

1 Introduction

According to List (1963) and Merriam (1964), most human communities have certain speech styles that are constrained to fit an external or imposed periodic intervals or beat, which manifests rhythmic patterns. This music-like feature makes poetry and nursery rhymes possible. However, apart from this artificially created artistic feature of speech rhythm, different languages, in their non-artistic forms, all have repetitive patterns that are at least intuitively detectable. Moreover, rhythm is among the first phonological features that children acquire. According to Karmiloff and Karmiloff-Smith (2001), “during the last trimester of intrauterine development, the fetus is known to be actively processing the sound of its mother’s speech” (p. 43). After being filtered through the amniotic fluid, the fetus can only recognize the “melody and rhythm of the language” (p. 43). After birth, the newborns are able to distinguish typologically (from the perspective of speech rhythm) different languages from their mother tongue (Nazzi, Bertoncini & Mehler, 1998; Nazzi, Jusczyk & Johnson, 2000; Bosch & Sebastian-Galles, 1997).

Early phonologists classify language rhythms into three typological categories; they are stress-timed languages, syllable-timed languages, and mora-timed languages (Abercrombie, 1967; Pike, 1945). Such classification is rested on the concept of isochrony, which will be discussed in more detail in

Section 2.

More recently, different approaches to speech rhythm have been proposed to explain what factors contributed to the perceived difference between syllable-timed languages and stress-timed languages, and various metrics concerning speech rhythm have been proposed (cf. Dellwo, 2004; Grabe & Low, 2002; Low, Grabe & Nolan, 2000; Low & Grabe, 1995; Ramus, Nespor & Mehler, 1999). These metrics measure the durations of vocalic intervals and intervocalic intervals of speech, and these metrics differ from different algorithms they adopt in which the durations of vocalic intervals and intervocalic intervals are manipulated in different ways. Details of these durational metrics are discussed in Section 3.

This dissertation aims to explore the rhythm of L2 English of Chinese EFL learners who are native speakers of Mandarin (born and have spent most of their lives in Beijing) and Cantonese (born and have spent most of their lives in Guangzhou, formerly known as Canton). Before that, the rhythm of both Mandarin and Cantonese, which has not been investigated as thoroughly as languages such as English, Dutch, French, Spanish and so on, is investigated, and compared with the rhythm of English.

Working hypotheses are proposed here that Mandarin is more

syllable-timed (or less stress-timed) than Cantonese (reasons for this hypothesis are discussed in Section 2), while English is the most stress-timed language of the three; and also the rhythm of L2 English is interfered by the rhythm of the first language. In other words, it is expected that the different metrical scores of L2 English is distributed between the scores of the participants' first languages (Mandarin and Cantonese) and native speakers' English.

In Section 2, early studies into speech rhythm and rhythmic typology are generally introduced. Section 3 reviews different durational metrics and previous studies in which such metrics are applied. Section 4 offers a brief review of Mandarin and Cantonese rhythm. Section 5 outlines rhythmic research in second languages. Section 6 introduces research questions and methodological issues. Sections 7 and 8 offer the results of the research and the discussion of the results and suggestions for further research were made. The last section gives the conclusion of the study.

2 Rhythmic Typology of Languages -- In Search of Isochrony

According to Roach (2000), rhythm involves “some noticeable event happening at regular intervals of time” (p. 134), and he likens speech rhythm with “the rhythm of a heart-beat, of a flashing light or of a piece of music” (p. 134). This definition depicts the isochronous feature of speech rhythm. Different languages are said to have different units for repetitive intervals, and the repetitive intervals are said to be of similar durations. If the isochronous intervals are defined by two stressed syllables, i.e. the stressed syllables occur at relatively regular intervals, regardless of whether they are separated by unstressed syllables or the number of unstressed syllables that separate them. In other words, the feet are equal in length. Languages with this stress-timing are called stress-timed languages. Languages such as English, Russian, Dutch and so on are stress-timed languages (Abercrombie, 1967; Pike, 1945).

Different from stress-timed languages, another type of languages has isochronous units as syllables. In such languages, no matter the syllables are stressed or unstressed, their durations are said to be relatively the same. Language with this syllable-timing are called syllable-timed languages. Typical syllable timed languages are the Roman languages, such as French and Spanish (Abercrombie, 1967; Pike, 1945).

In addition, there also exists another type of language called mora-timed language. Japanese is typically mora-timed (Bloch, 1942; Beckman, 1982; Port, Dalby & O'Dell, 1987). In Japanese, the special timing units, morae, are of relatively equal durations. Other mora-timed languages include Gilbertese and Luganda (Clark, Yallop & Fletcher, 2007).

Nevertheless, later empirical research failed to find sufficient evidence for the existence of isochrony, either in traditionally called stress-timed languages or syllable-timed languages. That is, from acoustic analysis, there is little evidence to back up the categorization of rhythm into stress-timing or syllable-timing. For instance, Dauer (1983) discovered that the durations between any successive stressed syllables are not the same, and the duration of feet is nearly in positive proportion to the number of unstressed syllables between the two stressed syllables. Moreover, researchers also discovered that the durations of unstressed syllables between stressed syllables are not significantly different from syllable-timed languages (Dauer, 1983; Roach, 1982). Therefore, the durations of feet in stress-timed languages are not equal in length as early researchers proposed.

Apart from stress-timed languages, such discrepancy also exists in traditionally called syllable-timed languages. For example, researchers

failed to find that syllables in French or Spanish are isochronous (de Manrique & Signorini, 1983; Pointon, 1980). Furthermore, Roach (1982) also discovered that the durations of syllables in typical syllable-timed languages like French, Yoruba, or Telegu are as varied as the syllable durations in typical stress-timed languages like English, Russian or Arabic.

Since evidence gained from instrumental research of rhythm breaks the isochronous patterns believed to exist in stress-timed languages and syllable-timed languages, Nespor (1990) even rejected to use the rhythmic categorization of stress-timing and syllable-timing, because there is no language that strictly belongs to either category. She argues that reported perceptual rhythmic differences between so-called stress-timed languages and syllable-timed language are due to “non-rhythmic phonological rules” that are idiosyncratic to different languages.

Although evidence supporting strict rhythmic categorization into stress-timed languages and syllable-timed languages is scant, it is conventional to continue referring the two types of languages as stress-timed and syllable-timed. Besides, perceptual studies also show that people can perceptually distinguish languages from different rhythm classes. Apart from that neonates are able to discriminate languages that have different rhythmic types from their first language (Nazzi, Bertoncini &

Mehler, 1998; Nazzi, Jusczyk & Johnson, 2000; Bosch & Sebastian-Galles, 1997), adults are also able to discriminate languages of different rhythm types. For example, by using synthesized speech in which intonational and segmental cues are eliminated, Ramus, Dupoux and Mehler (2003) discovered that adult participants could discriminate between stress-timed English and Dutch, and syllable-timed Spanish and Catalan. However, they failed to discriminate Polish, whose phonological structure makes its rhythm class different from either stress-timing or syllable-timing. The next section offers a brief discussion about it. Other rhythmic perceptual studies include Ramus & Mehler (1999) and Ramus, Nespore and Mehler (1999). The results of these studies all suggest that rhythmically different languages are perceptually distinguishable. Therefore, it is legitimate to retain the rhythmic classification. Other phonetic studies based on acoustic analysis suggest that the differences between stress-timed languages and syllable-timed languages originate from different duration in the vocalic portion and intervocalic (or consonantal) portion of the speech. Details of such studies are reviewed in the next section.

3 Rhythmic Typology of Languages Revisited – Durational Metrics of Rhythm

Failure to find isochrony pushes the research of rhythm to another paradigm. According to Dauer (1983), perceptual difference between stress-timed languages and syllable-timed languages arose from structural differences between the two types of languages. Such structural differences involve whether vowel reduction is permitted or not, and syllable structure and syllable duration.

3.1 Structural Differences between Stress- and Syllable-Timed Languages

3.1.1 Vowel Reduction

When discussing the contrasts between stress-timed languages and syllable-timed or mora-timed languages, a number of researchers refer to the phenomenon of vowel reduction (e.g. Delattre, 1969; Roach, 1982, Hoequist, 1983; Beckman, 1992; Ramus et al., 1999; Grabe & Low, 2002; White & Mattys, 2007a, 2007b). According to Dauer (1983, p. 57), vowel reduction can “maximize the difference between stressed and unstressed syllables in a stress-timed languages”. But for syllable-timed languages, they “do not regularly have reduced variants of vowels in unstressed position” (Dauer, 1983, p. 57).

Different languages have different patterns of phonological or phonetic vowel reduction in unstressed or unaccented syllables (Crosswhite, 2003 cited in Fletcher, 2010). Romance languages such as Spanish, French, or Italian do not have centralized vowels or schwa-like vowels for unstressed vowels, as Germanic languages like English do (Fletcher, 2010). Researchers (e.g. Koopmans-van Beinum, 1980; den Os, 1988, cited in Fletcher, 2010) compared vowel reduction between Germanic Dutch and Romance Italian, and found significant difference between the two languages.

However in terms of segment duration, although languages like Italian, French and Spanish do not have reduced or centralized unstressed vowels, the duration of unstressed or unaccented vowels are shorter and some formant undershoot was also observed (e.g. Delattre, 1969; den Os, 1988; Bertinetto & Fowler, 1989, cited in Fletcher, 2000). In addition, according to Delattre (1969, cited in Fletcher, 2000) stress-timed languages like English and German can “show greater differences in the degree of vowel centralization in stressed versus unstressed syllables than languages the rhythm divide like French and German” (p. 560). Moreover, “[r]educed variants of French and German vowels can be closer to their stressed or accented counterparts than unstressed vowels are to vowels in stressed syllables in English” (p. 560). To sum up, only the presence of vowel reduction alone is not sufficient to determine whether a language is

stress-timed or syllable-timed. Another factor that existed between stress-timed and syllable-timed languages is syllable structure. Section 3.1.2 gives an introduction.

3.1.2 Syllable Structure

According to Roach (1982) and Dauer (1983), syllable structure is another factor that contributes to the perception of stress-timed and syllable-timed languages. As Abercrombie (1967) suggested, stress-timed languages manifest a substantial degree of syllable duration variability, because of a greater number of permissible choices for syllable structures. A number of empirical studies supported this claim (e.g. Delattre, 1966; Crystal & House, 1990).

By comparing the maximum range of syllable duration variation under different stress conditions, utterance position, and syllable structure in four languages (stress-timed English and German, and syllable-timed Spanish and French), Delattre (1969) discovered that English has the most variation in duration across all factors, whereas Spanish has the least durational variation.

In addition, Crystal and House (1990) discovered that in American English syllable durations significantly vary across different syllable types. The

results of the study show that the duration of open syllables is significantly shorter than the duration of non-prepausal stressed CCVCC syllables, and the duration of unstressed open syllables has the shortest duration. Similar studies also include Hoequist (1983) and Fletcher (1991).

