Korean Journal of English Language and Linguistics, Vol 22, February 2022, pp. 153-168
DOI: 10.15738/kjell.22..202202.153

KOREAN JOURNAL OF ENGLISH LANGUAGE AND LINGUISTICS

ISSN: 1598-1398 / e-ISSN 2586-7474

http://journal.kasell.or.kr

Influence of Acoustic Cues on the Perception of Lexical Stress by Korean Learners of English and English Listeners*

Eunkyung Sung (Cyber Hankuk University of Foreign Studies)

Received: Jan. 10, 2022
Revised: Feb. 21, 2022
Accepted: Feb. 27, 2022

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 2021.

ABSTRACT

Sung, Eunkyung. 2021. Influence of acoustic cues on the perception of lexical stress by Korean learners of English and English listeners. *Korean Journal of English Language and Linguistics* 21, 153-168.

The purpose of the present study was to examine how listeners' native language affected the use of four different acoustic cues (duration, F0, intensity, and vowel quality) in identifying lexical stress in English. Native Korean and English listeners judged lexical stress placement in nonce words in which the first or the second syllable was varied systematically regarding the four acoustic cues. The results of proportions of trochaic responses and statistical analyses showed that all four cues were important predictors of the listeners' stress perception. However, the two listener groups performed differently when considering response patterns for four acoustic cues separately. Specifically, the Korean listeners were significantly more sensitive to F0 and intensity cues than the English listeners. There were no statistically significant differences between the two listener groups in terms of duration and vowel quality cues. Also, there were significant effects of interaction between acoustic cues and listener groups for duration and F0 cues. The Korean listeners relied more on F0 and less on duration than the English listeners. These findings imply that L2 listeners' perception of lexical stress is not entirely predicted by L1 prosody. Furthermore, for both groups of listeners the cue shift from iambic to noncontrastive stress patterns induced less changes in trochaic stress responses than that from noncontrastive to trochaic stress patterns. This preference for trochaic stress perception over iambic stress needs to be further investigated in the future study.

KEYWORDS

acoustic cues, duration, F0, intensity, vowel quality, lexical stress, perception, Korean listeners, English listeners

1. Introduction

Most languages have prosodic features such as word stress, intonation and tones. Languages also have different properties of lexical stress. In English and Spanish, stress is contrastive, and some words can vary in only the location of stress (e.g., éxport: noun, expórt: verb). Some languages such as French, Finnish, and Hungarian, however, do not contrast lexical items by stress patterns. For example, in French the lexical stress is positionally fixed, and the final non-schwa vowels are always stressed. It has also been known that nonnative speakers of English often revealed problems with vowel reduction related to production of unstressed syllables (Flege and Bohn 1989, Fokes and Bond 1989, Hammond 1986, Zhang et al. 2008, Zhang and Francis 2010).

Second language (L2) learners of English often experience difficulties in the perception of English lexical stress. These perceptual difficulties have traditionally been attributed to the prosodic structure of their first language (L1). Thus, French speakers showed more difficulty in discriminating words differing in stress pattern compared with English or Spanish speakers (Dupoux et al. 2001, Dupoux et al. 2008).

Furthermore, not only the presence or absence of contrastive stress but also acoustic cues associated with prosodic prominence affect L2 stress processing. Compared with the effects of acoustic cues on the production of L2 lexical stress, less has been known about how acoustic cues to contrast prosodic properties in L1 affect the perception of L2 stress. The effects of acoustic cues such as F0, duration and intensity on English stress perception have been investigated involving listener groups from different native languages. Montero (2007) compared the effects of acoustic cues on stress perception by three language groups (native French, Spanish, and English listeners). Frost (2011) focused on the relative importance of four acoustic cues by native English and French listeners. Chrabaszcz et al. (2017) examined how L1 affects listeners' weighting of acoustic cues in the perception of contrastive word stress involving native listeners of English, Russian and Mandarin. Wang (2008) demonstrated the use of acoustic cues in stress perception by Chinese learners of English. Zhang and Francis (2010) explored the function of vowel quality in native Mandarin speakers' perception of English lexical stress. Meng et al. (2020) compared perception patterns of different acoustic cues in English lexical stress between Mandarin and Cantonese speakers who learn English as a foreign language. Tremblay et al. (2021) investigated whether acoustic cue weightings were transferred from the native language to the second language involving Dutch L2 learners of English and native English listeners. A few studies have focused on Korean listeners' perception of acoustic cues in English lexical stress (Chung 2013, Kang and Kim 2016, Kim and Tremblay 2021). However, these studies did not examine all of the four acoustic cues (duration, F0, intensity, and vowel quality). The present study was conducted to compare the relative roles of the four acoustic cues in perceiving lexical stress between Korean learners of English and English listeners.

English lexical stress is displayed with multiple acoustic cues such as pitch (F0), duration, intensity, and vowel quality (Beckman and Pierrehumbert 1986, Fry 1958, Liberman 1975, Roach 2009). Compared with unstressed syllables, stressed syllables are higher in pitch, longer in duration, and stronger in intensity. Also, full (nonreduced) vowel quality is associated with stressed syllables. Whereas stressed vowels are produced in a more peripheral position, unstressed vowels are more centralized in acoustic F1–F2 space and reduced, resulting in a less distinct schwa-like vowel. In addition, 70% of English disyllabic content words have trochaic stress patterns (Cutler and Carter 1987).

Korean, however, does not have lexically contrastive stress (Jun 2005, 2014). Korean also does not have fixed stress at the word level like French or Finnish. According to Guion (2005) and Jun (2014), the domain of prosodic association for English stress is the lexical word whereas Korean pitch patterns are associated with the accentual phrase (AP). The default pitch contour of the AP in the Seoul Korean dialect is Low-High-Low-High

(LHLH). Thus, in Korean the pitch change does not signal any lexical shift and is associated with phrasal tones. On the contrary, in English pitch is one of the acoustic cues for lexical stress.

The purpose of this study was to investigate how listeners' native language affects the weighting of acoustic cues in the perception of contrastive lexical stress. Four acoustic cues (duration, F0, intensity and vowel quality) were manipulated for the judgment of English nonce word stress. Also, Korean listeners' response patterns in identifying lexical stress were compared with those of English listeners.

