

The Validity and Reliability of Rhythm Measurements in Automatically Scoring the English Rhythm Proficiency of Chinese EFL Learners

Jin Chen¹, Jianghao Lin² & Xinguang Li³

¹ Faculty of English Language and Culture, Guangdong University of Foreign Studies, Guangdong, China

² Financial Department, Guangdong University of Foreign Studies, Guangdong, China

³ Laboratory for Language Engineering and Computing, Guangdong University of Foreign Studies, Guangdong, China

Correspondence: Jin Chen, Faculty of English Language and Culture, Guangdong University of Foreign Studies, No. 2 Baiyun Avenue (N), Guangzhou, Guangdong, 510420, China. E-mail: chenjin_gdufs@163.com

Received: March 3, 2015 Accepted: April 7, 2015 Online Published: May 28, 2015

doi:10.5539/elt.v8n6p80 URL: <http://dx.doi.org/10.5539/elt.v8n6p80>

Abstract

This article aims to find out the validity of rhythm measurements to capture the rhythmic features of Chinese English. Besides, the reliability of the valid rhythm measurements applied in automatically scoring the English rhythm proficiency of Chinese EFL learners is also explored. Thus, two experiments were carried out. First, thirty students of English major and five native English speakers were selected to read ten English sentences. The participants were divided into four proficiency groups according to human scoring. Then seven previously proposed rhythm measurements were investigated in four proficiency groups. One-way ANOVA results showed that five rhythm measurements were valid to distinguish different English rhythm patterns among four proficiency groups. Based on the valid measurements, an experiment of automatic scoring for English rhythm proficiency was also conducted through statistical technique Multiple Regression. The correlation coefficient between the autoscores and the scores made by experienced teachers reached 0.866. The result showed a high reliability of the objective evaluation for English rhythm proficiency of Chinese EFL learners.

Keywords: automatic assessment, rhythm measurements, reliability, validity

1. Introduction

Acoustic-phonetic rhythm measurements have successfully identified rhythmic features of languages in L1 studies. Recently, researchers begin to apply these measurements in L2 studies and find some of them are valid to recognize the rhythmic characteristics of second languages. Furthermore, some researchers try to employ the valid measurements in autoscoring the EFL learners' rhythm proficiency. Up to now, however, few empirical studies have probed into Chinese English and Chinese EFL learners. Thus, this article aims to find out the validity of rhythm measurements to capture the rhythmic features of Chinese English. Besides, the reliability of the valid rhythm measurements applied in automatically scoring the rhythm proficiency of Chinese EFL learners is also explored.

1.1 Rhythm

Rhythm is one of the three aspects of prosody, along with stress and intonation. According to Zhang (2002), rhythm refers to the basic recurrence of elements or features in alternation with opposite or different elements or features. And speech rhythm is essentially a tendency for the stressed syllables to occur at more or less regular intervals of time.

Every language has its own characteristic rhythm. Initiated by Pike (1945) and Abercrombie (1967), languages of the world can be classified into three rhythmic categories, namely, stress-timed, syllable-timed and mora-timed from the perspective of human perception. These categories are defined on the hypotheses about units of equal duration. Roach (1982) believes that stress-timed languages exhibit more nearly equal intervals between stresses or rhythmic feet, syllable-timed languages display near isochrony between successive syllables and mora-timed languages have nearly isochronous mora.

Speech rhythm in English is said to be stress-timing. Wang (2002) claims that English rhythm is influenced by some factors like stress, linking, assimilation, elision, and weak forms. When it comes to Chinese rhythm, Gui (1985) believes that speech rhythm in Chinese is syllable-timing. Every word is read explicitly except few weak auxiliary words. Besides, the interval of time of each syllable is relatively equal. Due to the first language negative transfer, the English spoken by many Chinese EFL learners is an intermediate language whose rhythm tends to be more syllable-timed rather than stress-timed. The most distinctive characterization of Chinese English dwells on stress. Lin (2007) points out most Chinese EFL learners usually mistake the stress pattern in a word and they prefer to give equal stress to every English word in a sentence. Besides, many Chinese EFL learners are poor in weak forms and linking and thus the English they speak have more pauses and hesitation than that of the native speakers. To sum up, Chinese EFL learners are not good at the skills of adjusting English rhythm like stress, linking, assimilation, elision, and weak forms. Hence the English they speak is more syllable-timed whose syllable structures appear to be less complex than that of the native English speakers.

