

Mora-timed, stress-timed, and syllable-timed rhythm classes: Clues in English speech production by bilingual speakers

SHA LIU*  and KAYE TAKEDA

Fukuoka University, Japan

Received: February 20, 2021 • Revised manuscript received: August 3, 2021 • Accepted: August 9, 2021

Published online: September 10, 2021

© 2021 The Authors



ABSTRACT

A growing body of literature suggests that the world's languages can be classified into three rhythm classes: mora-timed languages, stress-timed languages, and syllable-timed languages. However, scholars cannot agree on which rhythmic measures discriminate rhythm classes most satisfactorily and whether the speech rate factor should be considered. In this study, we analyze speech production by bilingual speakers, and compare their production with that of monolingual speakers and ESL speakers. Our rhythmic metric measure results show that when speech rate is taken into consideration, a combination of the two metric measures for vowels, *Varco* ΔV and *vocalic nPVI*, is most reliable in discriminating different rhythm classes, while consonants do not seem effective, whether the speech rate factor is included or not.

KEYWORDS

bilingual, ESL, monolingual, rhythm class, rhythmic metric measure

1. INTRODUCTION

Abercrombie (1965, 1967) claims that the world's languages fall into three rhythmic groups: mora-timed languages, stress-timed languages, and syllable-timed languages (see also Pike 1946; Ladefoged 1975; Ramus et al. 1999). According to Ramus et al. (1999, 266), “rhythm type is correlated with the speech segmentation unit in any given language.” For example, English is a

* Corresponding author. E-mail: liusha@fukuoka-u.ac.jp

typical stress-timed language and speakers of English segment speech in feet, while Japanese is a representative mora-timed language and speakers of Japanese segment speech in morae. Following this claim, clear differences can be expected between the rhythm of English produced by native speakers of English and the rhythm of English produced by native speakers of Japanese who speak English as a second language (Japanese ESL speakers). This is because the rhythm of English produced by Japanese ESL speakers is expected to be influenced by the rhythm of Japanese. Numerous studies have been carried out to explore the influence of the rhythm of a first language on that of a second language (Wenk 1985; Munro & Derwing 1995; Archibald 1998; White & Mattys 2007; Li & Post 2014; Ordin & Polyanskaya 2015). Most agree that there is influence, which at least partly proves the validity of rhythm classification. However, scholars cannot agree on which rhythmic measures demonstrate the influence most satisfactorily. Another interesting topic includes using rhythmic measures to examine speeches by bilingual speakers. It seems that results from bilingual speakers should be between results from monolingual native speakers and those from second language speakers. However, very few studies have tackled this topic. Even fewer studies have compared rhythms of speeches by different bilingual speakers.

In this paper, we focus on bilingual speakers and compare their English speech production with that of monolingual native speakers of English and English speech production by ESL speakers. For a comprehensive view, we include monolingual English speakers, English-Japanese bilingual speakers, English-Mandarin bilingual speakers, Japanese ESL speakers, and Mandarin ESL speakers. This is because English, Japanese, and Mandarin each belong to a different rhythm class: English is an example of a stress-timed language, Japanese a mora-timed language, and Mandarin a syllable-timed language. Our aim is to compare the rhythms of these five groups of speakers, examine the validity of rhythm classification, and find the most discriminative metric measures in distinguishing rhythm classes. We not only focus on differences between the monolingual English speakers, the bilingual speakers, and the ESL speakers, but also on differences between the English-Japanese bilingual speakers and the English-Mandarin bilingual speakers. The reason for this is that the English-Japanese bilingual speakers and English-Mandarin bilingual speakers may show differences in English speech production in terms of rhythmic measures due to influences of the rhythms of Japanese and Mandarin, respectively, although the differences should not be overwhelmingly large.

The paper is organized as follows. Section 2 reviews rhythmic metric measures. Section 3 presents details in our experiment. Sections 4 through 6 give related analyses of the experiment results, examine the validity of rhythm classification, and find the most discriminative metric measures in distinguishing rhythm classes. And Section 7 outlines the conclusion of this paper.

2. RHYTHM CLASS HYPOTHESIS AND RHYTHMIC METRIC MEASURES

According to Abercrombie (1967, 96), rhythm is “the periodic occurrence of some sort of movement,” which produces “an expectation that the regularity of succession will continue.” Rhythm also exists in speech (Steele 1775; Pike 1946; Abercrombie 1965, 1967). For example, Pike (1946) and Abercrombie (1965, 1967) claim that stress-timed languages have roughly equal foot durations and syllable-timed languages have roughly equal syllable durations, where *foot* is



“the time interval between two stress beats” in English (Gut 2009, 160). Based on this assumption, Pike (1946) and Abercrombie (1965, 1967) further propose a dichotomy of the world’s languages into stress-timed and syllable-timed languages, and they argue that the dichotomy of languages is categorical. Typical examples of stress-timed languages include Dutch, English, German, etc. Representative examples of syllable-timed languages include French, Mandarin, Spanish, and so on. Bloch (1950), Han (1962), and Ladefoged (1975) propose a third type of rhythm, mora-timed languages, which is mainly exemplified by Japanese. Ladefoged & Johnson (2015, 261) suggest that “[a] mora is a unit of timing, in the sense that each mora . . . ha[s] approximately the same duration.” McCawley (1968, 133–134; 1978, 114) claims that the mora is “the unit of phonological distance” in Japanese and can only be defined as “something of which a long syllable consists of two and a short syllable consists of one.” According to Kubozono (2015), overlap exists between the mora and the syllable. To exemplify, in the word ‘Nagoya’, “each mora corresponds to a syllable” (Kubozono 2015, 11). Kubozono (2015, 11) also points out that the mora and the syllable do not always overlap: they “often fail to overlap” in many Sino-Japanese words and loan words from English and other Western languages. This is because some morae cannot form a syllable on their own. Such morae include (a) the second half of a long vowel, (b) the second half of a diphthong, (c) a moraic nasal or a coda nasal, and (d) a moraic obstruent or the first half of a geminate consonant (Kubozono 2015). Take the word /nɪp.pon/ ‘Japan’ as an example: the underlined /p/ is a moraic obstruent and the underlined /n/ is a moraic nasal (Kubozono 2015). They both are morae, but they cannot form syllables on their own.

In stress-timed languages, vowels in unstressed syllables are usually reduced and shortened (Abercrombie 1967; Bolinger 1986; Nord 1986; Gimson 1989; van Bergem 1993; Moon & Lindblom 1994; Kreidler 2004). Articulatorily and acoustically, the process of vowel reduction in stress-timed languages leads to centralization of related vowels (Lindblom 1963; Janson 1979). The most common reduced vowel is the schwa /ə/ (Chomsky & Halle 1968; Kreifler 2004; Ladefoged & Johnson 2015). Stressed vowels are produced with greater intensity and duration than reduced vowels, and thus are more perceptually salient to listeners (Flemming 2009; Harrington 2010). In contrast, syllable-timed languages usually do not have reduced vowels in unstressed positions and tend to give syllables approximately equal prominence (Firth 1948; Dauer 1983; Auer 1993; Dankovičová & Dellwo 2007). In other words, syllables in syllable-timed languages are more similar to each other. Therefore, the vowel reduction not only makes stressed syllables more salient, but also makes the durational differences between stressed and unstressed syllables more prominent in stress-timed languages than in mora-timed and syllable-timed languages (Pike 1946; Abercrombie 1967; Ladefoged 1975; Nord 1986; van Bergem 1993).