According to Dauer (1983), the perception of syllable-timing is also enabled by a great degree of syllabic regularity. By analyzing the corpora data, he found that approximately 70 percent of Spanish syllables were open, while only 44 percent of English syllables were open. Other studies on syllabic regularity in syllable-timed and stress-timed languages include Fletcher (1991, French data), Nespor (1990, Italian and Dutch data). In conclusion, syllable-timed languages tend to have simpler and more regular syllable structures, and stress-timed languages tend to have more complex syllable structures.

3.1.3 Towards a Gradient Continuum of Rhythmicity

As mentioned above, aspects like vowel reduction and syllable structure may all contribute to the perceived rhythmic differences among languages. Vowel reduction may or may not be present in certain languages, and even such feature exists in some languages, degrees of vowel reduction may vary. Besides, different languages may have different complexity of syllable structures. Taking both vowel reduction and syllable structure into

consideration, it is difficult to strictly label a certain languages as stress-timed or syllable-timed.

To deal with this problem, Dauer (1987) suggested that speech rhythm can be viewed as a gradient continuum. Different languages, depending on their structural features, can said to be on different positions of the continuum. For example, English can be regarded to be towards the stress-timed pole of the continuum, and Spanish can be said to be towards the syllable-timed pole of the continuum. The rhythmicity of other languages can vary between the two extreme ends of the continuum. For example, Polish is a language with complex syllable structure, but has low degree of vowel reduction. On the contrary, Catalan has simpler syllable structure, but has reduced vowels in unstressed syllables. Thus, the two languages are said to be intermediate languages (Nespor, 1990; Ramus et al., 1999). Since late 1990's, various metrics have been devised to facilitate rhythm studies. The most influential metrics are summarized in the next section.

3.2 Durational Metrics of Speech Rhythm

Durational metrics of speech rhythm can be roughly categorized into two groups. The first group of metrics measure the duration of vocalic and intervocalic intervals (e.g. Ramus et al., 1999; Grabe & Low, 2002; Dellwo, 2004). The second group of metrics measure either the duration of vowels

(e.g. Low & Grabe, 1995), or the durational contrast of successive syllables (e.g. Deterding, 2001). The first metrics group is more preferred, and reasons are discussed in detail in Fletcher (2010, pp. 564-565). In the following subsections, metrics of Interval Measures and Pairwise Variability Indices are summarized.

3.2.1 Interval Measures

Ramus, Nespore and Mehler (1999) developed the following interval measures:

- %V : the proportion of vocalic intervals in the utterance;
- ΔV : the standard deviation of vocalic intervals within the utterance;
- ΔC : the standard deviation of consonantal intervals within the sentence.

In their study, languages of typical rhythmic classes (stress-timed languages: English and Dutch; syllable-timed languages: Spanish, Italian and French; intermediate languages: Polish and Catalan; mora-timed language: Japanese) were analyzed. They found that %V and ΔC were directly related to syllable structures, and could be applied to classify the rhythmic classes of these languages. If a language has more complex syllable structures, then there will be more variability in the number of consonants that occur in a syllable.

As a result, the duration of syllables can vary from one another, thus the standard deviation of consonantal intervals (i.e. ΔC) should be high. Moreover, if a language has more reduced vowels, it should have lower proportion of vocalic intervals (i.e. %V), compared with languages that do not allow vowel reduction. Therefore, a more stress-timed language has higher ΔC and lower %V; whereas a more syllable-timed language has lower ΔC and higher %V.

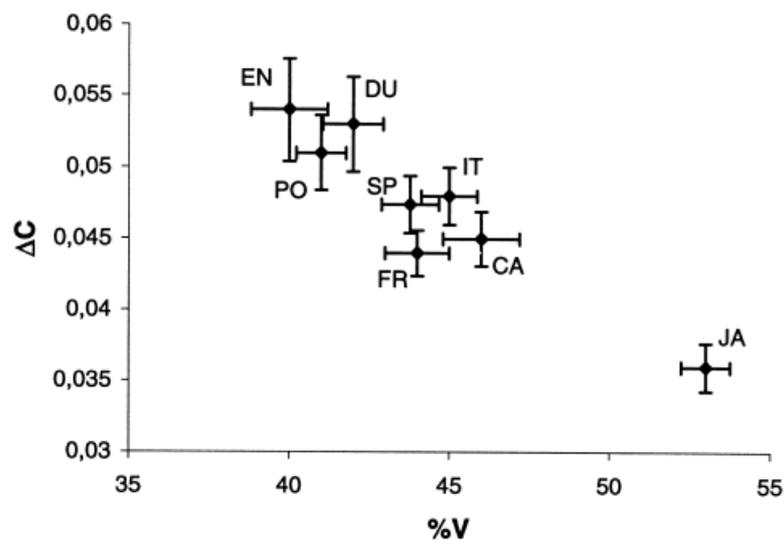


Figure 1 Distribution of languages using %V and ΔC . Error bars represent +/- standard error. (from F. Ramus, M. Nespors & J. Mehler, 1999, *Correlates of linguistic rhythm in the speech signal*, *Cognition*, 73, 265-292, p. 273).

In Figure 1, stress-timed languages English, Dutch, and Polish (which is also regarded as intermediate language) cluster in the top left of the plot. In the meanwhile, syllable-timed languages Spanish, French, Italian and Catalan (which is in a strict sense classified as an intermediate language) cluster in

the bottom right of the plot. And mora-timed Japanese showing the highest %V and lowest ΔC is in the bottom-most corner of the plot. In general, %V and ΔC are negatively correlated in deciding the rhythmic class of a language.

Another score in the interval metrics, ΔV (the standard deviation of vocalic intervals), was also used in Ramus et al. (1999). But it failed to show significant differences between the rhythmic classes. The reason for this is that vocalic intervals variation in a language is influenced by other factors, such as contrastive vowel length and lengthening in specific context, such as at the end of a prosodic phrase. Nevertheless, although ΔV does not have perfect relations with rhythm, this metric can still show some phonological properties of languages. According to Ramus and his colleagues' study (1999), although the scores of ΔC and %V of Polish are close to stress-timed languages like English and Dutch, but the Polish ΔV score is much lower than that of English and Dutch, suggesting that Polish has idiosyncratic features that make it rhythmically different.

3.2.2 Pairwise Variability Indices

Another set of durational metrics is often used in studies of speech rhythm. They are Pairwise Variability Indices (abbreviated as PVI). Different from interval measures, PVI measures the rhythmic contrast in successive

intervals (vocalic or intervocalic) (Grabe & Low, 2002; Low & Grabe, 1995; Low, Grabe & Nolan, 2000). PVI has an advantage over interval measure as discussed in Low et al. (2000). The authors (Low et al., 2000) devised two hypothetical languages. According to Figure 2, although the mean and the standard deviation of vocalic duration from the mean are identical in both Language A and Language B, overall patterns of vocalic duration are quite different. The perceptual rhythms of the two languages are different, but their interval measures scores are the same.

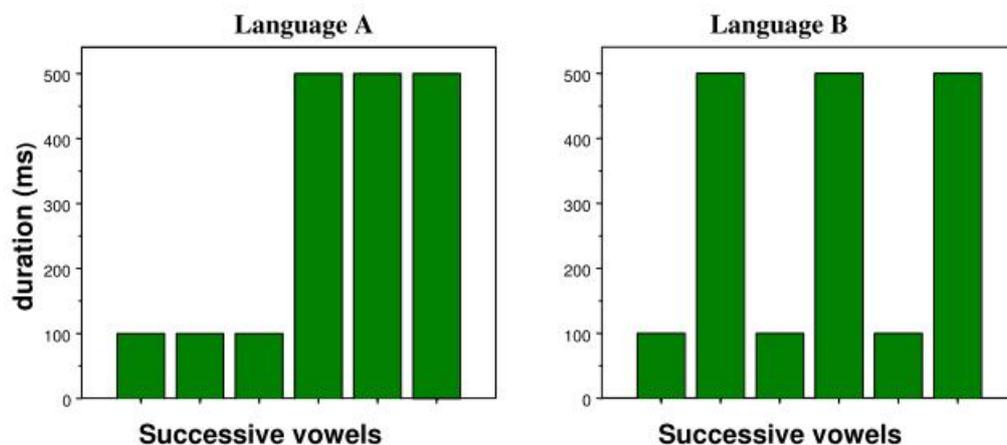


Figure 2 Successive vocalic durations of two hypothesized languages A and B. (from E. L. Low, E. Grabe & F. Nolan, 2000. Quantitative characterization of speech rhythm: Syllable-timing in Singapore English. *Language and Speech*, 4, 377-401).

In order to solve problems like this, researchers (Low et al., 2000; Grabe & Low, 2002) developed the metrics of Pairwise Variability Indices (PVI). They are designed to capture the successive durational difference between any two adjacent intervals in an utterance. Therefore, it has been suggested that

PVI measures and interval measures can be applied complementarily, especially with languages of mixed or intermediate rhythm classes, such as Polish and Catalan (Low et al., 2000).

PVI scores can be either normalized for speech rate or not. Normally, PVI measures for vocalic intervals are normalized (nPVI-V), whereas scores for intervocalic intervals are not normalized and raw PVI scores (rPVI-C) are directly applied (Low & Grabe, 2002). Rationale for normalization is discussed in Low & Grabe (2002), “an intervocalic interval is compositional. It can contain several different segmental units, and these may be subject to different speech rate effects. But the majority of [...] vocalic intervals consists of a single vowel that is stretches or compressed when speech rate changes” (see Section 2.4 of their paper).

Scores for rPVI-C and nPVI-V are computed according to the following formulae:

$$rPVI - C = \left[\sum_{k=1}^{m-1} |d_k - d_{k+1}| / (m-1) \right]$$

$$nPVI - V = 100 \times \left[\sum_{k=1}^{m-1} \left| \frac{d_k - d_{k+1}}{(d_k + d_{k+1})/2} \right| / (m-1) \right]$$

where m is the number of intervals in an utterance, and d is the duration of the k th interval.

By utilizing rPVI-C and nPVI-V measures, Grabe and Low successfully separated previously rhythmically classified languages, such as stress-timed British English, Dutch, and German, as well as syllable-timed French and Spanish, and mora-timed Japanese. And the results are clearly displayed in Figure 3.