1.2 Rhythm Measurements

Phoneticians have been recently interested in research on acoustic-phonetic measurements of rhythmic structure of languages with an aim to allow the tendency towards stress- or syllable-timing to be derived from the measurements. For the convenience of description, seven rhythm measurements successfully proposed in L1 studies can be classified into three kinds of measurements according to their measuring method. They are raw interval measurements (RIM), rate-normalized interval measurements (NIM) and Pairwise Variability Indices (PVI).

Based on the observation that stress-timed languages have a more complex and variable syllable structure than syllable-timed languages, Ramus, Nespors, and Mehler (1999) propose three measurements by measuring temporal characteristics of vocalic and consonantal intervals. Thus, three raw interval measurements (RIM)—the proportion of vocalic intervals (%V) (Note 1), the standard deviation of the vocalic (ΔV) and consonantal (ΔC) intervals are calculated. Among them, the results of %V are predicted to be larger in syllable-timed languages than in stress-timed languages while the results of ΔV and ΔC are the opposite.

Later, some studies (Barry, Andreeva, Russo, Dimitrova, & Kostadinova, 2003; Dellwo & Wagner, 2003) have found that ΔC varied considerably by speech rate at least in some languages including English and German. If this is the case, speech rate normalization (Note 2) of target utterances seems urgent when ΔC and ΔV are to be used. Hence, Dellwo (2006) puts forward the rate-normalized version of consonantal variability, that is, VarcoC. Soon, VarcoV, the rate-normalized standard deviation of vocalic interval duration is also added by White and Mattys (2007) to fill up the Rate-normalized Interval Measurements (NIM) inventory. The calculation formulas are presented as follows:

Calculation formula of VarcoC:

$$\text{VarcoC} = \frac{\Delta C \times 100}{\text{meanC}} \quad (1)$$

Calculation formula of VarcoV:

$$\text{VarcoV} = \frac{\Delta V \times 100}{\text{meanV}} \quad (2)$$

And their results are the same as those of ΔV and ΔC , predicted to be smaller in syllable-timed languages than in stress-timed languages.

Additionally, based on the observation that stressed and unstressed vowels in languages employing stress rhythm vary widely in duration whereas the durations of vowels in syllable rhythm languages vary less, Low, Grabe, and Nolan (2000) thus introduce Pairwise Variability Index (nPVI). Later, Grabe and Low (2002) add another measurement to their Pairwise Variability Index, based on the variability of consonantal intervals (rPVI). The calculation formulas are presented as follows:

Calculation formula of nPVI (d_k : duration of kth interval; m: number of intervals):

$$\text{nPVI} = 100 \times \left(\sum_{k=1}^{m-1} \left| \frac{d_k - d_{k+1}}{d_k + d_{k+1}} \right| \right) / (m - 1) \quad (3)$$

Calculation formula of rPVI:

$$rPVI = \frac{(\sum_{k=1}^{m-1} |d_k - d_{k+1}| D)}{(m-1)} \quad (4)$$

Due to factors like frequent vowel reductions and linking within words etc., the Pairwise Variability Indices (PVIs) are predicted to be larger in stress-timed languages than in syllable-timed languages.