Another feature of stress-timed languages is that they usually have more complex onsets and codas than syllable-timed and mora-timed languages. This is because syllable-timed and mora-timed languages more commonly have open syllables. To exemplify, “[t]he largest English syllable is CCCVCCCC,” with the most common syllable being CVC (McLeod 2010, 55). The most common syllable structure in Japanese and Mandarin is CV (Vance 1987; Otake 1990; Avery & Ehrlich 1992; Riney & Anderson-Hsieh 1993; Kubozono 1999; Duanmu 2000, 2016). In addition, successive morae are considered near-equal in duration (Bloch 1950; Han 1962; Ladefoged 1975; Shirai & Abe 2017). In syllable-timed languages, all syllables are thought to be isochronic (Abercrombie 1965, 1967; Pike 1946; Ladefoged 1975). Thus, mora-timed languages



are more similar to syllable-timed languages than to stress-timed languages (Grabe & Low 2002; Kubozono 2015).

The proportion of CV syllables in Japanese is even higher than that of French or Spanish, two syllable-timed languages (Otake 1990). Since French and Spanish have higher proportions of CV syllables than English, the ranking in terms of CV syllables is Japanese > French and Spanish > English, where > means higher than. The ranking is in conformity with the tendency of syllable structure to become simpler in stress-timed languages, to syllable-timed languages, and then to mora-timed languages (Nespor et al. 2011). Correspondingly, vowels take less time in stress-timed languages than in syllable-timed languages. In a similar vein, vowels occupy less time in syllable-timed languages than in mora-timed languages.

The rhythm class hypothesis has been investigated extensively in many different languages since its proposal. The studies of Ramus et al. (1999) and Grabe & Low (2002) are perhaps the most discussed. These two studies adopt different approaches to rhythmic metrics. We discuss their respective metric measures in the following subsections §2.1 and §2.2.

2.1. Vocalic and intervocalic intervals: rhythmic metric measures in Ramus et al. (1999)

Ramus et al. (1999) segment speech into vowels and consonants, and calculate vocalic and intervocalic intervals. They mainly focus on the three measures, %V, ΔV , and ΔC . The measure %V is the proportion of vocalic intervals within a sentence; ΔV refers to the standard deviation of the duration of vocalic intervals within each sentence; and ΔC is the standard deviation of the duration of intervocalic intervals within each sentence. With reference to eight languages, Ramus et al. (1999, 287) report that %V and ΔC “are congruent with the notion of rhythm classes.” For example, according to the authors, English has lower %V than French, because English has reduced vowels and French does not. In addition, English has higher ΔC , because English has more complex onset and coda structures than French. The differences between English and French in terms of %V and ΔC are in line with the supposition that English is a typical stress-timed language and French is a representative syllable-timed language.

One controversial issue not referred to in Ramus et al. (1999) is the speech rate factor. Barry et al. (2003) state that both ΔV and ΔC are inversely related to speech rate. Dellwo (2006) thus uses a normalized metric *Varco* ΔC , which is the standard deviation of intervocalic interval duration divided by the mean consonant duration. Dellwo (2006) claims that *Varco* ΔC discriminates better than ΔC between English and French. However, White & Mattys (2007, 520) argue that *Varco* ΔV appears to be “more reliable and discriminative than raw metrics,” while *Varco* ΔC “appear[s] to eliminate linguistically-interesting variation.” Since there are controversies over the speech rate factor, we will take a comprehensive look and take %V, ΔV , ΔC , *Varco* ΔV , and *Varco* ΔC all into consideration.

2.2. The pairwise variability index: rhythmic metric measures in Grabe & Low (2002)

Grabe & Low (2002) have adopted the *Pairwise Variability Index* (PVI) to speech rhythm class study. They (2002, 519) state that PVI measures “the durations of vowels and the duration of intervals between vowels (excluding pauses) in a passage of speech” and then calculates “the level of variability in successive measures.” Grabe & Low (2002) also claim that speech rate should be taken into consideration for the PVI calculation of vocalic intervals, since speech rate



may affect their duration significantly. They term this *normalized PVI for vocalic intervals*, or *vocalic nPVI* for short. They also argue that normalization is not necessary for intervocalic intervals. Accordingly, they use the raw PVI for intervocalic intervals, or *intervocalic rPVI* for short. The following are the equations of rPVI and nPVI (Grabe & Low 2002, 519–520).

(1) Equations of rPVI and nPVI (Grabe & Low 2002, 519–520)

$$\begin{aligned} \text{a:} \quad & \text{rPVI} \\ & \text{rPVI} = \left[\sum_{k=1}^{m-1} |d_k - d_{k+1}| / (m-1) \right] \\ \text{b:} \quad & \text{nPVI} \\ & \text{nPVI} = 100 \times \left[\sum_{k=1}^{m-1} \left| \frac{d_k - d_{k+1}}{(d_k + d_{k+1})/2} \right| / (m-1) \right]. \end{aligned}$$

In (1), m stands for “number of intervals . . . and d is the duration of the k th interval” (Grabe & Low 2002, 520). Due to vowel reduction in unstressed syllables, stress-timed languages should show more variability between two successive vocalic intervals than mora-timed and syllable-timed languages. In terms of intervocalic intervals, stress-timed languages are said to have more complex onset and coda structures, so they should show higher intervocalic interval variability. In addition, Grabe & Low (2002) claim that normalization should only be applied to vocalic intervals. Plainly, nPVI values of vocalic intervals and rPVI of intervocalic intervals are expected to be higher in stress-timed languages than in mora-timed and syllable-timed languages. The results reported in Grabe & Low (2002) are as expected for Dutch, English, and German, typical stress-timed languages, and as expected for French and Spanish, typical syllable-timed languages. Their results for Japanese, a mora-timed language, are similar to those for syllable-timed languages. However, White & Mattys (2007, 501) claim that, compared to rPVI and nPVI, Varco Δ V offers “the most discriminative analysis” after an examination of %V, Δ V, Varco Δ V, Δ C, Varco Δ C, intervocalic rPVI, vocalic rPVI, intervocalic nPVI, and vocalic nPVI.

2.3. Other approaches to rhythm classification

Ramus et al. (1999) in §2.1 and Grabe & Low (2002) in §2.2 both focus on speech production by native speakers to examine rhythm classes. Another way to approach rhythm classes is to examine the influence of a speaker’s first language (L1) on his or her second language (L2). If rhythm classification is tenable, influences of the rhythm of L1 on the rhythm of L2 can be expected. For example, Lin & Wang (2005) compare the English speech production of speakers of L1 English with that of Mandarin speakers of L2 English. The difference between %V values of L1 English and L2 English by Mandarin speakers is statistically significant, but the difference between Δ C values is not. Their (2005) finding suggests that vowels are better indicators of rhythm classification than consonants. Lin & Wang (2005) did not measure PVI values of vocalic intervals and intervocalic intervals, nor did they (2005) take mora-timed languages or speech rate into consideration.

