007) argues that words are stored in memory with detailed auditory information, including stress patterns, which may influence phonological processing. Nevertheless, we contend that real words should also be examined. Investigating L2 lexical stress perception using both real and nonce words offers a more comprehensive understanding of phonological processing and the learning mechanisms underlying stress perception. Therefore, the present experiment includes both real and nonce words to determine whether acoustic cues in stress perception operate differently across word types.

2. Literature Review

Research on the role of acoustic cues in English lexical stress perception has examined learners from diverse native language backgrounds to understand how their L1 influences their ability to recognize English stress patterns. Studies have specifically focused on Korean (Chung 2013, Kang and Kim 2016, Kim and Tremblay 2021, 2022, Lee 2022), Japanese (Chung 2016, Sugahara 2016), and Chinese listeners (Chrabaszcz et al. 2014, Wang et al. 2024) to assess their sensitivity to different acoustic cues when identifying English word stress. These studies aimed to explore how each group's native prosodic system affects their perception of English stress.

Chung (2013) examined English stress perception among two groups of Korean EFL learners: (1) speakers of Standard Korean, a non-stress, non-tonal language, and (2) speakers of the North Gyungsang dialect, a tonal language. Both groups relied heavily on pitch prominence rather than other acoustic cues, such as duration and intensity, when identifying English lexical stress. However, the North Gyungsang group relied significantly more on pitch than the Seoul Korean group, reflecting their language-specific phonological use of pitch at the lexical level. These findings suggest that the degree of reliance on pitch may vary depending on the extent to which pitch prominence is used in a learner's native phonological system. Kang and Kim (2016) further investigated how Korean learners of English perceived English lexical stress and the specific acoustic cues they depended on. Their study examined the roles of pitch, intensity, and duration as stress correlates. Results showed that Korean listeners, unlike native English speakers, relied heavily on pitch to identify stressed syllables while showing limited sensitivity to intensity and duration. This reliance on pitch was attributed to L1 transfer, as Korean prosody primarily uses pitch for segmental contrasts and phrase-level prominence, rather than for lexical stress.

¹ This study excludes the vowel quality cue to focus on the independent effects of each acoustic cue and to more accurately analyze L1 transfer effects. Vowel quality often co-varies with duration (Beckman and Edwards 1994, Beckman and Pierrehumbert 1986, Gay 1978), making it difficult to isolate its influence. Furthermore, Priva and Strand (2023) found that vowels tended to shift to lower F_1 values when their duration is shorter, which further complicates the analysis of vowel quality as a separate cue.

Kim and Tremblay (2021) investigated how Korean listeners' sensitivity to prosodic cues in their L1 influenced their ability to perceive lexical stress in English. The study compared two groups of Korean speakers: Gyungsang Korean (GK) speakers, who use pitch accents to distinguish word meanings, and Seoul Korean (SK) speakers, whose language lacks lexical pitch accents. The results showed that GK listeners were significantly better at identifying English lexical stress than SK listeners. This advantage is attributed to the transfer of pitch sensitivity from their L1 to L2, as pitch serves as a prominent cue for stress in English. These findings support the Cue-Weighting Transfer Hypothesis, which proposes that listeners transfer prosodic cues from their L1 to L2 contexts, depending on the cue prominence in their native language. Kim and Tremblay (2022) extended this research by examining how Seoul Korean listeners, whose L1 relied on intonational cues to distinguish segmental contrasts, performed in identifying English lexical stress compared to French listeners. The results indicated that Seoul Korean listeners were more accurate at identifying English lexical stress than their French counterparts. This advantage is attributed to the transfer of cue-weighting strategies from Korean, where pitch plays a central role in segmental and intonational contrasts. In contrast, French listeners struggled with English stress perception because French prosody does not use pitch in a similar way for segmental distinctions². These findings further support the Cue-Weighting Theory, which suggests that listeners rely on the most salient acoustic cues in their L1 when learning L2 contrasts.

Lee (2022) explored how Korean learners of English perceived lexical stress and the specific acoustic cues they relied on. The study examined the role of vowel quality, pitch, intensity, and duration in stress perception, comparing responses between Korean learners of English and native English speakers. The results revealed that both groups relied heavily on vowel quality to identify stressed syllables. However, Korean learners showed a greater reliance on pitch and intensity than on duration, unlike native English speakers, who prioritized duration as a key cue. This pattern reflects the influence of L1 prosodic transfer from Korean, where pitch and intensity play more prominent roles in higher-level prosody, such as intonation and phrase-level prominence, compared to duration.

Japanese is a pitch-accent language, meaning that the pitch pattern of a word (the rise and fall in f_0) can affect its meaning. Sugahara (2016) examined how Japanese listeners perceived English lexical stress compared to native English and Korean listeners. The results showed that both Japanese and Korean listeners had difficulty in correctly identifying stress patterns when pitch cues were ambiguous. In contrast, English listeners were less affected due to their reliance on additional cues like duration and vowel quality. Similarly, Chung (2016) investigated the role of pitch in perceiving English lexical stress by Japanese speakers. The study revealed that Japanese listeners were highly accurate in identifying stress when the stressed syllable had a higher pitch than the unstressed one. However, their performance dropped significantly when the stressed syllable had a lower pitch, indicating a strong reliance on pitch cues due to the prosodic structure of their native language. Comparisons with Korean speakers showed that Japanese listeners' dependency on pitch was higher than that of Seoul Korean speakers but lower than that of North Gyungsang Korean speakers. Since North Gyungsang Korean utilizes f_0 to distinguish lexical meaning, similar to Mandarin Chinese, speakers of this dialect demonstrated an even greater reliance on f_0 cues when perceiving English stress than Japanese speakers.