1.3 Previous Empirical Studies on Rhythm Measurements

The above rhythm measurements have been successfully applied into the monolingual studies (like Grabe & Low, 2002; Dellwo, 2006; White & Mattys, 2007; Ramus et al., 1999; etc.) to identify different first languages. However, it seems that they are not so uniform to differentiate the non-native languages. Low, Grabe, and Nolan (2000) found nPVI rather than rPVI could differentiate Singapore English and British English. Stockmal, Markus, and Bond (2005) reported that ΔC and rPVI may significantly distinguish the language Latvian spoken by native speakers and by Russian Latvian learners of different proficiencies while %V, ΔV and nPVI showed no significant difference among these groups. White and Mattys (2007), with native English speakers and Spanish learners of English as their participants, revealed that measurements like %V and VarcoV were more useful for non-native speech rhythm detection than ΔV , ΔC , nPVI and rPVI. The diverse results of all empirical studies show that rhythm measurements to distinguish different non-native rhythm characteristics are not unified. Vowel-based measurements appear to be more suitable in detecting the rhythm features of Singapore English, Spanish English while consonant-based measurements more effective in capturing that of Russian Latvian. However, thus far, few empirical studies focus on the non-native English spoken by Chinese EFL learners except Chen and Wang (2013). But this issue should be investigated, because, on one hand, the result of the rhythm measurements on the non-native rhythms is not so uniform. On the other hand, more and more Chinese people have begun to learn English nowadays, so it would be valuable to study the characteristics of English rhythm among this increasingly enlarged group.

Furthermore, some phoneticians recently try to employ acoustic-phonetic rhythm measurements in autoscoring the quality of EFL learners' oral language. Chung, Jang, W. Yun, I. Yun, and Sa (2008) first makes use of the rhythm measurements to autoscore the pronunciation accuracy of English speech produced by Korean learners of English. On this basis, another Korean researcher Jang (2008) further improves the experiment and clearly suggests an autoscoring method—Multiple Regression for English oral proficiency. The results of their experiments are not so convincing because the characteristics of rhythm is not enough to reflect the proficiency of English pronunciation. But these attempts illustrate the possibility of rhythm measurements for automatic scoring. Given that no empirical study has explored the autoscoring for Chinese EFL learners, this article tries to improve the existing experimental method as to autoscoring the English rhythm proficiency of Chinese EFL learners. And it may provide a theoretical foundation for Computer-Assisted Language Learning System for English Rhythm of Chinese EFL learners.

In order to achieve the above purposes, the current study is going to provide answers to the following questions:

- 1) What rhythm measurements successfully proposed in L1 studies is valid to capture the English rhythm patterns of Chinese EFL learners with different proficiencies?
- 2) How reliable are the valid rhythm measurements in automatically scoring the English rhythm proficiencies of Chinese EFL learners?

2. Method

2.1 Participants

Thirty students of English major from four different grades in the Faculty of English Language and Culture (FELC), Guangdong University of Foreign Studies were selected. There were 23 female students and 7 male students. In order to provide a stress-timed rhythm baseline, five native speakers from Britain were also chose. Two of them were male and the other three were female.

2.2 Instruments

The reading materials (see Appendix A) were ten sentences whose numbers of words ranged from 3 to 9. Considering the students' English level, the selected sentences had no infrequent words, no complex sentence structures. All of the sentence type was declarative sentence, which was widely used in spoken English. In order to properly exhibit the characteristics of English rhythm, each sentence was made to include at least one factor that can adjust the rhythm like linking, elision, assimilation and weak form.

In addition, the present study made use of the phonetic research tool Praat to segment the recordings of the 35 subjects. Mathematical software Matlab and Excel were also used to calculate the results of the measurements based on the segmented information. Finally, the data were analyzed through SPSS 17.0 and Matlab as well.

2.3 Procedures

2.3.1 Recording

The subjects were asked to read ten sentences. They were given time to read the sentences before the recordings were made. The non-native recordings were made in a recording studio and the native recordings were made in a quiet room. All the recordings were of good quality and had little noise.

2.3.2 Human Scoring

The English utterances spoken by Chinese EFL learners are not the same rhythm pattern. A research conducted by Feng (2010) discovered that Chinese EFL learners with different proficiencies of rhythm had different rhythmic patterns. The better English rhythms the Chinese EFL learners have the more stress-timed rhythm they will display. So it is necessary to divide all the learners into several representative groups according to their rhythm proficiencies before the experiment.