2. Perceptual Cues to English Lexical Stress

The presence of multiple cues in English lexical stress has led to little consensus regarding perceptual prominence. Sluijter and van Heuven (1996) indicated that F0 was the least reliable acoustic cue compared with duration, intensity, or vowel quality for the perception of English word stress. Okobi (2006) suggested that syllable duration was a strong cue for stress perception. It has also been known that intensity appears to be less effective in indicating stress than duration and pitch differences (Mattys 2000, Morton and Jassem 1965, van Heuven and Menert 1996). In terms of vowel quality, Fry (1965) noted that the vowel quality was outweighed by fundamental frequency, duration, and intensity differences. Conversely, Beckman (1986) pointed out that vowel quality was at least a stronger cue than intensity.

Only a few studies have explored cross-language perceptual patterns of multiple acoustic cues in lexical stress (Chrabaszcz et al. 2017, Chung 2013, Kang and Kim 2016, Kim and Tremblay 2021, Tremblay et al. 2021, Zhang and Francis 2010). Chrabaszcz et al. (2017) examined how listeners' native language affected their weighting of acoustics cues (vowel quality, pitch, duration, and intensity) when listeners perceived English word stress. In total, 45 native listeners of English, Russian, and Mandarin performed a stress identification task on nonce disyllabic words with fully crossed combinations of each of the four cues in both syllables. The results showed that the vowel quality was the strongest cue for all listener groups. Whereas pitch was the second cue for the English and Mandarin listeners, duration and intensity cues were employed by the Russian listeners to a greater extent compared with the English and Mandarin participants. In addition, cues were stronger in the iambic stress pattern than in the trochaic pattern. Although both English and Russian are stress languages, and Mandarin is a tonal language, the cue weighting patterns of the Mandarin listeners, but not the Russian listeners, is more similar to those of the English listeners. The authors indicated that in spite of prominent differences in language typology, similar patterns of perception could arise.

Zhang and Francis (2010) examined the function of vowel quality in native Mandarin speakers' perception of English lexical stress. The participants judged the placement of lexical stress when they heard synthesized tokens of the word *desert*, in which the first syllable *de-* was varied systematically in vowel quality and each of the other cues depending on the pair of cues in focus. The results showed that both English and Mandarin listeners weighted vowel quality more than the other cues, and vowel quality and duration were used as a combinational cue. However, different processing patterns were found between two listener groups. Mandarin listeners, but not English listeners, were affected by pitch contour conditions when they processed vowel quality and pitch cues. Zhang and Francis (2010) argued that these results could be explained by language-specific or cue-specific influence.

Chung (2013) and Kang and Kim (2016) examined the effect of pitch in Korean speakers' perception of English stress. In Chung (2013), only pitch was manipulated for the perception experiment. The results showed that the Korean speakers revealed difficulty in judging the stress pattern when stressed syllables had low pitch.

Chung (2013) suggested that pitch plays an important role in stress perception. However, these results cannot be generalized since other acoustic cues were not considered in the study. Kang and Kim (2016) examined the acoustic correlates of English stress in Korean L2 learners' perception. English nonce words were manipulated in terms of the three acoustic cues (i.e., duration, F0, and intensity). The results showed that F0 was the most reliable cue, which was followed by intensity. Duration did not affect the Korean speakers' perception of English word stress. The authors mentioned that the Korean speakers' strong reliance on pitch in detecting English lexical stress could be attributed to the use of pitch for the phonological and syntactic processes in Korean. Although there is no strong prosodic feature in the Korean lexicon, pitch still plays an important role in Korean phonology, particularly at the phrase and sentence levels (Jun 2014). In their study the effect of vowel quality was not explored.

Kim and Tremblay (2021) investigated whether listeners' use of acoustic cues to L2 lexical stress was transferred from phonological features of their native language. Gyeongsang-Korean (GK) and Seoul-Korean (SK) listeners as well as native English listeners processed the pitch accent contrast and the phonemic contrast. The two Korean listener groups were chosen since GK has lexical pitch accents although neither GK nor SK has lexical stress. The stimuli were manipulated such that stress would be represented by all acoustic cues (natural condition) or by f0 alone (f0-only condition). The results of processing Korean words showed GK listeners were more sensitive to pitch accent contrasts than SK listeners. Furthermore, the GK listeners outperformed the SK listeners in the processing of stress contrasts in L2. Kim and Tremblay (2021) suggested that L2 learners could transfer the use of a suprasegmental cue from a different phonological category (i.e., lexical pitch accents) to lexical stress in their L2.

Although acoustic cues used for English lexical stress are well documented, the role of acoustic cues in the perception of lexical stress has not been thoroughly investigated, specifically in L2 perception. Furthermore, the direct comparisons of lexical stress perception patterns between native Korean and English listeners regarding four acoustic cues are scarce. The present study examined how four acoustic cues are used differently in perceiving lexical stress between English and Korean listeners. English is a stress language, and these four acoustic cues are used to implement lexical stress, whereas Korean does not have lexical stress. This study attempted to address the following questions: How do native Korean listeners and English listeners weight each of the four acoustic cues in comparison to other cues to stress perception? Are there any differences between native Korean and English listeners with regard to each of the four acoustic cues in perceiving English lexical stress involving nonce words?

3. Methods

3.1 Participants

The native Korean listener group consisted of twenty one adults (four male and seventeen female) who were recruited from a university in Seoul. The participants' ages ranged from 20 to 35 years old, with the average being 27 years old. None of the Korean listeners were from the Gyeongsang area since the Gyeongsang Korean dialect has lexical pitch accents, with f0 contours being an important cue to lexical contrast. The Korean listeners were undergraduate or graduate students, and their majors were not related to English literature or English education. None of them had stayed in English-speaking countries for more than six months. Their English

proficiency levels were considered to be low-intermediate or low, based on a self-report and a voice recording of a short English paragraph.