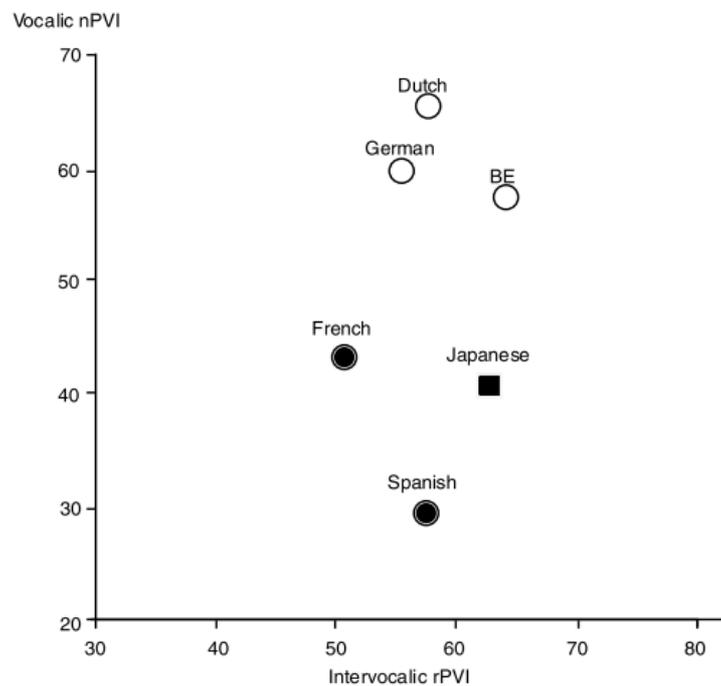


Figure 3 PVI profiles from prototypical stress-timed languages English, Dutch and German, syllable-timed languages French and Spanish, and mora-timed language Japanese. ○ = stress-timed; ● = syllable-timed; ■ = mora-timed (from E. Grabe & E. L. Low, 2002, *Durational variability in speech and the rhythm class hypothesis*. In C. Gussenhoven & N. Warner (eds.), *Laboratory Phonology 7* (pp. 515-543), Berlin: Mouton de Gruyter).

3.2.3 Speech Rate and Rate-Normalized ΔV and ΔC

The relationship between speech rate and rhythmic metrics is complicated. As Barry et al. (2003, cited in Fletcher, 2010) suggested, “languages tend to become ‘more syllable-timed’ at faster tempo, and that both ΔC and ΔV decrease as speaker tempo increases” (p. 566). Also, Ramus (2002, cited in Fletcher, 2010) suggested that “tempo must be controlled in corpora for measures like %V or ΔC to be useful” (p. 566). Research into speech rate and rhythmic measures are abundant, such as Barry et al. (2003), Dellwo and Wagner (2003), White and Mattys (2007a).

Under such background, two more speech rate normalized metrics were proposed by Dellwo and his colleagues (Dellwo et al., 2004; Dellwo, 2006, cited in Fletcher, 2010). They are rate normalized ΔC (i.e. VarcoC) and rate normalized ΔV (i.e. VarcoV). Here, “Varco” means variation coefficient. VarcoC and VarcoV are computed according to the following formulae:

$$VarcoC = \frac{\Delta C \cdot 100}{\bar{C}}$$

where ΔC , as in the interval measures, refers to the standard deviation of intervocalic intervals, and \bar{C} refers to the mean duration of intervocalic intervals; and

$$VarcoV = \frac{\Delta V \cdot 100}{\bar{V}}$$

where ΔV , as in the interval measures, refers to the standard deviation of vocalic intervals, and \bar{V} refers to the mean duration of vocalic intervals in an utterance.

3.2.4 Summary

In this section, different durational metrics aiming to categorize languages into different rhythmic types were reviewed. The interval measures (Ramus et al., 1999) compute the proportion or standard deviations of vocalic or intervocalic interval duration. The Pairwise Variability Indices (Low et al., 2000; Grabe & Low, 2002) measure the cumulative contrasts between successive vocalic intervals and intervocalic intervals. In addition, taking speech rate into consideration, speech rate normalized variation coefficients of ΔC and ΔV (VarcoC and VarcoV) were developed (Dellow et al., 2004; Dellow, 2006). In the next section, studies on the rhythm of English, Mandarin Chinese, and Cantonese using these durational metrics are reviewed.

4 Rhythm of English, Mandarin and Cantonese

4.1 Rhythm of English

English has long been regarded as a stress-timed language (Abercrombie, 1967; Pike, 1945). First of all, English allow complicated syllable structures, which is a characteristic of stress-timed languages. For example, the peak position can be filled in by various types of vowels, such as long/tense vowels or short/lax vowels. Besides, not only pure vowels can act as the nucleus, but also diphthongs and even triphthongs. For another thing, the onset position can be taken up by as many as three consonants (although they are constrained by phonotactic rules), and the coda position can be filled in by as many as four consonants (again, phonotactic rules constrain their forms of combination) (for detailed discussion of English syllable structure, see for example, Giegerich, 1992, and Roach, 2009). Thus, the duration of English syllables varies more compared with languages, such as Spanish and Italian.

Apart from syllable structure, English has apparent vowel reduction in unstressed syllables. And these unstressed vowels are frequently centralized as the schwa. In extreme cases, the vowel in an unstressed syllable is even dropped, leaving nasals such as [ŋ̥] or [m̥], and liquid such as [ɾ] to act as the nucleus. Thus, a large number of English words, especially function words have both strong forms and weak forms (c.f. Roach, 2009). Some

syllable-timed language, such as Spanish, to some extent, has reduced vowels, but according to Dauer (1983), these reduced vowels do not affect syllable length.

Durational metrics studies also proved the stress-timed rhythmicity of English, such as Ramus et al. (1999) (interval measures, such as %V, ΔC , and ΔV), Grabe and Low (2002) (both interval measures and pairwise variability indices, rPVI-C and nPVI-V), and White and Mattys (2007a) (interval measures, pairwise variability indices, and rate-normalized ΔC and ΔV , i.e. VarcoC and VarcoV).

4.2 Rhythm of Mandarin

Based on auditory impression, Mandarin Chinese is traditionally known as a syllable-timed language (Lin & Wang, 2007: Abstract). First of all, it has simpler syllable structure, compared with English. According to R. Cheng (1966, cited in Duanmu, 2007) and C. Cheng (1973, cited in Duanmu, 2007, p. 81), “a stressed syllable in SC [i.e. Standard Chinese, or Mandarin] can range from one to four sounds”, which include both the onset and the coda, as well as the nucleus. Thus, either the vocalic intervals or the intervocalic intervals have less variability in terms of duration.

Nevertheless, unlike other typical syllable-timed languages, Mandarin does

have reduced vowels in unstressed syllables. Often in an unstressed syllable, the original nasal coda is dropped, and the nucleus vowel is reduced, though nasalized under the influence of the dropped coda (Duanmu, 2007). Even though, durational metrics studies proved the syllable-timed rhythmicity of Mandarin, such as Low and Grabe (2002) (metrics applied were %V, ΔC , ΔV as well as rPVI-C and nPVI-V), Lin and Wang (2007) (%V, ΔC , ΔV , rPVI-C and nPVI-V were applied), Mok and Dellow (2008) (%V, ΔC , VarcoC, rPVI-C and nPVI-V as well as other indices of syllable durations were applied), and Mok (forthcoming) (metrics used are the same with Mok & Dellow (2008)).

4.3 Rhythm of Cantonese

Being a closely related language to Mandarin, Cantonese shares a lot of linguistic characteristics with Mandarin. Although a monolingual Cantonese speakers and a monolingual Mandarin speaker cannot understand each other, Cantonese is still considered as one of the many Chinese dialects in linguistic circle in China (Sun, 2006). Therefore, it is not surprising to have an impression of syllable-timing for Cantonese. In addition, Cantonese has all the features that a syllable-timed language is supposed to have. First of all, it has simple syllable structure. The coda and the onset positions allow only one consonant to the most, although the liquid /w/ can follow the stop (cf. Kao, 1971). Besides, vowel reduction is not common in Cantonese.

However, Cantonese has phonemic distinction between long and short vowels (cf. Kao, 1971); therefore duration in vocalic intervals can be more varied than a strict syllable-timed language.

Studies utilizing durational metrics offer positive evidence for the syllable-timed rhythmicity of Cantonese; for example, Mok and Dellow (2008) (%V, ΔC , VarcoC, rPVI-C and nPVI-V as well as other indices of syllable durations were applied), and Mok (forthcoming) (metrics used in Mok & Dellwo (2008) were also applied).

5 Studies on Second Language Rhythm

It is a known fact that during the process of second language acquisition or learning, first language has interfering effects upon the second language (cf. Odlin, 1989; Han & Odlin, 2006), especially in phonology. According to Gussenhoven and Jacob (2005), it is inevitable that learners “interpret the pronunciation of the words of the foreign language in terms of the phonological elements of their own” (p. 34). The rhythm of second language has attracted a number of researchers’ interests.

Before the durational metrics were proposed, Wenk (1985) investigated native speakers of syllable-timed French learning stress-timed English. He proposed that learners underwent a series of transitional stages starting from the rhythm of the first language to the rhythm of the second language. Wenk (1985) refers to the mother tongue French as a trailer-timed language (accented syllables located in the final position of the rhythmic groups), and English as a leader-timed language (accented syllables located in the opening position of rhythmic groups). The researcher discovered that the influences their first language have on the acquisition of second language rhythm depend on the proficiency of the learners, the phonetic environment, and the speech style (e.g. imitative reading, sentence-final word echoing, guided re-telling).

Research into second language rhythm using durational metrics also came into being as the rhythm metrics have been proposed.

A number of studies were conducted to investigate the rhythm of L2 English spoken by native speakers of Mandarin Chinese. Among these studies, the exploration of the rhythm of Singapore English is a special case. Although Singapore English is considered as one variant among world Englishes (see Kachru, Kachru & Nelson, 2009), considering the fact that Singapore English is inevitably affected by Chinese, it is still reasonable to treat it as a special kind of second language. By using interval measures and PVIs, researchers (Deterding, 2001; Low & Grabe, 1995; Low et al., 2000) discovered that Singapore English has the characteristics of syllable-timing. One factor, according to the above mentioned authors, that contribute the syllable-timing of Singapore English is that it has less degree of vowel reduction, compared with British English.

In the same manner, Jian (2004) also discovered that “reduced vowels in American English are more concentrated in the F1/F2 formant space than reduced vowels in Taiwan English, which are more dispersed” (in Abstract). Therefore, less reduced vowels in Taiwan English, like Singapore English, contribute to the syllable-timing of Taiwan English.

Other rhythmic research using durational metrics on second languages are also easy to find. For example, White and Mattys (2007a) discovered that VarcoV has the highest correlation with the ratings of accentedness of native speakers of Spanish who speak English. In this dissertation, the L2 English rhythm of learners whose first languages are Mandarin and Cantonese is investigated. Methodological issues of this study are described in the next section.

6 Methodology

6.1 Research Hypotheses and Research Questions

It has been proved by various researches that English is an epitome of stress-timed language, and Mandarin and Cantonese are syllable-timed languages. However, as for the rhythmicity of Mandarin and Cantonese, it is hypothesized that Cantonese is more syllable-timed than Mandarin, because apart from their simplicity in syllable structures, Mandarin has higher degrees in reducing vowels in unstressed syllables. Moreover, it is hypothesized that L2 English spoken by native speakers of Mandarin and Cantonese are less stress-timed than native English.

