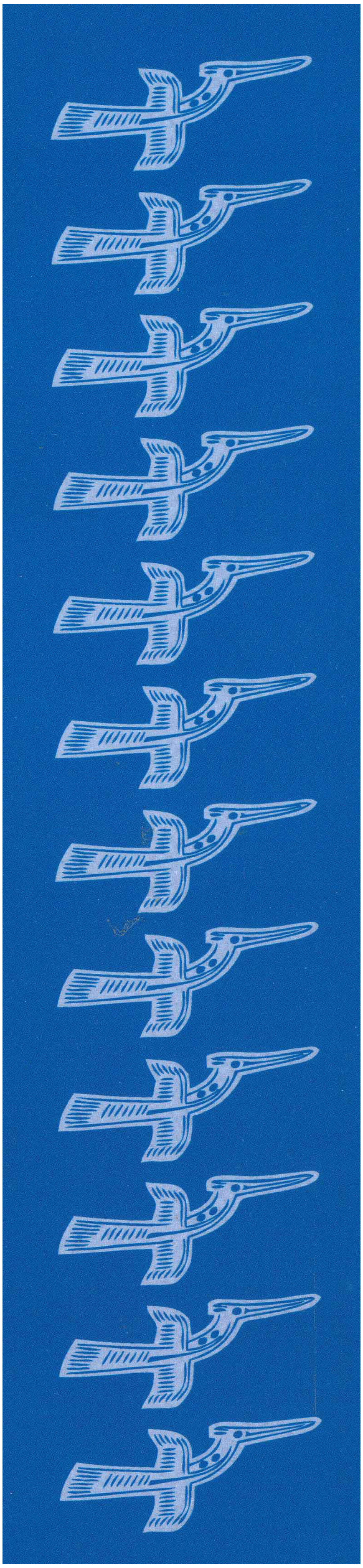
Mon-Khmer Studies

Volume 42



Author: YANIN Sawanakunanon
Title: *Segment timing in certain Austroasiatic languages: implications for typological classification*
Pages: 40-53
Date Received: 18/8/2013
Revised Text Accepted: 29/8/2013

Volume 42 Editors:
Paul Sidwell
Brian Migliazza

ISSN: 0147-5207
Website: http://mksjournal.org

Published by:
Mahidol University (Thailand)
SIL International (USA)

Segment timing in certain Austroasiatic languages: implications for typological classification

YANIN Sawanakunanon
Chulalongkorn University

Abstract
This study analyzed segment timing in Mon, Khmer, and Vietnamese. The speech data were segmented into vocalic, consonantal, voiced, and unvoiced intervals. The results showed that the variation of vocalic durations plays an important role in language classification. The different characteristics of vowels in each language led to different timing patterns. Khmer, a restructured language, has vowel length distinctions resulting in the highest variation of vocalic durations. Mon, a register language, follows Khmer with the distinction in phonation types. Vietnamese, a tonal language in which some tones co-occur with phonation, has the lowest variation of vocalic durations. It was noted that suprasegmental features had various levels of influence on segment timing.
Keywords: segment timing, speech rhythm, phonetics
ISO 639-3 language codes: vie, khm, mnw

1. Introduction

Studies investigating segment timing or segment duration have found that many factors affect segment timing. Some of those factors are syllable structure, segment position in the syllable, syllable position in word, phrase, or utterance, stress level, focus, sound environment, speech tempo, articulation process, as well as intrinsic duration of the segment itself (Botinis, Bannert, Fourakis, & Pagoni-Tetlow, 2002; de Jong, 1991, 2004; Greenberg, Carvey, Hitchcock, & Chang, 2003; Suomi, 2005; Warner & Arai, 2001).

There are also studies of segment timing that aim to classify languages according to temporal organization of segments in connected speech. Such studies developed from the study of speech rhythm. Rhythmic units in speech can be determined by the recurrence of stressed syllables, all syllables, or moras and the recurrence of such units are believed to be perceived as approximately equal in duration leading to a rhythmic pattern. It has been suggested that there are three types of speech rhythm: stress-timed, syllable-timed, and mora-timed, based on what units determine the rhythm. The classic examples of stress-timed languages are English and German. The main cited examples of syllable-timed languages are French, Spanish, and Italian (Pike, 1945). Finally, Japanese is a mora-timed language (Laver, 1994).

Acoustic studies, however, fail to support the theory that rhythmic units are perceptually equal in duration (Dauer, 1983; Luangthongkum, 1977; Roach, 1982; Sawanakunanon, 2002; Surinpiboon, 1985; Teeranon, 2000). Dauer (1983) suggested that phonological, phonetic, lexical, and syntactic factors, rather than the speaker's attempt to equalize inter-stress or inter-syllable intervals, may cause rhythmic differences. She further proposed three main differences between stress-timed and syllable-timed languages: the variation and complexity of syllable structure, the presence or absence of vowel reduction, and lexical stress. Stress-timed languages have more types of syllable structures, and those syllable structures are more complex than in syllable-timed languages. In addition, syllable weight plays some role in stress assignment. Heavy syllables tend to be stressed more than light ones.