The native English listener group was comprised of twenty one adults (eight male and thirteen female), who were undergraduate students, professors or English instructors, with the exception of one participant, who was a jazz pianist. Their ages ranged from 21 to 54 years old, with the average being 39 years old. The participants were from the United States, Australia, Canada, Ireland, and the United Kingdom. They had stayed in Korea from three months to 18 years (the average being 6 years). Their Korean proficiency levels were considered to be low based on a self-report, with the exception of one participant, who had a score of 4 on the TOPIK (Test of Proficiency in Korean), whose scale ranges from 1 (lowest) to 6 (highest). All of the Korean and English subjects were paid for their participation in an identification experiment. None of them reported any hearing problems, and all participants answered a questionnaire regarding their language background.

3.2 Materials and Design

This study follows the methods shown in Oh (2011) and Chrabaszcz et al. (2017). In Chrabaszcz et al. (2017) the fully crossed combinations of each of the four cues were used. In their study, two or more cues were manipulated and combined for each token, and many tokens had conflicting cues. In the present study, however, for each stimulus token one of the three cues (duration, F0, intensity) was trochaic (strong - weak), iambic (weak – strong), or noncontrastive, and the other two cues were noncontrastive. Vowel quality cues were combined with one of the three cues. For example, if pitch and vowel quality displayed trochaic stress, duration and intensity were noncontrastive. There were also tokens in which all four cues were noncontrastive.

The stimuli in the present study were modified natural recordings of the disyllabic nonce words. The nonce words with four vowel quality combinations (i.e., /kɑtɑ/, /kɑtə/, /kətɑ/, /kətə/) were produced by a phonetically trained male native speaker of American English from California. The words were produced in a carrier sentence 'I say _____.' five times and recorded using an Olympus LS-P4 in a soundproof studio. These nonce words were selected because they are phonotactically permissible in both English and Korean phonology although the vowels in the recorded tokens were qualitatively different than the similar vowels in Korean.

Next, the recordings were examined, and a set of two vowels (/ɑ/ and /ə/) that were acoustically most similar to each other were chosen. Acoustic cues of the two vowels were further modified to make them have very similar acoustic cue values regarding duration, F0 and intensity, and still sounded natural. As a result, acoustic parameters of the two basic vowels in each token have the same values except duration (duration /ɑ/ 347ms, /ə/ 340ms; F0 /ɑ, ə/ 86 ST (semitone); intensity /ɑ, ə/ 70Hz). The following table shows the values of the acoustic cues for these two basic vowels.

Table 1. Values of Acoustic Cues for the Two Vowels /ɑ/ and /ə/

	Duration (ms)	F0 (ST)	Intensity (dB)	F1 (Hz)	F2 (Hz)	F3 (Hz)
/ɑ/	347	86	70	722	1266	2396
/ə/	340	86	70	603	1585	2311

These two basic vowels were embedded into the four vowel quality combinations (i.e., /kɑtɑ/, /kɑtə/, /kətɑ/, /kətə/) by copying and pasting the vowels appropriately. These revised four tokens were manipulated to create three stimulus sets contrasting duration, F0 and intensity. All of the manipulations were made using Praat version 6.1.03 (Boersm and Weenink 2019).

The manipulation methods of duration, F0 and intensity cues were partially adopted from Oh (2011). The three steps of duration were set to be 0.5, 1, and 1.5 times of the two basic vowels. Thus, the duration of three stimulus set were 173.5, 347, 520.5ms for /ɑ/, and 170, 340, 510ms for /ə/. For F0 manipulation, the differences of F0 values were set to be ±5ST of the basic vowels.[1] The F0 values of three stimulus set were 81, 86, 91ST for both /ɑ/ and /ə/. In addition, the three steps of intensity were set to be 0.5 (-6.02dB), 1, and 1.5 (+6.02dB) of the basic vowels, resulting in 63.98, 70, 76.02dB. All final sound stimuli were the result of pitch synchronous overlap and add (PSOLA) resynthesis method in Praat. The following figure shows the basic token and the modified tokens of the vowel /ɑ/ in terms of duration, F0 and intensity.

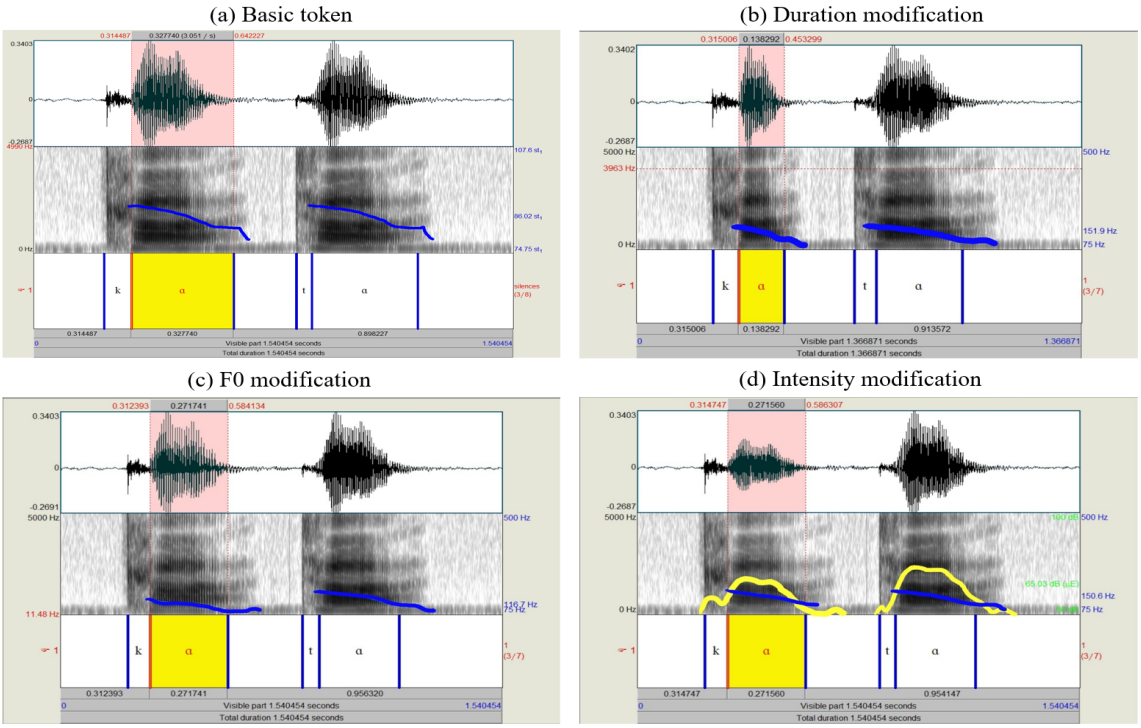

Figure 1. The Basic Token of /kɑtɑ/ (a), and the Modified Ones When the Acoustic Cue Value of the First Syllable /kɑ/ is Decreased in Duration (b), F0 (c), and Intensity (d)

