



Acoustic analysis of English lexical stress produced by Korean, Japanese and Taiwanese-Chinese speakers*

Ye-Jee Jung · Seok-Chae Rhee**

Abstract

Stressed vowels in English are usually produced using longer duration, higher pitch, and greater intensity than unstressed vowels. However, many English as a foreign language (EFL) learners have difficulty producing English lexical stress because their mother tongues do not have such features. In order to investigate if certain non-native English speakers (Korean, Japanese, and Taiwanese-Chinese native speakers) are able to produce English lexical stress in a native-like manner, speech samples were extracted from the L2 learners' corpus known as AESOP (the Asian English Speech cOrpus Project). Sixteen disyllabic words were analyzed in terms of the ratio of duration, pitch, and intensity. The results demonstrate that non-native English speakers are able to produce English stress in a similar way to native English speakers, and all speakers (both native and non-native) show a tendency to use duration as the strongest cue in producing stress. The results also show that the duration ratio of native English speakers was significantly higher than that of non-native speakers, indicating that native speakers produce a bigger difference in duration between stressed and unstressed vowels.

Keywords: English lexical stress, stressed/unstressed English vowels, prosody, foreign-accented speech

1. Introduction

As a representative example of stress-timed languages, English has some unique linguistic features that other languages do not have. One of these features is that English stress can determine a grammatical function and meaning of a word. For example, as Kondo (2009) exemplified in his study, there are some English words whose meanings change based on the position of primary stress: *Decrease* becomes a noun when its stress is on the initial syllable (/ˈdi:kri:s/) and it becomes a verb when it has stress on the

final syllable (/dɪˈkri:s/). As this example indicates, stress is an important suprasegmental component in English and incorrect use of stress can contribute to making English produced by non-native speakers sound foreign (Anderson-Hsieh *et al.*, 1992). There are three main acoustic correlates that indicate a realization of English stress: Duration, fundamental frequency (pitch, henceforth) and intensity (Archibald, 1992). Specifically, stressed syllables are usually produced with longer duration and higher pitch than unstressed syllables. In addition, stressed syllables are often louder than neighboring ones.

* This work was supported by Institute for Information & communications Technology Promotion(IITP) grant funded by the Korea government(MSIT). (R0126-15-1117, Core technology development of the spontaneous speech dialogue processing for the language learning)

** Yonsei University, scrhee@yonsei.ac.kr, corresponding author

Received 2 February 2018; Revised 7 March 2018; Accepted 28 March 2018

© Copyright 2018 Korean Society of Speech Sciences. This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0>) which permits unre-stricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Many non-native speakers of English fail to produce English stress as native English speakers do because their mother tongues do not have the same feature, causing a negative L1 transference on L2 (English) production. The study of McAllister *et al.* (2002) showed that the use of a feature utilized in the L2 but not in the L1 may cause a problem for L2 learners. In their experiment, native English speakers failed to produce and perceive Swedish vowel quantity (e.g. /ɛ:/ vs. /ɛ/). In Swedish, vowel duration is used to distinguish word meanings while English does not use vowel duration for the same function. This difference made native English speakers have difficulty learning and using the feature even after they learn Swedish phonological rules.

Using suprasegmental features in a wrong way can be an obstacle when non-native speakers communicate with native speakers. In other words, non-native speakers need to learn the correct use of suprasegmental features to have better communication with native speakers. The study of Anderson-Hsieh, *et al.* (1992) demonstrated using suprasegmental components proficiently play an important role in making better intelligibility and reducing foreign accent of non-native speeches. Additionally, Kang (2010) also showed that native speakers perceived non-native speeches as foreign-accented if non-native speakers did not use stress properly.

The current study aims to acoustically analyze English lexical stress realized by Korean, Japanese and Taiwanese-Chinese speakers in comparison to native English speakers. These three Asian languages were selected because their prosodic systems are all different from English prosodic system. Moreover, each language belongs to a different language group, respectively. Korean is a syllable-timed language where all syllables tend to have the similar vowel duration unlike English (Hong, 2012). Japanese is categorized as a pitch-accent language and only uses pitch to realize accent (Beckman, 1986). Finally, according to Moore & Jongman (1997), Chinese is a typical example of a tone language, which uses pitch to differentiate lexical meanings. The detailed explanations for each language will be provided below.