This study aims to answer the following questions:

- 1) How do English, Mandarin and Cantonese differ in terms of speech rhythm?
- 2) What are the rhythm class of the L2 English spoken by native speakers of Mandarin and Cantonese?
- 3) Why does the L2 English have such rhythmic features?

6.2 Participants

Eight native speakers of Mandarin and eight native speakers of Cantonese participated in the study, and completed all the reading aloud and recording tasks. However, only five participants' recordings in each group were taken

for further analysis. The informants who had been excluded for further analysis were due to poor recording quality, such as hoarse voice caused by pharyngeal infection, uncontrollably pausing too much at unnecessary positions, and the researcher's operating error.

Five Mandarin speakers were students majoring in the English language at a comprehensive university in Beijing. They were all in their third year of study, and therefore their proficiency of English can be regarded as above intermediate and even advanced. Except one participant who spent one of his adult years in Shanghai, all of them were born, and grew up in Beijing, and have spent most of their lives in Beijing. Hence, they speak typical Mandarin.

Similar to the participants in Beijing, five native speakers of Cantonese were also majoring in the English language at a comprehensive university in Guangzhou (formerly known as Canton). They were also in their third year of undergraduate study, and have above intermediated and even advanced proficiency of English. They were born, grew up, and have spent most of their lives in Guangzhou. Thus, they speak typical Cantonese.

The participants were all gender balanced, and their average age is 20 years. None of them reported to have speaking, hearing, or reading impairment.

All participants were paid 50 RMB yuan (approx. 5 pounds) for their participation.

6.3 Recording Procedure

The participants were recorded individually in a quiet room. Before recording started, the participants were given adequate time to familiarize themselves with the reading material, and practice the material if they felt the need to. They were required to read the material sentence by sentence at their normal speed. In case of stuttering, they were asked to record that problematic sentence again until they felt natural with that particular sentence. Moreover, they were encouraged to read the sentences fluently and reduce the number of unnecessary pauses. However, they could pause at the end of a prosodic phrase, which is normal in daily speech. They were required to read the Mandarin (or Cantonese) sentences (which will be introduced in section 6.5) first and the English sentences (also see section 6.5) next. The participants were asked to pause a little longer between sentences.

All the recordings were made by the Microtrack 24 / 96 solid state recorder with Audio Technica 8531 headset microphone. The sampling rate was 48 kHz. All the recordings were finally saved to the computer hard drive, and the sound files were carefully coded.

6.4 Control Data

Control data of native English speakers were obtained by date sharing request made to Arizona State University. After having completed a web-based online course “Protecting Human Research Participants” and obtained a certificate, the researcher was given access to sound files and data set used by White and Mattys (2007a). The control data were provided by native speaker of American English. Five of the speakers’ data were chosen at random to be used in this study. They were also sex balanced.

6.5 Reading Materials

English sentences were the ones used in White and Mattys (2007a). The mean number of syllables per sentence is 16.2. These sentences are listed below and the numbers of syllables are stated in parentheses.

- 1) The supermarket chain shut down because of poor management. (15)
- 2) Much more money must be donated to make this department succeed.
(17)
- 3) In this famous coffee shop they serve the best doughnuts in town. (15)
- 4) The chairman decided to pave over the shopping center garden. (17)
- 5) The standard committee met this afternoon in an open meeting. (17)

Mandarin and Cantonese sentences were constructed by analogy with the English sentences. Like the English sentences created by White and Mattys

(2007a), distribution of stressed and unstressed syllables was not controlled. Moreover, like the English sentences, approximants /w/, /j/, /r/ and /l/ were avoided in Mandarin and Cantonese sentences construction, since the boundary between an approximant and the following vowel is difficult to discern. Moreover, nasal [ŋ] is also avoided in the syllable onset of Cantonese sentences, because in this language [ŋ] and [l] are free variants of the same phoneme in the onset position. The average number of syllables per sentence is 16.6 for Mandarin, and 16.8 for Cantonese. The Mandarin and Cantonese sentences, together with their phonemic transcriptions are listed below with their numbers of syllables in parentheses. One thing needs to be mentioned about Mandarin rhotacized and non-rhotacized non-open central unrounded apical vowels is that they are transcribed as consonants in Duanmu (2007) or Lee and Zee (2003). Due to their clear vowel-like features (cf. Howie, 1976; Zhou & Wu, 1963), they are transcribed here by non-IPA symbols /ʌ/ and /ɿ/ used in Karlgren (1915-1926), which are also widely used by Chinese linguists (also see Pullum & Ladusaw, 1996).

Mandarin sentences:

1. 大姐今天早晨跟妈妈去这家超市买饺子。(17)

ta tɕiə tɕin t^hian tsau tɕən kən ma ma ts^hy tɕɿ tsia tɕ^hau ɕl mai tɕiao tsɿ

2. 他好想听大家唱那部电视剧的主题曲。(16)

t^ha hau ɕian t^hiŋ ta tɕia tɕ^haŋ na pu tian ɕl tsy tə tɕu ti ts^hy

3. 附近这家咖啡厅卖全市最好的芝士蛋糕。(17)

fu tɕin tɕɿ tɕia kʰa fei tʰiŋ mai tɕʰuan ʃl tsui hau tə tɕl ʃl tan kau

4. 校长决定将学校的足球场重新翻修。(15)

ɕiao tɕaŋ tsye tiŋ tɕiaŋ ɕye ɕiau tsu tɕʰəu tɕʰaŋ tɕʰuŋ ɕin fan ɕiu

5. 她跟同学说好今天早晨在肯德基门口见面。(18)

tʰa kən tʰuŋ ɕye ɕuo hau tɕin tʰian tsau tɕʰən tsai kʰən tɿ tɕi mən kʰou
tɕian mian

Cantonese sentences:

1. 我听朝同我阿妈去嗰间超市买饺子。(16)

ŋɔ: tɿŋ tsi:u tʰuŋ ŋɔ: a: ma: həy kɔ: ka:n tsʰi:u si: ma:i ka:u tsi:

2. 我好想听大家唱下寻晚部电视剧啲主题曲。(18)

ŋɔ: ho:u sœ:ŋ tʰɛ:ŋ ta:i ka: tsʰœ:ŋ ha: tsʰəm ma:n po:u ti:n si: kʰɛkʰ kɛ:
tsy tʰei kʰɔkʰ

3. 附近间咖啡店卖啲全部都係最好啲芝士蛋糕。(19)

fu kən ka:n ka: fɛ: ti:m ma:i kɛ: tsʰy:n po:u tu: hei tsøy ho:u kɛ: tsi: si:
ta:n ko:u

4. 主席决定啲将购物中心啲假山铺平。(15)

tsy: tsɿkʰ kʰytʰ tɿŋ tsɔ: tsœ:ŋ kʰeu mətʰ tsuŋ sɛm kɛ: ka: sa:n pʰo:u pʰiŋ

5. 我同我啲同学今朝係肯德基门口见面。(16)

ŋɔ: tʰuŋ ŋɔ: kɛ: tʰuŋ hɔkʰ tʰiŋ tsi:u hei haŋ takʰ kei mu:n həu ki:n mi:n

6.6 Segmentation and Measurements

The researcher identified and labeled vocalic intervals and intervocalic intervals by visual inspection of waveforms and wideband spectrograms displayed in Praat, which is also assisted by audio signals. After segmentation, the duration of vocalic and intervocalic intervals is measured using a Praat script. Since the speech data used by White and Mattys (2007a) are used as control data, the segmentation protocol used in this study is identical with the one applied in White and Mattys (2007a). As Wiget, White, Schuppler and Grenon (2010) discovered, little variation existed due to segmentation differences between measurers following an agreed protocol. The following paragraphs summarized the segmentation protocol.

The primary determiner of the placement of a vowel-consonant boundary is the end of the pitch period preceding a break in formant structure associated with a significant drop in waveform amplitude (White & Mattys, 2007a).

Additional standards for the placement of the vowel-consonant boundary in certain contexts are:

1. "Where the vowel offset was glottalized, a change in the shape of successive pitch periods, like lengthening or doubling;
2. "Before fricatives, the onset of visible friction;

3. "Before nasals, the appearance of nasal formant structure and a waveform amplitude minimum." (White & Mattys, 2007a, p. 507).

The consonant-vowel boundary is the beginning of the pitch period at the onset of vocalic formant structure, where this was associated with the appearance of pitch periods consistent with the body of the vowel (White & Mattys, 2007a). Aspiration after the stop release is segmented within the consonantal interval.

Following Grabe and Low (2002), silent pauses within sentences are excluded from the intervals. The durations of the intervals before and after the pauses are summed together.

7 Results

In this section, results of the study are reported. Differences among the three native languages (i.e. English, Mandarin and Cantonese) in terms of durational metrics were reported. Moreover, the differences between native English and L2 English are also compared. In Table 1, the means and standard errors for all the durational metrics of all groups of speakers are displayed.

Table 1. Means (standard errors) of durational metrics for all language groups.

| | Eng | Man | Can | Eng-Man | Eng-Can |
|------------|------------|------------|------------|----------------|----------------|
| ΔV | 46.1004 | 36.8405 | 51.3237 | 52.6955 | 47.6612 |
| | (1.40705) | (2.16055) | (.89498) | (3.12098) | (2.89827) |
| ΔC | 56.9879 | 34.5581 | 38.2973 | 59.0983 | 60.7381 |
| | (1.87970) | (2.20080) | (2.06989) | (1.91801) | (5.27124) |
| %V | 41.5554 | 51.5982 | 52.7256 | 44.6327 | 43.4504 |
| | (.83937) | (2.01049) | (.70332) | (.60266) | (2.33328) |
| VarcoV | 59.3842 | 35.6646 | 47.4435 | 51.7630 | 50.4609 |
| | (4.17688) | (2.77137) | (1.28325) | (1.89642) | (4.31546) |
| VarcoC | 48.9780 | 35.2209 | 38.8724 | 47.3238 | 47.7061 |
| | (2.16890) | (2.34678) | (1.85020) | (2.07840) | (1.64346) |
| nPVI-V | 66.7095 | 39.5189 | 55.1649 | 57.4366 | 58.8619 |
| | (2.41580) | (3.00890) | (1.46207) | (1.00518) | (4.63546) |
| rPVI-C | 67.5853 | 41.7336 | 46.6333 | 69.1565 | 75.4827 |
| | (.56329) | (2.37952) | (3.23636) | (3.33706) | (6.52581) |

In this table, Eng refers to native speakers' English; Man refers to native Mandarin; and Can refers to native Cantonese. In addition, Eng-Man refers

to L2 English spoken by speakers with Mandarin as their first language; and Eng-Can refers to L2 English spoken by native speakers of Cantonese.