The following table shows the acoustic cue values manipulated based on three levels. Level 1 indicates the acoustic cue values of the second syllable are stronger than those of the first syllable by either decreasing the cue values of the first syllable or increasing the cue values of the second syllable. Level 2 represents noncontrastive cue values between the two syllables. Level 3 manifests that that the acoustic cue values of the first syllable are stronger than those of the second syllable by either increasing the cue values of the first syllable or decreasing the cue values of the second syllable.

[1] Although a general unit used for pitch is Hz, semitone (ST) is considered a more important unit for pitch perception (Hart et al. 1990, Oh 2011).

Table 2. Three Levels of Acoustic Cue Values

Level	Vowel (ə ɑ)	Duration (ms)	F0 (semitone)	Intensity (dB)
	V1/V2	V1/V2	V1/V2	V1/V2
Level 1 (V1>V2)	/ɑ, ə/	/ɑ/ 520.5/347 347/173.5 /ə/ 510/340 340/170	/ɑ, ə/ 91/86 86/81	/ɑ, ə / 76.02/70 70/63.98
Level 2 (V1=V2)	ɑ/ɑ ə/ə	/ɑ/ 347/347 /ə/ 340/340	/ɑ, ə/ 86/86	/ɑ, ə/ 70/70
Level 3 (V1<V2)	/ə, ɑ/	/ɑ/ 347/520.5 173.5/347 /ə/ 340/510 170/340	/ɑ, ə/ 86/91 81/ 86	/ɑ, ə/ 70/76.02 63.98/70

For each of the four vowel quality combinations (/kɑtɑ/, /kɑtə/, /kətɑ/, /kətə/), 18 tokens were created [(duration 3 steps + pitch 3 steps + intensity 3 steps) x 2 syllables]. Therefore, the four basic tokens yielded a total of 72 tokens (18 x 4). In addition, filler tokens with the same manipulation process were included in the stimuli. The basic filler tokens were /subu/ and /tipi/, and 36 tokens were made based on these two tokens [(duration 3 steps + pitch 3 steps + intensity 3 steps) x 2 syllables x 2 words]. All of the experimental and filler tokens were repeated twice. Thus, in total, 144 experimental tokens [(duration 3 steps + pitch 3 steps + intensity 3 steps) x 2 syllables x 4 words x 2 repetitions] and 72 filler tokens (36 fillers x 2 repetitions) were used during the identification experiment. Furthermore, for a training session, 18 practice tokens were created with the basic token /gɛdɛ/ through the same manipulation process [(duration 3 steps + pitch 3 steps + intensity 3 steps) x 2 syllables].

3.3 Procedure

In order to test the effects of four acoustic cues on the perception of English lexical stress, the participants completed a forced-choice auditory identification task implemented through the PsychoPy program version 3.1.1 (Peirce 2007). The participants were tested with 216 tokens split into two blocks. The stimulus tokens in each block were randomly presented to each participant.

Before the main experiment began, the participants signed the consent form and filled out the questionnaire. Then they were given auditory instructions for the experiment. The participants also had a training session with 36 practice tokens (18 tokens x 2 repetitions) to familiarize them with the procedure and to make sure they understood the instructions correctly. The practice tokens were not used in the main task, and no feedback was given during the training session.

The participants heard a stimulus and were asked to click the corresponding buttons on a 5-point scale. Thus, with one press of a button, the participants could select the location of stress (first or second syllable stress) in disyllabic nonce words and the degree of confidence in their choice. The participants were given the option to select button 3 when they thought the two syllables in a word were equally stressed. The average running time for the main experiment was 15 minutes. The following figure manifests a computer screen with the auditory stimulus /kɑtɑ/.

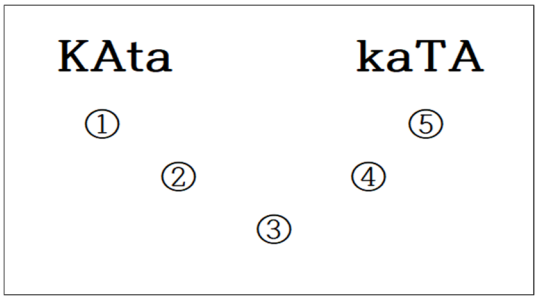

Figure 2. Screen Layout for the Identification Experiment

4. Results

Acoustic cue manipulations yielded three levels. Level 1 indicates stress on the first syllable (strong-weak: trochaic), level 2 represents noncontrastive stress across syllables, and level 3 denotes stress on the second syllable (weak-strong: iambic). Acoustic cues in stimulus tokens were coded such that each cue was either trochaic (1) or iambic (-1). A zero value was assigned to the cue that was noncontrastive, in which both syllables are stressed or both are unstressed. Thus, each of the four cues (duration, pitch, intensity, vowel quality) was coded as three values (1, 0, -1). Therefore, when pitch displayed a trochaic stress (level 1), and duration, intensity, and vowel quality were noncontrastive (level 2), those tokens were coded as 1 (0+1+0+0=1). On the contrary, when pitch displayed an iambic stress (level 3) and the other three cues were noncontrastive, those tokens were coded as -1 (0-1+0+0=-1). As previously mentioned, for each stimulus token one of the three cues (duration, F0, intensity) was trochaic, iambic or noncontrastive, and the other two cues are all noncontrastive. Vowel quality cues were represented in the four basic tokens (/kɑtɑ/, /kɑtə/, /kətɑ/, /kətə/), with /kɑtə/ coded as 1, /kɑtɑ/ and /kətə/ coded as 2, and /kətɑ/ coded as 3.