2. Literature Review

2.1. Acoustic characteristics of English stressed vowels

As mentioned above, English is a stress-time language where words have stress on, at least, one syllable and pitch accent is given to a syllable with primary stress (Beckman, 1986). The stress pattern of English is not fixed, indicating that it could be affected by various factors such as syllable structure and word class (Saha & Mandal, 2018). In other words, English stress is related to word class, in that the position of stress (either on the initial syllable or on the second syllable) determines a disyllabic word to be a noun or a verb. According to Guion (2005), disyllabic nouns are very likely to be stressed on their initial syllables while disyllabic verbs usually have stress on their final syllables.

English stress is manifested through three main acoustic cues, which are vowel duration, pitch, and intensity. Again, stressed vowels tend to have longer duration, higher pitch, and greater intensity (Archibald, 1992). Previous studies have tried to rank these acoustic cues to determine which cue is the strongest or the weakest one, but there is yet no consensus regarding the rank. For example, some previous studies (Fry, 1955; Bolinger, 1958; Beckman, 1986) asserted that pitch is the strongest stress cue than the other ones. Fry

(1955) also maintained that intensity is the least reliable acoustic cue because it can be easily influenced by external factors such as a recording environment while Bolinger (1958) regarded duration as the second-most important cue in realizing stress. By contrast, Sluijter & van Heuven (1996) did not agree with the previous studies (Fry, 1955; Bolinger, 1958; Beckman, 1986), asserting that the previous studies co-varied stress and accent. Instead, according to Sluijter & van Heuven (1996), vowel quality indicates stress most strongly and pitch is weaker than vowel quality, duration and even intensity.

2.2. Acoustic characteristics of Korean, Japanese and Taiwanese-Chinese vowels

Since the current study includes non-native speaker groups whose native languages are Korean, Japanese and Taiwanese-Chinese, acoustic characteristics of vowels of each language should be discussed. To begin with, Korean is identified as a syllable-timed language where syllables usually have identical duration (Hong, 2012). In addition, according to Kwon (2007), duration and intensity are not described as acoustic features that can differentiate a word meaning in Korean. Although vowel duration can be used to differentiate a word's meaning in some dialects of Korean (e.g., [nu:n] "snow" vs. [nun] "eye"), its function does not exist in modern standard Korean (Kim & Han, 1998).

Next, as a pitch-accent language, Japanese mainly uses pitch to realize accent (Kondo, 2009). Specifically, Kondo (2009) explained that Japanese accent is realized by a fall in pitch from an accented mora to the following mora. Mora is a basic unit of speech rhythm in Japanese and duration of a word or a phrase depends on the number of mora it has (Port *et al.*, 1987). In general, short syllables have one mora while long syllables consist of two morae (Tsujimura, 1996) but it does not mean that two morae are always acoustically twice longer than one mora (Beckman, 1982). Unlike modern standard Korean, duration can be used to differentiate lexical meanings in Japanese: *Kita* means "north" while *kiita* refers to the past form of "listen" (Tsushima, 2015). In other words, it can be concluded that Japanese prosody primarily uses pitch to indicate stress accent and duration to realize its mora system. However, intensity is not a meaningful cue to Japanese stress and there is no reduction for Japanese unstressed vowels.

Taiwanese-Chinese is one of the Chinese dialects that is spoken by the people living in Taiwan and also called Taiwanese Mandarin. According to Cheng (1985), Taiwanese-Chinese prosody is not largely different from standard Chinese prosody. Moreover, Ou (2010) explained that Taiwanese-Chinese is a language of lexical tone languages as standard Chinese is. Therefore, characteristics of standard Chinese prosody is going to be covered. Chinese is a tone language where there are four lexical tones: Tone 1 (high-level), tone 2 (high-rising), tone 3 (dipping) and tone 4 (high-falling) (Zhang *et al.*, 2008) and Chinese tone can distinguish word meanings like English stress. For instance, when *ma* is produced with tone 1, it means "mother" but its meaning becomes "horse" when produced with tone 3 (Na, 2013). The primary acoustic cue used to manifest Chinese tones is pitch (Liu & Samuel, 2004). However, it is hard to say that pitch is completely separated from duration or intensity: Tone 4 is typically longer than other tones and tone 3 also has relatively long duration and demonstrates a mid-syllable decrease in intensity (Zhang *et al.*, 2008). In fact, perceptual studies (Liu & Samuel, 2004; Whalen & Xu, 1992) have

proved that people are able to perceive Chinese tones by using other acoustic cues such as duration or intensity contour even when pitch information is absent.