What may also be enlightening is to examine speech production by bilingual speakers to look for potential influences of the rhythm of one language on the other. For example, Carter (2005) tests bilingual speakers of English and Spanish and finds that bilingual speakers show intermediate vocalic rPVI scores between the low vocalic rPVI scores of Spanish ESL speakers and the high vocalic rPVI scores of native English speakers. However, Carter (2005) does not take



speech rate into consideration, so he does not report scores of vocalic nPVI; nor does he report scores concerning intervocalic intervals.

2.4. Our study

It is still unclear which metric measures are the most discriminative and whether the speech rate factor should be considered. In addition, the research into mora-timed languages is not robust: previous studies are mainly concerned with stress-timed and syllable-timed languages. In this study, we examine all the three rhythm classes comprehensively. We adopt both the approach of Ramus et al. (1999) and that of Grabe & Low (2002). For each rhythmic measure, we calculate its values both with and without the speech rate factor. In other words, we measure %V, ΔV , Varco ΔV , vocalic rPVI, vocalic nPVI, ΔC , Varco ΔC , intervocalic rPVI, and intervocalic nPVI.

We take a different perspective to previous studies: although previous studies have mainly focused on native speakers or second language speakers, we analyze speech production by bilingual speakers, comparing their production with that of monolingual English speakers and ESL speakers. Since English, Japanese, and Mandarin are representatives of stress-timed, mora-timed, and syllable-timed languages, respectively, we take these three languages as exemplars of the respective rhythm classes.

Monolingual English speakers in this paper are defined as native speakers of English who only command English: they cannot produce speech in another language; nor can they comprehend another language (see e.g., Snow & Hakuta 1992; Mack 1997; Ellis 2007). Bilingual speakers in this paper are defined as those who have acquired two languages in their infancy and can produce fluent and effective speech in both languages (see e.g., Haugen 1953; Weinreich 1953). Since this paper focuses on English, Japanese, and Mandarin, the paper gathers English-Japanese bilinguals and English-Mandarin bilinguals. Our definition of a bilingual speaker is not as strict as that of Bloomfield (1933) as a perfect user of two languages; however, it is much stricter than that of MacNamara (1967) who includes anyone who has a minimal competence in listening, reading, speaking, or writing a language other than his/her native language. ESL speakers in this paper are those who have not learned English in their early childhood, who received English education at school, and have not lived in any English-speaking country for longer than one month (see e.g., Jenkins 2000; Mitchell & Myles 2004; Kormos 2006). These speakers have an intermediate proficiency in English.

2.5. Our hypotheses

Each ordinal number in Table 1, 1st, 2nd, etc., shows the ranking of a group in terms of a particular metric measure. To exemplify, 5th in the upper left corner means that the monolingual group has the lowest %V value. The term *EM bilingual* in Table 1 stands for the English-Mandarin bilingual group; *EJ bilingual* refers to the English-Japanese bilingual group. *Mandarin ESL* stands for the group which is composed of native speakers of Mandarin who speak English as a second language. Similarly, *Japanese ESL* refers to the group composed of native speakers of Japanese who speak English as a second language.

Vowels take less time in stress-timed languages than in mora-timed and syllable-timed languages, thus English is expected to have the lowest value in terms of %V. Vowel reduction in stress-timed languages makes durational differences between stressed and unstressed vowels



Table 1. Our hypotheses

	Hypotheses				
	%V	ΔV or Varco ΔV	ΔC or Varco ΔC	Vocalic rPVI or Vocalic nPVI	Intervocalic rPVI or Intervocalic nPVI
Monolingual	5th	1st	1st	1st	1st
EM bilingual	4th	2nd	2nd	2nd	2nd
EJ bilingual	3rd	3rd	3rd	3rd	3rd
Mandarin ESL	2nd	4th	4th	4th	4th
Japanese ESL	1st	5th	5th	5th	5th

greater, English therefore should have the highest ΔV or Varco ΔV value, and the highest vocalic rPVI or vocalic nPVI value. English is claimed to have a more complex onset and coda structure than Japanese and Mandarin. Thus the monolingual group should have the highest ΔC or Varco ΔC value, and the highest intervocalic rPVI or intervocalic nPVI value. We hypothesize that the results of rhythmic measures for the EM bilingual group and the EJ bilingual group should be intermediate between those of the monolingual group and those of the two ESL groups. We also hypothesize that the results of the two bilingual groups should be closer to those of the monolingual group than to the two ESL groups. In terms of the two bilingual groups, we hypothesize that the results of the EM bilingual group are closer to the monolingual group than the EJ bilingual group, since Japanese has an even higher proportion of CV syllables than syllable-timed languages.

We will develop the idea noted in the previous paragraph step by step. For ease of exposition, we first focus on the EJ bilingual group. Secondly, we apply the conclusion drawn from the EJ bilingual group to the EM bilingual group, and examine whether the conclusion still holds.

3. THE FIRST EXPERIMENT

As noted in §2.5, we first focus on EJ bilingual speakers and compare their English speech production with that of monolingual speakers of English and of Japanese ESL speakers.

3.1. Subjects for the first experiment

Three monolingual native speakers of English, one male and two female, were recorded. The three monolingual native speakers of English (hereafter the monolingual group) were born and brought up in California. They were also residents of California at the time of recording. Three EJ bilingual speakers (hereafter the EJ bilingual group) were recorded. They were all born in the western part of Japan, moved to California in their infancy, and spent their formative years in California. They were also residents of Japan at the time of recording. We tried to enroll EJ bilingual speakers among Californian residents, but it was difficult to gather three EJ bilingual speakers in California. We limited speakers in the monolingual group and the EJ bilingual group to Californian English speakers to reduce potential influences from different English accents as



much as possible. Three Japanese ESL speakers (hereafter the Japanese ESL group) were also recorded. They were all born and brought up in the western part of Japan. Their accents are not notably different from standard Japanese. All speakers in the Japanese ESL group are from the same area of Japan. The aim is also to reduce possible influences of different Japanese accents.

All monolingual speakers and EJ bilingual speakers were around 30–35 years old at the time of recording and college graduates. All Japanese ESL speakers were college students and were just over 20 years old at the time of recording. Although the Japanese ESL speakers are not of exactly the same age as the monolingual speakers and the EJ bilingual speakers, the age difference is not considerably large.