The listeners' responses of 1 or 2 on a computer screen were coded as trochaic and the other responses (3, 4, 5) were coded as non-trochaic. First, the effectiveness of each cue on listeners' stress perception was examined for each group separately. As can be seen in Figure 3, the x-axis manifests the sum of positive, negative, or neutral values of each acoustic cue in relation to trochaic stress. A positive value indicates that cues go in the trochaic direction whereas a negative value means that cues go in the iambic direction.

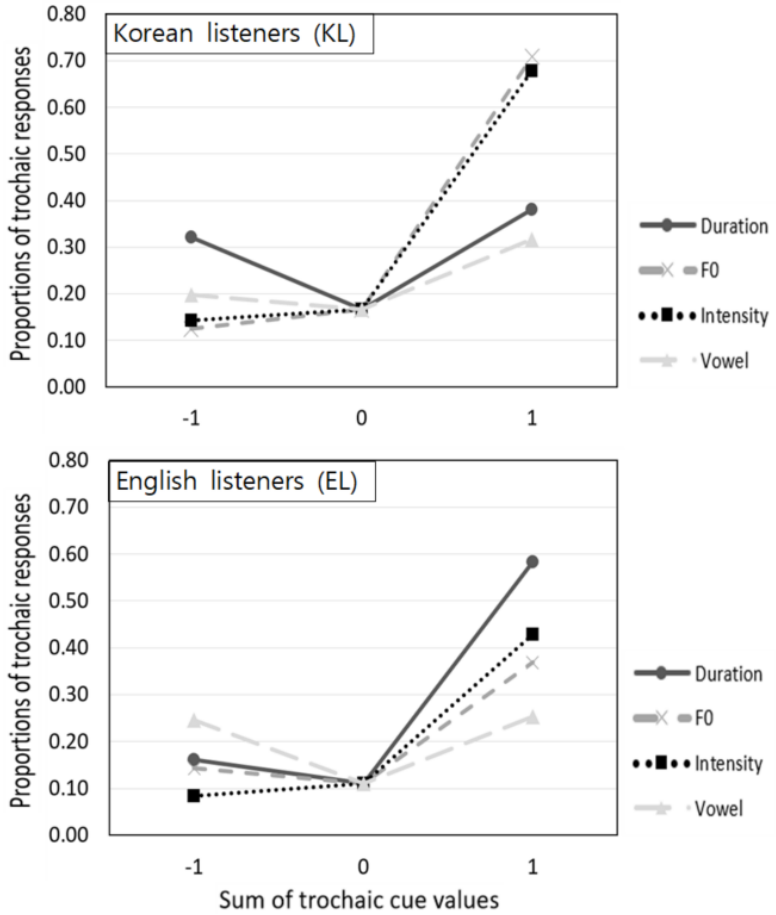

Figure 3. Proportions of Trochaic Responses (Responses of 1 or 2) for Native Korean and English Listeners as a Function of Duration, F0, Intensity, and Vowel Quality Manipulation

Figure 3 suggests a relationship between the cue values for trochaic stress and the corresponding responses of trochaic perception. In general, the listeners in the two language groups showed an increasing tendency to perceive stress as trochaic with higher values of cues. However, the two listener groups performed differently when considering response patterns for four acoustic cues separately. The Korean listeners were sensitive to F0 and intensity more than the other two cues. As the values of F0 and intensity cues in the first syllable increased, the more trochaic responses the Korean listeners showed. The similar pattern appeared in the results of vowel quality, but the slope was not as steep as those of pitch and intensity. On the other hand, the duration cue did not reveal consistent patterns. For the duration cue, the Korean listeners showed more trochaic stress responses when they heard the tokens with -1 value than when they heard the tokens with 0 value.

The English listeners were more sensitive to duration than to the other three cues. Similar response patterns were shown for the other cues, but the slopes for the other cues were more gradual than that for the duration cue. The Korean listeners responded that they heard trochaic stress when the stimuli had 1 value of F0 and intensity around 70% of the time. However, the English listeners' trochaic responses were around 40% with the same stimuli. In terms of duration cues, the opposite patterns were shown. That is, when the stimuli had 1 value of duration, the English listeners' trochaic responses were around 60% whereas those of the Korean listeners were less than 40%. The following figure shows the comparisons of the two listener groups in relation to each acoustic cue.

