A TEMPORAL FEATURE OF STRESSED SYLLABLES IN NATIVE ENGLISH AND KOREAN EFL SPEECH

Dokyung Kwon, Tae-Yeoub Jang

Hankuk University of Foreign Studies
dkkwon@hufs.ac.kr, tae@hufs.ac.kr

ABSTRACT

This study investigates English speech rhythm, focusing on the variations in the length of stressed syllables in English utterances of native speakers (NAT) and Korean EFL learners (EFL). Two corpora are used to measure the duration ratio of vowels in stressed syllables and compare it between NAT and EFL as well as among EFL groups divided by three proficiency levels. The rationale for using stressed vowels is that NAT tend to lengthen stressed syllables and shorten unstressed ones while EFL do not in their English speech due to L1 interference. Thus, it is hypothesized that NAT would show the highest ratio value which would fall as proficiency levels of EFL decrease. The statistical results indicate that the ratio is useful in capturing speaker group differences. Although there is room for further investigation, this study contributes to efforts in developing auto-scoring systems of English speech by presenting a temporal feature.

Keywords: L2 rhythm, rhythmic feature, vowel duration, stress, auto-scoring, proficiency level

1. INTRODUCTION

Prosody has a significant effect on the production of L2 speech. In particular, L2 speakers whose native language, as Korean, does not utilize a phonemic stress system fails to implement the correct speech rhythm when producing utterances of a stress-based target language like English. Empirical studies have shown that proper stress placement plays a crucial role in improving intelligibility and/or comprehensibility of L2 speech. (e.g., [3] and [6]). Thus, discovering rhythmic features based on stress will facilitate explanation of a variety of phenomena related to L2 acquisition.

When examining English L2 rhythm produced by Korean learners, focusing on the temporal aspect of syllables, or specifically vowels, could lead to meaningful findings. This is because English rhythm is closely associated with vowel reduction, where vowels lacking in stress are produced with shorter durations unlike Korean vowels that do not show such variations in length depending on characteristics of stress. Based on this notion, Kwon [7] compares the duration of vowels between English natives and Korean learners, targeting specifically those in unstressed positions. She found that unstressed vowels in English utterances of Korean learners were longer than those in native speech. One limitation of this study is that it looks only at differences in raw duration, failing to take account of other possible factors involved such as the speech rate effect.

Instead of employing raw measures, investigating proportional differences of vowel durations could be more effective. For instance, Jung & Rhee [5] examine English utterances of L2 speakers (Korean, Japanese, and Taiwanese) using the duration ratio of stressed vowels to unstressed vowels. Nakamura [8] have also found the ratio to be effective in characterizing English L2 rhythm of Japanese. However, only a few studies have looked at the measure of vowel durations across different proficiency levels. Galaczi et al. [4] provide a thorough analysis on the speakers of different L1 language backgrounds, though investigating a relatively small number of tokens for each language group. By utilizing various methods to calculate the ratio of stressed vowel duration, and by taking into account the proficiency levels of Korean L2 speakers, this study provides a comprehensive analysis of duration ratio as a prosodic feature. It is expected that native speakers would show the highest ratio of stressed vowels unlike Korean EFL learners who are assumed to be less apt at distinguishing between stressed and unstressed vowels due to differences in the rhythmic properties of the two languages. It is predicted that the effect of this difference would interact with the speakers' levels of proficiency, with those from lower levels showing lower ratio values as compared to those of higher levels.

2. DATA AND METHOD

2.1. Corpus

This study utilizes two corpora to examine whether stressed vowel ratios can be used to distinguish native speakers of English (NAT) from Korean EFL learners (EFL) as well as the EFL group divided by proficiency levels. The first corpus is PRAWN_dB

developed by Chung et al. [2]. The data consist of English words, phrases, and sentences read by 20 native speakers and 60 Korean learners. From the total of 16,160 tokens, 1,760 tokens were extracted for the analysis. These tokens come from 22 declarative sentences that are 7 to 19 syllables (6 to 13 words) in length, with 440 tokens from native speakers and 1,320 from Korean learners.

