Check for updates

Acquisition of English speech rhythm by Chinese learners of English at different English proficiency levels*

Jiaqi Zhang · Sook-hyang Lee**

Department of English Language and Literature, Wonkwang University, Iksan, Korea

Abstract

This study aims to investigate the rhythmic patterns in the English speech produced by Chinese learners of English who learn English as a foreign language (EFL learners). Utilizing interval-based rhythm metrics, namely, VarcoC, VarcoV, nPVI-C, nPVI-V, and %V, the study compared the rhythmic differences in English speech between ten native speakers from the United States and forty Chinese EFL learners from mainland China. A sentence elicitation task consisting of thirty picture prompts and corresponding thirty stimuli sentences with at least five vocalic and four consonantal intervals was conducted. Statistical results reveal that both Chinese advanced learners and beginners had significantly lower degree of stress-timed in their English speech, indicating that the acquisition of the L2 speech rhythm was influenced by the learners' L1 rhythmic pattern. In addition, the results also show that the Chinese advanced learners had significantly higher degree of stress-timed in their English speech than beginners and showed no significant difference with native speakers in VarcoC and nPVI-C. These results indicate that the direction of L2 speech rhythm development was from more syllable-timed to more stress-timed.

Keywords: English speech rhythm, L2 acquisition, rhythm metrics, durational variability

1. Introduction

Rhythm is defined as "the perceived regularity of prominent units in speech. These regularities may be stated in terms of patterns of stressed vs. unstressed syllables, syllable length (long vs. short) or pitch (high vs. low) - or some combination of these variables" (Crystal, 2011). In earlier period of study, Pike (1945) and Abercrombie (1967) developed a theory dividing all languages into those in which the stresses recur at approximately equal time intervals, and those in which the syllables recur at regular intervals. Within this approach, languages are classified into stress-timed language (in which there are almost equal intervals between stressed syllables), syllable-timed language (in which there are almost equal intervals between the successive syllables), and mora-timed language (in which the amount of time of each mora tends to be equal).

For a long time in the study of speech rhythm, English has traditionally been classified as a stress-timed language, in which

stressed syllables are longer than unstressed ones and tend to occur isochronously, with speech stream being like "morse-code". While Mandarin Chinese is a syllable-timed language, where stressed syllables acoustically are not perceived stronger than unstressed syllables. The syllables in Mandarin Chinese, whether stressed or not, tend to occur with equal intervals of time, just like "machine-gun" (Prator & Robinet, 1985). However, the isochrony in either stress-timed rhythm pattern or in syllable-timed rhythm pattern is not in complete symmetry. In other words, a distinct typological dichotomy of stress-timed vs. syllable-timed does not exist. As proposed by the Rhythm Class Hypothesis (Grabe & Low, 2002), the degree of stress-timed or syllable-timed could be different among different stress-timed languages or syllable-timed languages. And between the so-called "stress-timed" and "syllable-timed" languages there is no clear boundary providing a clear categorical distinction. Besides, what could not be neglected is that there is even considerable overlap between the stress-timed and the syllable-timed group and hitherto unclassified languages (Grabe & Low, 2002). So it is more precise to say that English is more stress-timed than Mandarin Chinese or that Mandarin Chinese is more syllable-timed than English, rather than in a distinct typological dichotomy manner. But no matter which classification manner is used, we can conclude that English and Mandarin Chinese are different in speech rhythm and consequently acquiring native-like English rhythm becomes a challenge for most of Chinese EFL learners.

Besides, syllable structure also plays an essential role in influencing speech rhythm. English and Mandarin Chinese has fairly different syllable structures. In English, syllable structure is more complicated with the allowance of consonant clusters while in Mandarin Chinese, syllable structure is simpler and consonant clusters are strictly prohibited. Additionally, in Mandarin Chinese, there is a prohibition of the occurrence of coda consonant (except for a small number of structures ended with nasal /n/ or /ŋ/), while English allows coda consonants and even consonant clusters in coda positions (CVC, CVCC, CVCCC, etc.). Therefore, in Mandarin Chinese, the strict phonotactic constraints lead to predominantly CV (consonant-vowel) syllables, which inevitably would lower the difference in duration of adjacent syllables. The structural differences in syllable between English and Mandarin Chinese might also become an obstacle in the acquisition of English speech rhythm by Chinese EFL learners.

During the past several decades, a plenty of studies have investigated the problems or difficulties that L2 learners are confronted with while acquiring the speech system of a second language. Neurological maturation has often been mentioned as a hindrance in L2 acquisition for learners after puberty. It will get even harder for learners to learn new forms of speech once they have passed a critical period (Lenneberg, 1967; Patkowski, 1990; Scovel, 1988), since neurological maturation may reduce neural plasticity (Lenneberg, 1967; Penfield, 1965) and cause a diminution in ability to add or modify sensorimotor programs for producing L2 sounds (Sapon, 1952). Thus L2 learners would have much more difficulty in acquiring native-like target language after the age of twelve due to the greater effect of L1 interference. In addition, L2 experience has proven to be a factor that might influence the acquisition of L2 speech (Bohn & Flege, 1990, 1992). With the increase of L2 experience,

proficiency of target language will also be improved.