7.1 Analyses of Durational Metrics for First Languages

7.1.1 Interval Measures

One-way analysis of variance (one-way ANOVA) was run for the metrics of ΔV , ΔC and %V among the groups of native Mandarin, native Cantonese, and native English, to see if significant difference existed among them. There was a significant effect of the three groups of first languages on the scores of ΔV [$F(2, 12) = 21.667, p < 0.0001$]. Post hoc tests (Tukey HSD) showed that the score of ΔV for Mandarin ($M = 36.8405$) is significantly lower than that for English ($M = 46.1004$) [$p < 0.0001$] and Cantonese ($M = 51.3237$) [$p = 0.004$]. However, the difference between English and Cantonese on ΔV is not significant [$p = 0.088$].

For the scores of ΔC , there was a significant effect of the three groups of first languages [$F(2, 12) = 34.215; p < 0.0001$]. Post hoc tests (Tukey HSD) showed that significant difference does not exist between Mandarin ($M = 34.5581$) and Cantonese ($M = 38.2973$) [$p = 0.429$]. However, the score of English ($M = 56.9879$) is significantly higher than the scores for Mandarin ($M = 34.5581$) and Cantonese ($M = 38.2973$) [$p < 0.0001$].

For the scores of %V, there was a significant effect of the three groups of first languages [$F(2, 12) = 21.646, p < 0.0001$]. Post hoc tests (Tukey HSD) showed that significant difference does not exist between Mandarin ($M = 51.5982$) and Cantonese ($M = 52.7256$) [$p = 0.821$]. However, the score of English ($M = 41.5554$) is significantly lower than that of Mandarin ($M = 51.5982$) and Cantonese ($M = 52.7256$) [$p < 0.0001$].

In addition, for rate normalized standard deviation of vocalic interval duration VarcoV, a significant effect of the three groups of first languages also exists [$F(2, 12) = 15.761, p < 0.0001$]. Post hoc tests (Tukey HSD) showed that significant difference between Mandarin ($M = 35.6646$) and Cantonese ($M = 47.4435$) exists, but is only marginal [$p = 0.04$]. Similarly, the difference between Cantonese ($M = 47.4435$) and English ($M = 59.3842$) is also marginal [$p = 0.038$]. However, the difference between Mandarin ($M = 47.4435$) and English ($M = 59.3842$) is significant [$p < 0.0001$].

For rate normalized standard deviation of intervocalic interval duration VacroC, there exists a significant effect of the three groups of first languages [$F(2, 12) = 11.174, p = 0.002$]. Post hoc tests (Tukey HSD) showed that significant difference does not exist between Mandarin ($M = 35.2209$) and Cantonese ($M = 38.8724$) [$p = 0.469$]. However, the scores of VacroC are significantly different between Mandarin ($M = 35.2209$) and English ($M =$

48.9780) [$p = 0.002$], as well as between Cantonese ($M = 38.8724$) and English ($M = 48.9780$) [$p = 0.015$].

To sum up, for all metrics of interval measures, including rate normalized metrics, English is significantly different from either Mandarin or Cantonese, but no significant differences were found between Mandarin and Cantonese. In the next section, the analyses of pairwise variability indices (PVI) among the three first languages are reported.

7.1.2 Pairwise Variability Indices

One-way analysis of variance (one-way ANOVA) was run for the metrics of nPVI-V and rPVI-C among the groups of native Mandarin, native Cantonese, and native English, to see if significant difference existed among them. There was a significant effect of the three groups of first languages on the scores of nPVI-V [$F(2, 12) = 34.379, p = 0.0001$]. Post hoc tests (Tukey HSD) showed that significant differences between nPVI-V scores of Mandarin ($M = 39.5189$) and Cantonese ($M = 55.1649$) exist ($p = 0.002$). Moreover, the difference between nPVI-V scores of Mandarin ($M = 39.5189$) and English ($M = 66.7095$) was also significant ($p < 0.0001$). Similarly, the difference between nPVI-V scores of Cantonese ($M = 55.1649$) and English ($M = 66.7095$) was also significant ($p = 0.013$).

For the metric of rPVI-C, there was a significant effect of the three groups of first languages [$F(2, 12) = 34.379, p < 0.0001$]. Post hoc tests (Tukey HSD) showed that significant difference does not exist between rPVI-V scores of Mandarin ($M = 41.7336$) and Cantonese ($M = 46.6333$) [$p = 0.335$]. However, significant difference between Mandarin ($M = 41.7336$) and English ($M = 67.5853$) exists [$p < 0.0001$]. In a similar manner, significant difference [$p = 0.0001$] also exists between Cantonese ($M = 46.6333$) and English ($M = 67.5853$) on rPVI-C scores.

In all, as for PVI measures, nPVI-V distinguishes English, Mandarin and Cantonese from each other; while rPVI-C distinguishes English from Mandarin and Cantonese. However, significant difference between Mandarin and Cantonese on rPVI-C was not found. In section 7.1, the results of analyses of durational metrics for all first languages were reported. In section 7.2, the results of analyses of durational metrics for second languages will be reported.

7.2 Analyses of Durational Metrics for Second Languages

7.2.1 Mandarin vs. English-Mandarin vs. English

In this section, differences among native Mandarin, English spoken by native speakers of Mandarin (English-Mandarin) and native English are compared on all the metrics scores.

7.2.1.1 Interval Measures

One-way analysis of variance (one-way ANOVA) was run for the metrics of ΔV , ΔC and %V among the groups of native Mandarin, English spoken by L1 speakers of Mandarin (English-Mandarin), and native English, to see if significant difference existed among them. There was a significant effect of the three groups on the scores of ΔV [$F(2, 12) = 11.613, p = 0.002$]. Post hoc tests (Tukey HSD) showed that the score of ΔV for Mandarin ($M = 36.8405$) is significantly lower than that for English-Mandarin ($M = 52.6955$) [$p = 0.001$]. Nevertheless, the difference between English-Mandarin ($M = 52.6955$) and native English ($M = 46.1004$) was not significant [$p = 0.156$]. Since significant difference has already been found between native Mandarin and English, and results have been reported in section 7.1; therefore comparisons between native languages will not be reported in section 7.2, and this also holds for other metrics.

For the scores of ΔC , there was a significant effect of the three groups (native English, English-Mandarin, and English) [$F(2, 12) = 46.027, p < 0.0001$]. Post hoc tests (Tukey HSD) showed that significant difference existed between Mandarin ($M = 34.5581$) and English-Mandarin ($M = 59.0983$) [$p < 0.0001$]. However, the difference between English-Mandarin ($M = 59.0983$) and native English ($M = 56.9879$) is not significant [$p = 0.743$].

For the scores of %V, there was a significant effect of the three groups (Mandarin, English-Mandarin, native English) [$F(2, 12) = 15.543, p < 0.0001$]. Post hoc tests (Tukey HSD) showed that significant difference between Mandarin ($M = 51.5982$) and English-Mandarin ($M = 44.6327$) exists ($p = 0.007$). Nevertheless, difference between native English ($M = 41.5554$) and English-Mandarin ($M = 44.6327$) was not significant [$p = 0.257$].

In addition, for rate normalized standard deviation of vocalic interval duration VarcoV, there exists a significant effect of the three groups (Mandarin, English-Mandarin, native English) [$F(2, 12) = 15.316, p < 0.0001$]. Post hoc tests (Tukey HSD) showed that significant difference between Mandarin ($M = 35.6646$) and English-Mandarin ($M = 51.7630$) exists [$p = 0.008$]. However, significant difference between English-Mandarin ($M = 51.7630$) and native English ($M = 59.3842$) does not exist [$p = 0.230$].

For rate normalized standard deviation of intervocalic interval duration VacroC, there exists a significant effect of the three groups (Mandarin, English-Mandarin, and native English) [$F(2, 12) = 11.646, p = 0.002$]. Post hoc tests (Tukey HSD) showed significant difference exists between Mandarin ($M = 35.2209$) and English-Mandarin ($M = 47.3238$) [$p = 0.006$]. However, native English ($M = 48.9780$) and English-Mandarin ($M = 47.3238$) are not

significantly different from each other [$p = 0.858$].

In conclusion, section 7.2.1.1 reported the results of interval measures including rate normalized metrics compared among native Mandarin, L2 English spoken by native speakers of Mandarin, and native English. Significant differences were found between native Mandarin and L2 English; however, differences between native English and L2 English for all metrics of interval measures were not significant.

7.2.1.2 Pairwise Variability Indices

One-way analyses of variance (one-way ANOVA) were conducted for the metrics of nPVI-V and rPVI-C among the groups of native Mandarin, native English, and L2 English spoken by native speakers of Mandarin, to see if significant difference existed among them. There was a significant effect of the three groups (Mandarin, English-Mandarin, and native English) on the scores of nPVI-V [$F(2, 12) = 36.049, p < 0.0001$]. Post hoc tests (Tukey HSD) showed that significant differences between nPVI-V scores of native Mandarin ($M = 39.5189$) and English-Mandarin ($M = 57.4366$) exists [$p < 0.0001$]. However, the difference between English-Mandarin ($M = 57.4366$) and native English ($M = 66.7095$) is only marginal [$p = 0.036$].

For the metric of rPVI-C, there was a significant effect of the three groups

(native Mandarin, English-Mandarin, and native English) [$F(2, 12) = 41.565$, $p < 0.0001$]. Post hoc tests (Tukey HSD) showed that significant difference exists between Mandarin ($M = 41.7336$) and English-Mandarin ($M = 69.1565$) [$p < 0.0001$]. However, significant difference does not exist between English-Mandarin ($M = 69.1565$) and native English ($M = 67.5853$) [$p = 0.889$].

In general, both nPVI-V and rPVI-C distinguish native Mandarin from L2 English spoken by L1 Mandarin speakers. However, significant difference between native English and L2 English was not found on the rPVI-C metric; and the difference between native English and L2 English on the nPVI-V metric was only marginally significant.

7.2.2 Cantonese vs. English-Cantonese vs. English

In this section, differences among native Cantonese, L2 English spoken by native speakers of Cantonese (English-Cantonese) and native English are compared on all the metrics scores.

7.2.2.1 Interval Measures

One-way analysis of variance (one-way ANOVA) was conducted for the metrics of ΔV , ΔC and $\%V$ among the groups of native Cantonese, English spoken by L1 speakers of Cantonese (English-Cantonese), and native English, to see if significant difference existed among them. For the scores of

ΔV , the effect of the three groups (Cantonese, English-Cantonese, and native English) was not significant [$F(2, 12) = 1.929, p = 0.188$]. And a post hoc test (Tukey HSD) failed to find significant differences among the three groups.