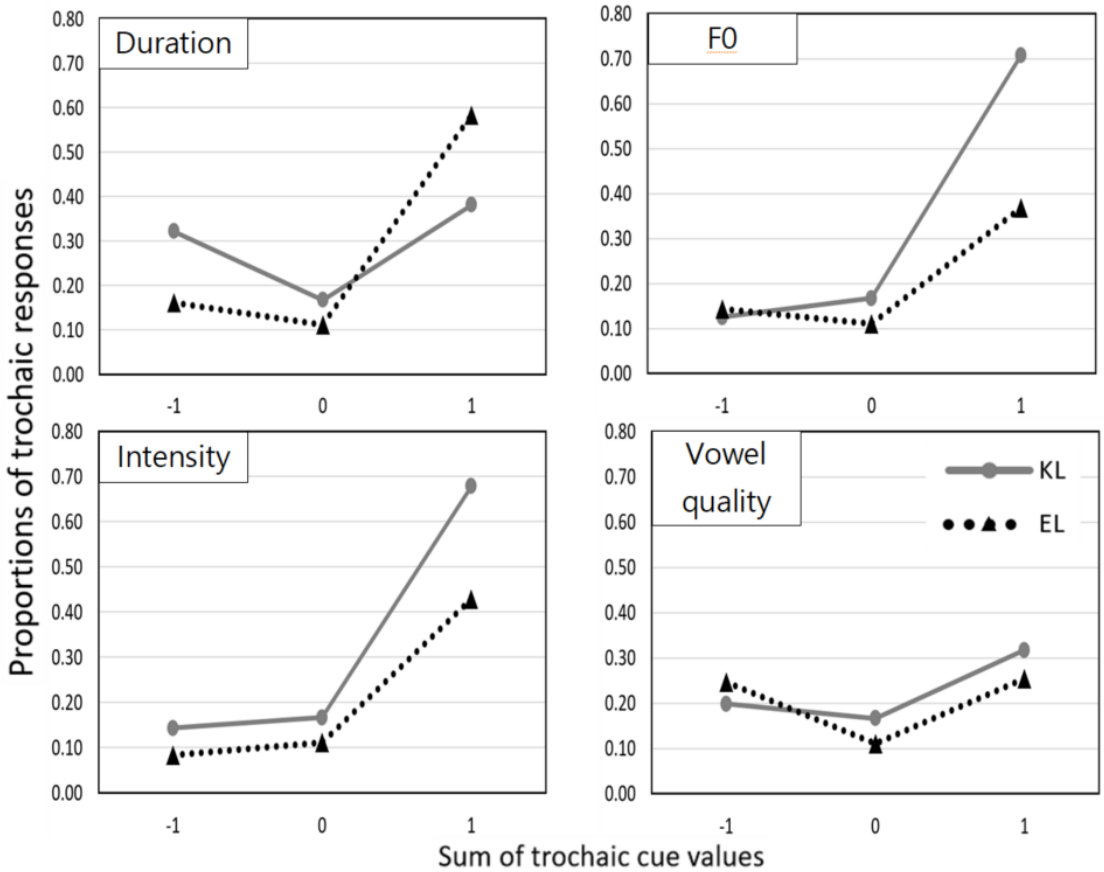

Figure 4. Comparisons of the Two Listener Groups' Proportions of Trochaic Responses (Responses of 1 or 2) as a Function of Duration, F0, Intensity, and Vowel Quality Manipulation

When comparing the two listener groups' trochaic responses for each acoustic cue, in general the Korean listeners' trochaic response proportions were higher than those of the English listeners. Specifically, the Korean listeners were more sensitive to the change of F0 and intensity cues than the English listeners. When F0 and intensity changed from 0 to 1, the Korean listeners' trochaic responses more dramatically rose compared with those of the English listeners. On the contrary, when duration changed from 0 to 1, the English listeners' response proportions more noticeably rose than those of the Korean listeners. Compared with the three other cues, the vowel quality cue had weaker effect on stress perception for both listener groups. In addition, in general when acoustic cues shifted from -1 to 0, the change of trochaic stress response proportions was not obvious. That is, when the stress pattern varied from iambic (V1<V2) to noncontrastive pattern (V1=V2), the trochaic stress response proportions did not change accordingly. However, when the stress pattern changed from noncontrastive to trochaic pattern, the listeners' trochaic stress responses markedly increased.

In order to examine the effects of the four acoustic cues, a logistic regression analysis was conducted for each listener group separately using the SPSS software package (ver. 22). The dependent variable was percentage of listeners' binomial responses (trochaic or non-trochaic). The results of binary logistic regression analyses for the Korean listeners showed a significant effect of F0 (Wald's $X^2(1) = 193.271$, $p < .0001$), intensity (Wald's $X^2(1) = 200.338$, $p < .0001$) and vowel quality (Wald's $X^2(1) = 8.77$, $p < .01$) on stress identification. However, there was no significant effect of duration (Wald's $X^2(1) = .007$, $p > .05$). These statistical results indicate that F0,

intensity, and vowel quality cues significantly contributed to the Korean listeners' stress identification performance, while the duration cue was relatively neglected.

The results of logistic regression analyses for the English listeners showed a significant effect of duration (Wald's $X^2(1) = 136.672$, $p < .0001$), F0 (Wald's $X^2(1) = 31.483$, $p < .0001$) and intensity (Wald's $X^2(1) = 110.053$, $p < .0001$) on stress identification. However, there was no significant effect of vowel quality (Wald's $X^2(1) = 2.436$, $p > .05$). That is, the English listeners' trochaic responses were significantly affected by duration, F0, and intensity cues, but the vowel quality cue was much less influential.

The two listener groups' binomial responses (trochaic or non-trochaic) were compared using a generalized linear (logistic) mixed-effects model (GLMM) for each acoustic cue separately. Random effects were participants and items, and fixed effects included the four cues (duration, F0, intensity, or vowel quality) and native language of the listeners. For each acoustic cue, three levels of trochaic cue values (-1, 0, 1) were included.

Table 3. The Effects of Listener Group and Acoustic Cues

Cue	Group			Cue			Group x Cue		
	df1, df2	*f*	*p*	*df1, df2*	*f*	*p*	*df1, df2*	*f*	*p*
Duration	1 1,170	.519	>.05	2 1,170	56.734	<.0001***	2 1,170	13.706	<.0001***
F0	1 1,170	4.225	<.05*	2 1,170	90.099	<.0001***	2 1,170	9.810	<.0001***
Intensity	1 1,170	4.593	<.05*	2 1,170	103.644	<.0001***	2 1,170	2.278	>.05
Vowel quality	1 1,002	.059	>.05	2 1,002	12.963	<.0001***	2 1,002	1.835	>.05

*** < .001, ** < .01, * < .05

As shown in Table 3, all four cues affected the two listener groups' stress identification ($p < .0001$ for all four acoustic cues). However, the significant effect of language group varied across acoustic cues. There were significant effects of F0 and intensity between the two listener groups ($p < .05$ for both cues). The Korean listeners significantly more sensitive to the change of F0 and intensity cues than the English listeners. There were no significant effect of duration and vowel quality between the two listener groups. Furthermore, there were significant effects of interaction between acoustic cues and language groups for duration and F0 cues ($p < .0001$ for both cues). The two listener groups' response patterns to duration and F0 cues were not consistent with each other. Additional post-hoc analyses in each level of trochaic cue values of duration and F0 revealed that the two listener groups' responses were significantly different when the trochaic cue value is 1 ($p < .01$ for duration, and $p < .0001$ for F0). The English listeners outperformed the Korean listeners when the duration cue value was 1 (trochaic pattern), whereas the Korean listeners outperformed the English listeners when the F0 cue value was 1 (trochaic pattern).