2.3. English lexical stress realized by EFL learners

Many previous studies have investigated how EFL learners realize English lexical stress, and what factors have effects on the realization. (Fokes & Bond, 1989; Flege & Bohn, 1989; Lee *et al.*, 2006). For instance, the number of syllables had an effect on realization of English lexical stress. In Fokes & Bond (1989)'s study, five non-native English speakers whose mother tongues were all different (Farsi, Japanese, Spanish, Hausa and Chinese) read the two-, three-, and four-syllable words, which contained the same syllable (for example, *compete* and *competition*). Then, vowel durations were measured using spectrograms. All native speakers demonstrated a consistent pattern in vowel durations: The stressed vowels were always produced with longer duration in all conditions. However, the non-native speakers' patterns were not the same as native speakers, resulting in not making difference in duration between unstressed and stressed vowels. The non-native speakers showed the most difficulty regarding duration in four-syllable words. It was assumed that non-native speakers' wrong patterns were associated with their native languages.

The results of Flege & Bohn (1989)'s study proved that English lexical stress is more problematic than English stress placement to non-native speakers. In particular, stress placement refers to the position of primary stress and both languages have free stress, opposed to fixed-stress languages such as Polish. The subjects (seven English and seven Spanish speakers) read aloud English word pairs, which was derived from the same morpheme (e.g. *application* vs. *apply*). The dependent variables measured in the study were vowel duration, vowel intensity, and stress placement. The results demonstrated that the Spanish speakers had less difficulty dealing with stress placement, meaning that they were able to know which vowel to stress. For instance, both speaker groups gave stress to the first vowel in *able* but not to the first vowel in *ability*. Furthermore, they also made a differentiation in vowel reduction and intensity between stressed vowels and unstressed vowels in most cases, though some Spanish speakers did not. The overall results suggested English lexical stress placement is less problematic than vowel reduction or intensity.

Age of acquisition also has influence on manifesting stress in a nativelike manner. Lee *et al.* (2006) studied early and late Korean and Japanese-English bilinguals and examined if they can produce unstressed English vowels as native speakers. They measured pitch, intensity and duration. 20 Korean-English bilinguals, 20 Japanese-English bilinguals and 10 native English speakers participated in the study. The participants were asked to produce 19 English words in a carrier phrase. The findings demonstrated that both bilinguals succeeded in producing high accuracy of pitch, meaning that they were nativelike in pitch. In terms of duration, they showed the difference: Korean bilinguals did not show nativelike production of duration while Japanese bilinguals were nativelike in the production of duration. They were also different in intensity feature because Korean bilinguals were not able to produce nativelike intensity unlike Japanese bilinguals. In general, these findings suggested that some phonetic features can be acquired and produced at nativelike levels even when they do not exist in the L1 and there is an effect of the learner's age on L2 production.

Additionally, the results indicated that even though some acoustic features do not exist in L1, L2 learners were able to learn them.

English reduced vowels could be a problem to EFL learners with high English proficiency. Kwon (2007) investigated English vowels produced by advanced-level Korean speakers. In the study, the proficiency of Korean speakers was strictly manipulated so that only those who are fluent in English were included. The reason for strict controlling English proficiency was to ensure that acoustic properties shown by Korean speakers is due to the effect of their L1, not due to their lack of English competence. When it comes to materials, Kwon used monosyllabic function words such as *to* or *of*, which were embedded in the reading passage. Then, she measured duration and pitch and compared each acoustic feature of Korean speakers to that of English speakers.

The advanced-level Korean speakers showed significantly longer duration for unstressed vowels than native speakers and the reason was assumed to be their L1, which lacks a property of reducing duration for unstressed vowels. Next, the pitch of unstressed vowels of Korean speakers was significantly lower than that of English speakers and there were two possible explanations: First, Korean speakers reported that they felt nervous about the recording and second, the individual difference between Korean and English speakers might result in pitch difference.

The study of Zhang *et al.* (2008) demonstrated that even though non-native speakers succeed in realizing English lexical stress as native speakers, they may still sound foreign. In their study, 10 Mandarin speakers and 10 English speakers participated and seven pairs of disyllabic words were used for stimuli. Each word pair was composed of a noun and a verb, which differed only in stress placement (e.g. *Object* vs. *obJECT*). First, the Mandarin speakers read target words embedded in context and frame sentence. Next, they read only target words in isolation and before the second recording, the experimenter explained English lexical stress rule (where to put stress for nouns and verbs) so that the Mandarin speakers knew the correct position of stress in the second recording. The results demonstrated that native Mandarin speakers were able to realize English lexical stress in a similar manner compared to native English speakers. Namely, they used longer vowel duration, higher pitch, and greater intensity for stressed syllables. However, five native English listeners who majored in linguistics evaluated that acoustic cues used by native Mandarin speakers were less acceptable than those of native English speakers. That is because even though both groups manifested stress in a similar way, Mandarin speakers produced English stress syllables with higher pitch than the native English speakers, influenced by their L1, which made Mandarin-accented English less native.