Mandarin Chinese is a tone language, where pitch contours are essential for distinguishing words. Chrabaszcz et al. (2014) explored how native speakers of English, Mandarin, and Russian used acoustic cues—vowel quality, pitch, duration, and intensity—to perceive word stress in disyllabic nonwords. The study found that all groups

² In Seoul Korean, the VOT of lenis and aspirated stops has gradually merged over time. As a result, this contrast is now primarily signaled by f_0 of the following vowel (Kang 2014, Silva 2006). In contrast, French primarily relies on VOT as the main cue for distinguishing stop consonants, with F_0 playing only a marginal role in this contrast (Serniclaes 1987).

relied most on vowel quality. English and Mandarin speakers utilized pitch as the second most important cue, whereas Russian speakers focused more on duration and intensity while largely ignoring pitch. Interestingly, despite English and Russian both being stress languages, Mandarin listeners exhibited stress perception patterns more similar to English speakers. The findings suggest that L1 prosodic features influence stress perception in L2, and that even with notable differences in language structures, similar perception patterns may still arise. Wang et al. (2024) investigated how native English and Mandarin Chinese listeners perceived English lexical stress, focusing on their use of acoustic cues such as pitch and vowel quality. The results revealed that native English listeners primarily relied on vowel quality to identify stressed syllables, while Mandarin listeners, whose native language is tonal and distinguishes lexical meanings through pitch variation, utilized both vowel quality and pitch to a similar extent. These findings suggest that Mandarin speakers incorporated pitch as an additional cue in English stress perception, likely due to their sensitivity to pitch variations in their L1.

Overall, research on the perception of English lexical stress by speakers of various native languages demonstrates that an individual's L1 prosodic system plays a significant role in shaping their L2 stress perception. Languages such as Korean, Japanese, and Chinese utilize pitch for different prosodic functions, which influences learners from these backgrounds to rely more heavily on pitch cues when perceiving English lexical stress compared to native English speakers. The present study aims to confirm the findings of previous studies and to determine whether the effects of L1 transfer are seen in real words as well as nonce words.

Although recent studies have investigated the role of acoustic cues in L2 lexical stress perception, most have focused exclusively on either real words or nonce words in their experiments. While some research has examined lexical stress perception using both types of stimuli, few studies have specifically addressed the influence of acoustic cues across different lexical contexts. This gap highlights the need for further studies to explore how listeners process acoustic cues in both familiar (real words) and unfamiliar (nonce words) contexts to better understand the interaction between lexical knowledge and acoustic sensitivity in L2 stress perception.

Studies that have examined speech perception patterns while comparing nonce words to real words include the following. Caspers (2009) investigated how native Dutch speakers and L2 learners with various L1 backgrounds perceived Dutch word stress by utilizing both real and nonsense Dutch words. The results revealed that L2 learners with fixed stress patterns in their L1 (e.g., Polish and Finnish) faced greater difficulty in perceiving Dutch stress compared to learners with no fixed stress patterns in their L1 (e.g., French and Chinese). Additionally, while both native speakers and L2 learners performed slightly better with real words than nonsense words, the difference was not substantial.

Other studies have also investigated the role of word familiarity by including both real and nonce words to explore segmental contrast perception in L2 learners (Sawada 2019, Yun 2014). Yun (2014) conducted identification tasks in which Korean listeners were asked to perceive English phoneme contrasts using both real and nonce words. The findings indicated that identification accuracy and reaction times differed between the two word types, suggesting that word familiarity influenced L2 phoneme perception. Specifically, participants were more accurate and faster in identifying phoneme contrasts when using real words compared to nonce words. Similarly, Sawada (2019) found that learners were faster at recognizing highly familiar words than less familiar words. The study concluded that familiarity, rather than frequency, was a better predictor of L2 speech processing. This suggests that learners rely more on internalized knowledge derived from exposure rather than on frequency counts observed in native language texts.

Previous studies suggest that familiarity with real words generally improves L2 learners' perception of phonemes, words, and lexical stress, leading to higher accuracy and faster processing compared to nonce words. However, these findings may not fully extend to the present study, which focuses on how acoustic cues such as

duration, pitch, and intensity influence L2 learners' perception of English lexical stress. This study specifically examines the role of L1 prosody in shaping L2 stress perception and investigates whether listeners' reliance on acoustic cues differs between real words and nonce words. By analyzing the interaction between lexical status and acoustic sensitivity, this research aims to provide deeper insights into how L2 learners process stress patterns in varying lexical contexts.

3. Method

3.1 Participants

Twenty-one Korean learners of English (4 males and 17 females) and 21 native English speakers (8 males and 13 females) participated in a perception experiment. All of the Korean learners were from Seoul or Gyeonggi Province and were standard Korean speakers. The Korean learners were undergraduate or graduate students at a university in Seoul. Their majors ranged from humanities, social sciences, and design, but none were English-related. None of the learners had lived in an English-speaking country for more than six months, and they had been learning English for a minimum of six years and a maximum of 17 years. In the questionnaire about their background, two of the Korean learners of English reported their TOEIC scores as being around 700. The other Korean learners of English stated that they did not have official scores for their English proficiency. Based on their responses to the questionnaires and recordings of short English paragraphs, their English proficiency was determined to be low-intermediate or low.

Native speakers of English were graduate students, university professors, or English instructors. The participants were from the United States, Australia, Canada, and the United Kingdom. Native speakers had lived in Korea for varying lengths of time, from 3 months to 14 years, and had a low level of Korean proficiency based on the questionnaires they filled out. All participants from the two groups reported no hearing difficulties and were compensated for their participation in the experiment.

3.2 Materials

3.2.1 Nonce words

In this study, we created stimuli by adapting the experimental methods of Chrabaszcz et al. (2014) and Oh (2011). The stimuli consisted of modified natural recordings of two disyllabic nonce words, each with identical vowels in both syllables (i.e., /kata/ and /kətə/). These words were spoken by a male native English speaker from California who had phonetic training, and the recordings were made in a soundproof studio using an Olympus LS-P4 recorder. Each word was recorded five times at a normal speech rate using the carrier sentence "I say _____" to ensure consistency in speech rate during recording. To measure and manipulate acoustic cues (i.e., duration, f₀, intensity), we used *Praat* (v. 6.1.03) (Boersma and Weenink 2019). Details on the measurement of these acoustic cues for two recorded vowels are provided below.