For the current analysis, the Korean speakers were divided into three proficiency levels. Basically, each speaker's level was defined by his/her self-assessed confidence, but all the speech tokens were re-examined by an experimenter to adjust the level when an obvious misclassification was observed. Table 1 below summarizes the number of subjects in each level.

Table 1: Number of Korean speakers in each proficiency level from the PRAWN data.

Proficiency level	Number of speakers
High	14
Mid	33
Low	13
Total	60

Another corpus was used for a more reliable source of proficiency ratings. This was the AESOP (Asian English Speech cOrpus Project) corpus, developed to study variations in Asian EFL speech [9]. It contains read speech and semi-spontaneous responses from EFL speakers from Hong Kong, Taiwan, Japan, Korean, Thailand and Vietnam. A total of 600 tokens (200 speakers X 3 sentences per speaker from 6 types of sentences) of Korean EFL speakers' utterances were provided with files of auto-segmented phones and words in TextGrid format of Praat [1].

Table 2: Number of tokens in each proficiency level from the AESOP data.

Proficiency level	Number of Tokens
High	159
Mid	383
Low	57
Total	600

Fluency ratings for each token were also provided, where four trained raters evaluated the English utterances of Korean speakers on a scale of one to five, with greater numbers indicating better performance. If the average values were less than three the tokens were considered to be at the low level, while values less than four were classified at the middle level, and the rest were grouped as high. The number of tokens in each level is described in Table 2.

2.2. Calculation

Each stressed syllable in content words is defined as with either primary or secondary stress. No syllables in function words are considered to be stressed assuming that their prominence is minimal. It is also important to note that this study uses phonologically pre-defined concept of stress. Although there are many variations to stress depending on the context, fixed information had to be used for the calculation of the durational ratio in a large set of data.

There are three types of stressed vowel ratios used in the analysis. They can be described by the following formula:

(1) $Stressed\ vowel\ ratio = \frac{\sum_1^n V_n}{X}$

V_n: Duration of n^{th} stressed vowel

X: Measures of comparison

For a number of stressed vowels existing in an utterance, the sum of their durations is divided by X that may refer to the total utterance duration (S2T), the total duration of unstressed vowels in an utterance (S2U), or the total duration of vowels (S2V). The purpose of using three different features is to examine their behaviors in different contexts and to find the most robust one in capturing the rhythmic differences between the speech of natives and Korean learners.

3. RESULTS

3.1. Results from the PRAWN Data

The difference between the ratios of NAT and the entire EFL in the PRAWN data was examined first. An analysis of the means indicates that NAT shows higher ratio values than EFL, just as originally hypothesized. The distribution of the ratios by the speaker groups is provided in Table 3.

Table 3: Means (Standard deviations) of NAT and EFL in the PRAWN data.

Ratio	NAT	EFL
S2T	0.16 (0.06)	0.15 (0.07)
S2U	1.53 (0.97)	1.23 (1.74)
S2V	0.55 (0.16)	0.48 (0.18)

All the ratios of NAT are higher than those of EFL, with S2U showing the largest difference in means between the groups (0.301) as it varies the most, followed by S2V (0.066) and S2T (0.014). The differences analyzed with the permutation test were found to be meaningful for S2T ($Z = -3.908$, $p < 0.001$), S2U ($Z = -6.799$, $p < 0.001$), and S2V ($Z = -7.030$, $p < 0.001$). Moreover, fitting and comparing linear mixed-effects (LME) models yielded a significant result for S2T (χ^2 (1) = 7.262, p = 0.007), S2U (χ^2 (1) = 9.087, p = 0.002), and S2V (χ^2 (1) = 36.751, $p < 0.001$).