3.2. Recording

All speakers were given the text from the PAC project, *Christmas Interview of a Television Evangelist*, three weeks before their recordings to become familiar with it.¹ They were instructed to practice the passage in their normal voice and at a rate that they felt natural and comfortable until they could read the passage fluently. They were also instructed that they should pause between sentences and repeat the whole sentence if they made a mistake. The ideal place to record appears to be a sound-proof room at our university. However, the physical distance prevented us from doing so. In addition, recent studies have shown that smartphone recording is acceptable for acoustic analysis (see e.g. [Maryn et al. 2017](#); [Oliveira et al. 2017](#); [Wu 2017](#); [van der Woerd et al. 2020](#)). Thus, recordings were made on the second author's iPhone 6 in a quiet room. The original format of the recordings was m4a. They were later converted to the wav format for acoustic analysis on Praat.

3.3. Segmentation and analysis

The first author segmented and labelled the recorded speeches. This procedure was carried out on speech waveforms and wideband spectrograms generated on Praat. The guidelines laid out in [Peterson & Lehiste \(1960\)](#), [Grabe & Low \(2002\)](#), and [White & Mattys \(2007\)](#) were generally followed. Pauses between intonation phrases were excluded from the analysis. The segmental boundaries were identified generally by taking spectral transitions into consideration. For example, the segmental sequence /do/ was divided into two segments, as the vocalic segment had a clearer formant structure in the spectrogram compared with the voiced obstruent. To take another example, the segmental boundary of /li/ was determined based on the observation that the amplitude of /l/ was lowered due to the approximant articulation, with the result that the spectrogram became grayer. A vocalic interval was the stretch of a speech signal between the vowel onset and the vowel offset. As a result, a vocalic interval might stretch over more than one syllable and even across word boundaries. An intervocalic interval was the stretch of the speech signal between a vowel offset and the onset of the next vowel. After the segmentation, the durations of each vocalic interval and each intervocalic interval were calculated.

¹PAC stands for *La Phonologie de l'Anglais Contemporain: Usages, Variétés et Structure* in French or *The Phonology of Contemporary English: Usage, Varieties and Structure* in English.



4. ENGLISH-JAPANESE BILINGUAL GROUP RESULT ANALYSES

In this section, we discuss our results for vowels and consonants respectively for easy understanding. We use group means in the following part of the paper. Firstly, let us turn to the rhythmic measures for vowels, including %V, ΔV , Varco ΔV , vocalic rPVI, and vocalic nPVI.

4.1. Results for vocalic intervals

We give the results for vowels in Fig. 1. The values of ΔV , Varco ΔV , vocalic rPVI, and vocalic nPVI are multiplied by 100.

If there is little or no influence of Japanese rhythm on the English speech production by the speakers in the EJ bilingual group, the results in terms of all five metric measures from the EJ bilingual group are expected to be the same as or close to those of the monolingual group. The only metric measure that shows a tendency towards this direction is %V: 39.11% for the EJ bilingual group is close to 39.92% for the monolingual group. However, the remaining four metric measures show large differences between the two groups. Therefore, it seems that the English speech production by the EJ bilingual group does show influences of Japanese rhythm. The next question is which metric measures discriminate the monolingual group and the EJ bilingual group satisfactorily. An interesting pattern can be noticed with a comparison of results in terms of ΔV and vocalic rPVI and results in terms of Varco ΔV and vocalic nPVI. Results in terms of ΔV and vocalic rPVI show that the monolingual group has the highest values and the EJ bilingual group has the lowest, with the Japanese ESL group intermediate between the two. On the other hand, results in terms of Varco ΔV and vocalic nPVI show that the monolingual group has the highest values, the EJ bilingual group has intermediate values, and the Japanese ESL group has the lowest values. The difference between the metric measures ΔV and vocalic rPVI, and the metric measures Varco ΔV and vocalic nPVI is that ΔV and vocalic rPVI do not take speech rate in consideration, while Varco ΔV and vocalic nPVI take speech rate into account. The results in terms of ΔV and vocalic rPVI appear to suggest that the Japanese ESL speakers realize vocalic interval variation in a more similar way to the monolingual English speakers than the EJ bilingual speakers. This does not concur with the subjective perception of

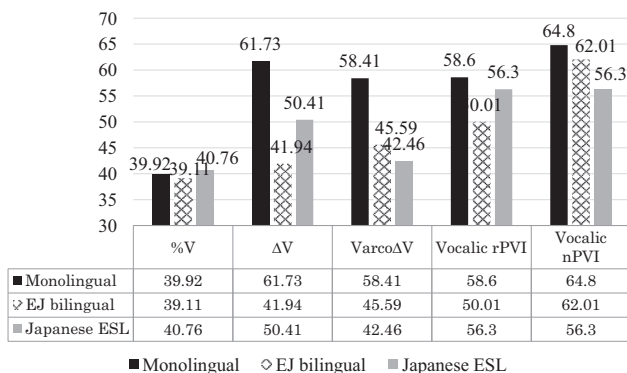


Fig. 1. Mean results of the speakers in the EJ bilingual group in terms of vocalic intervals



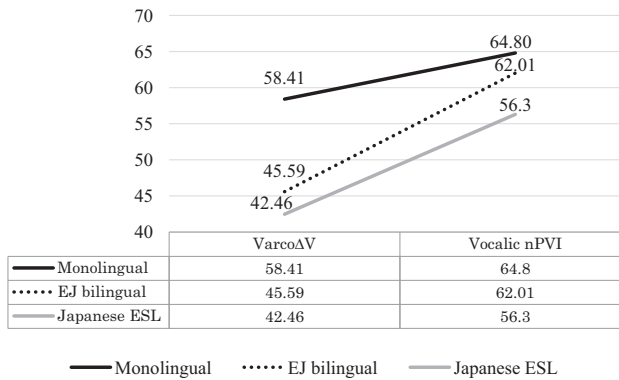


Fig. 2. Varco Δ V and vocalic nPVI (EJ bilingual group)

differences between EJ bilingual speakers' abilities in English and Japanese ESL speakers' abilities in English, especially in consideration of the fact that the EJ bilingual speakers in this study moved to California in their infancy and spent their formative years in California, while the Japanese ESL speakers in this study have not lived in an English-speaking country for more than one month. The metric measures Varco Δ V and vocalic nPVI seem to capture the hypothesized differences between groups better than Δ V and vocalic rPVI. Combining the results of Varco Δ V and vocalic nPVI from the three groups, it is clear that metric measures for vowels need to take speech rate into consideration. The results of %V are not completely in conformity with our hypothesis: the EJ bilingual group has the lowest value. However, the difference between the monolingual group and the EJ bilingual group in terms of %V appears to be marginal. Fig. 2 can present the discrimination more clearly.

In terms of Varco Δ V and vocalic nPVI, the values of the EJ bilingual group are intermediate between those of the monolingual group and those of the Japanese ESL group. There is no crossing between any two lines. One point that must be emphasized is that speech rate needs to be taken into consideration for vowels. This appears to be due to the fact that speech rate has a major influence on vowel reduction and thus on results of metric measures for vowels. It is claimed by Ramus et al. (1999) and Ramus (2002, 2003) that a combination of Δ V and either Δ C or %V is useful in discriminating languages of different rhythm classes. However, our results appear to argue against this claim.