Table 1. Acoustic Cue values for Vowels /a/ and /ə/

	duration (ms)	f0 (ST/Hz)	intensity (dB)	F1 (Hz)	F2 (Hz)	F3 (Hz)
/a/	347	86/143.6	70	722	1266	2396
/ə/	340	86/146.29	70	603	1585	2311

The two target vowels were manipulated to create three sets of stimuli, each contrasting in duration, f0, and intensity. For the duration manipulation, the values were set to 173.5ms, 347ms, and 520.5ms for /a/ and 170ms, 340ms, and 510ms for /ə/. Regarding pitch, the f0 values were adjusted by ± 5 semitones (ST) from the original vowels, resulting in pitch levels of 81ST, 86ST, and 91ST for both /a/ and /ə/. Although pitch is traditionally measured in Hz, semitones are considered more perceptually relevant units (Behrman 2007)³. For intensity, three levels were set at 0.5 times (-6.02dB), 1 time (baseline, 70dB), and 1.5 times (+6.02dB) of the original vowel intensity, resulting in values of 63.98dB, 70dB, and 76.02dB. All final stimuli were synthesized using the Pitch Synchronous Overlap and Add (PSOLA) method in *Praat*. The specific adjustment values for /kata/ are described below.

Table 2. Manipulation of Acoustic Cue Values in Nonce Words

cue	Steps	1	2	3
	Duration (ms)		173.5	347
f0 (ST)		81	86	91
Intensity (dB)		63.98	70	76.02

The experimental stimuli in this study consisted of 36 target tokens, created by combining three steps of duration, pitch, and intensity cues across two syllables and two words [(3 duration steps + 3 pitch steps + 3 intensity steps) x 2 syllables x 2 words]. Additionally, 36 filler tokens were included. Each token was repeated twice, resulting in a total of 144 tokens. The filler tokens were generated using the base words /subu/ and /tipi/. For the training session, 18 practice tokens were created using the base word /gɛdɛ/, following the same manipulation process.

3.2.2 Real words

The real word stimuli consisted of three minimal pairs where word stress shifts depending on whether the word is used as a noun or a verb. The selected pairs were: *Object* - *obJECT*, *CONsole* - *conSOLE*, and *CONverse* - *conVERSE*.⁴ These words were recorded following the same process used for the nonce stimuli. Subsequently, the acoustic features of each vowel in these word pairs were measured as follows.

³ *Praat* was used to convert the frequency unit Hz into semitones (ST).

⁴ Analyzing the frequency of each word by part of speech based on the COCA (Corpus of Contemporary American English), it was found that object (V) occurs 37,387 times and object (N) occurs 71,566 times. The frequency of console (V) is 9,459 times, console (N) is 46,747 times, and the frequency of converse (V) and converse (N) is 623 and 2,954 times, respectively.

Table 3. Acoustic Cues Values for Real Words: *Object, Console, Converse*

		duration (ms)	f0 (ST)	intensity (dB)	F1 (Hz)	F2 (Hz)	F3 (Hz)
OBject	/a/	171	85	61.44	1031.04	1461.92	2465.54
	/i/	147	81	56.57	678.84	177.42	2553.9
obJECT	/ə/	89	83	56.69	1353.07	2361.24	3001.37
	/ɛ/	183	86	61.15	882.06	1704.47	2485.7
CONsole	/a/	134	87	61.86	928.56	1172.12	1172.12
	/o/	117	81	56.13	481.3	909.91	909.91
conSOLE	/a/	49	86	56.01	675.89	1829.2	1829.2
	/o/	104	89	60.72	569.3	1002.31	1002.31
CONverse	/a/	122	87	62.4	953.81	1269.92	2428.11
	/ə/	115	79	56.34	461.7	1496.15	1808.96
conVERSE	/ə/	62	87	61.15	258.36	1572.94	2440.24
	/ɜ/	123	88	63.21	722.05	1470.62	2077.87

The three acoustic cues of real words were manipulated at three steps, following a similar method used for the nonce words. In this case, we adjusted the acoustic cues based on whether the first or second syllable was stressed. Specifically, if the first syllable was stressed, as in *OBject*, the second syllable was left unchanged, while the acoustic cues of the first syllable *OB* were adjusted across three steps. Conversely, if the second syllable was stressed, as in *obJECT*, the first syllable remained untouched, and the acoustic cues of the second syllable *JECT* were adjusted across three steps. This manipulation values for *Object* and *obJECT* are summarized in Table 4.

Table 4. Manipulation of Acoustic Cues for the Real Word *Object*

cue \ Steps	OB			ject	JECT			
	1	2	3	-	-	1	2	3
Duration (ms)	85.5	171	256.5	147	89	91.5	183	274.5
f0 (ST)	80	85	90	81	83	81	86	91
Intensity (dB)	55.42	61.44	67.46	56.57	56.69	55.13	61.15	67.17

The experimental stimuli for real words included 54 real words, which were designed using a combination of 3 steps of duration, pitch, and intensity across 2 syllables and 3 different words [(3 steps of duration + 3 steps of pitch + 3 steps of stress) x 2 syllables x 3 words]. In addition, 36 filler words were included. Each of these stimuli was repeated twice, resulting in a total of 180 experimental tokens.

3.3 Procedure

To investigate the effects of three acoustic cues on the perception of lexical stress in English, a forced-choice auditory identification task was conducted using *Psychopy* version 3.1 (Peirce 2007). The experiment consisted of two parts: a nonce word perception task and a real word perception task. The stimuli for both tasks were divided into two blocks and presented in a randomized order. Before starting the experiment, participants were informed about the procedure both verbally and through a consent form, followed by a questionnaire to collect background information. Additionally, a practice session with 36 tokens was conducted to familiarize

participants with the task. These tokens included some acoustic manipulations but were not used in the main experiment.

The main experiment was conducted in a quiet laboratory setting. Participants sat in front of a computer, where they viewed words and numbers on the screen while listening to auditory stimuli through a headset. Using a mouse, they clicked on the numbers to indicate their responses. The test used a 5-point Likert scale with responses ranging from 1 (indicating strong stress on the first syllable) to 5 (indicating strong stress on the second syllable). Buttons 1 and 5 represented stronger confidence levels, with button 1 indicating higher confidence in trochaic stress than button 2, and button 5 indicating higher confidence in iambic stress than button 4. If participants perceived equal stress on both syllables, they could select button 3. The nonce word experiment lasted approximately 15 minutes, while the real word experiment took about 20 minutes, for a total duration of 35 minutes. Below is an example of the experimental screens presented to participants.