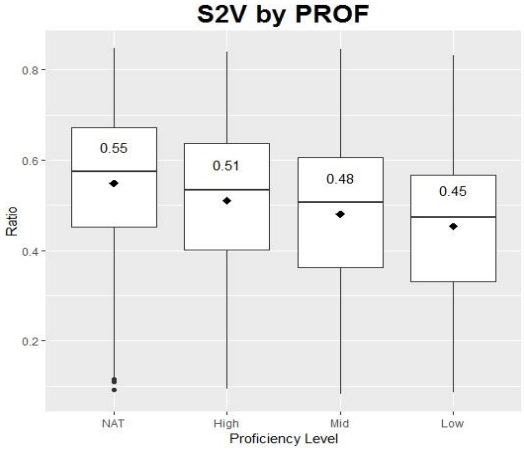
Korean EFL speakers were grouped into the three proficiency levels of high, mid, and low. These groups, along with NAT, are referred to as the factor PROF in the analysis of the data from PRAWN, and were compared across each of the ratios.

Table 4: Means (Standard deviations) of PROF in the PRAWN data.

Ratio	High	Mid	Low
S2T	0.16 (0.07)	0.15 (0.07)	0.13 (0.06)
S2U	1.36 (0.96)	1.17 (0.85)	1.04 (0.76)
S2V	0.51 (0.18)	0.48 (0.17)	0.45 (0.17)

As indicated in the above table, the mean values of the ratios all decreased as the proficiency level went from high to low. Boxplots showing the distribution of S2V for each group is provided below to visualize this falling pattern.

Figure 1: Boxplots of PROF (NAT, High, Mid, Low) with mean values of S2V labeled.



The results were verified using the Fisher-Pitman permutation test and fitting the LME model. The results of the permutation test indicate that the difference is significant for S2T ($\chi^2 = 23.2$, $p < 0.001$), S2U ($\chi^2 = 60.682$, $p < 0.001$), and S2V ($\chi^2 = 61.463$, $p < 0.0001$). The results of the fitted LME model show that there is a significant effect of PROF for S2T (χ^2 (3) = 13.992, p = 0.003), S2U (χ^2 (3) = 51.289, $p < 0.001$), and S2V (χ^2 (3) = 55.47, $p < 0.001$).

The post hoc tests were performed on both results from the permutation test and LME model analysis. For all three types of ratios used in the permutation test, the comparison results between each group are meaningful except for S2T between NAT and EFL-high, and between EFL-mid and EFL-low. S2T's power as a discriminative feature may be offset by the variations found in the total utterance duration.

In the LME analysis, significance was found between NAT and EFL-mid, between NAT and EFL-low, and between EFL-high and EFL-low for S2T. Conversely, the results of S2U and S2V are similar in that the differences between all of the groups were meaningful except for the pair EFL-mid and EFL-low.

3.2. Results from the AESOP Data

The factor marked as PROF differs across the AESOP and PRAWN data. I evaluated the level of proficiency for individual speakers in the extracted data from PRAWN, while individual utterances in AESOP were rated by four different evaluators. Also, there were no native participants represented in the AESOP data, so only three groups are compared: high, mid, and low. The distributions of the ratios exhibited by the three groups are shown in Table 5.

Table 5: Means (Standard deviations) of PROF in the AESOP data.

Ratio	High	Mid	Low
S2T	0.20 (0.06)	0.20 (0.05)	0.21 (0.05)
S2U	1.12 (0.79)	0.93 (0.51)	0.87 (0.38)
S2V	0.48 (0.15)	0.45 (0.13)	0.44 (0.12)

The mean values of S2T show a different pattern in that the low group's S2T is the highest, and that of the mid group is the lowest. This may suggest that the effect of S2T is offset by the fact that the low level speakers speak more slowly with longer durations. S2U and S2V exhibit a tendency similar to that of the previous results, showing the ratios decreasing with the level of proficiency, although the amount of decrease is much smaller.

The results from the Fisher-Pitman permutation test indicate a significant effect for S2U ($\chi^2 = 11.67$, $p = 0.002$), but not for S2T ($\chi^2 = 4.361$, $p = 0.114$) or S2V ($\chi^2 = 4.365$, $p = 0.108$). Fitting to an LME model leads to different outcomes where the differences between the groups are found to be meaningful for S2U (χ^2 (2) = 31.482, $p < 0.0001$) and S2V (χ^2 (2) = 34.411, $p < 0.0001$), but not for S2T (χ^2 (2) = 3.325, $p = 0.190$). The results of the two statistical analyses can differ since the LME model controls for random effects while the permutation test does not. In this sense, fitting LME models is advantageous for returning more reliable results.

