

ARTICLE

Evaluating Korean learners' English rhythm proficiency with measures of sentence stress

Ho-Young Lee¹ and Jieun Song^{2,*}

¹Seoul National University and ²University College London

*Corresponding author. Email: jieun.song@ucl.ac.uk

(Received 05 June 2018; revised 22 May 2019; accepted 24 May 2019)

Abstract

Previous research has suggested that the production of speech rhythm in a second language (L2) or foreign language is influenced by the speaker's first language rhythm. However, it is less clear how the production of L2 rhythm is affected by the learners' L2 proficiency, largely due to the lack of rhythm metrics that show consistent results between studies. We examined the production of English rhythm by 75 Korean learners with the rhythm metrics proposed in previous studies (pairwise variability indices and interval measures). We also devised new sentence stress measures (i.e., accentuation rate and accentuation error rate) and investigated whether these new measures can quantify rhythmic differences between the learners. The results found no rhythm metric that significantly correlated with proficiency in the expected direction. In contrast, we found a significant correlation between the learners' proficiency levels and both measures of sentence stress, showing that less-proficient learners placed sentence stress on more words and made more sentence stress errors. This demonstrates that our measures of sentence stress can be used as effective features for assessing Korean learners' English rhythm proficiency.

Keywords: accentuation; non-native speech rhythm; rhythm metrics; sentence stress; automatic rhythm scoring

Spoken languages have been classified into stress-timed, syllable-timed, and mora-timed languages. English and Dutch have been regarded as stress-timed languages; French, Italian, and Spanish as syllable-timed languages (Abercrombie, 1967; Lloyd James, 1940; Pike, 1945); and Japanese as a mora-timed language (Benguerel, 1999; Bloch, 1950; Han, 1962; Hoequist, 1983a, 1983b; Ladefoged, 1975; Port, Dalby, & O'Dell, 1987). This traditional classification was based on the assumption that the unit of rhythm (i.e., foot in stress-timed languages, syllable in syllable-timed languages, and mora in mora-timed languages) occurs at regular intervals. Although the isochrony of the rhythm units has not been empirically supported in later studies (e.g., Arvaniti, 2012; Dauer, 1983; Roach, 1982; Wenk & Wiolland, 1982), these rhythm types are still believed to be perceptually

© Cambridge University Press 2019. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

distinguishable even by infants (Nazzi, Bertoncini, & Mehler, 1998; Ramus, Dupoux, & Mehler, 2003).

In order to quantify such rhythmic differences, a variety of rhythm metrics have been developed. Ramus et al. (1999) proposed three interval measures: %V (the proportion of vocalic intervals in a sentence), ΔV (the standard deviation of vocalic intervals), and ΔC (the standard deviation of consonantal intervals). They showed that because stress-timed languages permit more complex syllable structures and have vowel reduction in unstressed syllables, they can have lower %V and higher ΔV and ΔC compared to syllable-timed languages. Another widely used rhythm metric is the pairwise variability index (PVI hereafter) proposed in Low, Grabe, and Nolan (2000). The PVI measures the level of variability in syllable duration by calculating the average difference in duration between two successive syllables, with stress-timed languages expected to have higher PVI scores than syllable-timed languages. Grabe and Low (2002) proposed raw PVI for consonantal intervals (rPVI-C) measured without speech rate normalization, and normalized PVI for vocalic intervals (nPVI-V) to control for speech rate variation in vocalic intervals. In addition, rate-normalized standard deviations, VarcoV and VarcoC, were developed in Dellwo (2006).

These rhythm metrics have been widely used for various languages. Despite the popularity, there have been large discrepancies between studies. For example, different studies have classified the same languages into different rhythm types, or some studies have failed to find significant differences even between prototypical languages such as English and French (e.g., Baltazani, 2007; Grabe & Low, 2002; White & Mattys, 2007; see Arvaniti, 2012, for a review). Arvaniti (2012) tested the performance of the rhythm metrics (ΔC , %V, rPVI, nPVI, VarcoC, and VarcoV) on English, German, Greek, Italian, Spanish, and Korean. Her results showed that the metric scores were affected by various types of “noise” in the data such as elicitation methods (e.g., read or spontaneous speech), the complexity of syllable structure, and interspeaker variation. This suggests that the rhythm metrics may not be robust enough to quantify rhythmic differences between languages.

The aim of the present study was to find reliable metrics that can quantify rhythmic characteristics of non-native speakers. Previous research has also investigated the rhythm of non-native speech using those rhythm metrics or acoustic correlates of stress (i.e., pitch, intensity, duration, or vowel quality). Non-native speakers whose native language has been described as a syllable-timed or mora-timed language (e.g., Malay, Vietnamese, Spanish, and Japanese) were shown to speak English with characteristics of their first language rhythm (e.g., Carter, 2004, 2005a, 2005b; Mochizukisudo & Fokes, 1985; Mochizukisudo & Kiritani, 1991; Nguyen, 2003; White & Mattys, 2007). Similarly, Singapore English and Hong Kong English were found to have smaller variability in syllable durations than British English as measured by the rhythm metrics, due to the influence of Singaporean Mandarin and Cantonese, respectively (Deterding, 2001; Low et al., 2000; Setter, 2006).