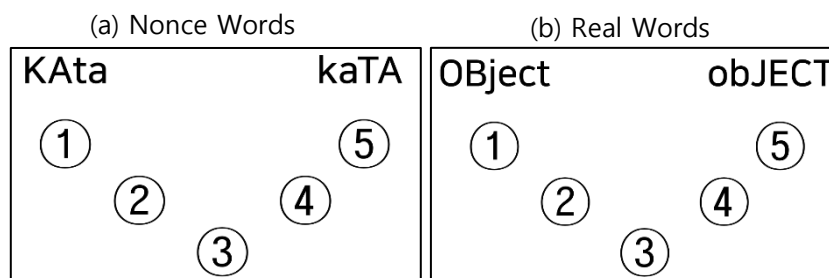


Figure 1. Screen Layouts for the Identification Experiment

3.4 Data Analysis

Each token in the nonce word stimuli was manipulated in three ways: (1) stress on the first syllable (strong-weak, or trochaic), (2) non-contrastive stress across syllables (equal stress or no clear distinction), and (3) stress on the second syllable (weak-strong, or iambic). In each token, one acoustic cue—duration, pitch, or intensity—was adjusted to indicate stress, while the other two cues remained noncontrastive. The real-word stimuli were manipulated in a similar manner, but with less distinct differences between the acoustic values of the first and second syllables compared to nonce words. For nonce words, identical vowels were used in both syllables to maintain uniformity prior to manipulation. In contrast, for real words, only the stressed syllable's acoustic cues were adjusted, leaving the unstressed syllables unchanged from the naturally recorded forms.

To assess how listeners perceived stress based on these manipulated cues, the percentages of three response types—trochaic, noncontrastive, and iambic—were calculated for each cue within each listener group. Responses were coded as trochaic if participants selected options 1 or 2 on the screen, as noncontrastive when selecting 3, and as iambic when choosing 4 or 5. Figures 2 and 3 illustrate how the two listener groups (Korean and English) responded to variations in duration, f_0 , and intensity.

4. Results

4.1 Nonce Words

The following boxplots illustrate the percentages of trochaic, noncontrastive, and iambic responses by two listener groups across variations in duration, f0, and intensity cues in nonce words. The x-axis represents acoustic cue levels (1, 2, and 3), where level 1 corresponds to trochaic stress, level 2 to noncontrastive stress, and level 3 to iambic stress patterns. The y-axis shows the percentage of each response type for each listener group across the three acoustic cues.

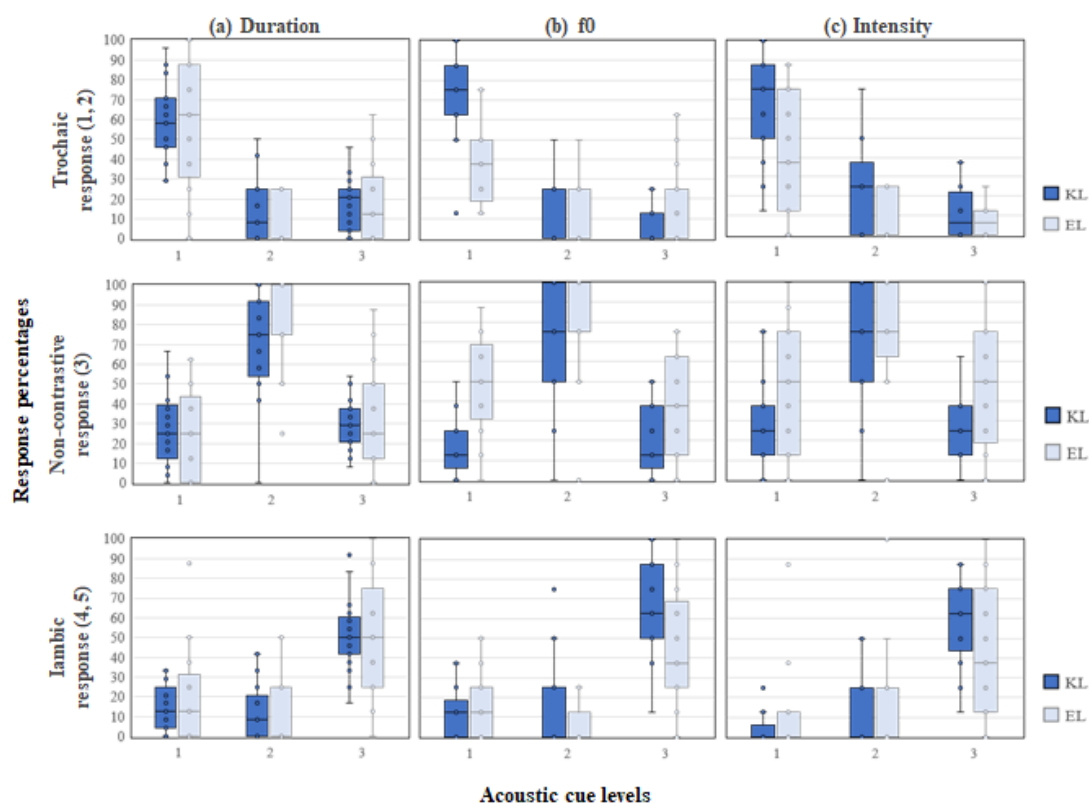


Figure 2. Percentages of Trochaic, Noncontrastive, and Iambic Responses by Korean and English Listeners across Duration, F0, and Intensity Cue Levels in Nonce Words

Figure 2 illustrates distinct patterns in how Korean listeners (KL) and English listeners (EL) perceived stress patterns across three acoustic cues. Both listener groups showed similar sensitivity to duration variations, though English listeners were slightly more sensitive than Korean listeners. However, Korean listeners exhibited greater sensitivity to f0 and intensity cues compared to English listeners. Specifically, Korean listeners' trochaic responses showed a steeper decline as f0 and intensity cues transitioned from level 1 to 2. Additionally, Korean listeners' iambic responses increased more steeply from cue level 2 to 3 than those of English listeners. These findings indicate that Korean listeners relied more on f0 and intensity cues than English listeners. Moreover,

both groups showed a peak in noncontrastive responses at cue level 2 across all acoustic cues, suggesting that they perceived balanced cue values as neutral or ambiguous.