Developing a native-like rhythmic pattern is one of the most essential aspects in language acquisition. In the previous studies it was found that there were cross-linguistic developmental similarities found in L1 rhythm acquisition. And the direction of speech rhythm development is from syllable-timed to stress-timed. Payne et al. (2012) found that age had main effects on %V and VarcoV. In child speech of English, Spanish and Catalan, six-year-old children have higher VarcoV scores than two-year-old children. Grabe et al. (1999) found that even though pairwise variability grew faster in child speech of English than that of French, nPVI would increase with age in child speech in the two languages. As for the speech of bilingual children, they also go from a feature of more syllable-timing towards stress- timing, either they belong to rhythmically similar languages (Whitworth, 2002), or they belong to rhythmically different languages (Bunta & Ingram, 2007). Of course, different from cross-linguistic developmental similarities, language-specific rhythmic developmental patterns also exist. Payne et al. (2012) observed that compared with Spanish and Catalan, English two-year-olds show higher degree of stress-timed rhythm patterns. Bunta & Ingram (2007) found that: 1. English speech produced by English- Spanish bilingual children demonstrated a higher degree of syllable-timing than that produced by monolingual English children of the same age. 2. Simultaneous English-Spanish bilinguals had more features of stress-timed in their English speech than in their Spanish speech. 3. English-Spanish bilingual children produced less equal V intervals than monolingual Spanish children in Spanish. Besides, Ordin & Polyanskaya (2015) observed that even though German and French learners of English demonstrated a similar developmental pattern of speech rhythm in the acquisition of L2 English, German learners of English achieved a closer degree of durational variability to English, whereas French learners of English showed significantly lower degree of durational variability and stress- timing than native British speakers, even in the advanced learners group.

The same developmental trend of rhythm acquisition (from more syllable-timed to more stress-timed) can also be observed in L2 acquisition. Li & Post (2014) found that advanced learners of English produced English speech with higher degree of durational variability than learners at intermediate level no matter whether their first language is stress-timed German or syllable-timed Chinese. In addition, Stockmal et al. (2005) proposed that durational variability of consonantal intervals also significantly increased with L2 acquisitional progresses going on. Using a longitudinal design, Ordin & Polyanskaya (2014) observed in their research that English speech produced by adult L2 learners first showed more features of syllable-timing and lower degree of durational variability before they finally developed into the target language rhythmic patterns, indicating that learners of English whose native languages are syllable-timed would develop their English rhythm patterns from more syllable-timed to increasingly more stress-timed with the increasing length of residence. Ordin et al. (2011) found that durational variability of vocalic and consonantal intervals correlates with progress in English proficiency in learners whose L1 belongs to stress-timed rhythmic pattern. Furthermore, Ordin & Polyanskaya (2015) revealed that durational variability of English speech produced

by both German and French learners of English increased with L2 acquisition progressing, indicating again that the developmental direction of L2 English speech rhythm is from more syllable-timed to more stress-timed in spite of whether the L1 of the learner is rhythmically similar to or different from the target language.

To make a summary of what we have found on speech rhythm:

1. English and Mandarin Chinese have different rhythmic patterns. In English, stressed syllables tend to have longer durations than unstressed ones and the unstressed syllables might be reduced. Besides, the intervals between two stressed syllables tend to be equal. While in Mandarin Chinese, each syllable tends to take up the same amount of time and generally there is a lack of reduced vowels.
2. English and Mandarin Chinese have different syllable structures. Syllable structure in English is more complicated (CV, CVC, CVCC, CVCCC, CCV, CCVC, CCVCC, CCCVCC, etc.) than that in Mandarin Chinese (CV, CVn, CVŋ).
3. Acquisition of L2 rhythm is influenced by learners' L1 rhythmic pattern.
4. The direction of speech rhythm development is from syllable-timed to stress-timed in both L1 and L2 acquisition.

Based on the analyses on differences between English and Chinese speech rhythmic patterns and syllable structures, as well as the findings in previous studies on speech acquisition, we predict that Chinese EFL learners would show a characteristic of syllable-timed in their production of English (lower durational variability and higher %V than native speakers) and that the advanced learners would show more stress-timed (higher durational variability and lower %V) features than the beginners but not reach a native-like level. Rhythm metrics, as used in previous studies, are also utilized in the present study.

The present study is improved and extended from the early version presented in Seoul International Conference on Speech Science 2019 (SICSS, 2019). Firstly, we improved the data in the study by excluding those from beginners who were later regarded as having significantly lower English proficiency than other members of the same group, and then added the same amount of data from those who were closer to the average English proficiency of the whole group. That is, the English proficiency in the beginners group was further controlled at a closer level. Secondly, statistical analysis model was improved. Two-way ANOVA with subject group as fixed factor, 30 target sentences as random factor and rhythmic metrics (VarcoV, VarcoC, nPVI-V, nPVI-C, and %V) as the dependent variable was adopted and thought to be a better model. Thirdly, interpretation of the results was more elaborate and profound than before and the introduction and method parts were also revised and supplemented.

2. Method

2.1. Subjects

Fifty speakers were recruited to participate in the experiment, ten of whom were native speakers from the United States, twenty were advanced Chinese learners, and twenty were beginners from mainland China. The entirety of the ten native speakers (who were gender-balanced - five males and five females) speak GA and have no speech or hearing disorders, and they range from 24 to 29 years old. The twenty Chinese advanced learners (ten males and ten females) were English majors studying in the Hebei Normal University for Nationalities in mainland China and had been studying English as a foreign language for a minimum of 12 years. In regards to their proficiency in the English language, they had passed TEM 4 (Test for English Majors Band 4 in mainland China) with scores higher than 60. The twenty beginners (ten males and ten females) were all middle school students in The Third Middle School of Longhua County in Hebei Province, China, and had been studying English as a foreign language for a maximum of five years.

The advanced learners and beginners were divided based on their lengths of English learning (LEL) and scores in several recently taken English tests that potentially reflected their proficiency in the language. The reference standard for advanced learners was TEM 4 (Test for English Majors Band 4 in mainland China), and for beginners it was the Middle-School Entrance Examination. Detailed information on the two subject groups is presented in Tables 1 and 2.