However, it remains uncertain whether the rhythm metrics can be used to evaluate the rhythmic characteristics of non-native learners with varying levels of proficiency. For example, Stockmal, Markus, and Bond (2005) showed that ΔC and PVI-C can differentiate low-proficiency and high-proficiency Russian learners

of Latvian. In this case, the authors suggested that the high variability of the rhythm scores among nonproficient learners was due to their imperfect fluency rather than the influence of their native language (L1). In contrast, Chen and Zechner (2011) did not find any significant correlation between rhythm scores and proficiency levels for non-native speakers of English whose L1 was Mandarin, although adding the rhythm metrics to their automatic English speech scoring system significantly improved the performance of the system (i.e., higher agreement between machine-predicted scores and human scores) compared to a model that was only based on nonrhythm features.

It also remains debated how much speech rhythm contributes to the perception of non-native speech. For example, some studies have found that fluency-related characteristics such as pause duration, pause frequency, and speech rate are stronger predictors of proficiency or foreign accentedness than rhythm-related characteristics such as the degree of stress-timing (e.g., Trofimovich & Baker, 2006; Iwashita, Brown, McNamara, & O'Hagan, 2008). Similarly, Jenkins (2000) excluded English stress-timed rhythm from her *lingua franca* core by describing it as a "non-core feature," whereas nuclear stress was described as one of the most important factors affecting intelligibility in communication among non-native speakers ("core feature") due to its role in conveying important or new information in the sentence. Although quantifying the relative contribution of rhythm to the perception of second language (L2) speech compared to other aspects of speech is beyond the scope of the present study, it is important to establish stress metrics that can quantify non-native speakers' rhythm proficiency reliably.

The present study investigated the production of speech rhythm by Korean learners of English. Some studies have suggested that Korean learners' English rhythm is more syllable timed than stress timed due to the influence of their L1 rhythm (e.g., Jang, 2008; Kim, Flynn, & Oh, 2007; Lee & Kim, 2005). However, it has been debated in previous research whether Korean is a stress-timed or syllable-timed language, with some studies suggesting that it does not belong to any rhythm class (see Jeon 2015, for a brief review).¹ Korean rhythm has some characteristics of a stress-timed language in that Korean permits syllable codas and syllable weight plays an important role in shaping Korean rhythm patterns (cf. Lee, 1990). However, Korean can also be more similar to a syllable-timed language in that Korean stress is not easily perceived due to the lack of vowel reduction, which is one of the most important distinctive properties of stress-timed languages (Dauer, 1983). Lee (2011) suggested that low-proficiency Korean learners tend to place sentence stress on most words in a sentence including grammatical words and use strong vowels in unstressed syllables, which gives the impression of syllable-timed rhythm, or even word-timed rhythm. In contrast, the rhythm of high-proficiency speakers sounds closer to stress-timed rhythm, suggesting that the rhythm of non-native speakers with the same L1 background can vary depending on their proficiency.

It remains unclear whether the rhythm metrics are suitable for quantifying the differences in rhythm between Korean learners of English. Jang (2008) found significant correlations between Korean learners' English proficiency and some of the rhythm metrics, such as %V, VarcoV, and nPVI-V, but only for a subset of sentences, whereas other metrics, such as the proportion of the duration of

function words in a sentence,² articulation rate, and the number of silence intervals, significantly correlated with proficiency for most of the sentences that were tested. Although the limitations of the rhythm metrics have been previously reported (e.g., Arvaniti, 2012; Wiget et al., 2010), we investigated whether the rhythm metrics can capture rhythmic differences between Korean learners of English in a large-scale study and compared these with the metrics that we will propose in this study.

It should be noted that the rhythm metrics described above are only derived from durational patterns of speech.³ However, Fuchs (2014) has developed nPVI-V metrics that calculate the degree of variability in loudness and simultaneous variability in loudness and duration of vocalic intervals. While his new metrics were shown to be effective in showing rhythmic differences between Indian and British English, very few studies have used acoustic correlates other than duration to quantify rhythmic characteristics. Temporal characteristics measured by the previous rhythm metrics can only partially account for perceived rhythm because English rhythm is associated with the alternation of stressed and unstressed syllables. That is, the rhythmic beat of stressed syllables is manifested not only in prolonged duration but also in higher (or sometimes lower) pitch and stronger intensity in English.

Rhythmic beats that are perceived on the utterance level are referred to as sentence stress, whose placement was called “accentuation” in Gimson (1980). For example, grammatical words do not usually bear sentence stress when produced in a sentence. A syllable receiving sentence stress and any following unstressed syllables comprises a foot in English (Abercrombie, 1967). Placing sentence stress on most words in a sentence including grammatical words can thus lead to more syllable-timed rhythm patterns. Sentence stress in this paper is also different from “pitch accent,” which is the stress involving salient pitch prominence that is caused by an important intonational event in intonational phonology (cf. Pierrehumbert, 2000; Silverman et al., 1992). Sentence stresses cause the impression of stress-timed rhythm whereas pitch accent serves to lead a pitch pattern (e.g., head and nuclear tone in the British tradition) in intonation. In Jassem’s (1999) prosodic model, sentence stress, which was termed “tertiary accent,” was clearly distinguished from word stress (“potential for accent”) and pitch accent (“primary or secondary accents”).

Furthermore, because one of the functions of sentence stress is to mark semantically important words, analyzing the speaker’s sentence stress assignment may measure what is perceived as speech rhythm more accurately than simply calculating the degree of temporal variation between intervals (i.e., the previous rhythm metrics). That is, it is also important to assess which words and syllables received sentence stress in a sentence.