To investigate the effects of these acoustic cues, listener groups, and their interactions, a multinomial logistic regression model was conducted using R (R Core Team 2024). This model was chosen because the dependent variable had three categories (trochaic, noncontrastive, and iambic), and it accounted for individual variability as a random effect. The independent variables included two listener groups (Korean and English) and three levels for each acoustic cue (duration, f0, intensity). The dependent variable consisted of three listener responses: trochaic, noncontrastive, and iambic. The model tested both main effects and interactions of group and cue. In this analysis, each response category was compared to a reference category. In the present study, the noncontrastive response served as the reference for interpreting results. The results of these statistical analyses are summarized in Table 5.

Table 5. The Effects of Listener Group and Acoustic Cues in Nonce Words

Variable	Trochaic response				Iambic response			
	Estimate	Std.Error	z value	p-value	Estimate	Std.Error	z value	p-value
(Intercept)	-2.108	0.290	-7.264	<.001***	-5.334	0.397	-13.443	<.001***
L1	0.644	0.411	1.568	0.117	0.632	0.549	1.152	0.249
Duration	0.212	0.132	1.611	0.107	0.035	0.151	0.231	0.818
f0	1.189	0.156	7.627	<0.001***	2.505	0.195	12.835	<0.001***
Intensity	0.996	0.157	6.359	<0.001***	2.441	0.203	12.054	<0.001***
L1 x Duration	0.875	0.201	4.355	<0.001***	1.739	0.238	7.306	<0.001***
L1 x f0	-0.779	0.212	-3.684	<0.001***	-1.234	0.257	-4.805	<0.001***
L1 x Intensity	-0.201	0.216	-0.929	0.353	-0.815	0.275	-2.969	<0.01**

The statistical analysis of nonce words revealed that f0 and intensity cues had a significant impact on trochaic responses ($p < 0.001$ for both). While duration alone did not significantly influence trochaic responses, its effect varied depending on the listener group (L1), indicating a significant interaction between duration and L1. Additionally, a significant interaction was observed between L1 and f0, suggesting that the influence of f0 on trochaic perception differed between Korean and English listeners. As illustrated in Figure 2, English listeners were more sensitive to variations in duration, whereas Korean listeners relied more on f0 and intensity cues. However, the interaction between L1 and intensity was not statistically significant.

Similarly, iambic responses were strongly influenced by f0 and intensity, both showing highly significant effects ($p < 0.001$). Duration alone did not significantly predict iambic responses unless considered in combination with listener group characteristics. Further interaction analyses revealed that L1 moderated the effects of duration, f0, and intensity, suggesting that lexical stress perception is shaped not only by acoustic properties but also by linguistic experience. Overall, these findings highlight the crucial role of f0 and intensity as primary cues for stress perception, while listener group differences influenced these patterns through interactions, particularly with duration and f0.

4.2 Real Words

The following boxplots display the percentages of trochaic, noncontrastive, and iambic responses by the two listener groups across variations in duration, f0, and intensity cues in real words. The x-axis displays the acoustic cue levels (1, 2, and 3), with level 1 indicating trochaic stress, level 2 representing noncontrastive stress, and level 3 corresponding to iambic stress patterns. The y-axis illustrates the percentage of each response type for

both listener groups across the three acoustic cues.

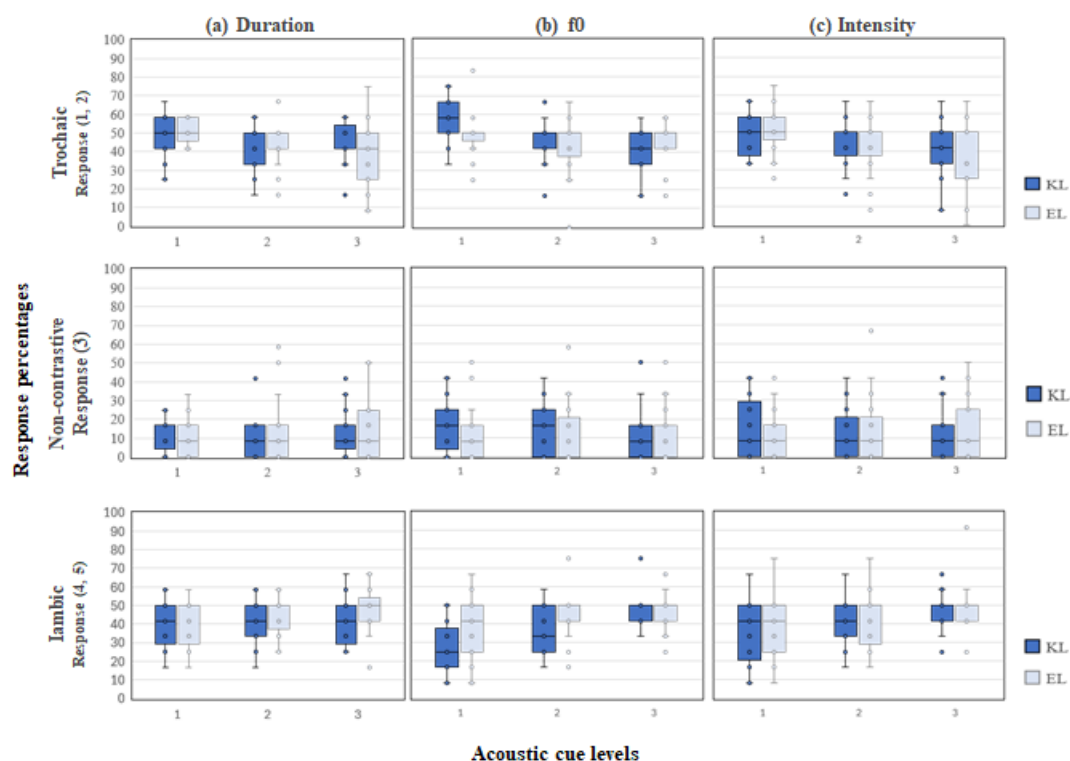


Figure 3. Percentages of Three Responses (Trochaic, Noncontrastive, Iambic) by Korean and English Listeners as a Function of Duration, F0, and Intensity in Real Words

The boxplots in Figure 3 illustrate the effects of duration, f0, and intensity on the perception of trochaic, noncontrastive, and iambic responses for Korean listeners (KL) and English listeners (EL). For trochaic responses, variations in duration had a stronger impact on English listeners compared to Korean listeners. English listeners' responses showed a slight decline from cue level 2 to 3, whereas Korean listeners exhibited a more consistent pattern. Conversely, f0 had a greater influence on Korean listeners, as their responses declined more steeply from cue level 1 to 2. Both groups showed a relatively similar trend in response to intensity, with only slight decreases observed from cue level 1 to 3.