For the scores of ΔC , there was a significant effect of the three groups (Cantonese, English-Cantonese, native English) [$F(2, 12) = 12.176, p = 0.001$]. Post hoc tests (Tukey HSD) manifested that significant difference between English-Cantonese ($M = 60.7381$) and native Cantonese ($M = 38.2973$) exists [$p = 0.002$]. However, difference between English-Cantonese ($M = 60.7381$) and native English ($M = 56.9879$) was not significant [$p = 0.728$].

For the scores of $\%V$, a significant effect of the three groups (Cantonese, English-Cantonese, native English) exists [$F(2, 12) = 16.136, p < 0.0001$]. Post hoc tests (Tukey HSD) showed that significant difference exists between native Cantonese ($M = 52.7256$) and English-Cantonese ($M = 43.4504$) [$p = 0.002$]. But the difference between English-Cantonese ($M = 43.4504$) and native English ($M = 41.5554$) is not significant [$p = 0.650$].

Furthermore, for rate normalized standard deviation of vocalic interval duration VarcoV , there exists no significant effect of the three groups (Cantonese, English-Cantonese, native English) [$F(2, 12) = 3.066, p = 0.084$]. Post hoc tests (Tukey HSD) failed to find significant differences among the

three groups.

For rate normalized standard deviation of intervocalic interval duration *VacroC*, there exists a significant effect of the three groups (Cantonese, English-Cantonese, and native English) [$F(2, 12) = 8.393, p = 0.005$]. Post hoc tests (Tukey HSD) showed that significant differences between *VarcoC* scores of native Cantonese ($M = 38.8724$) and English-Cantonese ($M = 47.7061$) exists [$p = 0.017$]. However, significant difference does not exist between English-Cantonese ($M = 47.7061$) and native English ($M = 48.9780$) [$p = 0.885$].

In summary, standard deviation of vocalic interval duration, both raw and rate normalized, failed to find significant differences among the groups of native Cantonese, native English, and L2 English of Cantonese speakers. However, the other metrics of interval measures find significant differences between Cantonese and L2 English, but failed to find significant differences between native English and L2 English.

7.2.2.2 Pairwise Variability Indices

One-way analyses of variance (one-way ANOVA) were conducted for the metrics of *nPVI-V* and *rPVI-C* among the groups of native Cantonese, native English, and L2 English spoken by native speakers of Cantonese, to see if

significant differences existed among them. A significant effect of the three groups of native Cantonese, native English and L2 English on the scores of nPVI-V was not found [$F(2, 12) = 3.539, p = 0.062$]. And post hoc tests (Tukey HSD) failed to find significant differences among the three groups.

However, a significant effect of the three groups (native Cantonese, native English, and English-Cantonese) exists on the scores of rPVI-C [$F(2, 12) = 12.493, p = 0.001$]. Post hoc tests (Tukey HSD) showed that significant difference exists between Cantonese ($M = 46.6333$) and English-Cantonese ($M = 75.4827$) [$p = 0.001$]. However, significant difference was not found between English-Cantonese ($M = 75.4827$) and native English ($M = 67.5853$) [$p = 0.409$].

In general, nPVI-V does not distinguish among the groups of native Cantonese and English, and L2 English by Cantonese speakers. However, rPVI-C distinguishes native Cantonese and L2 English, but also failed to distinguish L2 English from native English.

7.2.3 English-Mandarin vs. English-Cantonese vs. English

In this section, differences among the three types of English (i.e. native English, L2 English spoken by both Mandarin and Cantonese speakers) on all the durational metrics will be compared.

7.2.3.1 Interval Measures

One-way analysis of variance (one-way ANOVA) was conducted for the metrics of ΔV , ΔC and %V among the groups of L2 English of Cantonese native speakers (English-Cantonese), L2 English of speakers of Mandarin (English-Mandarin), and native English, to see if significant difference existed among them. For the scores of ΔV , the effect of the three groups (English-Mandarin, English-Cantonese, and native English) was not significant [$F(2, 12) = 1.771, p = 0.212$]. For the scores of ΔC , there was also no significant effect of the three groups (English-Mandarin, English-Cantonese, and native English) [$F(2, 12) = 0.303, p = 0.744$]. In terms of %V, the effect of the three groups (English-Mandarin, English-Cantonese, and native English) was not significant either [$F(2, 12) = 1.110, p = 0.361$]. Post hoc tests (Tukey HSD) failed to find significant differences among the three groups for the scores of ΔV , ΔC , and %V.

As for rate normalized standard deviations of the duration of vocalic intervals and intervocalic intervals (VarcoV and VarcoC), significant effect of language groups (English-Mandarin, English-Cantonese, and native English) was not found either [for VarcoV, $F(2, 12) = 1.757, p = 0.214$; for VarcoC, $F(2, 12) = 0.192, p = 0.828$]. Moreover, Post hoc tests (Tukey HSD) failed to find significant differences among the three groups for the scores of VarcoC and

VarcoV.

7.2.3.2 Pairwise Variability Indices

One-way analyses of variance (one-way ANOVA) were run for the metrics of nPVI-V and rPVI-C among the groups of native English, L2 English by L1 Mandarin speakers, and L2 English by native speakers of Cantonese, to see if significant differences existed among them.

A significant effect of the three groups (native English, L2 English by L1 Mandarin speakers, and L2 English by native speakers of Cantonese) on the scores of nPVI-V was not found [$F(2, 12) = 2.640, p = 0.112$]. Furthermore, a significant effect of the three groups was also absent on the scores of rPVI-C [$F(2, 12) = 0.970, p = 0.407$]. Post hoc tests (Tukey HSD) failed to find significant differences among the three groups for the scores of nPVI-V, and rPVI-C.

In conclusion, for all the durational metrics, including interval measures and pairwise variability indices, significant differences were not found among the groups of native English, L2 English by Mandarin speakers and L2 English by Cantonese speakers.

8 Discussion

8.1 Rhythm of First Languages

8.1.1 Interval Measures

As has been reported in section 7.1.1, English is significantly higher than both Mandarin and Cantonese (as is shown in Figure 4) in terms of ΔC , which is enabled by complex syllable structure of English. As for Mandarin and Cantonese, they are not significantly different on ΔC , since both of their syllable structures are simpler. However, in terms of ΔV , Mandarin is significantly different from English and Cantonese, but there is no significant difference between English. One possible explanation for this is that both Cantonese and English have phonemic distinction between long and short vowels, which caused higher standard deviation of vocalic intervals duration. As for %V, English is significantly lower than Mandarin and Cantonese, which may be caused by both its complex syllable structure and reduced vowels.

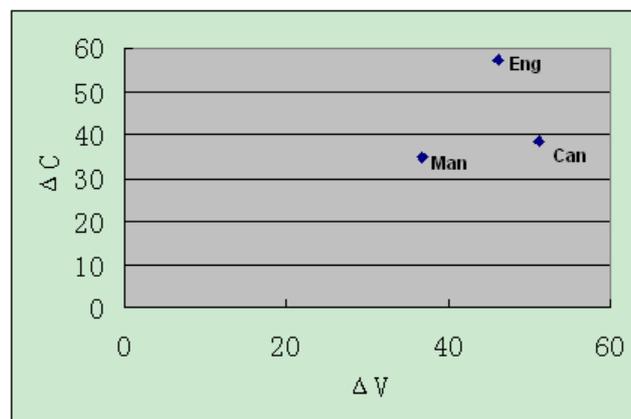


Figure 4 Distribution of English, Mandarin and Cantonese over the $\Delta C \times \Delta V$ plane.

As for rate normalized ΔV , namely VarcoV, English is significantly higher than Cantonese, which is again higher than Mandarin, although the significance between Mandarin and Cantonese as well as the difference between English and Cantonese is only marginal [Man vs. Can: $p = 0.04$; Eng vs. Can: $p = 0.038$]. As for rate normalized ΔC , i.e. VarcoC, English is significantly higher than both Mandarin and Cantonese; whereas the difference between Cantonese and Mandarin is not significant. Such results suggested that vocalic intervals duration measure is sensitive to speech rate, which is in line with Grabe & Low (2002).

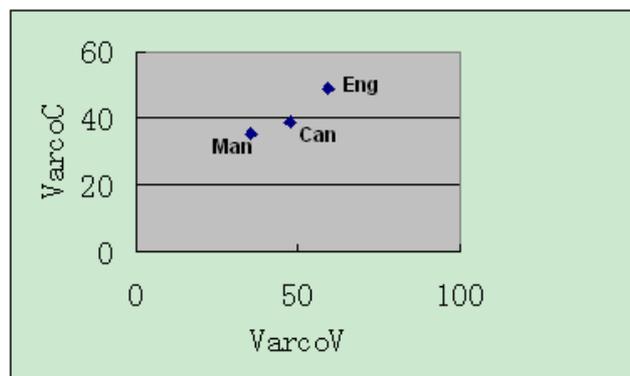


Figure 5 Distribution of English, Mandarin and Cantonese over the VarcoC \times VarcoV plane.

In general, when synthesizing the results of both ΔV , ΔC and $\%V$, as well as VarcoV and VarcoC metrics, English is more isolated from both Mandarin and Cantonese, and is more stress-timed as been proved by a myriad other studies. At the same time, Mandarin and Cantonese tend to cluster, and have lower scores of almost all metrics, suggesting their syllable-timed

rhythmicity. Between Cantonese and Mandarin, Cantonese is significantly higher than Mandarin in terms of ΔV and V_{acroV} , showing a less syllable-timed rhythmicity than Mandarin, which does not conform to the hypothesis made in section 6.1 that Cantonese is more syllable-timed than Mandarin because Mandarin has higher degrees in reducing vowels in unstressed syllables. One possible explanation for such disconformity is that the sentence materials are supposed to be read aloud by the participants; therefore they were more inclined to use “citation forms” while reading; thus vowels in unstressed syllables were not reduced to a significant extent.

8.1.2 Pairwise Variability Indices

As has been reported in section 7.1.2, English is significantly higher than Cantonese and Mandarin in terms of nPVI-V, and in the meanwhile, Cantonese is higher than Mandarin for the same metric (see also Figure 6). Conforming to the results of other studies, English shows typical stress-timing in terms of nPVI-V, because not only does it have phonemically distinguished long and short vowels, but also a high degree of vowel reduction. Therefore, the difference between successive vocalic intervals duration is greater. As for Cantonese which has phonemically different long and short vowels, it is natural to have higher nPVI-V than Mandarin, which does not have long-short vowel distinction. Moreover, in carefully reading tasks, participants were not likely to reduce unstressed

vowels; therefore Mandarin has the lowest nPVI-V.

As for rPVI-C, English has the highest scores too, due to its complex syllable structure. Both Mandarin and Cantonese have simpler syllable structures; therefore no significant significance exists between them.