As mentioned above, low-proficiency Korean learners tend to place sentence stress on most words in a sentence including grammatical words, whereas high-proficiency learners tend to place sentence stress mostly on content words as native speakers do. To verify this observation using reliable measures, we propose two sentence stress metrics: “accentuation rate” (the ratio of the number of accented words to the total number of test words) and “accentuation error rate” (the ratio of the number of sentence stress errors to the total number of words in test sentences). In the present study, low-proficiency learners were expected to speak English with higher accentuation and accentuation error rates than high-proficiency learners.

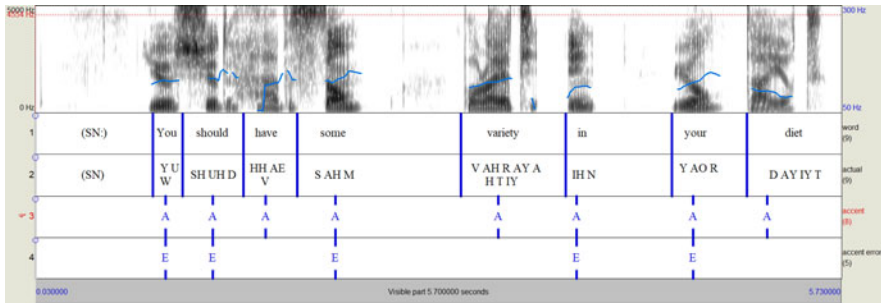


Figure 1. Sample of sentence stress annotation.

To sum up, the aim of the present study was to find robust rhythm measures to validate our hypothesis that speech rhythm of Korean learners of English varies depending on their proficiency levels. This study thus investigated whether scores of our sentence stress metrics and the rhythm metrics (i.e., interval measures and PVIs) correlated with the learners' proficiency levels. We used 525 sentences that were extracted from a large sentence corpus of Korean learners of English. The present study enhanced the reliability of the rhythm measurements by using a large number of sentences and speakers, compared to previous studies, which often used a very small number of sentences but with more controlled syllable structure (e.g., Ramus, Nespor, & Mehler, 1999; White & Mattys, 2007).

Methods

Materials

In order to carry out the rhythm analyses, we used the Korean Learners' English Accentuation Corpus (KLEAC), which was made to develop an automatic sentence stress prediction, detection, and feedback system (Lee et al., 2017). This database consists of recordings of 5,500 English sentences read by 75 Korean learners, who were middle school students aged between 13 and 14. The sentences were made up of words and sentence structures that were appropriate for the students, as they were originally developed as the training corpus of an automatic speech scoring system for secondary school students in Korea.⁴ The speakers of the KLEAC were learning English as an L2 both at school and at a private English institute in South Korea at the time of testing, where the average length of learning was 7–8 years. Seven phonetically trained Korean labelers marked sentence stress on the stressed syllables imposed on the recorded sentences. The annotations were partly crosschecked between the labelers at an early stage, and they had regular meetings to discuss problematic cases that they had faced in the labeling process. The labelers showed very strong interrater agreement rates (e.g., Fless's κ was .868; see Lee et al., 2017, for further details).

Sentence stress errors were also marked on a separate tier as shown in Figure 1. They occurred when a grammatical word was stressed by mistake, when a content word was not stressed, or when stress fell on the wrong syllable. However, it should be noted that some grammatical words like demonstratives (e.g., "this" and "that")

tend to receive sentence stress while some content words like relative adverbs (i.e., “when,” “why,” and “where”) do not (see Kingdon, 1958, for details).

Sentences of the KLEAC database were also different from the materials used in previous rhythm studies (e.g., Arvaniti, 2012; Low et al., 2000). Specifically, syllable complexity, or the distribution of stressed and unstressed syllables within a sentence, was controlled in some of the previous studies to manipulate the durational variability of successive intervals. In contrast, the sentences used in the present study were not contrived as in the previous studies. Instead, 7 sentences were randomly selected from each of the 75 speakers (i.e., a total of 525 sentences; 97 unique sentences), which allowed for a greater number of natural sentences to be used. This large set of sentences can better represent the phonological structures of English (e.g., syllable complexity or the number of unstressed and stressed syllables), thereby increasing the reliability of the metric scores without being too sensitive to characteristics of individual sentences. Average sentence length was approximately 10 syllables.

Speaking proficiency rating task

Seven native English speakers performed the speaking proficiency rating task. They were not phonetically trained annotators, but they were all familiar with Korean-accented English and were living in South Korea at the time of testing. The raters were instructed to listen to seven sentences from each speaker and assess their overall speaking proficiency level on a 5-point scale, with 5 being the most intelligible and nativelike and 1 being the least intelligible and nativelike. The raters were given practice trials at the beginning of the task and were allowed to listen to the sentences as many times as they needed.

The proficiency scale used in this study is simplified from the 6-point Interagency Language Roundtable scale, which is the standard grading scale for measuring speaking proficiency in the United States (see <http://www.govtllr.org/skills/ILRscale2.htm>). Because all the Korean speakers had learned English for 7–8 years at school, we omitted the speaking 0 level (*no proficiency*). As the raters assessed the learners’ overall speaking proficiency, the ratings can reflect not only accentedness and intelligibility (e.g., Munro & Derwing, 1995) but also fluency (e.g., Tavakoli & Skehan, 2005).