Noncontrastive responses were significantly lower across all three acoustic cues and for both listener groups compared to trochaic and iambic responses. This pattern is likely due to differences in stimulus manipulation between nonce and real words. In nonce words, identical vowels were presented when the acoustic cue was set to level 2, whereas levels 1 and 3 involved manipulating either the first or second vowel. In contrast, real-word stimuli involved naturally recorded speech with only the stressed vowel modified, while the unstressed vowel remained unchanged. Consequently, even at cue level 2, the acoustic properties of the two vowels in real words were not entirely uniform, potentially leading to a reduced occurrence of noncontrastive responses.

For iambic responses, differences between the two groups were also apparent. English listeners demonstrated a slight increase in responses as the duration cue progressed from levels 2 to 3, whereas Korean listeners'

responses remained relatively stable across all three levels. Similar to trochaic responses, this suggests that duration exerted a stronger influence on English listeners' perception of lexical stress. However, Korean listeners were more responsive to variations in f0, showing greater sensitivity to this cue compared to English listeners. Both groups exhibited a comparable pattern in response to intensity, with only a slight rise observed from cue level 1 to 3.

Overall, the results for real words suggest that duration, pitch, and intensity all play a role in stress perception, but none of the cues dominated. Korean listeners appeared more attuned to f0 for both trochaic and iambic patterns, while English listeners were more sensitive to duration for these response types. Additionally, the influence of acoustic cues was less pronounced in real words compared to nonce words, indicating that listeners rely more on acoustic cues in unfamiliar lexical contexts. Constraints associated with manipulating acoustic properties in real words may have also contributed to the reduced sensitivity observed in familiar lexical contexts. The table below displays the results of statistical analyses for real words. A multinomial logistic regression model was applied, using the noncontrastive response as the reference category for comparison and interpretation.

Table 6. The Effects of Listener Group and Acoustic Cues in Real Words

Variable	Trochaic response				Iambic response			
	Estimate	Std.Error	z value	p-value	Estimate	Std.Error	z value	p-value
(Intercept)	-1.682	0.438	-3.841	<0.001***	-1.009	0.184	-5.496	<0.001***
L1	-1.119	0.667	-1.679	0.093	0.403	0.261	1.548	0.122
Duration	0.070	0.141	0.495	0.620	0.078	0.095	0.818	0.413
f0	0.039	0.141	0.277	0.782	0.485	0.098	4.943	<0.001***
Intensity	-0.033	0.144	-0.227	0.820	0.285	0.095	2.988	<0.01**
L1 x Duration	0.363	0.213	1.701	0.089	0.141	0.136	1.040	0.298
L1 x f0	-0.131	0.211	-0.620	0.535	-0.334	0.136	-2.448	<0.05*
L1 x Intensity	0.301	0.218	1.381	0.167	-0.071	0.135	-0.526	0.599

For trochaic responses involving real words, none of the acoustic cues showed statistically significant effects on stress perception, as indicated by their p-values ($p > 0.05$ for all three cues). However, there was a marginally significant interaction between L1 (listener group) and duration ($p = 0.089$), suggesting that English and Korean listeners may differ in their sensitivity to duration cues, with English listeners potentially relying more on duration. Other interactions, including L1 x f0 and L1 x intensity, did not yield significant results, indicating no strong differences between listener groups for these cues in trochaic responses.

For iambic responses, both f0 ($p < 0.001$) and intensity ($p < 0.01$) had statistically significant effects, indicating that these cues played an important role in stress perception. English and Korean listeners exhibited differences in sensitivity to these cues. Specifically, the interaction between L1 and f0 was significant ($p < 0.05$), suggesting that Korean listeners were more influenced by f0 than English listeners. In contrast, the interactions involving L1 x duration and L1 x intensity were not significant, implying no substantial group differences for these cues. Overall, the results emphasize that f0 and intensity cues were more influential in iambic responses, particularly for Korean listeners.

When comparing nonce words and real words, the results indicate that listeners were more sensitive to acoustic cues in nonce words than in real words. In nonce words, f0 and intensity had strong effects on stress perception ($p < 0.001$) in both trochaic and iambic responses. In contrast, real words showed weaker effects overall. Only f0 ($p < 0.001$) and intensity ($p < 0.01$) were significant in iambic responses, while duration remained non-significant in both response types.

These findings suggest that in the absence of lexical familiarity, listeners relied more heavily on acoustic information. Conversely, in real words, listeners likely drew on lexical knowledge, reducing their dependence on acoustic cues. Furthermore, although the influence of acoustic cues was generally stronger for nonce words than for real words in both groups, consistent patterns were evident regardless of lexical status. These findings suggest that Korean and English listeners processed stress cues differently, reflecting differences in their phonological systems and highlighting cross-linguistic variations in stress perception strategies.

5. Discussion and Conclusion

The present study has two primary objectives. First, it investigates the influence of L1 prosody on the perception of lexical stress in an L2 context. Second, it examines whether listeners' reliance on acoustic cues differs when processing real words versus nonce words.

The results of this study revealed clear differences in how Korean and English listeners relied on acoustic cues, such as duration, f_0 , and intensity, to perceive lexical stress. These findings align with previous research indicating that L1 prosodic patterns significantly influence L2 stress perception. For nonce words, f_0 and intensity were the strong cues in processing lexical stress. There were strong interaction effects between language and duration, language and f_0 , and language and intensity. For Korean listeners, f_0 and intensity were the primary cues used to perceive stress, while duration had little impact on their perception of stress in nonce words. In contrast, English listeners relied more on duration than Korean listeners.