From previous studies, it becomes clear that English lexical stress realization of non-native speakers is distinguishable from that of native English speakers, contributing to making non-native speakers' English sound foreign. In addition, the prosodic system of non-native speakers' L1 has an effect when producing L2 (English) prosody. However, previous studies did not focus on the hierarchy of acoustic cues. Therefore, the current study not only investigates how non-native speakers realize English stress but also concentrates on the hierarchy of acoustic cues to examine which cue is most frequently used by each speaker group.

3. Research Questions and hypotheses

The research question for the current study is as follows:

Are non-native speakers able to realize English lexical stress the same as native speakers? In other words, are they able to produce unstressed vowels with shorter duration, lower pitch and weaker intensity?

The hypotheses of the above research question are based on the findings of Lee *et al.*'s study (2006), Kwon's study (2007) and Zhang *et al.*'s study (2008). One thing to note is that Zhang *et al.* (2008) used native Mandarin speakers as subjects while the current study used data from Taiwanese-Chinese speakers. However, since Taiwanese-Chinese has a phonology system similar to native Mandarin, which is spoken in the mainland of China (Cheng, 1985), it would be acceptable to set the hypothesis for Taiwanese-Chinese speaker, following the results of Zhang *et al.*'s study (2008). The hypotheses for each speaker group are as follows:

- (1) The English unstressed vowels produced by Korean speakers would have lower pitch than stressed ones. However, the unstressed vowels and the stressed vowels would not be different in intensity and duration.
- (2) The English unstressed vowels produced by Japanese speakers would have lower pitch, shorter duration, and weaker intensity than the unstressed ones.
- (3) The English unstressed vowels produced by Taiwanese-Chinese speakers would also demonstrate lower pitch, shorter duration, and weaker intensity than the unstressed ones.

4. Method

4.1. AESOP corpus

All speech samples used for the current study were extracted from English L2 learners' corpus named AESOP. A brief introduction of the AESOP corpus will be provided based on the book edited by Tono *et al.* (2012). The AESOP project started from the fact that Asian English offers plentiful variations in pronunciation, lexicon, and grammar and studying these variations will help to develop teaching and learning methods for those who speak EFL. It was launched in 2008 and its ultimate goal is developing an English speech corpus of Asian language speakers including Taiwan, Japan, Korean, Hong Kong, Thailand and Vietnam. The AESOP corpus "consists of reading tasks and semi-spontaneous responses to questions" (Kondo *et al.*, 2015). There are six tasks in the corpus and all of them are reading sentences which include target words. In addition to reading tasks, there are also two kinds of semi-spontaneous responses, which are computer-prompted dialogue and picture-description task.

For the current study, speech samples of four speaker groups - native English, Korean, Japanese and Taiwanese-Chinese speakers - were extracted from the AESOP corpus and all speakers were male-restricted. Each speaker read 15 sentences, yielding in 600 (15 sentences * 40 speakers) speech samples in total. The non-native speakers were selected randomly and four native English male speakers were additionally recruited since there were only six native English males in the corpus data.

4.2. Materials

From 15 sentences (for the full list of sentences, see Appendix) produced by 40 speakers (10 speakers for each group), disyllabic words were extracted for the measurement, regardless of part of speech. There were 17 disyllabic words but the word *written* was excluded because the second syllable of the word contains a syllabic consonant, instead of a vowel. As a result, the current study analyzed 16 disyllabic words, which were *allowed*, *although*, *any*, *birthday*, *evening*, *fancy*, *faster*, *learning*, *morning*, *party* (appeared twice), *picture*, *taking*, *visit*, *window* and *woman*. Among them, *allowed* and *although* have its primary stress on the second syllable while the rest of them have its primary stress on the initial syllable.

4.3. Measurement and analysis

The speech samples were analyzed by using Praat (Boersma & Weenink, 2016; version 6.0.14) in terms of vowel duration (in seconds), pitch (in hertz) and intensity (in decibel). Vowel duration was measured based on spectrogram and pitch was measured by using the "get pitch" function from the program. Likewise, intensity was also measured by using the "get intensity" function.