Table 1. Means of age, LEL and TEM 4 scores for Chinese advanced learners (Standard deviations are in parentheses)

Age	LEL	Test for English Majors Band 4
21.6 (0.73)	12.75 (0.99)	65.35 (3.84)

LEL, lengths of English learning.

Table 2. Means of age, LEL and Middle-School Entrance Examination scores for beginners (Standard deviations are in parentheses)

Age	LEL	Middle-School Entrance Examination
13.75 (0.54)	4.2 (0.4)	76.65 (6.37)

LEL, lengths of English learning.

As highlighted in Tables 1 and 2, the average LEL for the advanced learners is 12.75, while for the beginners it is 4.2. An independent samples t-test shows that there is a significant difference ($p<.001$) in LEL between the two learner groups, indicating that the Chinese advanced learners have significantly longer exposure time to English than the beginners' group. Furthermore, in terms of the advanced learners, all members have passed TEM 4 with an average score of 65.35, meaning that their English proficiency is higher than the average level of English majors nationwide and far higher than the average level of college students in the country. As for the beginners, their average score in their Middle-School Entrance Examination is 76.65. The Middle-School Entrance Examination is a teenager-oriented test consisting of simple questions in the skills of English vocabulary and grammar, as well as straightforward tests of listening, reading, and writing ability (an excerpt from the Middle-School Entrance Examination, Hebei Textbook Edition, is presented in Appendix 1).

On the other hand, TEM 4 is explicitly aimed at English majors, and comprehensively tests examinee's English listening, reading, writing, vocabulary, and grammar skills (an excerpt of TEM 4 from 2018 is shown in Appendix 2). The TEM 4 is

administered by the National Advisory Committee for Foreign Language Teaching (NACFLT) on behalf of the Higher Education Department, Ministry of Education, People's Republic of China. As a criterion-referenced test, the TEM 4 is designed in strict accordance with the teaching requirements outlined in the Chinese National College English Teaching Syllabus for English Majors and has been through several revisions during its development since 1991. Now it has grown into a well-established English language test in China. Previous studies have proposed that TEM 4 is a reasonably reliable and valid test which is set at an appropriate (difficulty) level as defined in the test specification (Jin & Fan, 2011; Tian, 2015). And TEM 4 score is also found significantly correlated with examinee's learning effort and motivation (Chen, 2012). According to statistics, the national pass rate for TEM 4 in mainland China in 2018 was just 51.79%. Even English majors may have some difficulty passing TEM 4. We can thus conclude that the TEM 4 exam is far more challenging than the Middle-School Entrance Examination, and even middle-school students who achieved top scores in the Middle-School Entrance Examination could struggle with the TEM 4. So, we conclude that there is a substantial difference in English proficiency between the advanced learners' group and beginners group in our study.

2.2. Materials and Procedures

In terms of the materials used in the research by Ordin & Polyanskaya (2015), in the present study, 30 picture prompts, and the corresponding 30 stimuli sentences containing at least five vocalic and five consonantal intervals were selected and presented to the speakers in Microsoft PowerPoint on separate slides. Elicited sentences in the experiment are demonstrated in Appendix 3. At first, the speakers observed the images accompanied by the corresponding descriptive sentences. After that, they were required to remember the sentences. They were allowed to flip through the slides and view each slide for as long as they felt necessary. Afterward, they were asked to look at only the pictures and recall the corresponding sentences from memory. Then, they were required to say the corresponding sentences out loud into a Zoom H1 recorder. Verbal prompts were given to the speakers to assist them in producing the target sentences if they said any phrases that were different from the originals. For instance, if the subject said, "The boy is going to school" while the target sentence was actually "The children are going to school," the researcher would say: "Yes, you are right. The boy is going to school. But if you look at the picture again more carefully, you will notice that some other children are also going to school. Could you please take a look at the picture again and say what you are looking at out loud again?" Mostly, one example of verbal prompting was enough for the subjects to say the correct sentence. In some sporadic cases, when speakers were still unable to replicate the correct target sentence, second-time verb prompts were given. By doing so, we ensured that each subject produced the correct target sentence as we wanted and that the recording procedure was more similar to daily communication, which happens spontaneously.

2.3. Measures and statistical analyses

Rhythm can be defined as patterns of change throughout certain types of speech intervals: vocalic intervals and consonantal intervals. Within this framework, rhythm metrics, providing a

workbench of speech rhythm as durational variability, are put forward to capture systematicity in patterns of durational variability in a speech and to ensure that speech rhythmic patterns can be measured objectively and compared across or within languages. Generally, stress-timed languages have higher durational variability than syllable-timed languages. The global metrics ($\triangle$, Varco) reflect the degree of difference between the duration of each single vocalic or consonantal interval and the mean duration of the corresponding interval over the entire utterance. And the local metrics (rPVI, nPVI) assess the durational differences between successive intervals, demonstrating the durational variation of vocalic and consonantal intervals in a pairwise manner. Rhythm metrics utilized for measuring durational variability in the present study are normalized metrics: VarcoV, VarcoC, nPVI-V, and nPVI-C, as presented in Table 4. Raw metrics, which is demonstrated in Table 3, such as $\triangle$V, $\triangle$C, rPVI-V, and rPVI-C, are excluded, as they are under the influence of speech tempo and vary in values when speech tempo changes (Dellwo, 2006; Grabe & Low, 2002).

Besides, %V is also included in the present study. Even though %V is not an indicator of durational variability, it has proven to be an efficient indicator of speech rhythmic pattern and robust to the variation of speech tempo (Ordin & Polyanskaya, 2014; Ramus et al., 1999; Wiget et al., 2010). Generally, stress-timed languages have lower %Vs than syllable-timed languages.