The present study adopted the scoring approach of Absolute Scales proposed by Diekerson (1997). This approach means that the scoring is based on one standard regardless of the students' different learning years of English, improvements etc. According to Wang (2002), stress is the basis of English rhythm. And English rhythm is well embodied in other factors like linking, assimilation, elision, and weak forms. Thus a standard for scoring the rhythm proficiency was drawn up (see Table 1).

Table 1. Standard for rhythm evaluation

Levels	Standard
A	Accurate stress; to be adept with pronunciation techniques like linking, assimilation, elision, and weak forms; to exhibits the stress-timed rhythm well.
B	One or two inaccurate stress; to use linking, assimilation, elision, and weak forms such techniques sometimes; to be influenced by syllable-timed rhythm to some extent.
C	Poor stress; to adjust English rhythm withoutany techniques; quite similar to the syllable-timed rhythm.

The recordings of 30 Chinese EFL learners were scored by one Chinese experienced university English teacher and one British university teacher. Two teachers separately gave an overall score to each student according to their English rhythm presented in the ten sentences the students have recorded. Although an evaluation standard has been put forward, it is still necessary to test the reliability of the scores in case two teachers may have inconsistent opinions towards the standard. Thus, Pearson correlation and T-test were conducted by SPSS 17.0.

The correlation between the scores made by two teachers significantly reached .693, indicating that two teachers had a consistent standard when scoring. Besides, a pair-samples t-test was run to see whether there was significant difference between the evaluations of two teachers, with the teachers as two groups and the scores as the variables. The result of the t-test showed that the difference between two teachers was not significant, $t(29) = 1.682, p = .103$. Thus the scores were basically reliable for the following data analysis.

In order to make a parallel comparison, three proficiency groups, namely level A, level B and level C were designed to consist of five speakers respectively as to have the same number with the group of native speakers. Hence, fifteen non-native samples, each of whom had the same scale evaluated by two teachers, were selected to fall on three proficiency groups proportionally.

2.3.3 Segmenting and Calculating

The recordings were segmented by Praat. Every consonant and vowel in sentences was segmented and then the segmented information was saved in Excel. Next, the results of the rhythm measurements for every subject were calculated by Excel and Matlab based on the segmented data.

2.3.4 Analyzing the Data

After the data were collected, one-way ANOVA was conducted by using SPSS 17.0 to find out the valid

measurements which can distinguish the native and non-native English rhythm patterns.

With the valid measurements found in the results of experiment one, an experiment of automatic scoring was also performed by the statistical technique called Multiple Regression. This technique has been widely applied in automatic assessment for essay writing (Page, 1994; Page & Petersen, 1995; etc.) and began to be employed in automatic scoring for oral English by Jang (2008). Its equation can be represented as follows,

$$y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_nx_n \quad (x_n: \text{features}) \quad (5)$$

In the present study, the valid measurements were regarded as features to score the overall rhythm ability. Besides, the estimation of coefficients was conducted by Partial Least Squares in the mathematical software Matlab. Hence, twelve samples in four groups were randomly chosen as the training data in order to estimate the regression coefficients (β s).

The rest of the eight samples from the four groups were used as the testing data and applied in the Multiple Regression. After the autoscores for these eight samples were calculated, the Person correlation was conducted to see the correlation between the autoscoring results and the human scores.

3. Results and Discussion

3.1 Verification of the Valid Measurements to Distinguish Different Rhythm Proficiencies

3.1.1 Verification of Raw Interval Measurements (RIM)

The average values according to the Ramus measurements for the four groups are given in Table 2. It shows that the percent vocalic interval of native speakers is lower than that of Level B and Level C while higher than that of Level A. As English has more vowel reduction than that of Chinese, the percent vocalic intervals in native speakers are predicted to be less than those of Chinese EFL learners who are influenced by first language transfer and have a more syllable-timed rhythm. Thus, %V doesn't conform to the previous prediction. However, ΔC and ΔV accord with the prediction as the mean values are largest in native speakers, decreasing as the proficiency level gets lower.

Table 2. Average values of RIM for four groups of speakers

Subjects	N	%V (%)	ΔC	ΔV
Native speakers	5	50.73	0.077	0.094
Level A	5	49.70	0.068	0.066
Level B	5	52.70	0.046	0.060
Level C	5	55.15	0.044	0.056

Note. %V = percent vocalic intervals, ΔC = SD of consonantal intervals, ΔV = SD of vocalic intervals.