There were some exceptional cases where the speech samples were not able to be analyzed because some non-native speakers misread the words. For example, one of Korean speakers read the word *fancy* in a wrong way, without pronouncing the second vowel of the word. When this error was found, the speech sample was excluded from the analysis.

Finally, for the dependent variable, the current study used ratio, instead of values of duration, pitch, and intensity to see to what extent native speakers and non-native speakers varied each acoustic cue to indicate stress (Zuriazq & Sereno, 2007). In particular, when it comes to duration, the ratio was measured by dividing vowel duration of the stressed syllable by vowel duration of the unstressed syllable. Applying the same methodology, the ratio of pitch and intensity were calculated.

5. Results

Before analyzing the results of each acoustic cue, the overall descriptive statistics will be presented in <Table 1>. Native speakers and non-native speakers demonstrated the identical tendency: duration was the strongest stress cue, with both pitch and intensity being weaker. For instance, the intensity ratio of Taiwanese-Chinese speakers was 1.00, meaning that they did not use intensity properly to manifest stress. Next, the ratio of duration of native speakers was higher than non-native speakers, indicating that native speakers made a bigger difference in duration than non-native speakers. When it comes to pitch or intensity, the difference between native speakers and non-native speakers was not statistically significant.

Table 1. Descriptive statistics of the ratio of stressed to unstressed vowels for each acoustic cue

	Duration	Pitch	Intensity
English (N=10)	1.675	1.057	1.026
Korean (N=10)	1.396	1.049	1.035
Japanese (N=10)	1.363	1.030	1.025
Taiwanese (N=10)	1.222	1.010	1.000

N, the number of speakers for each group

5.1. Duration

<Table 2> shows vowel duration ratio of each speaker group. Native speakers showed shorter duration in both stressed and unstressed vowels than non-native speakers. This result is in line with previous studies (e.g. Kwon, 2007), which have proved that one of common characteristics of non-native English is longer duration than native English. However, since the current study only deals with the ratio of each acoustic cue, duration was not statistically verified.

As for the ratio of stressed to unstressed vowels, the ratio of native speakers was higher than that of non-native speakers. The higher ratio of native speakers indicates that even though non-native speakers also succeeded in making a difference in duration of two vowels, the difference was smaller than that of native speakers. Moreover, for a statistical verification, one-way ANOVA was conducted and the results are summarized in <Table 3>.

Table 2. Average duration ratio of each speaker group

	ENG	KOR	JPN	TWN
Duration of stressed Vs (s)	0.082	0.089	0.086	0.094
Duration of unstressed Vs (s)	0.049	0.064	0.063	0.077
Ratio	1.68	1.40	1.36	1.22

ENG, Native; KOR, Korean; JPN, Japanese; TWN, Taiwanese-Chinese

Table 3. Results of one-way ANOVA test of duration ratio

	DF	Mean square	F-value	p-value
Between groups	3	5.506	20.158	< 0.001
Within groups	616	0.273		
Total	619			

ANOVA, analysis of variance; DF, degree of freedom

As the above table indicated, there was a significant effect of speaker group on duration ratio ($F(3, 616) = 20.158, p < 0.001$). In addition, the results of post-hoc Tukey test demonstrated there was a significant difference between native speakers and non-native speakers while there was no significant difference among non-native speakers.

5.2. Pitch

Unlike duration, for all speakers, pitch was not a strong acoustic cue when realizing English lexical stress, as presented in <Table 4>. Every speaker group showed the ratio slightly higher than 1, implying that the initial syllable and the second syllable were almost

the same in terms of pitch (Native: 1.06, Korean: 1.05, Japanese: 1.03 and Taiwanese-Chinese: 1.01). The results of one-way ANOVA test demonstrated (<Table 5>) that there was no significant effect of speaker group on pitch ratio ($F(3, 636) = 1.532, p = 0.205$). In other words, native speakers and non-native speakers were not significantly different in pitch ratio.

Table 4. Average pitch ratio of each speaker group

	ENG	KOR	JPN	TWN
Pitch of stressed Vs (hz)	129.20	128.52	134.33	122.99
Pitch of unstressed Vs (hz)	121.50	122.40	130.84	121.52
Ratio	1.06	1.05	1.03	1.01

ENG, Native; KOR, Korean; JPN, Japanese; TWN, Taiwanese-Chinese

Table 5. Results of one-way ANOVA test of pitch ratio

	DF	Mean square	F-value	p-value
Between groups	3	0.072	1.532	0.205
Within groups	616	0.047		
Total	619			

ANOVA, analysis of variance; DF, degree of freedom

5.3. Intensity

Similar to pitch, <Table 6> shows that intensity was not a strong cue in manifesting English lexical stress, for both native and non-native speakers. The ratio of native speakers, Korean speakers, Japanese speakers and Taiwanese-Chinese speakers were 1.03, 1.04, 1.03 and 1.00, respectively. Namely, although it is widely known that stressed vowels are usually produced with greater intensity, both native and non-native speakers did not make a large intensity difference when producing stressed vowels.