Table 3. Raw rhythm metrics

	Metrics name	Description
Raw metrics	$\triangle$ (delta)	Standard deviation of X intervals
	rPVI	$\left[\sum_{k=1}^{m-1} \left| d_k - d_{k+1} \right| / (m-1) \right]$

$\triangle$ (delta) is the standard deviation in the duration of a particular speech interval (e.g., $\triangle$V: for vocalic intervals; $\triangle$C: for consonantal intervals) (Ramus et al., 1999).

rPVI is the raw pairwise variability index for a specific type of speech interval, which in other words, is the mean of the durational differences between successive intervals (e.g., rPVI-V: for vocalic intervals; rPVI-C: for consonantal intervals) (Grabe & Low, 2002).

Table 4. Normalized rhythm metrics

	Metrics name	Description
Normalized metrics	Varco	$\triangle$X×100/mean(X)
	nPVI	$100 \times \left(\sum_{k=1}^{m-1} \left| \frac{d_k - d_{k+1}}{(d_k + d_{k+1})/2} \right| / (m-1) \right)$

Varco is the standard deviation in the duration of certain speech intervals split by their mean duration (e.g., VarcoV=$\triangle$V/meanV; VarcoC=$\triangle$C/meanC) (Dellwo, 2006).

nPVI is the normalized pairwise variability index for certain types of speech intervals (e.g., nPVI-V: for vocalic intervals; nPVI-C: for consonantal intervals). Normalization is achieved by using the mean duration of the corresponding intervals, which decreases the influence of the speech tempo variation on

the measure of durational variability (Grabe & Low, 2002).

%V is the percentage of vocalic intervals within the sentence, which is reported to be robust to impact speech tempo (White & Mattys, 2007; Wiget et al., 2010).

The VarcoV, VarcoC, nPVI-V, nPVI-C, and %V of each target sentence was calculated and compared among the three experimental groups. The whole recording procedure yielded 1,500 tokens in total (50 speakers×30 sentences).

All the sentences were then labeled on the TextGrid window of the Praat by three tiers. In the process of labeling, the three annotation tiers include the Orthographic Tier, the Consonant and Vowel Tier, and the Word Tier. Each target sentence was segmented into vocalic and consonantal intervals. The silence part was labeled as "SL." The vocalic and consonantal ranges were categorized under the guidance of the standard proposed by Peterson & Lehiste (1960). We followed the specific segmentation procedure outlined by Payne et al. (2012). An example of sentence segmentation is shown in Figure 1.
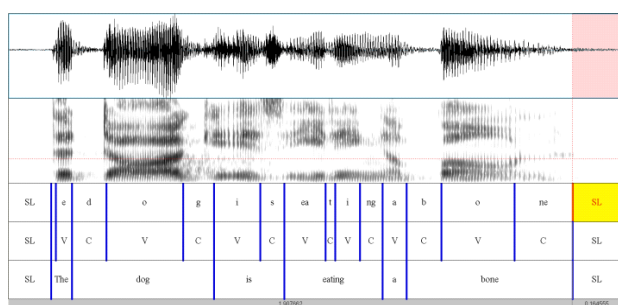

Figure 1. Example of sentence segmentation in Praat

The statistical analyses were executed in SPSS 22. To investigate how the rhythm measures of Chinese EFL learners (Varco, nPVI, %V) modify with increasing proficiency in their L2 English speech and how they are different to that of native speakers, a series of two-way ANOVAs with subject group as fixed factor, 30 target sentences as random factor and rhythmic metrics (VarcoV, VarcoC, nPVI-V, nPVI-C, and %V) as the dependent variable, were conducted.
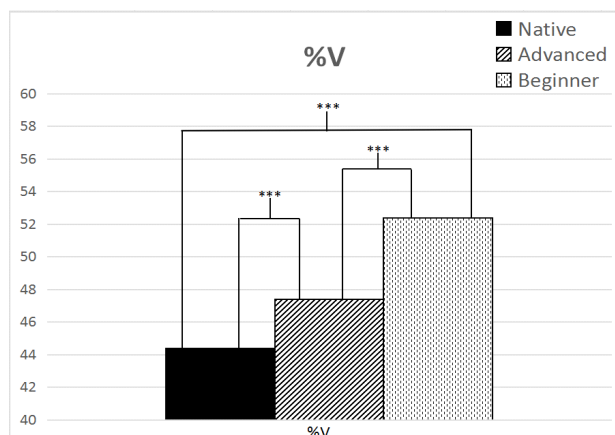
3. Results

Table 5 and Figures 2~6 show that the degree of stress-timing is highest in English produced by native speakers, followed by Chinese advanced learners, and then beginners. To investigate whether the differences in rhythm metrics among the three groups of speakers are significant, a series of two-way ANOVAs are performed and the statistical results demonstrate that the differences between the three groups are significant for all the five rhythm metrics: %V ($F_{(2, 58)}$=350.101, $p<.001$), VarcoV ($F_{(2, 58)}$=206.231, $p<.001$), nPVI-V ($F_{(2, 58)}$=257.74, $p<.001$), VarcoC ($F_{(2, 58)}$=69.462, $p<.001$) and nPVI-C ($F_{(2, 58)}$=165.372, $p<.001$). The results indicate that the English speech rhythmic patterns not only significantly differ between Chinese EFL learners and native speakers, while also develop significantly as a function of proficiency in groups of advanced learners and beginners. It is also observed that there is no significant difference between the 30 target sentences for all

the five rhythm metrics and that the interaction of the subject group (fixed factor) and 30 target sentences (random factor) is also not significant, which means that differences in rhythm metrics between the three subject groups were not significant between the 30 target sentences.