This reliance on pitch and intensity is consistent with studies by Chung (2013), Kang and Kim (2016), and Kim and Tremblay (2021, 2022). These studies showed that Korean learners of English tended to transfer their L1 prosodic strategies to English, using pitch as a primary cue due to its role in marking prosodic boundaries in Korean. Similarly, Lee (2022) also found that Korean listeners utilized pitch and intensity more than duration, indicating that they could attend to pitch and intensity cues in L2 due to their experience with these cues in L1. In Korean, f_0 plays a crucial role at both segmental and post-lexical levels. At the segmental level, f_0 helps distinguish stop categories, particularly when the voice onset time (VOT) values overlap for word-initial stops. Specifically, vowels following lax stops exhibit significantly lower f_0 values compared to those following aspirated stops (Kang 2014, Silva 2006). Additionally, the Korean prosodic structure relies on phrasal tone patterns, where prominence is marked at the edges of prosodic units rather than at the word level (Jun 2005, 2014).

The findings of this study support the Cue-Weighting Transfer Hypothesis proposed by Kim and Tremblay (2021, 2022), indicating that listeners apply perceptual weighting strategies from their L1 to L2 contexts. According to Kim and Tremblay (2021, 2022), the process of interpreting lexical stress in L2 is unlikely to depend solely on whether lexical stress distinctions exist in the L1. Instead, prosodic cues used for processing one phonological feature in L1 can be transferred to the interpretation of a different phonological feature in L2. Sugahara (2016) and Wang et al. (2024) also showed that Japanese and Mandarin Chinese speakers similarly transferred their native prosodic strategies when processing L2 lexical stress.

Unlike Korean listeners, English listeners primarily relied on duration to perceive stress, with f_0 being the least significant cue. Strong interaction effects were observed between language and duration, language and f_0 , and language and intensity in iambic responses of nonce words. However, the interaction effect between language and intensity was not observed in trochaic responses. This finding aligns with previous studies (Beckman and Edwards 1994, Beckman and Pierrehumbert 1986, Roach 2009, Sluijter and van Heuven 1996,

Yu and Andruski 2010). Sluijter and van Heuven (1996) demonstrated that English stress patterns were primarily marked by duration and intensity, rather than pitch. Yu and Andruski (2010) also identified duration as the most consistent and reliable cue for native English listeners when perceiving stress, while intensity played a comparatively minor role. This preference for duration is further supported by the close relationship between stress placement and syllable weight in English, where syllables with long vowels often receive primary stress (Roach 2009).

The present study examined how acoustic cues influenced the perception of lexical stress in both nonce words and real words. The results showed that nonce words elicited a stronger reliance on acoustic cues than real words for both Korean and English listeners. For nonce words, significant effects of f_0 and intensity were observed, along with interaction effects between language and each of the three cues (i.e., f_0 , duration, and intensity) in iambic responses. In trochaic responses, similar patterns were observed, but the interaction effect between language and intensity was not observed.

In contrast, the influence of acoustic cues in real words was generally weaker. Only in iambic responses f_0 and intensity showed strong effects, and a marginally significant interaction effect between language and f_0 appeared. There were no acoustic cue or interaction effects in trochaic responses. These findings suggest that both Korean and English listeners relied less on purely acoustic information when processing real words. This implies that lexical familiarity can reduce the dependence on bottom-up acoustic cues for both L1 and L2 listeners. For real words, pre-existing lexical knowledge may enable listeners to apply top-down processing strategies. This finding aligns with previous research (Port 2007, Vitevitch et al. 1997), which suggests that real words may not effectively reveal underlying phonological processes because they are stored in memory with detailed auditory information, including stress patterns. Additionally, the present study identified limitations in manipulating acoustic cues in real words compared to nonce words, as efforts were made to ensure that the real-word stimuli sounded as natural as possible. Vowel quality may have also influenced listeners' responses, particularly since reduced vowels were used for unstressed syllables. The real-word stimuli in this experiment were heteronyms, whose pronunciation varies depending on their function as nouns or verbs due to differences in stress placement. Participants may have interpreted these words based on the form that was more familiar to them, potentially influencing their responses.

In sum, the current results align with previous studies, confirming that the perception of lexical stress in L2 English is strongly influenced by the prosodic features of listeners' L1. Specifically, Korean listeners, whose native language relies on pitch to indicate phrasal prominence and segmental distinctions of stop categories, consistently relied on these same cues when perceiving English lexical stress. In contrast, English listeners primarily depended on duration cues. Furthermore, the comparison between nonce words and real words revealed that listeners relied more heavily on acoustic cues when processing nonce words than real words. This study offers valuable insights into the interaction between L1 prosody and L2 stress perception.

These findings have practical implications for L2 pronunciation teaching, particularly for Korean learners of English. Emphasizing duration as a stress marker in instructional materials and training could help reduce L1 transfer effects. Activities that focus on lengthening stressed syllables or contrasting short and long vowel sounds could enhance learners' awareness of English stress patterns. Additionally, the present findings suggest the importance of lexical context in stress perception. Instructors should incorporate both familiar and unfamiliar words in stress training exercises. Using nonce words can help learners focus on acoustic cues without interference from existing lexical knowledge. Once learners develop sensitivity to these cues, real words can be gradually introduced to reinforce stress patterns in meaningful contexts.

The Korean participants in the present study were all learners of English with low-intermediate or low proficiency levels. Future research should examine how learners of different proficiency levels use acoustic cues to recognize lexical stress in both nonce and real words. Future studies should also consider vowel quality cues, which were not addressed in the present study. Moreover, participants in this study were more sensitive to acoustic cues in iambic responses than in trochaic responses. This raises the question of whether Korean learners of English exhibit a general preference for iambic stress. Future research is essential to determine whether such a bias is pervasive and to explore its implications for L2 learning.