Table 6. Average intensity ratio of each speaker group

	ENG	KOR	JPN	TWN
Intensity of stressed Vs (db)	70.386	66.920	63.865	66.874
Intensity of unstressed Vs (db)	68.524	64.840	62.126	66.523
Ratio	1.03	1.04	1.03	1.00

ENG, Native; KOR, Korean; JPN, Japanese; TWN, Taiwanese-Chinese

<Table 7> demonstrates the summary of the results of one-way ANOVA test of intensity ratio. It turned out the effect of speaker group on intensity ratio was not significant ($F(3, 636) = 1.675, p = 0.171$). There was no significant difference between native and non-native speakers in intensity ratio.

Table 7. Results of one-way ANOVA test of intensity ratio

	DF	Mean square	F-value	p-value
Between groups	3	0.040	1.675	0.171
Within groups	616			
Total	619			

DF, degree of freedom

6. Discussion & conclusion

The current study aims to investigate how native and non-native English speakers realize lexical stress and if there is any difference between native and non-native speakers. The hypothesis set for Korean speakers was they would fail to realize English lexical stress as native speakers, not making a difference in vowel duration and intensity. By contrast, Japanese and Taiwanese-Chinese speakers were expected to manifest English lexical stress as native speakers, using longer duration, higher pitch, and greater intensity for stressed vowels. The results of the current study demonstrated that the hypothesis set for Korean speakers was not correct. In fact, it turned out that Korean speakers were able to use both vowel duration and intensity to realize English stress, even though the ratio of vowel duration was higher (1.4) than that of intensity (1.04). Several assumptions can be made regarding the result.

First, Korean speakers might be proficient enough to acknowledge lexical stress rule of English and they might have tried to pronounce words in a native-like manner. In other words, they would not have been negatively influenced by their mother tongue when producing English. Even though Korean does not have lexical stress, they could have learned the feature through acquisition. Secondly, as previous studies (Elder *et al.*, 2005; Zuraiq & Sereno, 2007; Na, 2013) demonstrated, vowel duration is the strongest cue to native English speakers. Therefore, when Korean learners perceive English, vowel duration could be more salient cue than the other ones and this might have an effect on Korean learners' production of English lexical stress. However, additional studies should be conducted to find the exact reason.

The noteworthy thing is that both native and non-native speakers used vowel duration as the strongest and pitch as the second strongest cue in manifesting English lexical stress. For all speaker groups, intensity was the weakest cue. Especially, when it comes to Taiwanese-Chinese speakers, the ratio of intensity was 1.00, indicating that they did not make intensity difference between stressed and unstressed vowels. The result is partially in accordance with Fry (1955) and Beckman & Pierrehumbert (1986): Fry (1955) maintained that intensity is the weakest and the least reliable acoustic cue in realizing English stress, which was also true in the current study. At the same time, according to Beckman & Pierrehumbert (1986), both duration and intensity are the most reliable acoustic correlates. However, in the current study, all speakers showed a strong tendency to show the most distinguishable difference in duration, but not in pitch or intensity. Therefore, the current study shows again that the hierarchical order of acoustic cues used in English stress is subject to change, rather than fixed.

One possible explanation for the strong tendency towards using duration as the strongest stress cue is that some stressed vowels analyzed in the current study are inherently longer than unstressed vowels. For example, in the word *allowed* and *taking*, stressed vowels from each word are diphthongs (/aʊ/ and /eɪ/, respectively) and it is natural that they have longer duration than monophthongs. In addition, as Peterson & Lehiste (1960) said, duration of a vowel is also affected by its surroundings. When a vowel is followed by voiced consonants, its duration becomes longer than when it is followed by voiceless consonants. In the materials for the current study, there were some stressed vowels followed by voiced consonants (e.g. *fan* from *fancy* and *par* from *party*) and this phonetic environment might have had an influence on duration of

stressed vowels.