In order to explore whether the mean results of the rhythm measurements that fit the prediction between these four groups are statistically different, One-Way ANOVA was run. The mean results of consonant interval variability (ΔC) were significantly different, $F(3, 16) = 6.133$, $p = .006$, showing that the difference between different proficiency groups was statistically significant. Additionally, the vocalic interval variability (ΔV) also reached statistical significance, $F(3, 16) = 15.055$, $p = .000$, indicating that there was a significant difference among the four groups.

To sum up, three measurements under the Raw Interval Measurements, except the percent vocalic interval, are consistent with the prediction. Moreover, both the consonantal interval variability and the vocalic interval variability are statistically significantly different among four proficient levels. Therefore these two measurements are valid to capture different rhythm proficiencies.

3.1.2 Verification of Rate-normalized Interval Measurements (NIM)

The results of the rate-normalized interval measurements, Varco ΔC and Varco ΔV are given in Table 3. It seems that the values of these two measurements are the same with the prediction, highest in native speakers and getting lower with the proficient levels declining. Then the two measurements were submitted to One-way ANOVAs with proficiency group as the between factor.

Table 3. Average values of NIM for four groups of speakers

Subjects	N	Varco Δ C	Varco Δ V
Native speakers	5	67.66625	63.49606
Level A	5	66.03409	51.11762
Level B	5	61.86792	49.74016
Level C	5	57.22211	46.69629

Although the lower proficiency groups showed the lower values on the measure of the measurement Varco Δ C, the difference among different groups did not reach statistical significance, $F(3, 16) = 1.686, p = .210$. However, the measurement Varco Δ V seems valid. On one hand, it is consistent with the prediction. On the other hand, the differences did reach statistical significance between four levels, $F(3, 16) = 5.162, p = .011$.

From the above data analysis, we can see that Varco Δ V instead of Varco Δ C under Rate-normalized Interval Measurements is effective to distinguish the rhythm patterns of different proficient levels.

3.1.3 Verification of Pairwise Variability Indices (PVI)

The Grabe measurements nPVI and rPVI are given in Table 4. The values of nPVI and rPVI were predicted to be higher in stress-timed rhythm than syllable-timed rhythm. As shown in Table 4, the native speakers, namely, the highest level, showed the higher values of nPVI and rPVI than those of Chinese EFL learners. And with the levels of the learners going down, the values decrease correspondingly. Next, the two descriptive parameters were submitted to One-way ANOVAs with the speaker group as the between factor.

Table 4. Average values of PVI for four groups of speakers

Subjects	N	nPVI	rPVI
Native speakers	5	79.17819	0.07499
Level A	5	60.84869	0.06735
Level B	5	59.94142	0.051594
Level C	5	56.92467	0.041832

With $F(3, 16) = 4.435, p = .019$, the values of nPVI were statistically different among groups. In addition, the result $F(3, 16) = 5.815, p = .007$, indicated that the other measurement rPVI, the variability of consonantal intervals, was significant higher for native speakers and lower for the Chinese EFL learner.

In summary, both the two measurements nPVI and rPVI are in keeping with the previous prediction and their average values for four groups are significant different. Hence, the Grabe measurements are useful to capture the different rhythm patterns among four proficient levels.

3.2 The Reliability of the Valid Rhythm Measurements in Automatically Scoring

3.2.1 Autoscoring of Rhythm by the Multiple Regressions

From the above data analysis, Δv , Δc , Varco Δv , nPVI, rPVI were found to be valid to distinguish the different English rhythm patterns among four proficient groups. These five measurements were regarded as five features to score the overall rhythm ability. Then the equation can be established as soon as the regression coefficients (β s) were estimated. Thus twelve samples were as training data and the calculation was run in the mathematical software Matlab. The estimated results were $\beta_1 = -38.679$, $\beta_2 = -21.036$, $\beta_3 = -9.216$, $\beta_4 = 0.013$, $\beta_5 = -3.556$, $\beta_0 = 5.529$. Therefore the Multiple Regression equation for rhythm autoscoring can be illustrated as:

$$\text{Rhythm Score} = 5.529 + (-38.679) \cdot \Delta v + (-21.036) \cdot \Delta c + (-9.216) \cdot \text{Varco} \Delta v + 0.013 \cdot \text{nPVI} + (-3.556) \cdot \text{rPVI} \quad (6)$$