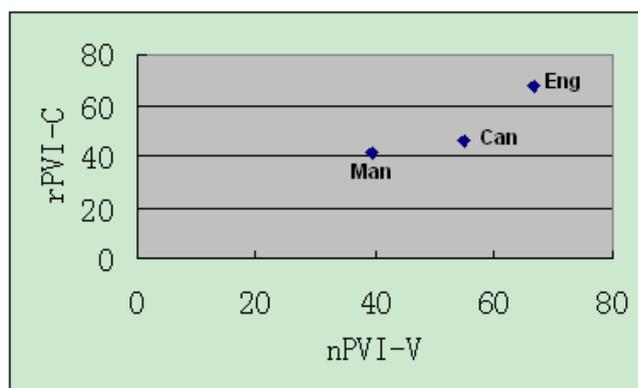


Figure 6 Distribution of English, Mandarin and Cantonese over the nPVI-V × rPVI-C plane.

To sum up, for both metrics of pairwise variability indices, English has the highest degree of stress-timing, as is also shown in the plane of Figure 6, English is isolated from both Mandarin and Cantonese. As for Mandarin and Cantonese, Cantonese has a less degree of syllable-timing, judging from its significant higher nPVI-V. Similar to interval measures, this does not conform to the hypothesis either, and explanation for this is partly because of less reduced unstressed vowels in Mandarin, and partly due to long-short vowel distinction in Cantonese.

8.2 Rhythm of Second Languages

8.2.1 Interval Measures

As has been reported in section 7.2.1.1, Mandarin is significantly lower than both English and English-Mandarin in terms of both ΔV and ΔC . However, the difference between English and English-Mandarin is not significant in terms of both ΔV and ΔC . According to Figure 7, the two types of English, namely native English and L2 English cluster together, and have a higher degree of stress-timing; whereas, Mandarin is isolated, showing a typical syllable-timed rhythm. Such results suggest that the L2 English speakers with Mandarin as first language have acquired the rhythm of English to a large extent. To be more specific, they acquired the long-short vowel distinction, and reduced vowels in weak forms; all these ensure a higher standard deviation of vocalic intervals duration. Moreover, they also acquired the syllable structure of English, particularly consonant clusters in the onset of the coda positions. Epenthesis of vowels between elements in a consonant cluster is avoided, which is different from what a lot of Chinese who began to learn English do. They always break consonant clusters with added vowels to conform to the simpler Chinese syllable structure. As for %V, native Mandarin is significantly different from English and English-Mandarin, while no significant difference exists between English and English-Mandarin, which also suggests that the participants have acquired the rhythm of English to a large extent.

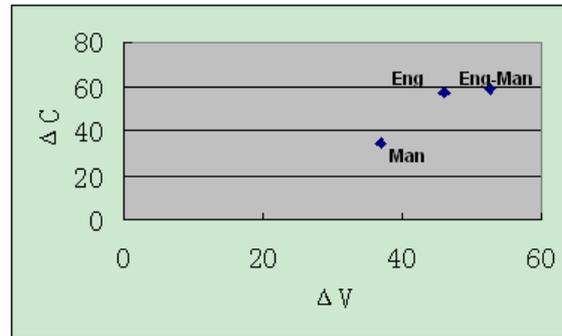


Figure 7 Distribution of Mandarin, English and English-Mandarin over the $\Delta C \times \Delta V$ plane.

As for Cantonese learners of English, they also manifested a high degree of acquisition of English rhythm. As has been reported in section 7.2.2.1, Cantonese is not significantly lower than English and English-Cantonese in terms of ΔC , and the difference between English and English-Cantonese is not significant. Nevertheless, from Figure 8 it is apparent that English and English-Cantonese cluster, showing their shared rhythmic class as stress-timed. Also from the results of %V measures, English and English-Cantonese cluster, and both types of English are significantly different from Cantonese. This further suggests that English and English-Cantonese belong to the same rhythmic category.

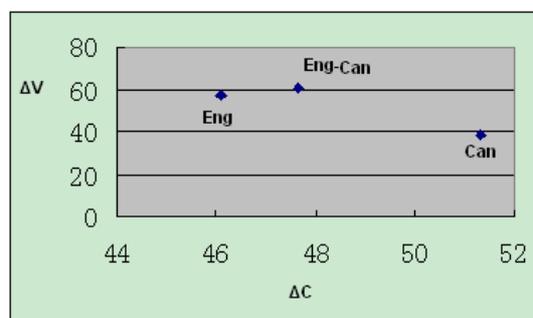


Figure 8 Distribution of Cantonese, English and English-Cantonese over the $\Delta C \times \Delta V$ plane.

As for rate normalized metrics like VarcoV and VarcoC, Mandarin is significantly lower than both native English and English-Mandarin; whereas the difference between English and English-Mandarin is not significant for both VarcoV and VarcoC. Such results are similar to the results of ΔV and ΔC measures, suggesting that speech rate does not influence rhythmic measures in this case. However, as Figure 10 shows, although English-Cantonese and Cantonese seem to cluster, the difference between English-Cantonese and English on VarcoC is not significant. The similar VarcoV between Cantonese and English-Cantonese may suggest that speech rate plays a role in categorizing speech rhythm.

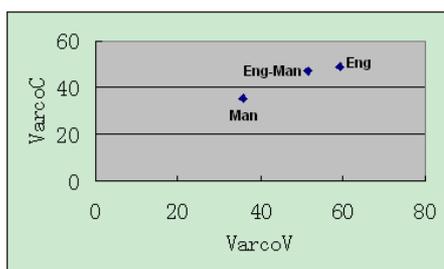


Figure 9

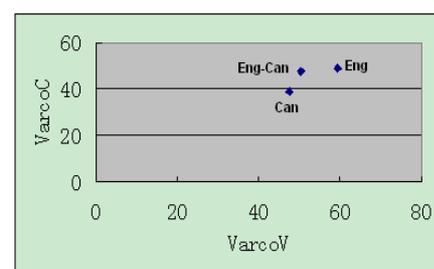


Figure 10

Figures 9 & 10 Distributions of English, Mandarin/Cantonese and English-Mandarin/Cantonese over the VarcoC \times VarcoV plane.

In general, metrics of interval measures successfully distinguished syllable-timed Mandarin and Cantonese for both native English, and L2 English. This may suggest that the participants whose L1's are Mandarin and Cantonese have acquired the stress-timing of English rhythm.

8.2.2 Pairwise Variability Indices

Similar to the results of interval measures, English-Mandarin and English-Cantonese are significantly different from the L1's, but are clustered with native English on the measures of nPVI-V and rPVI-C. As Figures 11 and 12 show, English-Mandarin and English-Cantonese cluster with native English. This implied that both groups of native speakers of Mandarin and Cantonese broke away from the rhythmic constraints of the first languages, and have acquired the stress-timed rhythm of English.

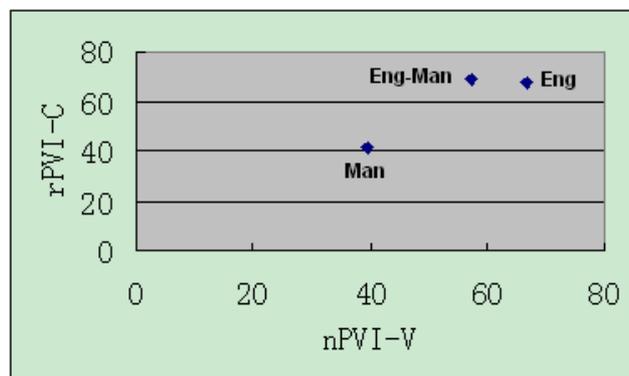


Figure 11 Distribution of English, Mandarin and English-Mandarin over the nPVI-V \times rPVI-C plane.

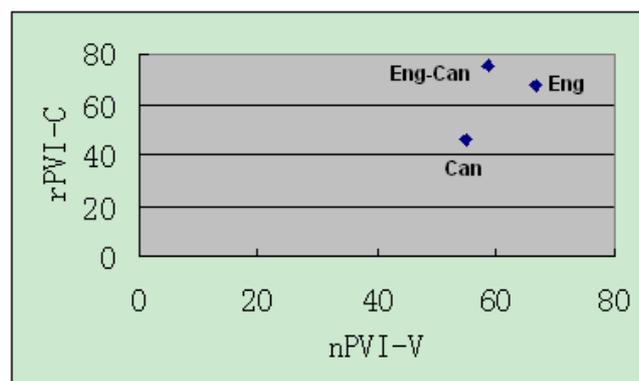


Figure 12 Distribution of English, Cantonese and English-Cantonese over the nPVI-V \times rPVI-C plane.

8.2.3 Summary

As the results show, L2 English spoken by native speakers of Mandarin and Cantonese are significantly different from native Mandarin and Cantonese for all the measures of durational metrics. Moreover, both types of L2 English (English-Mandarin and English-Cantonese) cluster with English on almost all metrics, manifesting the rhythmicity of stress-timing. Also as has been reported in section 7.2.3, native-English, English-Mandarin and English-Cantonese are not significantly different from each other on all scores of durational metrics. This implies that the L2 learners of English have acquired the stress-timing rhythm of English to a large extent. Or to be more specific, they acquired the complex syllable structure of English, and vowel reduction, as well as long-short vowels of English.

Nevertheless, this does not mean that the rhythm of English-Mandarin and English-Cantonese is identical with the rhythm of native. A number of scholars (Barry et al., 2009; Arvaniti, 2009; Kohler, 2009) questioned the reliability of durational metrics of rhythm, and advocating other methods for the study of rhythm. For example, Kohler (2009) proposed a paradigm shift to broader issues like perception of rhythm and communication function of rhythm. Therefore, if the rhythm of English-Mandarin and English-Cantonese is studied using other research paradigms, their

distinctiveness from native English may be found. Moreover, the L2 participants of this study are all post-intermediate learners of English, and only one modality of reading (i.e. sentence reading) may all contribute to the results.

8.3 Suggestions for Further Research

First of all, L2 English learners of different proficiency could be enrolled, so that the trajectory of rhythm acquisition may be obtained. For example, beginners might break English consonant clusters with epenthesis vowels, resulting in more syllable-timed rhythm.

Secondly, different modalities of reading could be implemented, such as sentence reading, passage reading and story narration. For example, in passage reading, vowel lengthening at the end of a prosodic phrase may be very common, and in story narration, the participants may speak in a very natural way in their L1, so that reduced vowels in unstressed Mandarin syllables may be common. Moreover, in L2 sentence reading, the participants may be very careful about the way in which the sentences are supposed to be read, resulting in a very native-like rhythmic pattern. However in story narration, they may not have as much mental capacity to process how the rhythm should be like, thus features deviant from native English could be discovered.