There are some limitations in the current study. To begin with, since the current study extracted speech samples from the L2 learners' corpus, the personal data of each speaker was not fully accessible. However, it is important to ensure that all speakers speak a standard language to exclude possible effects arising from speaking dialects. Secondly, non-native speakers' English proficiency was not strictly controlled. It is desirable to control English proficiency of non-native speakers to the similar level since non-native speakers with high English proficiency is less likely to be affected by their L1 in a negative way (Elder *et al.*, 2005). Additionally, materials should be manipulated so that their phonetic environment has a minimum effect on acoustic features to obtain more precise results. For instance, as Kwon (2007) and Na (2013) said in their studies, a vowel followed by *r* is usually co-articulated with it, making it harder to clarify a segment boundary between the vowel and the consonant. In addition to phonetic environment, prosodic condition also needs to be controlled more strictly. That is because some words used in the study (e.g. *although* or *any*) do not usually receive stress in the sentences since they are function words. Therefore, it is necessary to ensure that every target word receives stress within the sentences.

The current study has its own implication that it has tried to compare native and non-native English speakers who speak different mother tongues, by using the L2 learners' corpus called AESOP. The results of the study would be helpful for teachers who teach EFL learners coming from various linguistic backgrounds. Furthermore, investigating suprasegmental features of non-native English would be also advantageous to EFL learners in that wrong use of suprasegmental features play an important role in making their English sound foreign and awkward. Since the AESOP corpus includes multiple Asian-language speakers such as Vietnamese or Thai, it would be also interesting to investigate their English to promote a better understanding of Asian-accented English.

References

- Anderson-Hsieh, J., Johnson, R., & Koehler, K. (1992). The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure. *Language Learning*, 42(4), 529-555.
- Archibald, J. (1992). Transfer of L1 parameter settings: Some empirical evidence from polish metrics. *Canadian Journal of Linguistics*, 37(3), 301-340.
- Beckman, M. (1982). Segment duration and the 'mora' in Japanese. *Phonetica*, 39(2-3), 113-135.
- Beckman, M. (1986). *Stress and non-stress accent*. Dordrecht: Walter de Gruyter (Vol. 7).
- Beckman, M., & Pierrehumbert, J. (1986). Intonational structure in Japanese and English. *Phonology Yearbook*, 3, 255-309.
- Boersma, P., & Weenink, D. (2016). Praat: doing phonetics by computer. Version 6.0.21. Retrieved from <http://www.praat.org/> on January 16, 2018.
- Bolinger, D. (1958). A theory of pitch accent in English. *Word*, 14(2-3), 109-149.
- Cheng, R. (1985). A comparison of Taiwanese, Taiwan Mandarin, and Peking Mandarin. *Language*, 352-377.
- Elder, C., Golombek, P. Nguyen, T., & Ingram, J. (2005). Vietnamese acquisition of English word stress. *TESOL Quarterly*, 39(2),