5. Discussion and Conclusion

The purpose of the present study was to examine how listeners' native language affected the use of four different acoustic cues (duration, F0, intensity, and vowel quality) in identifying lexical stress in English. Native

Korean and English listeners participated in a stress identification task involving manipulated nonce word tokens. The results of proportions of trochaic responses and statistical analyses indicated that all four cues were important predictors of the listeners' stress perception. Specifically, the Korean listeners were sensitive to F0, intensity, and vowel quality cues, but not to the duration cue. The English listeners demonstrated a distinctively different pattern. They were sensitive to duration, F0, and intensity, but not to the vowel quality cue. When comparing the identification performance between the Korean listeners and English listeners, the Korean listeners were significantly more sensitive to F0 and intensity cues than the English listeners. There were no statistically significant differences between the two listener groups in terms of duration and vowel quality cues. Moreover, significant effects of interaction between acoustic cues and listener groups were revealed in terms of duration and F0 cues. The Korean listeners depended more on F0 and less on duration than the English listeners when identifying lexical stress.

The present results that pitch and intensity were the strongest cues for the Korean listeners were partially consistent with those of Kang and Kim (2016). Kang and Kim (2016) examined acoustic correlates of English stress in Korean L2 learners' perception involving duration, pitch and intensity cues. They found that pitch and intensity have positive correlations, and duration has a negative correlation with the response for Korean L2 learners. Although F0 and intensity are not used to indicate lexical prominence in Korean, the Korean listeners were sensitive to F0 and intensity variations in perceiving English stress. In Korean F0 plays an important role at the segmental and post-lexical levels. At the segmental level, F0 is used to distinguish between stop categories. Since the VOT values overlap between the word-initial stops, lax stops are differentiated from aspirated and tense stops based on F0 differences (Kang 2014, Kim et al. 2002, Silva 2006). In addition, Korean prosodic structure is known to be characterized by phrasal tone patterns, and the prominence is marked by the edge of prosodic units (Jun 2005, 2014). Jun (2005, 2014) pointed out that Korean phrasal tones are categorized as an Accentual Phrase (AP), which is realized as pitch or F0. Furthermore, Kim and Tremblay (2021) found that compared to Seoul Korean (SK) listeners, Gyeongsang Korean (GK) listeners showed higher accuracy in detecting English lexical stress. Kim and Tremblay (2021) argued that GK listeners transferred the use of F0 cues from the processing of pitch accents in the GK dialect to the processing of L2 lexical stress. The authors also suggested that Korean listeners' perception of stress was not modulated by duration or intensity cues, but by F0 cues. In the present study, although the Korean participants were not GK listeners, they were able to use F0 in perceiving lexical stress. The Korean listeners' heavy reliance on F0 in the present study could be ascribed to an important role of F0 in Korean segmental and post-lexical levels. This finding suggests that acoustic cues used in one phonological phenomenon in L1 can be transferred to another phonological phenomenon in L2.

The current results showed that intensity was an important cue for both Korean and English listeners. Further, the Korean listeners were more sensitive to the intensity cue than the English listeners. These results were not entirely consistent with those of the previous research. Kang and Kim (2016) manifested that intensity had statistically significant correlations with the responses of the Korean listeners, and that intensity was also the most reliable cue for the English listeners. However, in their study stress perception patterns were not directly compared between the two listener groups. Although intensity is not used in any level of Korean phonology, the Korean listeners utilized the intensity cue in perceiving lexical stress in the present study.

For the Korean listeners, the effect of duration was the smallest among the four acoustic dimensions investigated. Although both duration and intensity are not used for any phonological contrast in Korean, the Korean listeners were sensitive to intensity changes, but not to durational differences. These results were consistent with those in Kang and Kim (2016) in which only five out of 41 participants responded to the

difference of duration. On the other hand, in the current results, duration was the most prominent cue for the English listeners. The strong effects of duration on the native English speakers' perception of stress were also reported in Wang (2008).

In the present study the effect of vowel quality cue was weaker than the other cues in stress identification. The English listeners did not attend to the vowel quality cue, which is not consistent with previous studies (Chrabaszcz et al. 2017, Zhang and Francis 2010). Chrabaszcz et al. (2017) found that vowel quality was the strongest cue for the English listeners. Zhang and Francis (2010) also manifested that both English and Mandarin listeners consistently weighted vowel quality more than other cues. However, the vowel quality manipulations in the present study were different from those in previous studies. In Chrabaszcz et al. (2017) fully crossed combinations of each of the four cues (duration, F0, intensity and vowel quality) were used, and consequently many of the tokens have conflicting cues. Zhang and Francis (2010) used synthesized tokens of the word *desert*, in which formant frequencies were manipulated along with each of the other cues (duration, F0 and intensity) depending on the pair of cues in focus. Thus, in the previous studies all the stimulus tokens were manipulated with at least two acoustic cues at the same time. However, in the present study, the basic two vowel (/ɑ/ and /ə/) that had very similar acoustic values and still sounded natural were chosen and embedded in the four vowel quality combinations (/kɑtɑ/, /kɑtə/, /kətɑ/, /kətə/). The manipulation method of the vowel quality cue is different from that of the other three cues. For each of the four vowel quality combination, one of the three acoustic cues (duration, F0, intensity) was trochaic, iambic or noncontrastive, and the other two cues are noncontrastive. Moreover, although the two vowels /ɑ/ and /ə/ were distinctive in terms of spectral feature as shown in Table 1, the characteristics of the vowel /ə/ used in the stimuli were not the same with those of the schwa used in natural speech. Therefore, compared to the other cues, vowel quality could be less salient as a stress cue in the current study. Both the native and L2 listeners were not able to utilize the vowel quality cue in perceiving English lexical stress.

Another finding in this study was that both listener groups' trochaic response proportions noticeably increased when the stress pattern changed from noncontrastive (V1=V2) to trochaic pattern (V1>V2). However, when the stress pattern varied from iambic (V1<V2) to noncontrastive (V1=V2), there were no notable changes in the listeners' trochaic responses. In other words, even when the stimuli included an iambic stress pattern, the listeners often responded that they heard noncontrastive or trochaic stress. These response patterns may have related to the dominance of trochaic stress in English. According to van Heuven and Menert (1996) and Chrabaszcz et al. (2017), the dominance of trochaic stress (about 70% of disyllabic content words) in English biases listeners to perceive stress on the first syllable even in words with exactly the same sound segments, identical pitch, intensity and duration values across the two syllables. In the present study, cue changes from iambic to noncontrastive patterns led to less trochaic stress response changes than those from noncontrastive to trochaic patterns. In order to confirm this tendency of preference for trochaic stress perception over iambic stress more studies must be undertaken.