Calculation of rhythm metrics

The interval measures used in this study were %V, ΔV , ΔC , VarcoV, and VarcoC. The pairwise variability indices (i.e., rPVI-V, rPVI-C, nPVI-V, and nPVI-C) were also used. The definition of each metric is provided in Table 1.

To calculate these rhythm metrics, vocalic and consonantal intervals of the sentences were initially autosegmented with a forced aligner that extracts the vocalic nucleus of each syllable (Mertens, 2004), and they were then manually modified by two phonetically trained labelers. This was performed following the conventions used in previous studies on the rhythm metrics, such as Ramus et al. (1999) and Grabe and Low (2002). For example, when more than one vowel appeared consecutively, they were labeled as one vocalic interval. This also applied when segmenting consonantal intervals. Pauses were excluded from the analysis. The rhythm metrics were calculated using the software

Table 1. Rhythm metrics used in this study

Rhythm metrics	
%V	Proportion of vocalic intervals in a sentence
ΔV	Standard deviation of vocalic interval duration
ΔC	Standard deviation of consonantal interval duration
VarcoV	Rate-normalized standard deviation of vocalic interval duration
VarcoC	Rate-normalized standard deviation of consonantal interval duration
rPVI-V	Raw pairwise variability index for vocalic intervals
rPVI-C	Raw pairwise variability index for consonantal intervals
nPVI-V	Rate-normalized pairwise variability index for vocalic intervals
nPVI-C	Rate-normalized pairwise variability index for consonantal intervals

Correlatore 2.1 (Mairano & Romano, 2010), and the values of each rhythm metric were averaged across sentences for each speaker in order to perform correlation analyses between their proficiency scores and rhythm metric values.

Calculation of accentuation and accentuation error rates

Sentence stress was also analyzed using the same 525 sentences (i.e., 7 sentences per speaker) from the KLEAC database. As explained, sentence stress and sentence stress errors were annotated in the corpus. In this study, the total number of words in the sentences, the number of sentence stresses imposed, and the number of sentence stress errors were counted for each speaker to calculate the rate of accentuation and accentuation errors as shown in the equations in (1).

(1) Accentuation and accentuation error rates

a. Accentuation rate

$$= \text{Total number of sentence stresses} / \text{Total number of words} \times 100 (\%)$$

b. Accentuation error rate

$$= \text{Total number of sentence stress errors} / \text{Total number of words} \times 100 (\%)$$

We conducted correlation analyses to see if individual speakers' accentuation and accentuation error rates correlated with their proficiency scores. All statistical analyses in this study were performed in R.

Results

Correlation between rhythm metrics and proficiency levels

Correlation analyses were performed for each speaker's average score on each rhythm metric and their average proficiency score. The distribution of the individual speakers' average proficiency scores is displayed in Figure 2. Even though fewer

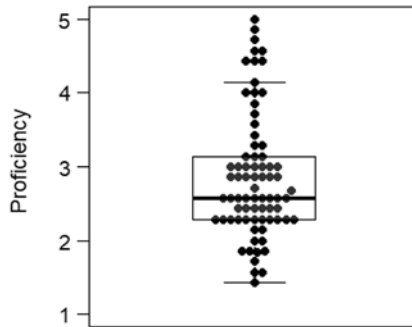


Figure 2. Boxplot showing the distribution of average proficiency scores of 75 learners. Each individual dot represents an average proficiency score of each learner.

speakers received high scores over 3–4 (i.e., the data was positively skewed to some degree), their proficiency scores covered a wide range on the scale (range: 1.429 ~ 5, median: 2.571, mean: 2.866). The results of the correlation analysis showed that out of nine rhythm metrics, four rhythm metrics (ΔV , ΔC , rPVI-V, and rPVI-C) were significantly correlated with proficiency scores, as shown in Table 2. However, as displayed in Figure 3, there were negative relationships between the rhythm metrics and proficiency. That is, speakers who received higher proficiency scores (i.e., more nativelike) had lower rhythm scores, more similar to those of a syllable-timed language. This was the opposite of our expectation. However, it should be noted that the significant correlations were only found with metrics that did not control for speech rate variation, suggesting that the higher rhythm scores of low-proficiency learners were probably related to their slower speech rate. Other metrics failed to show any significant correlation with proficiency.

Correlation between accentuation and accentuation error rates and proficiency levels

Correlation analyses were also carried out for individual learners' accentuation and accentuation error rates and proficiency scores. As shown in Figure 4, there was a significant negative relationship between learners' accentuation rates and proficiency scores (Spearman's $\rho = -.57$, $p < .001$), meaning the higher the learner's proficiency score (i.e., more nativelike), the fewer sentence stresses they produced. Similarly, there was a significant negative correlation between learners' accentuation error rates and proficiency scores (Spearman's $\rho = -.60$, $p < .001$), which means that the higher the learner's proficiency score was, the fewer accentuation errors they made. That is, the perceived level of Korean learners' overall speaking proficiency was highly correlated with their accuracy in producing sentence stress, which is closely related to rhythmic patterns of speech.