4.2. Results for intervocalic intervals

As shown in Fig. 3, in terms of all four metric measures for intervocalic intervals, the results for the monolingual group and those for the Japanese ESL group are relatively close to each other. The most unexpected result is that the EJ bilingual group has the lowest values in terms of all four metric measures. This does not concur with the subjective perception of differences between EJ bilingual speakers' abilities in English and Japanese ESL speakers' abilities in English. The results of metric measures for intervocalic intervals in Fig. 3 cannot clearly discriminate between groups.



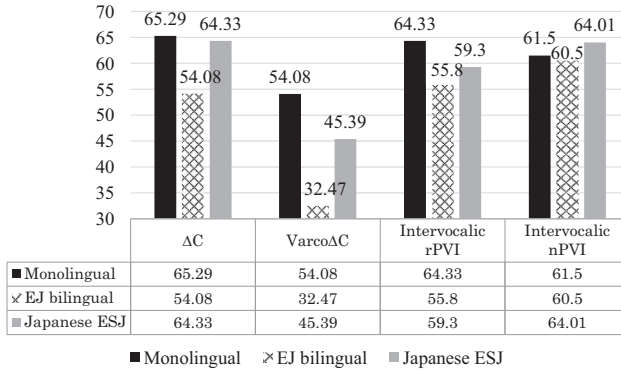


Fig. 3. Mean results of the speakers in the EJ bilingual group in terms of intervocalic intervals

4.3. Section summary

Our result shows that metric measures for vowels, both in terms of syllable durations and the pairwise variability index, taking speech rate into consideration, are more reliable and more effective at discriminating different rhythm classes. What is interesting is that metric measures for consonants, both in terms of syllable durations and pairwise variability indexes, do not appear effective in discriminating different rhythm classes. This conclusion still holds even after the speech rate factor is taken into consideration. The most effective way to examine the validity of the conclusion here is perhaps to test it against a syllable-timed language. Thus we turn to Mandarin in the next section.

5. RESULTS FROM THE ENGLISH-MANDARIN BILINGUAL GROUP

The experiment for the English-Mandarin group was carried out in 2016. The preliminary results were presented and published in 2017 and 2019, respectively (Liu & Takeda 2017, 2019). In the following, we give a brief overview of those results.

5.1. Subjects, recording, segmentation, and analysis

Three bilingual speakers of English and Mandarin (henceforth the EM bilingual group), one male and two female, were recorded. They were born in China and moved to California as infants. They were also residents of California at the time of recording.

Three native speakers of Mandarin who speak English as a second language (henceforth the Mandarin ESL group) were recorded. They were all born and brought up in the central part of China. Their accents are not markedly different from standard Mandarin.

All EM bilingual speakers were around 30–35 years old at the time of recording and college graduates. All Mandarin ESL speakers were college students and were just over 20 years old at the time of recording. Procedures of recording, segmentation, and analysis are the same as those for the first experiment in §3.2 and §3.3.



5.2. Results in Liu & Takeda (2017, 2019)

English is claimed to have both full and reduced vowels and is expected to have more durational variabilities than Mandarin. Thus the Mandarin ESL group is expected to have the highest %V value, the lowest ΔV or Varco ΔV value, and the lowest vocalic rPVI or vocalic nPVI value. If influences of Mandarin rhythm exist, the EM bilingual group is expected to show intermediate results in terms of all the metric measures just noted. We give the results for vowels in Fig. 4. The values of ΔV , Varco ΔV , vocalic rPVI, and vocalic nPVI are multiplied by 100.

The results of %V are not completely in conformity with our hypothesis: the EM bilingual group has the lowest value. However, the difference between the monolingual group and the EM bilingual group in terms of %V appears to be marginal. The EM bilingual group has intermediate results in terms of Varco ΔV and vocalic nPVI: the influence of Mandarin rhythm seems to be in the English speech production by the EM speakers. Results in Fig. 4 echo the conclusion of §4 that speech rate needs to be taken into account for measures of vocalic intervals. We show the results concerning intervocalic intervals in Fig. 5.

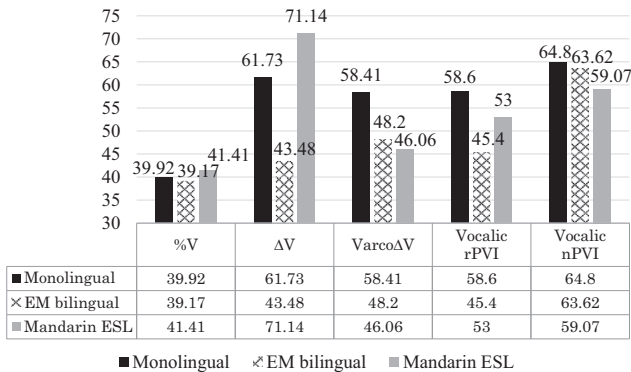


Fig. 4. Mean results of the speakers in the EM bilingual group in terms of vocalic intervals

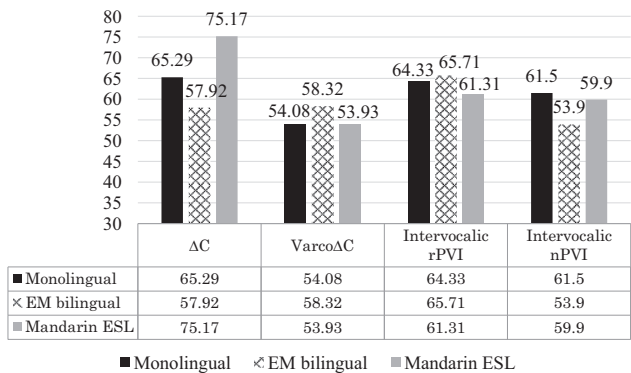


Fig. 5. Mean results of the speakers in the EM bilingual group in terms of intervocalic intervals



All values in Fig. 5 have been multiplied by 100. In terms of all the four metric measures, the EM bilingual group has the most extreme values, which appear to echo the results in Fig. 3 that metric measures for intervocalic intervals cannot clearly discriminate between different groups.

Similar to the results in §4, we find that a combination of Varco Δ V and vocalic nPVI can discriminate the three groups satisfactorily. Thus far, we have separated EJ bilingual speakers from EM bilingual speakers. In the next section, we compare the two bilingual groups, and examine whether the conclusions we have drawn from §4 and this section still hold. We also discuss the statistical analysis results in §6.

6. A COMPARISON OF RESULTS FROM ALL GROUPS

We list our experiment results in Table 2 along with our hypotheses in Table 1 to give a clear comparison between them. In consideration of the unreliability of rhythmic measures for consonants, we omit hypotheses concerning consonants in this section. In addition, we have shown that the factor of speech rate needs to be considered for measures of vowels, thus we omit the measures of Δ V and vocalic rPVI in Table 2.

Similar to Table 1, each ordinal number in Table 2, 1st, 2nd, etc., shows the ranking of a group in terms of a particular metric measure. It would be too complex to compare the results from all five groups simultaneously. For ease of understanding, we first make a comparison between the two bilingual groups; second, a comparison between the two ESL groups; third, a comparison between all groups; and finally a review of the statistical analysis results for both vowels and consonants.