3.2.2 The Correlation of the Autoscores and Human Scoring Results

The rest eight samples among four levels were used as testing data which was shown in Table 5. For the convenience of calculation, the proficiency scales were turned into numbers. Then “Native speaker”, “A”, “B”, “C” were changed into “4”, “3”, “2”, “1” respectively.

Table 5. Testing data for automatic scoring of rhythm ability

Subjects	Δv	Δc	Varco Δv	nPVI	rPVI	Human Scores
1	0.045012	0.029763	40.09857	48.52905	0.03186	4(Native speaker)
2	0.066473	0.058724	50.45465	64.00033	0.05924	3(A)
3	0.057459	0.036081	47.45698	61.6113	0.03962	3(A)
4	0.060254	0.053838	41.81648	47.92657	0.04888	3(A)
5	0.054624	0.0412	45.67525	52.77497	0.04389	3(A)
6	0.061036	0.043911	47.81754	58.27072	0.04133	3(A)
7	0.091559	0.106096	61.4282	75.10155	0.10149	2(B)
8	0.066169	0.035369	60.94218	74.17538	0.06914	3(A)

These raw data were applied into the equation for rhythm autoscoring. And the autoscores for eight samples were 3.941 (4), 2.672 (3), 3.515(4), 2.789 (3), 3.379 (3), 3.167 (3), 0.771 (1), and 3.341 (3) successively.

In order to find out the reliability of the autoscoring results, the Pearson correlation was run to the scores calculated by computer and made by the perception of experienced teachers. The correlation between autoscores and scores made by teachers was statistically significant ($p = .005$), indicating that the objective evaluation is related to the subjective evaluation and its reliability reached.866.

4. Conclusion

Some implications are illuminated in the present study. On one hand, the study provides an acoustic-phonetic evidence to reveal that the English syllable structures of Chinese EFL learners are different from those of native speakers. The syllable structures of students with lower rhythm proficiency are less complex than those of the higher proficient and native speakers. In order to improve the Chinese EFL learners' English rhythm, it is important for teachers to inoculate students with the pronunciation rules of consonant and vowel like linking, assimilation, elision, and weak forms etc. and urge them to practice in such ways. On the other hand, the high reliability of the autoscoring for English rhythm proficiency run by Multiple Regression makes a modest contribution to the study of Computer-Assisted Language Learning System for English Rhythm. To some extent, it offers a phonetic theoretical foundation for the objective evaluation system for English rhythm of Chinese EFL learners.

The limitations of the present study cannot be neglected. The participants in the study are insufficient. First, each group only contains five participants. This may not be enough to represent the characteristics of every proficiency level. Second, only twelve samples are used as training data for Multiple Regression. This is obviously quite small data to build a robust model. Third, the gender of the participants in the study is not equally the same. More female samples are collected and less male participants are included.

Acknowledgments

We would like to express our gratitude to the editor and the anonymous reviewers for their helpful suggestions and comments. This research was financially supported by an MOE (Ministry of Education in China) Project of Humanities and Social Sciences (13YJAZH046), a Guangzhou Science Technology Project (2014J4100025), and a grant from Guangdong Natural Science Foundation (2014A030313576).

References

- Abercrombie, D. (1967). *Elements of general phonetics*. Edinburgh: Edinburgh University Press.
- Barry, W. J., Andreeva, B., Russo, M., Dimitrova, S., & Kostadinova, T. (2003). Do rhythm measures tell us anything about language type? In M. J. Solé, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetics Sciences* (pp. 2693-2696). Universitat Autònoma de Barcelona, Barcelona.
- Chen, J., & Wang, G. Z. (2013). The validity of rhythm measurements in evaluating the rhythm proficiencies of Chinese EFL learners. *Foreign Languages in China*, 10(1), 54-58.
- 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, 61-72.