Discussion and conclusion

It has previously been observed that rhythmic characteristics in English speech produced by Korean learners vary widely depending on their English proficiency, with more proficient learners producing more nativelike, stress-timed rhythm (e.g., Lee,

Table 2. Results of the correlation analyses performed between proficiency scores and rhythm metrics

Rhythm metrics	Spearman correlation coefficient (ρ value)
%V	$\rho = -.03$ ($p = .8103$)
ΔV	$\rho = -.35$ ($p = .0023$)
ΔC	$\rho = -.46$ ($p < .001$)
VarcoV	$\rho = .04$ ($p = .751$)
VarcoC	$\rho = -.03$ ($p = .794$)
rPVI-V	$\rho = -.35$ ($p = .0024$)
rPVI-C	$\rho = -.49$ ($p < .001$)
nPVI-V	$\rho = -.03$ ($p = .7732$)
nPVI-C	$\rho = -.09$ ($p = .4543$)

2011). However, previous studies have not been able to quantify the rhythm of non-native speakers using reliable measures. The aim of this study was to find metrics that are robust enough to validate this observation. The results showed that Korean learners' rhythmic beat placement as measured by accentuation and accentuation error rates was more nativelike as their speaking proficiency increased.

It thus appears that our method of evaluating non-native speakers' English rhythm using sentence stress metrics can appropriately quantify rhythmic differences between non-native speakers with varying levels of speaking proficiency. The present study also demonstrates that rhythmic characteristics in non-native speech are closely related to judgments of the speaker's speaking proficiency by native listeners, in contrast to what some previous studies have found, such as Iwashita et al. (2008). As the assessment of speaking proficiency in this paper was based on read speech, not on spontaneous conversational speech, it inevitably had some limitations in evaluating the individual learners' speaking proficiency. According to Pearson's technical report on Versant English Test (Pearson Education, 2011), however, there was a high degree of correlation between the scores of the Versant English Test, which is automatically assessed mostly with read and repeated speech, and those of other well-established speaking tests of English, which are evaluated with conversational speech. Hence, basing the assessment on read speech did not appear to pose a serious problem in carrying out this study. In addition, our measure of speaking proficiency can be linked to the degree of foreign accentedness or intelligibility, so it would be interesting to investigate how non-native production of rhythm or sentence stress affects accentedness and intelligibility in future studies.

In contrast, despite using a large set of sentences (i.e., 525 sentences) to calculate the rhythm metrics reliably in the present study, the pairwise variability indices and interval measures did not show the rhythmic differences between Korean learners. The negative correlation found between learners' proficiency levels and some of the rhythm metrics was unexpected. It is unlikely that the rhythm of low-proficiency learners was more similar to that of native speakers. Given that the negative correlation was only found in raw variability indices or interval measures that do not control for speech rate variation (i.e., rPVI-V, rPVI-C, ΔV , and ΔC), it seems that

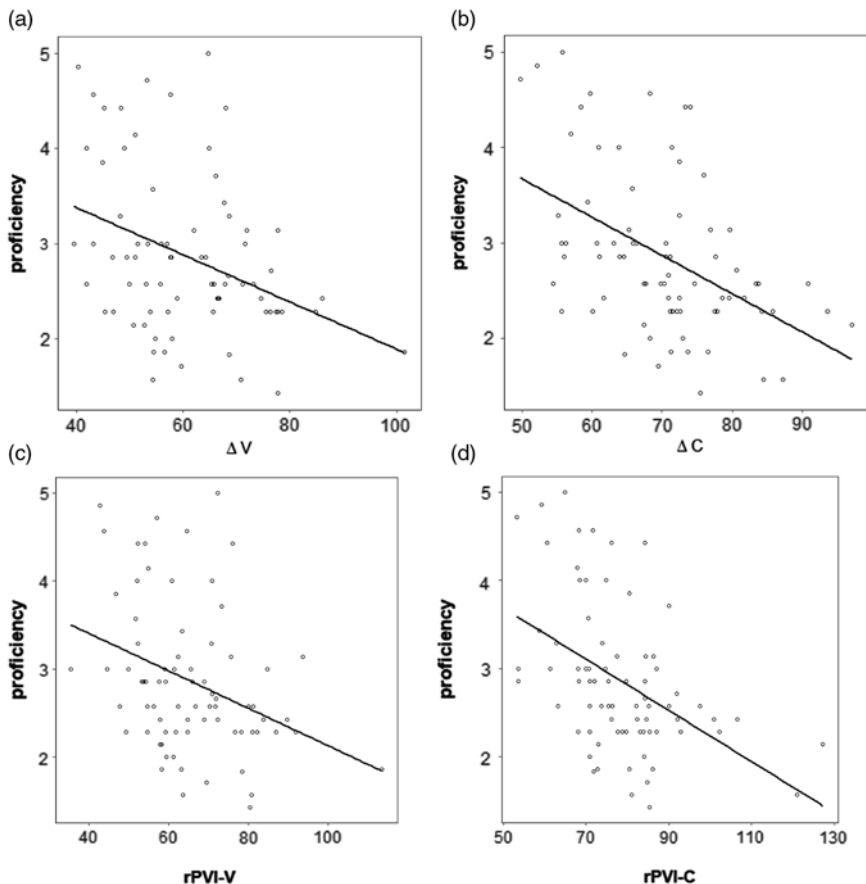


Figure 3. Scatterplots with regression lines showing the relationship between proficiency and (a) ΔV (b) ΔC (c) rPVI-V, and (d) rPVI-C. As the speaker's proficiency score increased, values of ΔV , ΔC , rPVI-V, and rPVI-C decreased.

the high metric scores of low-proficiency learners were driven by slower speaking rate or disfluencies in their speech. That is, long intervals in their speech might have increased the degree of absolute differences between intervals (Dellwo, 2006). Similar results have been found in previous studies due to slower speaking rates of non-native speakers (e.g., Jang, 2008; Lin & Wang, 2007).