6.1. The EM bilingual group and the EJ bilingual group

As presented in Table 2, the EJ bilingual group is expected to have the lowest Varco Δ V and vocalic nPVI values among the monolingual group, the EM bilingual group, and the EJ bilingual group.

Fig. 6 shows that the EM bilingual group has a higher Varco Δ V value and a higher vocalic nPVI value than the EJ bilingual group. This indicates that the English speech production of the EM and EJ bilingual groups shows influences of the rhythm of Mandarin and of Japanese, respectively.

Table 2. Our hypotheses and our results from all groups

	Hypotheses		Results			
	Varco Δ V	Vocalic nPVI		Varco Δ V		Vocalic nPVI
Monolingual	1st	1st	1st	58.41	1st	64.80
EM bilingual	2nd	2nd	2nd	48.20	2nd	63.62
EJ bilingual	3rd	3rd	4th	45.59	3rd	62.01
Mandarin ESL	4th	4th	3rd	46.06	4th	59.07
Japanese ESL	5th	5th	5th	42.46	5th	56.30



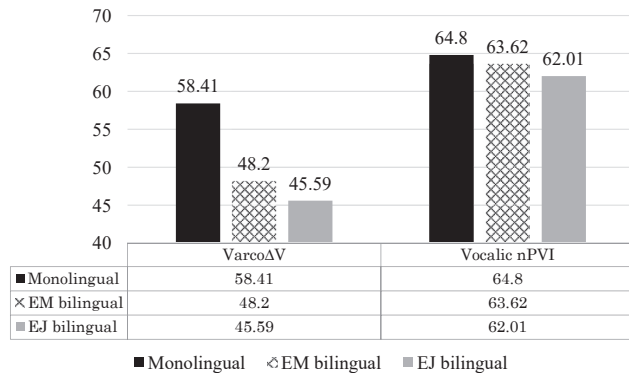


Fig. 6. VarcoΔV and vocalic nPVI (monolingual, EM bilingual, and EJ bilingual)

At the same time, the EM bilingual group and the EJ bilingual group do not show a large difference either in terms of VarcoΔV or vocalic nPVI, which is not unexpected. The one-way ANOVA test performed on GraphPad Prism version 8.0.0 for Windows (GraphPad Software, San Diego, CA; hereafter the GraphPad software) also shows that the difference in vowels between the EM bilingual group and the EJ bilingual group is not statistically significant ($P = 0.99$).

6.2. The Mandarin ESL group and the Japanese ESL group

The Mandarin ESL group is expected to have higher VarcoΔV and vocalic nPVI values than the Japanese ESL group due to the influence of their respective native languages. Our results are in line with our hypotheses, as shown in Table 2.

The two ESL groups are expected to have lower VarcoΔV and vocalic nPVI values than the monolingual group, the EM bilingual group, and the EJ bilingual group. In terms of vocalic nPVI, our result is in complete conformity with our hypothesis. There is one problem regarding VarcoΔV: the Mandarin ESL group has a slightly higher VarcoΔV value than the EJ bilingual group, although the difference is quite small. This also seems due to the influence of Japanese: Japanese is expected to have the lowest VarcoΔV value. The one-way ANOVA test performed on the GraphPad software shows that the difference in vowels between the Mandarin ESL group and the Japanese ESL group is statistically significant ($P < 0.01$).

6.3. A comparison of all groups

We use Fig. 7 to graphically illustrate results in Table 2. Fig. 7 shows that vocalic nPVI is notably effective in discriminating different groups. The dashed horizontal line can separate the two bilingual groups from the monolingual group. The black horizontal line can separate the two bilingual groups from the two ESL groups. The monolingual group and the two bilingual groups are above the two ESL groups due to their higher vocalic nPVI values. To be specific, the two bilingual groups occupy an intermediate position between the monolingual group and the two ESL groups in terms of vocalic nPVI. This suggests that the speakers in the two bilingual groups made unstressed vowels shorter than the speakers in the two ESL groups, but did not make the



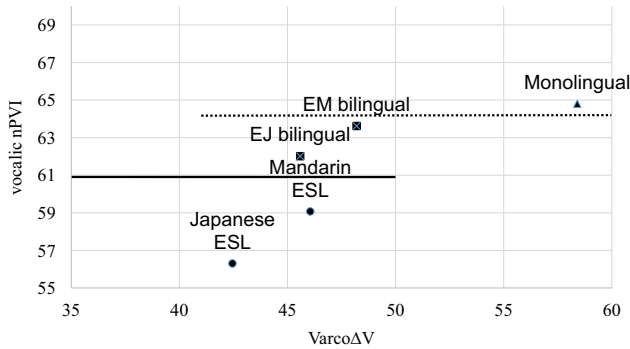


Fig. 7. VarcoΔV and vocalic nPVI (all groups)

contrast between stressed and unstressed syllables as great as the speakers in the monolingual group.

6.4. Statistical analysis results

We give all statistical analysis results for vocalic intervals in Table 3 and all results for inter-vocalic intervals in Table 4.

In Table 3, gray cells are where rows and columns intersect to have null or repeated results. For example, the cell at which the row *EM bilingual* and the column *EM bilingual* intersect is gray since this result is null. The cell at which the row *EJ bilingual* and the column *EM bilingual* intersect is gray because the same result has been shown in the cell at which the row *EM bilingual* and the column *EJ bilingual* intersect. The shaded cell shows the only unexpected result, details of which will be given in the next paragraph.

Table 3. Statistical analysis results for vocalic intervals (all groups)

	Monolingual	EM bilingual	EJ bilingual	MandarinESL
EM bilingual	$P = 0.08$			
EJ bilingual	$P = 0.34$	$P = 0.99$		
Mandarin ESL	$P < 0.01$	$P < 0.01$	$P < 0.01$	
Japanese ESL	$P = 0.39$	$P = 0.06$	$P = 0.02$	$P < 0.01$

Table 4. Statistical analysis results for intervocalic intervals (all groups)

	Monolingual	EM bilingual	EJ bilingual	MandarinESL
EM bilingual	$P = 0.10$			
EJ bilingual	$P = 0.23$	$P = 0.90$		
Mandarin ESL	$P = 0.16$	$P < 0.01$	$P = 0.00$	
Japanese ESL	$P = 0.00$	$P = 0.00$	$P = 0.00$	$P = 0.08$



The one-way ANOVA test performed on the GraphPad software shows that the differences in vocalic intervals generally follow our hypotheses. To exemplify, the differences in vocalic intervals between the EJ bilingual group and the Mandarin ESL group are expected to be large, which is supported by the statistical analysis result ($P < 0.01$). The only unexpected result is that the differences in vocalic intervals between the monolingual group and the Japanese ESL group are not statistically significant ($P = 0.39$), which is shown in the shaded cell. We are not yet sure how to interpret this. One possible explanation is that speakers were asked to read at a speed that they were comfortable with. Japanese ESL speakers read at a relatively slow speed. Our subjective judgement is that slow and comfortable speech speed may have helped Japanese ESL speakers arrive at this result. Another possible reason seems to be that the statistical analysis was carried out on raw data of vowels, without taking speech rate differences between groups into consideration. Recall that the two metric measures excluding speech rate, ΔV and vocalic rPVI, in Fig. 1 also show that the Japanese ESL group are closer to the monolingual group than the EJ bilingual group.