Table 5. Means (standard deviation) of rhythm metrics for ten native speakers, twenty Chinese advanced learners and twenty beginners
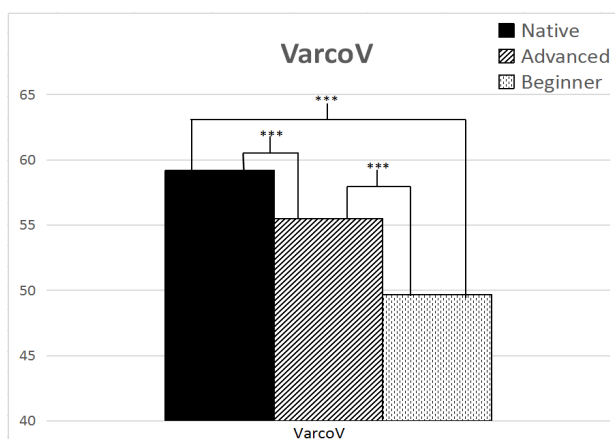
	American	Chinese (A)	Chinese (B)
%V	44.4	47.4	52.4
	(6.2)	(6.1)	(8.6)
VarcoV	59.2	55.5	49.7
	(5.6)	(6.9)	(8.7)
nPVI-V	65.5	60.3	55.1
	(5.9)	(8.3)	(6.8)
VarcoC	54.5	53.6	49.9
	(4.4)	(7.5)	(5.8)
nPVI-C	65.9	65.3	57.8
	(6.1)	(7.9)	(7.5)

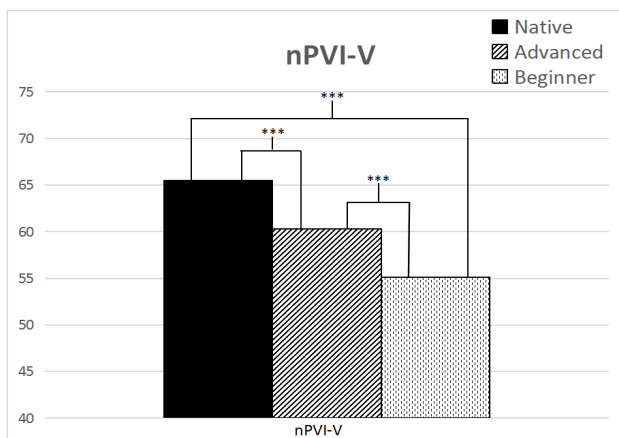A, advanced learners; B, beginners.

$^{***}p<.001$.

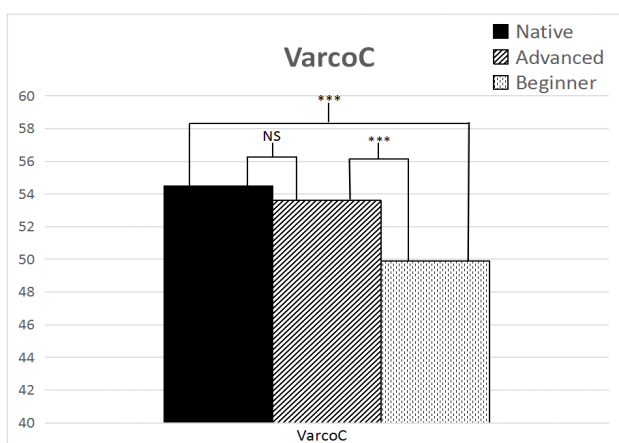**Figure 2.** %V of ten native speakers, twenty Chinese advanced learners and twenty beginners

$^{***}p<.001$.

Figure 3. VarcoV of ten native speakers, twenty Chinese advanced learners and twenty beginners

***p<.001.

Figure 4. nPVI-V of ten native speakers, twenty Chinese advanced learners and twenty beginners

***p<.001.

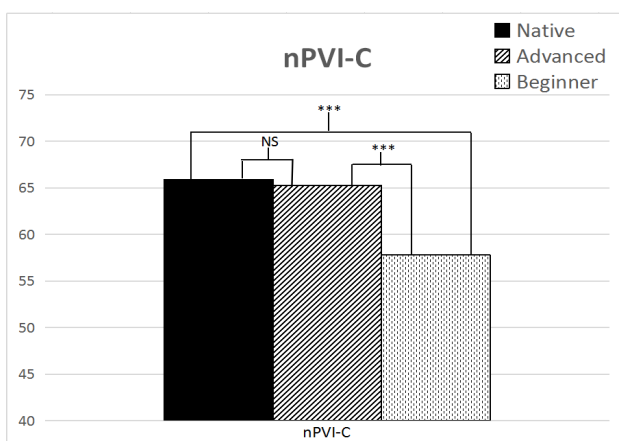**Figure 5.** VarcoC of ten native speakers, twenty Chinese advanced learners and twenty beginners.

***p<.001.

Figure 6. nPVI-C of ten native speakers, twenty Chinese advanced learners and twenty beginners

Table 6. *p*-values in Bonferroni post hoc tests for the rhythm metrics calculated in the English speech of native speakers, Chinese advanced learners and beginners

	N vs. CA	N vs. CB	CA vs. CB
%V	p<.001***	p<.001***	p<.001***
VarcoV	p<.001***	p<.001***	p<.001***
VarcoC	p=.095	p<.001***	p<.001***
nPVI-V	p<.001***	p<.001***	p<.001***
nPVI-C	p<.715	p<.001***	p<.001***

***p<.001.

N, native speakers; CA, Chinese advanced learners; CB, Chinese beginners.

To examine the precise differences in rhythm metrics among the three experimental groups, Bonferroni post hoc tests are conducted, and the results demonstrate that both Chinese advanced learners and beginners have significantly higher %V than native speakers (p<.001 and p<.001) (Table 6). This means that the proportion of vocalic intervals in Chinese EFL learners' speech is significantly higher than that of native speakers, indicating a lower degree of stress-timing in the speech of Chinese EFL learners. Besides, the difference in %V between advanced learners and beginners is also significant (p<.001) (Table 6). This means that beginners possess more vocalic intervals in their speech compared to advanced learners, and they show significantly more syllable-timed features in their English speech patterns than advanced learners.