One may argue that the rhythm metrics might have accurately captured Korean learners' rhythmic characteristics, which did not differ depending on proficiency. It is also possible to argue that rhythmic characteristics did not simply affect the perception of proficiency. However, such explanations would not be congruent with what we found with the sentence stress metrics; the placement of rhythmic beats by less-proficient Korean learners was more similar to that of syllable-timed rhythm or word-timed rhythm. Although it is difficult to determine the validity of the rhythm metrics solely based on the current findings, it is more likely that the rhythm

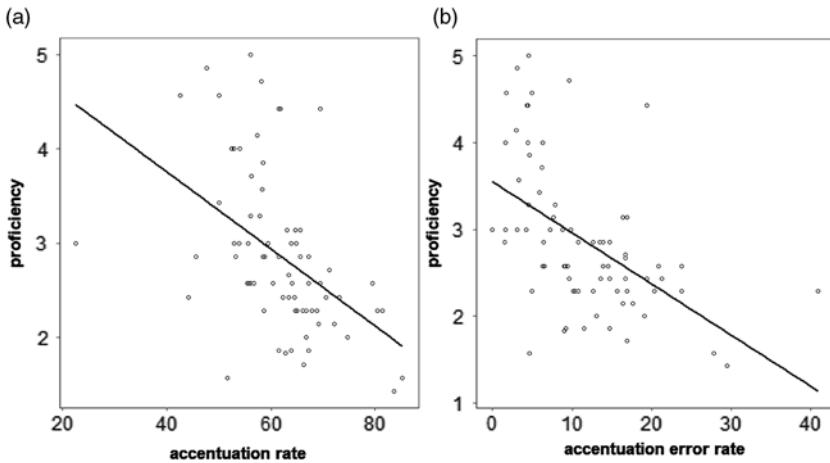


Figure 4. Correlation (a) between the proficiency level and the accentuation rate and (b) between the proficiency level and the accentuation error rate. As the speaker's proficiency score increased, the number of sentence stresses and sentence stress errors they produced decreased.

metrics were not very accurate at capturing rhythmic differences between non-native learners with different levels of proficiency, as well as those between different languages (e.g., Arvaniti, 2012).

The present study suggests that sentence stress can be an effective measure for evaluating Korean learners' English rhythm proficiency. Automatic English speech scoring systems have been developed to assess English learners' speaking proficiency levels (e.g., Bernstein, Van Moere, & Cheng, 2010; Blake, Wilson, Cetto, & Pardo-Ballester, 2008; Chandel et al., 2007; Chen & Zechner, 2011; Johnson, Kang, & Ghanem, 2016; Nielson, 2011; Teixeira, Franco, Shriberg, Precoda, & Sonmez, 2000; Zechner, Higgins, & Xi, 2007; Zechner, Xi, & Chen, 2011). However, most systems have heavily relied on fluency-related features like speech tempo and the number and duration of filled pauses, and rhythmic features have rarely been used except in Chen and Zechner (2011), where rhythm metrics were found to be useful in improving the accuracy of the system.

The results of this paper suggest that incorporating our sentence stress metrics into an automatic English speech scoring system may improve its accuracy. Lee et al. (2017) have developed an automatic sentence stress prediction, detection, and feedback system. This system analyzes learners' sentence stress placement using a detection model that is trained using acoustic, lexical, and syntactic features, then compares it with a reference generated by a prediction model, and offers feedback to the learners on the errors that they made. The accuracy of the prediction and detection of this system reached 96.6% and 84.1%, respectively. This level of accuracy is high enough to suggest that the system can automatically assess accentuation and accentuation error rates, and the system has been shown to be effective in improving learners' accentedness and rhythm (Lee et al., 2017). Furthermore, the results of the present study have important implications for the teaching of English pronunciation in general; focusing on lowering the number of sentence stress errors can be an efficient training method to help learners increase their rhythm proficiency.

Acknowledgments. This work was supported by the Seoul National University Research Grant in 2017. The KLEAC database we used for this study was developed in a research project funded by SK Telecom of South Korea. The authors would also like to thank Dr. Kiduk Yoon and other colleagues for their contribution to the development of the KLEAC database.

Notes

1. Cho (2004) has suggested that Korean is a mora-timed language by showing that listeners could not distinguish rhythmic patterns of Korean from those of Japanese in perception experiments.
2. This was measured in Jang (2008) because function words are normally reduced in English.
3. Similarly, Volin (2017) suggested replacing the term rhythm metrics with durational variation metrics.
4. The recorded sentences were taken from a training corpus that had been built by a Korean company, HCI Lab, to develop a SpeechRater (cf. Zechner et al., 2007) or Versant (cf. Blake et al., 2008) type automatic speech scoring system for secondary school students.