In sum, the current study demonstrated that both Korean and English listeners showed systematic response variations as a result of the manipulation of the four acoustic cues when they identified lexical stress of English nonce words. However, the responses of the Korean listeners were different from those of the English listeners in their relative reliance on the four cues. The Korean listeners showed sensitivity to F0, intensity and vowel quality, and less reliance on duration, whereas the English listeners were sensitive to duration, F0, and intensity. The significant perceptual differences between the two listener groups were shown in F0 and intensity cues. The Korean listeners utilized F0 and intensity more than the English listeners in perceiving lexical stress although intensity is not used in any phonological contrast in Korean. Furthermore, the Korean listeners relied less on

duration cues than the English listeners. These findings imply that L2 listeners' perception of lexical stress is not entirely tuned depending on the acoustic cues used in L1 prosody. In addition, both groups of listeners have shown that the cue shift from iambic to noncontrastive stress patterns induced less changes in trochaic stress responses than that from noncontrastive to trochaic stress patterns. This preference for trochaic stress perception over iambic stress needs to be further investigated in a future study involving more subjects and stress patterns in various contexts.

References

Beckman, M. E. 1986. *Stress and Non-stress Accent*. Dordrecht: Foris Publications.

Beckman, M. E. and J. B. Pierrehumbert. 1986. Intonational structure in Japanese and English. *Phonology* 3, 255-309.

Boersma, P. and D. Weenink. 2019. Praat: Doing phonetics by computer (Version 6.1.03) [Computer program]. Retrieved from www.praat.org

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, 383-398.

Cutler, A. and D. M. Carter. 1987. The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language* 2(3-4), 133-142.

Dupoux, E., S. Peperkamp and N. Sebastián-Gallés. 2001. A robust method to study stress "deafness". *The Journal of the Acoustical Society of America* 110(3), 1606-1618.

Dupoux, E., N. Sebastián-Gallés, E. Navarrete and S. Peperkamp. 2008. Persistent stress 'deafness': The case of French learners of Spanish. *Cognition* 106(2), 682-706.

Flege, J. E. and O. S. Bohn. 1989. An instrumental study of vowel reduction and stress placement in Spanish-accented English. *Studies in Second Language Acquisition* 11(1), 35-62.

Fokes, J. and Z. S. Bond. 1989. The vowels of stressed and unstressed syllables in nonnative English. *Language Learning* 39(3), 341-373.

Frost, D. 2011. Stress and cues to relative prominence in English and French: A perceptual study. *Journal of the International Phonetic Association* 41(1), 67-84.

Fry, D. B. 1965. The dependence of stress judgments on vowel formant structure. In X. Zwerner and W. Bethge, eds., *Proceedings of the 5th International Congress of Phonetics Sciences*, 306-311. Munster: Karger/Basel.

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

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.

Hammond, R. H. 1986. Error analysis and the natural approach to teaching foreign languages. *Lenguas Modernas* 13, 129-139.

Hart, J., R. Collier and A. Cohen. 1990. A perceptual study of intonation: an experimental-phonetic approach to speech melody. Cambridge: Cambridge University Press.

Jun, S. A. 2005. Korean intonational phonology and prosodic transcription. In S. 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. J. 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, M, R., P. S. Beddor and J. Horrocks. 2002. The contribution of consonantal and vocalic information to the perception of Korean initial stops. *Journal of Phonetics* 30(1), 77-100.

Liberman, M. Y. 1975. *The Intonational System of English.* Doctoral dissertation, Massachusetts Institute of Technology.

Mattys, S. L. 2000. The perception of primary and secondary stress in English. *Perception and Psychophysics* 62, 253-265.

Meng, Y., J. Zhang, S. Liu. and C. Wu. 2020. Influence of different acoustic cues in L1 lexical tone on the perception of L2 lexical stress using principal component analysis: An ERP study. *Experimental Brain Research* 238, 1489-1498.

Montero, D. 2007. *The Perception of the Acoustic Correlates of Stress: A Cross-linguistic Study on English, French and Spanish*. Unpublished master's thesis, University of Washington.

Morton, J. and W. Jassem. 1965. Acoustic correlates of stress. *Language and Speech*, 8, 159-181.

Oh, J. 2011. A study on prosodic cue determining syntactic moods in Korean. *Korean Linguistics* 51, 117-140.

Okobi, A. 2006. *Acoustic Correlates of Word Stress in American English.* Doctoral dissertation, Massachusetts Institute of Technology, Cambridge.

Peirce, J. W. 2007. PsychoPy - psychophysics software in Python. *Journal of Neuroscience Methods* 62(1-2), 8-13.

Roach, P. 2009. *English Phonetics and Phonology Fourth Edition: A Practical Course*. Ernst Klett Sprachen.

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.

Silva, D. J. 2006. Acoustic evidence for the emergence of tonal contrast in contemporary Korean. *Phonology* 23(2), 287-308.

Tremblay, A., M. Broersma, Y. Zeng, H. Kim, J. Lee and S. Shin. 2021. Dutch listeners' perception of English lexical stress: A cue-weighting approach. *Journal of the Acoustical Society of America* 149(6), 3703-3714.

van Heuven, V. J. and L. Menert. 1996. Why stress position bias? *The Journal of the Acoustical Society of America* 100(4), 2439-2451.

Wang, Q. 2008. *Perception of English Stress by Mandarin Chinese Learners of English*. Doctoral dissertation, University of Victoria.

Zhang, Y. and A. Francis. 2010. The weighting of vowel quality in native and non-native listeners' perception of English lexical stress. *Journal of Phonetics* 38(2), 260-271.

Zhang, Y., S. Nissen and A. L. Francis. 2008. Acoustic characteristics of English lexical stress produced by native Mandarin speakers. *Journal of the Acoustical Society of America* 123(6), 4498-4513.

Examples in: English
Applicable Languages: English
Applicable Level: Tertiary