Another two post hoc pairwise comparative analyses demonstrate that both of the two Chinese EFL learner groups have significantly lower VarcoV and nPVI-V than native speakers (Table 6), which means that the durational variability of the vocalic intervals in the English speech of the Chinese EFL learners is significantly lower than that of native speakers, and that the English speech of Chinese EFL learners demonstrates more features of syllable-timed than that of native speakers. Furthermore, as demonstrated in Table 6, the significantly higher VarcoV (p<.001) and nPVI-V (p<.001) of advanced learners over beginners also indicates a higher degree of stress-timing in the English speech of advanced learners than that of beginners.

In the measurement of VarcoC and nPVI-C, two indicators being able to demonstrate durational variability of consonantal intervals, post hoc analyses show that beginners have significantly lower VarcoC (p<.001) and nPVI-C (p<.001) than native speakers (Table 6). As for the comparison between the two non-native groups, Chinese advanced learners have significantly higher values in VarcoC (p<.001) and nPVI-C (p<.001) in their English speech than beginners do (Table 6). Also, what went beyond our expectation is the fact that for VarcoC and nPVI-C, no significant difference is found between advanced learners and native speakers (p=.095 for VarcoC; p=.715 for nPVI-C) (Table 6). Following Ordin & Polyanskaya (2015), we also propose a potential explanation that there are a more significant assimilation effect and co-articulation phenomena in native English speakers' speech, and this might lead to cluster simplification and thus a lowering of VarcoC and nPVI-C.

4. Discussion

In the present study, we analyzed and compared the English

rhythmic patterns of Chinese EFL learners and native speakers from the United States. It was discovered that there were significant differences in rhythm patterns between Chinese EFL learners and native speakers. Durational variability in the English speech of Chinese EFL learners was lower than that of native speakers, even in the advanced learners' group. This indicates that Chinese EFL learners have yet to sufficiently acquire the features of English rhythmic patterns, and when speaking English, they are not able to rid themselves of the rhythmic patterns of their L1 (Mandarin Chinese). Therefore their English speech shows a lower degree of stress-timing than native speakers. Also, Chinese EFL learners showed higher %V than native speakers, which is also a typical sign of a lower degree of stress-timing. Additionally, we also found that there were significant variations between advanced learners and beginners. The advanced learners had a higher degree of durational variability and more stress-timed features in their English speech than the beginners, which means that in the process of L2 English rhythm acquisition, Chinese EFL learners' speech rhythm developed from more syllable-timed patterns towards more stress-timed patterns.

Firstly, we found in the statistical results that both advanced learners and beginners showed significantly higher %Vs than native speakers. This could possibly be connected to the influence of their L1 Mandarin Chinese, which was a syllable-timed language with only CV structures (except for a small number of structures ending in /n,ŋ/). Consequently, when Chinese EFL learners spoke English, they tended to habitually insert additional /ə/-like vowels after coda consonants in their English speech to form CV structures that were similar to the syllable structures of their L1, leading to a rising proportion of vocalic intervals, thus increasing the value of %V. Based on the results of the present study alongside prior research (Ordin & Polyanskaya, 2014; Ramus et al., 1999; Wiget et al., 2010), we could arguably conclude that %V is an effective indicator of distinguishing different rhythmic patterns.

Additionally, we also found that in the VarcoV and nPVI-V measurements, both advanced learners and beginners had significantly lower values in these two indicators than native speakers, indicating that in their English speech, Chinese EFL learners were not capable of effectively lengthening the stressed vowels and shortening the unstressed vowels as native speakers were, and as a consequence of this, their English speech was perceived to be foreign-accented. We can also explain this as the interference of their L1 Mandarin Chinese. As in Mandarin Chinese, there is not as much durational contrast of vowels as there is in English, such as tense vs. lax vowels, long vs. short vowels, or stressed vs. unstressed vowels. Therefore, when Chinese EFL learners spoke English, they tended to transfer the aspect of lacking vowel durational contrast in Mandarin Chinese into their English speech production. This would effectively result in a higher degree of syllable-timed speech and a tendency of being isochronous in vocalic intervals, which would cause decreases in the values of VarcoV and nPVI-V.

As for VarcoC and nPVI-C, we found in the statistical results that the VarcoC and nPVI-C of beginners were significantly lower than those of native speakers. This indicated that in terms of the durational variability of consonantal intervals,

the beginners' group showed a lower durational variability than native speakers and more features of syllable-timed, suggesting that they had difficulty in grasping the rhythmic characteristics of English consonantal intervals. This might still be a result of the influence of the learners' first language. In Mandarin Chinese, syllable structure is simple, and consonant clusters are not allowed, while English has more complex syllable structures (CV, CCV, CCVC, CCVCC, CCCVCC, etc.). Therefore, when Chinese EFL learners speak in English, it may be difficult for them to grasp to a sufficient extent the sophisticated structural features of English consonants as well as the durational variability of consonantal intervals present in English speech. When speaking English, they would possibly either simplify the consonant clusters, or insert additional /ə/-like vowels in between the consonants clusters. Either way would inevitably lead to a lowering of VarcoC and nPVI-C.

Furthermore, what was beyond our expectation was that there was no significant difference in VarcoC and nPVI-C between advanced learners and native speakers. However, we would not necessarily conclude that the advanced learners had a good command of English rhythmic features because other indicators in the rhythm metrics such as VarcoV, nPVI-V, and %V revealed that the English rhythm pattern used by the advanced learners still failed to achieve a native-like level. They still possessed lower durational variability and a higher degree of syllable-timing than native speakers. Instead, we attribute this non-significant difference between advanced learners and native speakers to the more considerable effect of assimilation and co-articulation phenomena in native speakers' speech, which may lead to cluster simplification and thus a decrease in VarcoC and nPVI-C.