References

- Altmann, H. 2006. *The Perception and Production of Second Language Stress: A Cross-linguistic Experimental Study*. Doctoral dissertation, University of Delaware.
- Beckman, M. and J. Edwards. 1994. Articulatory evidence for differentiating stress categories. *Papers in Laboratory Phonology III: Phonological Structure and Phonetic Form*.
- Beckman, M. E. 1986. *Stress and Non-stress Accent*. Foris Publications.
- Beckman, M. E. and J. B. Pierrehumbert. 1986. Intonational structure in Japanese and English. *Phonology* 3, 255-309.
- Behrman, A. 2007. *Speech and Voice Science*. Plural Publishing, Inc.
- Boersma, P. and D. Weenink. 2019. Praat: Doing phonetics by computer (version 6.1.03) [Computer program]. Available online at www.praat.org
- Caspers, J. 2009. The perception of word stress in existing and non-existing Dutch words by native speakers and second language learners. *Linguistics in the Netherlands* 26(1), 25-38.
- Chrabaszcz, A., M. Winn, C. Y. Lin, and W. J. Idsardi. 2014. Acoustic cues to perception of word stress by English, Mandarin, and Russian speakers. *Journal of Speech, Language, and Hearing Research* 57(4), 1468-1479.
- Chung, Y. H. 2013. Effects of L1 prosodic system on the perception of English stress: a case study of Korean EFL learners. *Language Research* 49(2), 383-398.
- Chung, Y. H. 2016. Effects of pitch on the perception of English stress by Japanese speakers. *Korean Journal of Linguistics* 41(2), 341-358.
- Cutler, A. 2005. Lexical stress. In D. B. Pisoni and R. E. Remez, eds., *The Handbook of Speech Perception*, 264-289, Blackwell.
- Fry, D. B. 1958. Experiments in the perception of stress. *Language and Speech* 1, 126-152.
- Gay, T. 1978. Physiological and acoustic correlates of perceived stress. *Language and Speech* 21, 347-353.
- Guion, S. G. 2005. Knowledge of English word stress patterns in early and late Korean-English bilinguals. *Studies in Second Language Acquisition* 27(4), 503-533.
- Jun, S. A. 2005. Korean intonational phonology and prosodic transcription. In S-A Jun, ed., *Prosodic Typology: The Phonology of Intonation and Phrasing*, 201-229. Oxford: Oxford University Press.
- Jun, S. A. 2014. Prosodic typology: by prominence type, word prosody, and macro-rhythm. In S-A. Jun, ed., *Prosodic Typology II: The Phonology of Intonation and Phrasing*, 520-540. Oxford: Oxford University Press.
- Kang, H. and H. J. Kim. 2016. Acoustic correlates of English stress in Korean L2 learners' perception. *Korean Journal of English Language and Linguistics* 16(2), 169-196.

- Kang, Y. 2014. Voice Onset Time merger and development of tonal contrast in Seoul Korean stops: A corpus study. *Journal of Phonetics* 45, 76-90.
- Kim, H. and A. Tremblay. 2021. Korean listeners' processing of suprasegmental lexical contrasts in Korean and English: A cue-based transfer approach. *Journal of Phonetics* 87, 1-15.
- Kim, H. and A. Tremblay. 2022. Intonational cues to segmental contrasts in the native language facilitate the processing of intonational cues to lexical stress in the second language. *Frontiers in Communication* 7, 845430.
- Lee, G. 2022. Perceptual weighting on English lexical stress by Korean learners of English. *Phonetics and Speech Sciences* 14(4), 19-24.
- Liberman, M. Y. 1975. The intonational system of English. Doctoral dissertation, Massachusetts Institute of Technology.
- Oh, J. 2011. A study on prosodic cue determining syntactic moods in Korean. *Korean Linguistics* 51, 117-140.
- Peirce, J. W. 2007. PsychoPy - psychophysics software in Python. *Journal of Neuroscience Methods* 62(1-2), 8-13.
- Port, R. F. 2007. How are words stored in memory? Beyond phones and phonemes. *New Ideas in Psychology*, 25(2), 143-170.
- Priva, U. and E. Strand. 2023. Schwa's duration and acoustic position in American English. *Journal of Phonetics* 96, 101198.
- R Core Team. 2024. R: A language and environment for statistical computing (version 4.3.1) [Computer software]. Available online at <https://www.R-project.org>
- Roach, P. 2009. *English Phonetics and Phonology Fourth Edition: A Practical Course*. 4th edition. Cambridge: Cambridge University Press.
- Sawada, E. 2019. The effect of word familiarity and lexical neighbourhood on L2 speech perception. *ICU Working Papers in Linguistics (ICUWPL)* 6, 49-57.
- Serniclaes, W. 1987. *Etude expérimentale de la perception du trait de voisement des occlusives du français*. Doctoral dissertation, Université Libre de Bruxelles.
- Silva, D. J. 2006. Acoustic evidence for the emergence of tonal contrast in contemporary Korean. *Phonology* 23.2, 287-308.
- Sluijter, A. and V. Heuven. 1996. Spectral balance as an acoustic correlate of linguistic stress. *The Journal of the Acoustical Society of America* 100, 2471-2485.
- Sugahara, M. 2016. Is Japanese listeners' perception of English stress influenced by the antepenultimate accent in Japanese? Comparison with English and Korean listeners. *Doshisha Studies in English* 96, 61-111.
- Vitevitch, M. S., P. A. Luce, J. Charles-Luce, and J. D. Kemmerer. 1997. Phonotactics and Syllable Stress: Implications for the Processing of Spoken Nonsense Words. *Language and Speech* 40(1), 47-62.
- Wang, F., D. Deng, K. Tang, and R. Wayland. 2024. The Effect of pitch accent on the perception of English lexical stress: Evidence from English and Mandarin Chinese listeners. *Languages* 9(3), 87.
- Yu, V. Y. and J. E. Andruski. 2010. A cross-language study of perception of lexical stress in English. *Journal of Psycholinguistic Research* 39, 323-344.
- Yun, G. 2014. Korean listeners' perception of L2 English phoneme contrast. *Studies in Phonetics, Phonology, and Morphology* 20(2), 161-185.

Examples in: English

Applicable Languages: English

Applicable Level: Tertiary