No significance was found for S2T or S2V, while S2U distinguished between EFL-high and EFL-mid, and between EFL-high and EFL-low. S2T could not distinguish between any of the groups in PROF, whereas S2U and S2V behaved similarly in that meaningful results were found between all the pairs except EFL-mid and EFL-low. The same pattern was found in the analysis of the data from PRAWN, which makes the results more reliable.

4. DISCUSSION

The results generally support the hypothesis, indicating that these ratios can discriminate between speakers, especially between EFL and NAT. It is also to note that S2U seem to be distinguishing the speaker groups most effectively. Nevertheless, EFL-mid and EFL-low showed marginal differences between the groups in PROF. An instant impression from these results is that speakers of the lower levels do not differ in their pronunciation of vowels in terms of duration. In other words, speakers of both the low and mid-levels produce English vowels of comparable length, a characteristic that contrasts with the high level and native speakers, who place prominence on stressed syllables by lengthening their vowels. This would have to be verified by further research, as there still remains the possibility that the results from this study were affected by the data used, such as the restriction to solely auto-aligned information and the limited number of tokens for EFL-low.

Notwithstanding the above constraints, this study contributes to ongoing efforts to characterize the features of L2 rhythm by taking into account various methods to calculate the stressed vowel ratios and the proficiency level of Korean learners. The two sets of data used in this study differ in their definition of proficiency level, yet showed rather consistent results. Identifying the effect of sentences and using a larger data set for the EFL-low group could enhance the performance of the ratios in future work. In addition, the findings of this study have implications for auto-scoring systems of L2 speech where measurements and weightings of pronunciation features play an important role. This study showed that the stressed vowel ratios, calculated from auto-segmented data, could distinguish between the speaker groups, meaning that the result can be directly implemented into auto-scoring systems of Korean EFL speech. Utilizing the duration ratio of stressed vowels as a feature in such systems would lead to meaningful results.

5. ACKNOWLEDGEMENTS

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).

6. REFERENCES

[1] Boersma, P. 2001. Praat, a system for doing phonetics by computer. *Glot International* 5:9/10, 341-345.

[2] Chung, H., Jang, T. Y., Yun, W., Yun, I., Sa., J. 2008. A study on automatic measurement of pronunciation accuracy of English speech produced by Korean learners of English. *Language and Linguistics* 42, 165-196.

[3] Field, J. 2005. Intelligibility and the listener: The role of lexical stress. *TESOL Quarterly* 39, 399-423.

[4] Galaczi, E., Post, B., Li, A., Barker, F., Schmidt, E. 2017. Assessing second language pronunciation: Distinguishing features of rhythm in learner speech at different proficiency levels. In: T. Issacs, T., Trofimovich, P. (eds), *Second Language Pronunciation Assessment*, 157-182.

[5] Jung, Y. J., Rhee, S. C. 2018. Acoustic analysis of English lexical stress produced by Korean, Japanese, and Taiwanese-Chinese speakers. *Phonetics and Speech Sciences* 10(1), 15-22.

[6] Kang, O., Rubin, D. L., Pickering, L. 2010. Suprasegmental measures of accentedness and judgements of English language learner proficiency in oral English. *Modern Language Journal* 94, 554-566.

[7] Kwon, H. 2007. A phonetic analysis of Korean EFL learners' production of English reduced vowel. *SNU Working Papers in English Linguistics and Language* 6, 1-13.

[8] Nakamura, S. 2011. Characteristics of constrast between the stressed and the unstressed in rhythm units observed in duration structure in English speech by Japanese learners. *Pan-Pacific Association of Applied Linguistics* 15(1), 177-189.

[9] Visceglial, T., Tseng, C., Kondo, M., Meng, H., Sagisaka, Y. 2009. Phonetic aspects of content design in AESOP (Asian English Speech cOrpus Project). *2009 Oriental COCOSDA International Conference on Speech Database and Assessments* Urumqi, 60-65.