Finally, for the intra-group comparison of Chinese EFL learners, we found that whether it was on the rhythm metrics of vocalic intervals (%V, VarcoV, nPVI-V), or on the rhythm metrics of consonantal intervals (VarcoC, nPVI-C), advanced learners demonstrated significantly higher durational variability and a higher degree of stress-timed characteristics in English speech than beginners. As previous studies proposed, the developmental direction of speech rhythm pattern in L2 acquisition was from more syllable-timed to more stress-timed. The results of the present study may offer solid support for this finding.

5. Conclusion

This study investigated the rhythm patterns in the English speech produced by Chinese EFL learners. Utilizing the interval-based rhythm metrics, it compared the rhythmic differences in English speech between ten native speakers and forty Chinese EFL learners. Statistical results highlighted that there were significant differences in English rhythm patterns between native speakers and Chinese EFL learners and between Chinese advanced learners and beginners, providing evidence for findings in previous studies that the acquisition of the L2 rhythm was influenced by learners' L1 rhythmic pattern, and the direction of L2 speech rhythm development was from syllable-timed to stress-timed.

This study could be extended in the future by focusing on the English speech produced by non-native speakers speaking not only one specific type of syllable-timed language but also

speakers using other types of syllable-timed languages. For instance, Korean is also categorized into the "syllable-timed" side in the whole rhythmic classification continuum. According to the Rhythm Class Hypothesis (Grabe & Low, 2002), the degree of syllable-timed rhythm might vary among different syllable-timed languages. Besides, phonotactic differences in language might also lead to various features of speech rhythm. Thus, even though Mandarin Chinese and Korean are both categorized as syllable-timed languages, there are expected differences between the rhythmic patterns of English speech produced by Chinese and Korean EFL learners. Additionally, for pedagogical implications, in light of the present study, it is suggested that L2 English teachers could attempt to focus on enhancing the durational variability in students' English speech, thus creating a typical rhythmic pattern of English. Concentrating more on the features of connected speech and stronger on durational contrasts between stressed and unstressed syllables, long and short vowels or tense and lax vowels might be effective methods for L2 English teaching.

References

Abercrombie, D. (1967). *Elements of general phonetics*. Edinburgh, UK: Edinburgh University Press.

Bohn, O. S., & Flege, J. E. (1990). Interlingual identification and the role of foreign language experience in L2 vowel perception. *Applied Psycholinguistics*, *11*(3), 303-328.

Bohn, O. S., & Flege, J. E. (1992). The production of new and similar vowels by adult German learners of English. *Studies in Second Language Acquisition*, *14*(2), 131-158.

Bunta, F., & Ingram, D. (2007). The acquisition of speech rhythm by bilingual Spanish- and English speaking four- and five-year-old children. *Journal of Speech, Language, and Hearing Research*, *50*(4), 999-1014.

Chen, X. (2012). A linear regression analysis on TEM4. *Journal of Civil Aviation Flight University of China*, *6*, 020.

Crystal, D. (2011). *A dictionary of linguistics and phonetics*. Hoboken, NJ: John Wiley & Sons.

Dellwo, V. (2006). Rhythm and speech rate: A variation coefficient for deltaC. In P. Karnowski, & I. Szigeti (Eds.), *Language and language-processing* (pp. 231-241). Frankfurt, Germany: Peter Lang.

Grabe, E., Post, B., & Watson, I. (1999). The acquisition of rhythmic patterns in English and French. *Proceedings of the International Congress of Phonetic Sciences* (pp. 1201-1204). San Francisco, CA.

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, Germany: Mouton de Gruyter.

Jin, Y., & Fan, J. (2011). Test for English majors (TEM) in China. *Language Testing*, *28*(4), 589-596.

Lenneberg, E. H. (1967). The biological foundations of language. *Hospital Practice*, *2*(12), 59-67.

Li, A., & Post, B. (2014). L2 acquisition of prosodic properties of speech rhythm: Evidence from L1 Mandarin and German Learners of English. *Studies in Second Language Acquisition*

36(2), 223-255.

Ordin, M., & Polyanskaya, L. (2014). Development of timing patterns in first and second languages. *System*, *42*, 244-257.

Ordin, M., & Polyanskaya, L. (2015). Acquisition of speech rhythm in a second language by learners with rhythmically different native languages. *The Journal of the Acoustical Society of America*, *138*(2), 533-544.

Ordin, M., Polyanskaya, L., & Ulbrich, C. (2011). Acquisition of timing patterns in second language. *Proceedings of 12th Annual Conference of the International Speech Communication Association*. Florence, Italy.

Patkowski, M. S. (1990). Age and accent in a second language: A reply to James Emil Flege. *Applied Linguistics*, *11*(1), 73-89.

Payne, E., Post, B., Astruc, L., Prieto, P., & del Mar Varnell, M. (2012). Measuring child rhythm. *Language and Speech*, *55*(2), 203-229.

Penfield, W. (1965). Conditioning the uncommitted cortex for language learning. *Brain*, *88*(4), 787-798.

Peterson, G. E., & Lehiste, I. (1960). Duration of syllable nuclei in English. *The Journal of the Acoustical Society of America*, *32*(6), 693-703.

Pike, K. L. (1945). *The Intonation of American English*. Ann Arbor, MI: University of Michigan.

Prator, C. H., & Robinett, B. W. (1985). *Manual of American English Pronunciation for adult foreign students*. New York, NY: Holt, Rinehart & Winston.

Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, *73*(3), 265-292.