References

- Abercrombie, D. (1967). *Elements of general phonetics*. Edinburgh: Edinburgh University Press.
- Arvaniti, A. (2012). The usefulness of metrics in the quantification of speech rhythm. *Journal of Phonetics*, 40, 351–373.
- Baltazani, M. (2007). Prosodic rhythm and the status of vowel reduction in Greek. *Selected Papers on Theoretical and Applied Linguistics from the 17th International Symposium on Theoretical & Applied Linguistics*, 1, 31–43.
- Benguerel, A. (1999). Stress-timing vs syllable-timing vs mora-timing: The perception of speech rhythm by native speakers of different languages. *VARIA, Etudes & Travaux*, 3, 1–18.
- Bernstein, J., Van Moere, A., & Cheng, J. (2010). Validating automated speaking tests. *Language Testing*, 27, 355–377.
- Blake, R. J., Wilson, N. L., Cetto, M., & Pardo-Ballester, C. (2008). Measuring oral proficiency in distance, face-to-face, and blended classrooms. *Language Learning & Technology*, 12, 114–127.
- Bloch, B. (1950). Studies in colloquial Japanese IV: Phonemics. *Language*, 26, 86–125.
- Carter, P. M. (2004). *The emergence of Hispanic English in the Raleigh community: A sociophonetic analysis* (Unpublished master's thesis, North Carolina State University).
- Carter, P. M. (2005a). Prosodic variation in SLA: Rhythm in an urban North Carolina Hispanic community. *Penn Working Papers in Linguistics*, 11, 59–71.
- Carter, P. M. (2005b). Quantifying rhythmic differences between Spanish, English, and Hispanic English. In R. S. Gess & E. J. Rubin (Eds.), *Theoretical and experimental approaches to Romance linguistics* (pp. 63–75). Amsterdam: Benjamins.
- Chandel, A., Parate, A., Madathingal, M., Pant, H., Rajput, N., Ikbal, S., Verma, A. (2007). *Sensei: Spoken language assessment for call center agents*. Paper presented at the IEEE Workshop on Automatic Speech Recognition & Understanding (ASRU), Tokyo, Japan, December 9–13.
- Chen, L., & Zechner, K. (2011). *Applying rhythm features to automatically assess non-native speech*. Paper presented at the annual conference of the International Speech Communication Association, Florence, Italy, August 27–31.
- Cho, M.-H. (2004). Rhythm typology of Korean speech. *Cognitive Processing*, 5, 249–253.
- Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics*, 11, 51–62.
- Dellwo, V. (2006). Rhythm and speech rate: A variation coefficient for delta C. In P. Karnowski & I. Szigeti (Eds.), *Language and language-processing: Proceedings of the 38th linguistics colloquium, Piliscsaba 2003* (pp. 231–241). Frankfurt am Main, Germany: Lang.
- Deterding, D. (2001). The measurement of rhythm: A comparison of Singapore and British English. *Journal of Phonetics*, 29, 217–230.
- Fuchs, R. (2014). Integrating variability in loudness and duration in a multidimensional model of speech rhythm: Evidence from Indian English and British English. *Proceedings of Speech Prosody*, 7, 290–294.
- Gimson, A. C. (1980). *An introduction to the pronunciation of English*. London: Edward Arnold.

- Grabe, E., & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. In C. Gussenhoven, & N. Warner (Eds.), *Laboratory phonology* Vol. 7 (pp. 515–546). Berlin: de Gruyter.
- Han, M. S. (1962). The feature of duration in Japanese. *Onsei no kenkyuu*, **10**, 65–80.
- Hoequist, C. J. (1983a). Durational correlates of linguistic rhythm categories. *Phonetica*, **40**, 19–31.
- Hoequist, C. J. (1983b). Syllable duration in stress-, syllable- and mora-timed languages. *Phonetica*, **40**, 203–237.
- Iwashita, N., Brown, A., McNamara, T., & O'Hagan, S. (2008). Assessed levels of second language speaking proficiency: How distinct? *Applied Linguistics*, **29**, 24–49.
- Jang, T. Y. (2008). Speech rhythm metrics for automatic scoring of English speech by Korean EFL learners. *Malsori*, **66**, 41–59.
- Jassem, W. (1999). English stress, accent and intonation revisited. *Speech and Language Technology*, **3**, 33–50.
- Jenkins, J. (2000). *The phonology of English as an international language*. Oxford: Oxford University Press.
- Jeon, H.-S. (2015). Prosody. In L. Brown & J. Yeon (Eds.), *Handbook of Korean linguistics* (pp. 41–58). New York: Wiley-Blackwell.
- Johnson, D. O., Kang, O., & Ghanem, R. (2016). Improved automatic English proficiency rating of unconstrained speech with multiple corpora. *International Journal of Speech Technology*, **19**, 755–768.
- Kim, J., Flynn, S., & Oh, M. (2007). Non-native speech rhythm: A large-scale study of English pronunciation by Korean learners. *Studies in Phonetics, Phonology and Morphology*, **13**, 245–275.
- Kingdon, R. (1958). *The groundwork of English intonation*. London: Longman.
- Ladefoged, P. (1975). *A course in phonetics*. New York: Harcourt Brace Jovanovich.
- Lee, G. G., Lee, H. Y., Song, J., Kim, B., Kang, S., Lee, J., & Hwang, H. (2017). Automatic sentence stress feedback for non-native English learners. *Computer Speech and Language*, **41**, 29–42.
- Lee, H. Y. (1990). *The structure of Korean prosody* (Unpublished doctoral dissertation, University College London).
- Lee, H. Y. (2011). *Evaluation of Korean learners' English accentuation*. Paper presented at the 16th National Conference of the English Phonetic Society of Japan and the Second International Congress of Phoneticians of English, Kochi, Japan, November 5–6.
- Lee, O. H., & Kim, J. M. (2005). Syllable-timing interferences with Korean learners' speech of stress-timed English. *Speech Sciences*, **12**, 95–112.
- Lin, H., & Wang, Q. (2007). Mandarin rhythm: An acoustic study. *Journal of Chinese Language and Computing*, **17**, 127–140.
- Lloyd James, A. (1940). *Speech signals in telephony*. London: Pitman & Sons.
- Low, E. L., Grabe, E., & Nolan, F. (2000). Quantitative characterizations of speech rhythm: Syllable-timing in Singapore English. *Language and Speech*, **43**, 377–401.
- Mairano, P., & Romano, A. (2010). Un confronto tra diverse metriche ritmiche usando Correlatore. In S. Schmid, M. Schwarzenbach, & D. Studer (Eds.), *La dimensione temporale del parlato: Proceedings of the V National AISV Congress* (pp. 79–100). Torriana: EDK.
- Mertens, P. (2004). *The Prosogram: Semi-automatic transcription of prosody based on a tonal perception model*. Paper presented at the 2004 Speech Prosody Conference, Nara, Japan, March 23–26.
- Mochizukisudo, M., & Kiritani, S. (1991). Production and perception of stress related durational patterns in Japanese learners of English. *Journal of Phonetics*, **19**, 231–248.
- Mochizukisudo, Z. S., & Fokes, J. (1985). Non-native patterns of English syllable timing. *Journal of Phonetics*, **13**, 407–420.
- Munro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, **45**, 73–97.
- Nazzi, T., Bertoni, J., & Mehler, J. (1998). Language discrimination by newborns: Toward an understanding of the role of rhythm. *Journal of Experimental Psychology*, **24**, 756–766.
- Nguyen, T. A. T. (2003). *Prosodic transfer: The tonal constraints on Vietnamese acquisition of English stress and rhythm* (Unpublished doctoral dissertation, University of Queensland).
- Nielson, K. B. (2011). Self-study with language learning software in the workplace: What happens? *Language Learning & Technology*, **15**, 110–129.
- Pearson Education. (2011). *Versant English Test: test description and validation summary*. Palo Alto, CA.
- Pike, K. (1945). *The intonation of American English*. Ann Arbor, MI: University of Michigan Press.