- 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 Linguistic Colloquium* (pp. 231-241). Frankfurt: Peter Lang.
- Dellwo, V., & Wagner, P. (2003). Relations between language rhythm and speech rate. In M. J. Solé, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetics Sciences* (pp. 471-474). Universitat Autònoma de Barcelona, Barcelona.
- Diekerson, W. B. (1997). Grading progress: The role of absolute and relative scales. *Proceedings of the 6th National Conference on the Education and Employment of Graduate Teaching Assistants*.
- Feng, M. (2010). *A contrastive study of the acoustic features of the speech of Chinese English learners with good pronunciation and poor pronunciation* (Unpublished master's thesis). Sichuan International Studies University, Chongqing, China.
- Grabe, E., & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. In C. Gussenhoven, & N. Warner (Eds.), *Laboratory phonology 7* (pp. 515-546). Berlin: Mouton.
- Gui, C. K. (1985). *American English applied phonetics*. Shanghai: Shanghai Foreign Education Press.
- Jang, T. Y. (2008). Speech rhythm metrics for automatic scoring of English speech by Korean EFL learners. *Malsori Speech Sounds The Korean Society of Phonetic Sciences and Speech Technology*, 66, 41-59.
- Lin, Q. M. (2007). The communication function and teaching method of English pronunciation. *Journal of Primary and Secondary School Foreign Education*, 4, 1-5.
- Low, E. L., Grabe, E., & Nolan, F. (2000). Quantitative characterizations of speech rhythm: Syllable-timing in Singapore English. *Language and Speech*, 43, 377-401. <http://dx.doi.org/10.1177/00238309000430040301>
- Page, E. B. (1994). Computer grading of student prose, using modern concepts and software. *Journal of Experimental Education*, 62(2), 127-142. <http://dx.doi.org/10.1080/00220973.1994.9943835>
- Page, E. B., & Petersen, N. S. (1995). The computer moves into essay grading: Updating the ancient test. *Phi Delta Kappan*, 76, 561-565. <http://dx.doi.org/10.2307/20405396>
- Pike, K. (1945). *The intonation of American English*. Ann Arbor: University of Michigan Press.
- Ramus, F., Nespors, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, 73, 265-292. [http://dx.doi.org/10.1016/S0010-0277\(99\)00058-X](http://dx.doi.org/10.1016/S0010-0277(99)00058-X)
- Roach, P. (1982). On the distinction between "stress-timed" and "syllable-timed" languages. In D. Crystal (Ed.), *Linguistic controversies*. London: Arnold.
- Stockmal, V., Markus, D., & Bond, D. (2005). Measures of native and non-native rhythm in a quantity language. *Language and speech*, 48, 55-63. <http://dx.doi.org/10.1177/00238309050480010301>
- Wang, G. Z. (2002). *English pronunciation & intonation practice* (12th ed.). Beijing: Higher Education Press.
- White, L., & Mattys, S. L. (2007). Calibrating rhythm: First language and second language studies. *Journal of Phonetics*, 35, 501-522. <http://dx.doi.org/10.1016/j.wocn.2007.02.003>
- Zhang, F. T. (2002). *British English phonetics and phonology*. Chengdu: Sichuan University Press.

Notes

Note 1. "The proportion of vocalic intervals" refers to the proportion of the duration of all vowels in a sentence in the total duration of a sentence.

Note 2. Speech rate-normalized: speech rate varies from person to person. In order to eliminate this influencing factor, speech rate is usually normalized by dividing the same number into the raw data.

Appendix A

Reading Materials

- 1) He is thirty-five.
- 2) Six thousand five hundred patients.
- 3) They take up around two hours a day.

- 4) He works forty hours a week.
- 5) He has six weeks.
- 6) He earns about thirty nine thousand pounds a year.
- 7) The surgery's pleasant.
- 8) He enjoys getting to know people.
- 9) About eight minutes.
- 10) It's never boring!

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).