- 309-319.
- Flege, J., & Bohn, O. (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., & Bond, Z. (1989). The vowels of stressed and unstressed syllables in nonnative English. *Language Learning*, 39(3), 341-373.
- Fry, D. (1955). Duration and intensity as physical correlates of linguistic stress. *The Journal of the Acoustical Society of America*, 27(4), 765-768.
- Guion, S. (2005). Knowledge of English word stress patterns in early and late Korean-English bilinguals. *Studies in Second Language Acquisition*, 27(4), 503-533.
- Hong, S. (2012). A phonetic study on the production and training of English lexical stress by Korean speakers. *The Journal of Modern British & American Language & Literature*, 30(4), 99-122.
- Kang, O. (2010). Relative salience of suprasegmental features on judgments of L2 comprehensibility and accentedness. *System*, 38(2), 301-315.
- Kim, H., & Han, J. (1998). Vowel length in modern Korean: An acoustic analysis. In B. Park, & J. Yoon (Eds.), *Proceedings of the 14th International Conference on Korean Linguistics* (pp. 412-418). Seoul: Hanshin.
- Kondo, M. (2009). Is acquisition of L2 phonemes difficult? Production of English stress by Japanese speakers. *Proceedings of the 10th Generative approaches to Second Language Acquisition Conference (GASLA 2009)* (pp. 105-112). Somerville, MA: Cascadia Proceedings. Project, 2009.
- Kondo, M., Tsubaki, H., & Sagisaka, Y. (2015). Segmental variation of Japanese speakers' English: Analysis of "the North Wind and the Sun" in AESOP Corpus (< Feature Articles > Corpus-Based Data-Driven Speech Research in Asia). *Journal of the Phonetic Society of Japan*, 19(1), 3-17.
- Kwon, H. (2007). A phonetic analysis of advanced Korean EFL Learners' English reduced vowels. *Korean Journal of English Language and Literature*, 7(4), 547-564.
- Lee, B., Guion, S., & Harada, T. (2006). Acoustic analysis of the production of unstressed English vowels by early and late Korean and Japanese bilinguals. *Studies in Second Language Acquisition*, 28(3), 487-513.
- Liu, S., & Samuel, A. (2004). Perception of Mandarin lexical tones when F0 information is neutralized. *Language and Speech*, 47(2), 109-138.
- McAllister, R., Flege, J., & Piske, T. (2002). The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English and Estonian. *Journal of Phonetics*, 30(2), 229-258.
- Moore, C., & Jongman, A. (1997). Speaker normalization in the perception of Mandarin Chinese tones. *The Journal of the Acoustical Society of America*, 102(3), 1864-1877.
- Na, T. (2013). *Comparison of acoustic characteristics of English lexical stress between Korean and Mandarin speakers*. (Unpublished Master's Thesis). Yonsei University. Retrieved from RISS on January 20, 2018.
- Ou, S. (2010). Taiwanese EFL learners. *Concentric: Studies in Linguistics*, 36(1), 1-23.
- Peterson, G., & Lehiste, I. (1960). Duration of syllable nuclei in English. *The Journal of the Acoustical Society of America*, 32(6), 693-703.
- Port, R., Dalby, J., & O'Dell, M. (1987). Evidence for mora timing in Japanese. *The Journal of the Acoustical Society of America*, 81(5), 1574-1585.
- Saha, S., & Mandal, S. (2018). Phonetic realization of English lexical stress by native (L1) Bengali speakers compared to native (L1) English speakers. *Computer Speech & Language*, 47, 1-15.
- Sluijter, A., & Van Heuven, V. (1996). Spectral balance as an acoustic correlate of linguistic stress. *The Journal of the Acoustical Society of America*, 100(4), 2471-2485.
- Tono, Y., Kawaguchi, Y., & Minegishi, M. (2012). *Developmental and crosslinguistic perspectives in learner corpus research* (Vol. 4). Amsterdam: John Benjamins Publishing.
- Tsujimura, N. (1996). *An introduction to Japanese linguistics*. Hoboken: John Wiley & Sons.
- Tsushima, T. (2015). A case study: How do Japanese learners of English learn to use acoustic cues to differentiate stressed and unstressed syllables in pronunciation training?. *The Journal of Humanities and Natural Sciences*, 137, 33-57.
- Whalen, D., & Xu, Y. (1992). Information for Mandarin tones in the amplitude contour and in brief segments. *Phonetica*, 49(1), 25-47.
- Zhang, Y., Nissen, S., & Francis, A. (2008). Acoustic characteristics of English lexical stress produced by native Mandarin speakers. *The Journal of the Acoustical Society of America*, 123(6), 4498-4513.
- Zuraid, W., & Sereno, J. (2007). English lexical stress cues in native English and non-native Arabic speakers. *Ratio*, 1(2), 829-832.

• **Ye-Jee Jung**

Department of English Language and Literature
Yonsei University
50 Yonsei-ro, Seodaemun-gu
Seoul 03722, Korea
Tel: 02-2123-2300
Email: jungyeji718@yonsei.ac.kr
Fields of interest: Phonetics, Second Language Acquisition

• **Seok-Chae Rhee** corresponding author

Department of English Language and Literature
Yonsei University
50 Yonsei-ro, Seodaemun-gu
Seoul 03722, Korea
Tel: 02-2123-4483
Email: scrhee@yonsei.ac.kr
Fields of interest: Phonetics, Phonology

Appendix. Full List of the Sentences (Target Words in Bold)

1. I said supermarket five times.
2. I said elevator ten times.
3. Did he go to the hospital?
4. When will Bill be available?
5. **Although** Fred didn't have **any** experience, he had no trouble **learning** how to make a video.
6. When Sue left this **evening** for California, she said she would call me tomorrow.
7. If you want to learn Vietnamese, I think it will be easier than Japanese
8. No. Mary is **taking a morning** flight.
9. No. Video recordings are not **allowed**.
10. If the **birthday party** wasn't for Mary, then who was it for?
11. Jane saw a **picture** of the boy she was fond of.
12. John went to **visit** the **woman** he had written to.
13. I can run **faster** than you can.
14. He went to a **fancy dress party** as a guest, but what did he dress as?
15. Would you like a **window** seat or an aisle seat?