- Pierrehumbert, J. B.** (1980). The phonology and phonetics of English intonation (Doctoral dissertation, Massachusetts Institute of Technology).
- Port, R. F., Dalby, J., & O'Dell, M.** (1987). Evidence for mora-timing in Japanese. *Journal of the Acoustical Society of America*, **81**, 1574–1585.
- Ramus, F., Dupoux, E., & Mehler, J.** (2003). *The psychological reality of rhythm class: Perceptual studies*. Paper presented at the 15th International Congress of Phonetic Sciences, Barcelona, Spain, August 3–9.
- Ramus, F., Nespor, M., & Mehler, J.** (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, **73**, 265–292.
- Roach, P.** (1982). On the distinction between “stress-timed” and “syllable-timed” languages. In D. Crystal (Ed.), *Linguistic controversies* (pp. 73–79). London: Arnold.
- Setter, J.** (2006). Speech rhythm in world Englishes: The case of Hong Kong. *TESOL Quarterly*, **40**, 763–782.
- Silverman, K., Beckman, M., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., & Hirschberg, J.** (1992). *ToBI: A standard for labeling English prosody*. Paper presented at the International Conference on Spoken Language Processing, Alberta, Canada, October 13–16.
- Stockmal, V., Markus, D., & Bond, D.** (2005). Measures of native and non-native rhythm in a quantity language. *Language and Speech*, **48**, 55–63.
- Tavakoli, P., & Skehan, P.** (2005). Strategic planning, task structure, and performance testing. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 239–273). Amsterdam: Benjamins.
- Teixeira, C., Franco, H., Shriberg, E., Precoda, K., & Sonmez, M. K.** (2000). *Prosodic features for automatic text-independent evaluation of degree of nativeness for language learners*. Proceedings of 6th International Conference on Spoken Language Processing, Beijing, China, October 16–20.
- Trofimovich, P., & Baker, W.** (2006). Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition*, **28**, 1–30.
- Volin, J.** (2017). Appeal and disrepute of the so-called global rhythm metrics. *Acta Universitatis Carolinae Philologica*, **3**, 79–94.
- Wenk, B., & Wioland, F.** (1982). Is French really syllable-timed? *Journal of Phonetics*, **10**, 193–216.
- White, L., & Mattys, S. L.** (2007). Calibrating rhythm: First language and second language studies. *Journal of Phonetics*, **35**, 501–522.
- Wiget, L., White, L., Schuppler, B., Grenon, I., Rauch, O., & Mattys, S. L.** (2010). How stable are acoustic metrics of contrastive speech rhythm? *Journal of the Acoustical Society of America*, **127**, 1559–1569.
- Zechner, K., Higgins, D., & Xi, X.** (2007). *SpeechRater™: A construct-driven approach to scoring spontaneous non-native speech*. Paper presented at the Workshop on Speech and Language Technology in Education, Farmington, PA, October 1–3.
- Zechner, K., Xi, X., & Chen, L.** (2011). *Evaluating prosodic features for automated scoring of non-native read speech*. Paper presented at the 2011 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU), Big Island, HI, December 11–15.