The remaining question is why metric measures of consonants are not reliable in discriminating rhythm classes. We turn to statistical analysis results for consonants to look for possible hints.

Similar to Table 3, the gray cells in Table 4 have null or repeated results. There is only one result that is not in line with our hypotheses in Table 4: the result of $P = 0.16$ in the shaded cell at which the row *Monolingual* and the column *Mandarin ESL* intersect. Nevertheless, Table 4 shows us that statistical analysis results for consonants are generally as expected. In §4 and §5, we draw the conclusion that metric measures of consonants are not reliable in discriminating different rhythm classes. A possible explanation includes that consonants may not be strongly correlated with rhythm classes. To exemplify, one main characteristic of stress-timed languages is the reduction of unstressed vowels. Simply put, one main characteristic of stress-timed languages is correlated with vowels, and this seems to partly explain the lack of strong correlation between consonants and rhythm classes. This also partly explains why the statistical analysis results for consonants appear plausible, while the metric measure results do not. We will leave further explanation of this issue to future research.

7. CONCLUSION

We have chosen English, Japanese, and Mandarin as our focus in this paper. Our result shows that metric measures for vowels, whether in terms of syllable durations or the pairwise variability index, offer more reliable discrimination between different rhythm classes. Another prominent feature is that metric measures for vowels need to take speech rate into consideration. The reason seems to be that vowel reduction has a strong correlation with speech rate.

What is interesting is that measures for consonants, both in terms of syllable durations and pairwise variability indexes, do not appear effective in discriminating different rhythm classes. This conclusion still holds even after speech rate is taken into consideration. However, the statistical analysis results for consonants are generally as expected. Our research cannot satisfactorily explain why consonants are not strongly correlated with rhythm classes. One possible explanation is that the reduction of unstressed vowels, one main characteristic of stress-timed



languages, is correlated with vowels, not consonants. Further pursuit of this question will be for future study.

The conclusion for this paper was drawn from a relatively small number of participants, so some caution is necessary in interpreting it. In addition, this paper focused exclusively on California English. Whether a similar conclusion can be drawn if the focus was on other varieties of English is also a question that needs further research.

ACKNOWLEDGEMENTS

For help in getting this article to its final form, we are grateful to Professor Jacques Durand and Professor Daiki Hashimoto for their advice on acoustic analysis, to Professor Eiji Yamada and Professor Hajime Takeyasu for advice and discussion, to Professor David Farnell and Professor Stephen Howe for editing our paper. All remaining errors are our responsibility. This work was funded by JSPS Grant-in-Aid for Early-Career Scientists (KAKENHI-PROJECT-20K13072).

REFERENCES

- Abercrombie, David. 1965. *Studies in phonetics and linguistics*. London: Oxford University Press.
- Abercrombie, David. 1967. *Elements of general phonetics*. Edinburgh: Edinburgh University Press.
- Archibald, John. 1998. *Second language phonology*. Amsterdam: John Benjamins.
- Auer, Peter. 1993. Is a rhythm-based typology possible? A study of the role of prosody in phonological typology. *KonTRI Working Paper 21*. Konstanz: Universität Konstanz.
- Avery, Peter and Susan Ehrlich. 1992. *Teaching American English pronunciation*. Oxford: Oxford University Press.
- Barry, William J., Bistra Andreeva, Michela Russo, Snezhina Dimitrova and Tanja Kostadinova. 2003. Do rhythm measures tell us anything about language type? *Proceedings of the 15th International Congress of Phonetics Science*. 2693–2696.
- Bloch, Bernard. 1950. *Studies in colloquial Japanese IV: Phonemics*. *Language* 26. 86–125.
- Bloomfield, Leonard. 1933. *Language*. New York, NY: Holt, Rinehart & Winston.
- Bolinger, Dwight. 1986. *Intonation and its parts: Melody in spoken English*. Stanford, CA: Stanford University Press.
- Carter, Phillip M. 2005. Quantifying rhythmic differences between Spanish, English, and Hispanic English. *Amsterdam Studies in the Theory and History of Linguistic Science Series 4(272)*. 63–75.
- Chomsky, Noam and Morris Halle. 1968. *The sound pattern of English*. New York, NY: Harper&Row.
- Dankovičová, Jana and Volker Dellwo. 2007. Czech speech rhythm and the rhythm class hypothesis. In J. Trouvain and W. Barry (eds.) *Proceedings of the 16th International Congress of Phonetic Sciences*. 1241–1244.
- Dauer, Rebecca M. 1983. Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics* 11. 51–62.
- Dellwo, Volker. 2006. Rhythm and speech rate: A variation coefficient for delta C. In P. Karnowski and I. Szigeti (eds.) *Language and language processing: Proceedings of the 38th Linguistic Colloquium*. Frankfurt: Peter Lang. 231–242.
- Duanmu, San. 2000. *The phonology of standard Chinese*. New York, NY: Oxford University Press.