9 Conclusion

This study investigated the speech rhythm of L2 English (spoken by native speakers of Mandarin and Cantonese), as well as the rhythm of native English, Mandarin, and Cantonese. By using different durational metrics of speech rhythm, such as interval measures (ΔV , ΔC , %V, VarcoV and VarcoC) as well as pairwise variability indices (rPVI-C and nPVI-V), it has been found that both Mandarin and Cantonese deviate from English in almost all metrics, showing their distinct rhythm class as syllable-timed. However, different from what has been hypothesized, English-Mandarin and English-Cantonese failed to show features of syllable-timing. They cluster with native English on all durational metrics, showing a rhythmic class of being stress-timed. Therefore, the results implied that post-intermediate learners of English with mother tongues of Mandarin and Cantonese have avoided the negative rhythmic transfer from their native languages, and have acquired English rhythm to a great extent. However, the results do not suffice to prove that the participants manifested the English rhythm as identical with English native speakers, and suggestions for further research were made in the end.

References

- Abercrombie, D. (1967). *Elements of general phonetics*. Edinburgh: Edinburgh University Press.
- Arvaniti, A. (2009). Rhythm, timing, and the timing of rhythm. *Phonetica*, 66: 46-63.
- Barry, W. J., Andreeva, B., Russo, M., Dimitrova, S. & Kostadinova, T. (2003). Do rhythm measures tell us anything about language type? In M. J. Sol'e, D. Recasens & J. Romero (eds.), *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 2693-2696). Barcelona/Australia: Casual Productions.
- Barry, W. J., Andreeva, B., Koreman, J. (2009). Do rhythm measures reflect perceived rhythm? *Phonetica*, 66: 78-94.
- Beckman, M. E. (1982). Segment duration and the "mora" in Japanese. *Phonetica*, 39: 113-135.
- Beckman, M. E. (1992). Evidence for speech rhythm across languages. In Y. Tohkura, E. Vatikiotis-Bateson & Y. Sagisaka (eds.), *Speech perception, production and linguistic structure* (pp. 457-463). Amsterdam: IOS Press.
- Bertinetto, P. M. (1989). Reflections of the dichotomy stress versus syllable timing. *Revue de phonétique appliquée*, 91-3: 99-130.
- Bosch, L., Sebastian-Galles, N. (1997). "The role of prosody in infants' native-language discrimination abilities: the case of two phonologically

- close languages", In *EUROSPEECH-1997*, 231-234.
- Cheng, C. (1973). *A synchronic phonology of Mandarin Chinese*. The Hague: Mouton.
- Cheng, R. L. (1966). Mandarin phonological structure. *Journal of Linguistics*, 2: 135-262.
- Clark, J., Yallop, C. & Fletcher, J. (2007). *Introduction to Phonetics and Phonology*. Oxford: Blackwell.
- Crosswhite, K. (2003). Vowel reduction. In B. Hayes, R. Kirchner & D. Steriade (eds.), *Phonetically based phonology* (pp. 191-231). Cambridge: Cambridge University Press.
- Crystal, T. H. & House, A. S. (1990). Articulation rate and the duration of syllables and stress groups in connected speech. *Journal of Acoustical Society of America*, 88: 101-112.
- Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics*, 11: 51-62.
- Dauer, R. M. (1987). Phonetic and phonological components of language rhythm. *Proceedings of the 11th International Congress of Phonetic Sciences*, Tallinn, Estonia, 5, 447-450.
- Delattre, P. (1966). A comparison of syllable length conditioning among Languages, *International Review of Applied Linguistics*, 4: 183-198.
- Delattre, P. (1969). An acoustic and articulatory study of vowel reduction in four languages. *International Review of Applied Linguistics*, 7: 295-325.

-
- Dellwo, V. (2004). The Bonn tempo-corpus and Bonn tempo-tools: a database for the study of speech rhythm and rate. In *Proceedings of the 8th ICSLP*, Jeju Island, Korea.
- Dellwo, V. (2006). Rhythm and speech rate: a variation coefficient for deltaC. *Proceedings of the 38th Linguistic Colloquium*.
- Deterding, D. (2001). The measurement of rhythm: a comparison of Singapore English and British English. *Journal of Phonetics*, 29: 217-230.
- Duanmu, S. (2007). *The phonology of standard Chinese* 2nd edition. Oxford: Oxford University Press,
- Fletcher, J. (1991). Rhythm and final lengthening in French. *Journal of Phonetics*, 19: 193-212.
- Fletcher, J. (2010). The prosody of speech: timing and rhythm. In W. J. Hardcastle, J. Laver & F. E. Gibbon (eds.), *The Handbook of Phonetic Sciences* 2nd edition. Oxford: Wiley-Blackwell.
- Giegerich, H. J. (1992). *English phonology: an introduction*. Cambridge: Cambridge University Press.
- 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. 515-543). Berlin: Mouton de Gruyter.
- Gussenhoven, C. & Jacobs, H. (2005). *Understanding Phonology* 2nd edition. London: Hodder Arnold.
- Han, Z. & Odlin, T. (Eds.) (2006). *Studies of Fossilization in Second Language*

-
- Acquisition*. Clevedon: Multilingual Matters.
- Hoequist, C., Jr. (1983). Durational correlates of linguistic rhythm categories. *Phonetica*, 40: 19-31.
- Howie, J. (1976). *An acoustic study of Mandarin tones and vowels*. London: Cambridge University Press.
- Jian, H.-L. (2004). An acoustic study of speech rhythm in Taiwan English, In *INTERSPEECH-2004*, pp. 1265-1268.
- Kachru, B. B., Kachru, Y. & Nelson, C. (2009). *The Handbook of World Englishes*. Oxford: Wiley-Blackwell.
- Kao, D. L. (1971) *Structure of the syllable in Cantonese*. The Hague: Mouton.
- Karlgren, B. (1915-1926). *Etudes sur la phonologie chinois*. Leyde: E.-J. Brill.
- Karmiloff, K. & Karmiloff-Smith, A. (2001). *Pathways to language: from fetus to adolescent*. Cambridge, MA: Harvard University Press.
- Kohler, K. J. (2009). Rhythm in speech and language. *Phonetica*, 66: 26-45.
- Koopmans-Van Beinum, F. J. (1980). Vowel contrast reduction, an acoustic and perceptual study of Dutch vowels in various speech conditions. Ph.D. Dissertation, Academische Pers B.V., Amsterdam.
- Lee, W.-S. & Zee, E. (2003). Standard Chinese (Beijing). *Journal of the International Phonetic Association*, 33: 109-112.
- Lin, H. & Wang, Q. (2007). Mandarin rhythm: an acoustic study. *Journal of Chinese Language and Computing* 17: 127-140.
- List, G. (1963). The boundaries of speech song. *Ethnomusicology*, 7: 1-16.

-
- Low, E. L. & Grabe, E. (1995). Prosodic patterns in Singapore English. *Proceedings of the XIIIth International Congress of Phonetic Sciences*, Stockholm.
- Low, E. L., Grabe, E. & Nolan, F. (2000). Quantitative characterizations of speech rhythm: syllable-timing in Singapore English. *Language and Speech*, 43: 377-401.
- de Manrique, B. A. M. & Signorini, A. (1983). Segmental durations and the rhythm in Spanish. *Journal of Phonetics*, 11: 117-132.
- Merriam, A. (1964). *The anthropology of music*. Chicago: Northwestern University Press.
- Mok, P. P. K. (forthcoming). On the syllable-timing of Cantonese and Beijing Mandarin. To appear in *Chinese Journal of Phonetics*.
- Mok, P. P. K. & Dellwo, V. (2008). Comparing native and non-native speech rhythm using acoustic rhythmic measures: Cantonese, Beijing Mandarin and English. *Speech Prosody 2008*. Campinas, Brazil.
- Nazzi, T., Bertoni, J. & Jacques, M. (1998). Language discrimination by newborns: towards an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance*, 24: 756-766.
- Nazzi, T., Jusczyk, P.W., & Johnson, E.K. (2000). Language discrimination by English-learning 5-month-olds: effect of rhythm and familiarity. *Journal of Memory and Language*, 43: 1-19.
- Nespor, I. (1990). On the rhythm parameter in phonology. In I. Roca (ed.)

-
- Logical issues in language acquisition* (pp. 157-175). Dordrecht: Foris.
- den Os, E. (1988). Rhythm and tempo of Dutch and Italian: A contrastive study. Doctoral Dissertation, Rijksuniversiteit, Utrecht.
- Odlin, T. (1989). *Language transfer: cross-linguistic influence in language learning*. Cambridge: Cambridge University Press
- Pike, K. (1945). *The intonation of American English*. Ann Arbor: University of Michigan Press.
- Pointon, G. E. (1980). Is Spanish really syllable-timed? *Journal of Phonetics*, 8: 293-304.
- Port, R. F., Dalby, J. & O'Dell, M. (1987). Evidence for mora-timing in Japanese. *Journal of the Acoustical Society of America*, 29: 191-215.
- Pullum, G. & Ladusaw, W. A. (1996). *Phonetic symbol guide* 2nd edition. Chicago: The University of Chicago Press.
- Ramus, F. (2002). Acoustic correlates of linguistic rhythm: perspectives. *Proceedings of speech prosody 2002, Aix-en-Provence*, 115-120.
- Ramus, F. & Mehler, J. (1999). Language identification with suprasegmental cues: a study based on speech resynthesis. *Journal of the Acoustical Society of America*, 105: 512-521.
- Ramus, F., Nespors, M. & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, 73: 265-292.
- Ramus, F., Dupoux, E. & Mehler, J. (2003). The psychological reality of rhythm classes: perceptual studies. *Proceedings of the 15th International*

Congress of Phonetic Sciences, Barcelona.

- Roach, P. (1982). On the distinction between “stress-timed” and “syllable-timed” languages. In D. Crystal (ed.), *Linguistic controversies*. London: Edward Arnold.
- Roach, P. (2000). *English Phonetics and Phonology* 3rd edition. Cambridge: Cambridge University Press.
- Roach, P. (2009). *English Phonetics and Phonology* 4th edition. Cambridge: Cambridge University Press.
- Sun, C. (2006). *Chinese: A linguistic introduction*. Cambridge: Cambridge University Press.
- Wenk, B. J. (1985). Speech rhythms in second language acquisition. *Language and Speech*, 28: 157-175.
- White, L. & Mattys, S. L. (2007a). Calibrating rhythm: first and second language studies. *Journal of Phonetics*, 35: 501-522.
- White, L. & Mattys, S. L. (2007b). Rhythm typology and variation in first and second languages. In P. Prieto, J. Mascaró & M.-J. Solé (eds.) *Segmental and Prosodic Issues in Romance Phonology* (pp. 237-257). Amsterdam: John Benjamins.
- Wiget, L., White, L., Schuppler, B., Grenon, I., Rauch, O. & Mattys, S. L. (2010). How stable are acoustic metrics of contrastive speech rhythm? *Journal of Acoustical Society of America*, 127: 1559-1569.
- Zhou, D. & Wu, Z. (1963). *Putonghua fayin tu pu*. Beijing: Shangwu

Yinshuguan.