Sapon, S. M. (1952). An application of psychological theory to pronunciation problems in second language learning. *The Modern Language Journal*, *36*(3), 111-114.

Scovel, T. (1988). *A time to speak: A psycholinguistic inquiry into the critical period for human speech*. Rowley, MA: Newbury House.

Stockmal, V., Markus, D., & Bond, D. (2005). Measures of native and non-native rhythm in a quantity Language. *Language and Speech*, *48*, 55-63.

Tian, X. H., (2015). A study on the reliability and validity of TEM 4. *Teaching of Forestry Region*, *1*, 42-43.

White, L., & Mattys, S. L. (2007). Calibrating rhythm: First language and second language studies. *Journal of Phonetics*, *35*(4), 501-522.

Whitworth, N. (2002). Speech rhythm production in three German-English bilingual families. *Leeds Working Papers in Linguistics & Phonetics*, *9*, 175-205.

Wiget, L., White, L., Schuppler, B., Grenon, I., Rauch, O., & Mattys, S. L. (2010). How stable are acoustic metrics of contrastive speech rhythm? *The Journal of the Acoustical Society of America*, *127*(3), 1559-1569.

• **Jiaqi Zhang**

Graduate Student, Department of English Language and Literature
Wonkwang University
460 Iksan-daero, Sin-dong
Iksan-si, Jeollabuk-do, Korea
Tel: +82-10-8608-0659

Email: zjq496166366@gmail.com
Areas of interest: Phonetics

• **Sook-hyang Lee,** Corresponding author
Professor, Department of English Language and Literature
Wonkwang University
460 Iksan-daero, Sin-dong
Iksan-si, Jeollabuk-do, Korea
Tel: +82-10-3004-6913
Email: shlee@wku.ac.kr
Areas of interest: Phonetics

Appendix 1. Excerpt of Middle-School Entrance Examination of Hebei Textbook Edition

Mrs Green is going to give a birthday party for Mary (Mary is her daughter). Mary is going to be thirteen years old. Twenty friends of Mary's are going to come to the party. They are all girls. Mrs Green is getting ready for party. Mrs White is helping her.

"Let's make a pretty cake," Mrs White says to Mrs Green.

"Ok. Thank you very much."

Mrs Green is going shopping now. She's going to buy fruit for the party.

Mrs Green buys some fruit. She buys a lot of pears, orange, and bananas. Then she goes home.

It's three o'clock in the afternoon. Everything is ready. Now the first girl is arriving. The party is going to start in thirty minutes.

( ) 1. _____ is going to give a birthday party for Mary.
　　　A. Mrs White B. Mrs Green C. Mary D. Mary's friend

( ) 2. Mary is going to be _____ years old.
　　　A. twenty B. ten C. twelve D. thirteen

( ) 3. _____ are going to come to the party.
　　　A. Thirty boys B. Twenty girls C. Forty children D. Mrs Green and Mrs White

( ) 4. Mrs Green buys _____ for the party.
　　　A. a cake B. some fruit C. some bananas and apples D. a lot of oranges

( ) 5. The party is going to begin at _____ in the afternoon.
　　　A. two thirty B. three C. four D. three thirty

Appendix 2. Excerpt of TEM 4 in 2019

(9) Oatley agreed. "Alice Munro doesn't tell you what to think," he said. "You, yourself, have to make inferences about characters. And that's often what we're doing in our real-life conversations." Or at least that's what people could do.

(10) Reading literary fiction could also offer a way to "practice" your social skills and use them more in real life, according to another researcher not involved in the study. "It's like how pilots train in a flight simulator,' said Raymond Mar, an associate professor of psychology at York University, in Toronto, who has collaborated with Oatley. "This is a great study," Mar said of the new study. But he added that the overall research in this area is "still in its infancy" and one key question is whether literary fiction really is better than other fiction.

(11) Mar and his colleagues recently found that fans of romance novels tended to do best on tests of empathy. Unlike the current study, Mar's study did not test people after having them read different types of fiction. So it's possible that there is something else about romance-novel readers that makes them more understanding of others.

(12) Still, according to Mar, it's too early to tell people to trade in their Danielle Steel for Alice Munro, at least if the goal is boosting empathy.

(13) It's also possible that plays, movies or even TV shows could build your empathy muscles, according to Kidd. But reading may be special, he said, because it provides no visuals and you have to engage your imagination more.

(14) Everyone agreed that the findings suggest literature is important beyond entertainment or improving vocabulary. "There's a common belief that reading literature is frivolous, or not practical," Mar said. "But there's a growing body of evidence that it's important in skills that we need in our lives."

48. According to the new study, what kind of books are likely to help people better understand others' feelings?
　　A. Science fiction.
　　B. Romance novels.
　　C. Literary fiction.
　　D. Nonfiction.

Appendix 3. Elicited sentences used in the experiment

1. The dog is eating a bone.
2. The book is on the table.
3. The girl is eating an apple.
4. The ball is under the chair.
5. The boy is kicking the ball.
6. The man is catching a fish.
7. The children are watching the giraffe.
8. The boy is drinking juice.
9. The bread is on the table.
10. The boy is eating ice-cream.
11. The boy is under the tree.
12. The cat is drinking milk.
13. The girl is talking on the phone.
14. The boy is walking.
15. The baby is eating an orange.
16. The baby is crying.
17. The baby is sleeping.
18. The cat is chasing the mouse.
19. The man is walking.
20. The man is chasing the bus.
21. The baby is taking a bath.
22. The boy is riding a bike.
23. It's raining outside.
24. The girl is eating candy.
25. The children are going to school.
26. It's snowing outside.
27. The dog is chasing the cat.
28. The spoon is on the table.
29. The girl is chasing the boy.
30. The computer is on the table.