- Duanmu, San. 2016. Syllable structure. In R. Sybesma (ed.) *Encyclopedia of Chinese language and linguistics*, Volume 4. Leiden: Brill. 230–236.
- Ellis, Elizabeth. 2007. Monolingualism: The unmarked case. *Estudios de Sociolingüística* 7(2). 173–196.
- Firth, John R. 1948. Sounds and prosodies. *Transactions of the Philological Society* 47(1). 127–152.
- Flemming, Edward. 2009. The phonetics of schwa vowels. In D. Minkova (ed.) *Phonological weakness in English*. London: Palgrave Macmillan. 78–98.
- Gimson, Alfred Charles. 1989. *An introduction to the pronunciation of English*, 4th edn. London: Edward Arnold.
- Grabe, Esther and Ee Ling Low. 2002. Durational variability in speech and the rhythm class hypothesis. In C. Gussenhoven and N. Warner (eds.) *Laboratory phonology 7*. Berlin: Mouton de Gruyter. 515–546.
- GraphPad Prism. 2019. Version 8.0.0 for Windows. San Diego, CA: GraphPad Software. Computer software.
- Gut, Ulrike. 2009. *Non-native speech: A corpus-based analysis of phonological and phonetic properties of L2 English and German*. Frankfurt am Main: Peter Lang.
- Han, M.S. 1962. The feature of duration in Japanese. *Onsei no Kenkyuu* 10. 65–80.
- Harrington Jonathan. 2010. Acoustic phonetics. In W.J. Hardcastle, J. Laver and F.E. Gibbon (eds.) *The handbook of phonetic sciences*, 2nd edn. Chichester: Wiley Blackwell. 91–93.
- Haugen, Einar. 1953. *The Norwegian language in America: A study of bilingual behavior*. Philadelphia, PA: University of Pennsylvania.
- Janson, Tore. 1979. Vowel duration, vowel quality, and perceptual compensation. *Journal of Phonetics* 7. 93–103.
- Jenkins, Jennifer. 2000. *The phonology of English as an international language*. New York, NY: Oxford University Press.
- Kormos, Judit. 2006. *Speech production and second language acquisition*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Kreidler, Charles. 2004. *The pronunciation of English*. Oxford: Blackwell.
- Kubozono, Haruo. 1999. *Nihongo no Onsei [Japanese phonology]*. Tokyo: Iwanami.
- Kubozono, Haruo. 2015. Introduction to Japanese phonetics and phonology. In Haruo Kubozono (ed.) *Handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton. 1–40.
- Ladefoged, Peter. 1975. *A course in phonetics*. New York, NY: Harcourt Brace Jovanovich.
- Ladefoged, Peter and Keith Johnson. 2015. *A course in phonetics*, 7th edn. Stamford: Cengage Learning.
- Li, Aike and Brechtje Post. 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.
- Lin, Hua and Qian Wang. 2005. Vowel quantity and consonant variance: A comparison between Chinese and English. *Proceedings of Between Stress and Tone*. Leiden, June 2005.
- Lindblom, Björn. 1963. Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America* 35. 1773–1781.
- Liu, Sha and Kaye Takeda. 2017. Production of English by bilingual speakers: Any influence from different rhythm types. Presentation at PAC 2017 – Phonology and interphonology of contemporary English: From native corpora to learner corpora, Paris Nanterre University, Paris, France, September 28–30.
- Liu, Sha and Kaye Takeda. 2019. English speech production by bilingual speakers: Evidence for or against rhythm classification. *The Bulletin of Central Research Institute Fukuoka University, Series A: Humanities* 19(1). 35–44.



- Mack, Molly. 1997. The monolingual native speaker: Not a norm, but still a necessity. *Studies in the Linguistic Sciences* 27. 113–146.
- MacNamara, John. 1967. The linguistic independence of bilinguals. *Journal of Verbal Learning and Verbal Behavior* 6(5). 729–736.
- Maryn, Youri, Femke Ysenbaert, Andrzej Zarowski and Robby Vanspauwen. 2017. Mobile communication devices, ambient noise, and acoustic voice measures. *Journal of Voice* 31(2). 248.e11–248.e23.
- McCawley, James D. 1968. *The phonological component of a Grammar of Japanese*. The Hague: Mouton.
- McCawley, James D. 1978. What is a tone language. In V. Fromkin (ed.) *Tone: A linguistic survey*. New York, NY: Academic Press. 113–131.
- McLeod, Sharynne. 2010. Laying the foundations for multilingual acquisition: An international overview of speech acquisition. In M. Cruz-Ferreira (ed.) *Multilingual norms*. Berlin: Peter Lang. 53–71.
- Mitchell, Rosamond and Florence Myles. 2004. *Second language learning theories*, 2nd edn. London: Hodder Arnold.
- Moon, Seung Jae and Björn Lindblom. 1994. Interaction between duration, context, and speaking style in English stressed words. *Journal of the Acoustical Society of America* 96. 40–55.
- Munro, Murray J. and Tracey M. Derwing. 1995. Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech* 38. 289–306.
- Nespor, Marina, Mohinish Shukla and Jacques Mehler. 2011. Stress-timed vs. syllable-timed languages. In M. van Oostendorp, C. J. Ewen, E. Hume and K. Rice (eds.) *The Blackwell companion to phonology*. London: Blackwell Publishing. 1147–1157.
- Nord, Lennart. 1986. Acoustic studies of vowel reduction in Swedish. *Quarterly Progress and Status Report* 4. 19–36.
- Oliveira, Gisele, Gaetano Fava, Melody Baglione and Michael Pimpinella. 2017. Mobile digital recording: Adequacy of the iRig and iOS device for acoustic and perceptual analysis of normal voice. *Journal of Voice* 31(2). 236–242.
- Ordin, Mikhail and Leona Polyanskaya. 2015. Acquisition of speech rhythm in a second language by learners with rhythmically different native languages. *Journal of the Acoustical Society of America* 138. 533–545.
- Otake, Takashi. 1990. Rhythmic structure of Japanese and syllable structure. *IEICE Technical Report* 89. 55–61.
- Peterson, Gordon E. and Ilse Lehiste. 1960. Duration of syllable nuclei in English. *Journal of the Acoustical Society of America* 32(6). 693–703.
- Pike, Kenneth Lee. 1946. *The intonation of American English*, 2nd edn. Ann Arbor: University of Michigan Press.
- Ramus, Franck. 2002. Acoustic correlates of linguistic rhythm: Perspectives. *Proceedings of Speech Prosody*. 115–120.
- Ramus, Franck. 2003. The psychological reality of rhythm classes: Perceptual studies. *Proceedings of the 15th International Congress of Phonetic Sciences*. 337–342.
- Ramus, Franck, Marina Nespor and Jacques Mehler. 1999. Correlates of linguistic rhythm in the speech signal. *Cognition* 72. 1–28.
- Riney, Tim and Janet Anderson-Hsieh. 1993. Japanese pronunciation of English. *JALT Journal* 15(1). 21–36.
- Shirai, Katsuhiko and Masanobu Abe. 2017. *Recent progress in Japanese speech synthesis*. Forlag: Taylor & Francis.



- Snow, Catherine E. and Kenji Hakuta. 1992. The costs of monolingualism. In James Crawford (ed.) *Language loyalties: A source book on the official English controversy*. Chicago, IL: The University of Chicago Press. 384–394.
- Steele, Joshua. 1775. *An essay towards establishing the melody and measure of speech, to be expressed and perpetuated by peculiar symbols*. London: W. Bowyer and J. Nichols, for J. Almon.
- Van Bergem, Dick R. 1993. Acoustic vowel reduction as a function of sentence accent, word stress, and word class. *Speech Communication* 12. 1–23.
- Van der Woerd, Benjamin, Min Wu, Vijay Parsa, Philip C. Doyle and Kevin Fung. 2020. Evaluation of acoustic analyses of voice in nonoptimized conditions. *Journal of Speech, Language and Hearing* 63(12). 3991–3999.
- Vance, Timothy J. 1987. *An introduction to Japanese phonology*. Albany, NY: State University of New York Press.
- Weinreich, Uriel. 1953. *Languages in contact: Findings and problems*. The Hague: Mouton.
- Wenk, Brian J. 1985. Speech rhythms in second language acquisition. *Language and Speech* 28(2). 157–175.
- White, Laurence and Sven L. Mattys. 2007. Calibrating rhythm: First language and second language studies. *Journal of Phonetics* 35. 501–522.
- Wu, Qunli. 2017. An iPhone-based binaural recorder for sound quality analysis. *Sound & Vibration* 51(4). 16–17.

Open Access. This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium for non-commercial purposes, provided the original author and source are credited, a link to the CC License is provided, and changes – if any – are indicated.