Vowel reduction is found in stress-timed languages. While vowel reduction is conditioned by phonetic factors in such languages, it seems to be conditioned by phonological environment in languages with syllable-timed rhythm. Moreover, syllable-timed languages do not regularly have reduced variants of vowels in unstressed position. Most stress-timed languages have lexical or word-level stress realized by phonetic characteristics such as high pitch, greater length, loudness, and full vowel quality, which make stressed syllables prominent. It could be argued, therefore, that all syllables tend to be equally prominent in syllable-timed languages. However, some languages have a mixture of characteristics from both rhythmic classes. For example, Catalan, with syllable structures similar to those of Spanish, could be classified as a syllable-timed language, yet it has

YANIN Sawanakunanon. 2013. Segment timing in certain Austroasiatic languages: implications for typological classification. *Mon-Khmer Studies* 42:40-53

vowel reduction. In contrast, Polish, which has a great variety of syllable structures, has no vowel reduction (Nespor, 1990). Although Dauer's proposal may not hold true for every language, it provides alternative ways of explaining factors determining speech rhythm.

Most languages studied in this aspect are European languages and some major eastern languages. Only a few Southeast Asian languages are attested (Grabe and Low, 2002; Romano, Mariano, and Calabro, 2011). Austroasiatic languages, however, have never been analyzed in this fashion. In this study, segment timing of Burmese Mon, Surin Khmer, and Hanoi Vietnamese will be examined. Burmese Mon is a register language in which phonation types of its vowels are phonemically distinctive. Surin Khmer is a restructured language. Its vowel system lost phonation distinction and resulted in phonemic difference in vowel length and various vowel qualities. Unlike the other two languages, Vietnamese is a tonal language.

The three languages have different dominant phonetic and phonological characteristics. In terms of syllable structures, Mon and Khmer are rich in sesquisyllabic words. The stress pattern of light and heavy syllables in Mon and Khmer, thus, could be similar to stress-timed languages and different from Vietnamese which is considered a monosyllabic language. As for vowels, Surin Khmer has vowel length distinction. Vietnamese has one pair of short and long vowel but Mon has none. Suprasegmental features in the three languages are also different. There are six tones in Vietnamese and two of which co-occurred with phonations. Phonation also plays an important role in Mon vowels resulting in two sets of vowels with different phonation types. In this paper, we will see how and whether Mon, Khmer, and Vietnamese, whose phonetic characteristics affect segment duration in different ways, can be classified by their segment timing pattern by following the analyses of the three language classification models explained below in §2.

2. Language classification models

Besides a number of studies in speech rhythm from phoneticians, there are some works in psycholinguistics dedicated to speech rhythm as well. Psycholinguistic studies of speech segmentation reveal infants' ability to determine word boundaries by using rhythmic cues, which are stressed syllables in stress-timed languages, syllables in syllable-timed languages, and moras in mora-timed languages. Adults continue using this ability in second-language acquisition (Mehler, Dommergues, Fraunfelder, & Segui, 1981). Moreover, infants' ability to discriminate languages with different rhythm classes and the ability to group languages with the same type of rhythm suggest that there must be some characteristics in common between languages in the same group which differentiate them from another group (Mehler & Christophe, 1995; Nazzi, Bertoncini, & Mehler, 1998; Ramus & Mehler, 1999). This raises the question as to what those common characteristics are. With the assumption that infants perceive vowels better than consonants because of the higher energy and duration of vowels and that they perceive speech as successions of these high energy sounds (vowels) alternating with noise (consonants), resynthesized speech which replaced all vowels by /a/ and all consonants by /s/ was used in a language discrimination experiment (Ramus & Mehler, 1999). Their results on language discrimination with the use of the resynthesized speech supported the findings of the experiment with natural speech.

Ramus, Nespor, and Mehler (1999) then developed an acoustic model of rhythmic classification. This model incorporates three parameters derived from the duration of vocalic and intervocalic intervals, which are intervals of successive vowels and consonants respectively[1]. These parameters are the proportion of vocalic intervals in the sentence (%V), the standard deviation of the duration of vocalic intervals within each sentence (ΔV), and the standard deviation of the duration of intervocalic intervals within each sentence (ΔC). They found that %V and ΔC show the grouping of languages which supports the theory of three types of speech rhythm. In their study, two languages which had never been classified by speech rhythm were tested. Polish has complex syllable structures, and yet does not have vowel reduction which is claimed to be a characteristic of stress-timed languages. Catalan, on the contrary, has vowel reduction but simple syllable structures. The model groups Polish with English and Dutch while Catalan is grouped with Spanish, Italian, and French. This result suggests that languages in the study are grouped by the variation and complexity of syllable structure, not the existence of vowel reduction.

[1] To illustrate, Ramus et al. (1999) gives an example of the phrase 'next Tuesday on' which can be transcribed as /nɛkstjuzdeiɔn/. The three vocalic intervals from this phrase are the intervals consisting of /ɛ/, /u/, and /eiɔ/. The four intervocalic (or consonantal) intervals are the intervals consisting of /n/, /kstj/, /zd/, and /n/.

However, the use of vocalic and intervocalic intervals raises some questions. How can infants or adult listeners distinguish between a nasal consonant, which is a part of an intervocalic interval, and a nasal vowel, which is a part of a vocalic interval? Should syllabic consonants and glides be included in vocalic or intervocalic intervals? In their study, Galves, Garcia, Duarte, and Galves (2002) found evidence that infants perceive speech signals on the basis of sonority and obstruency. The criterion used to determine sonorant and obstruent sounds in their study is neither articulatory nor phonological but based purely on the acoustic properties of speech. Steiner (2003), using a sonority hierarchy, classified sounds into eight groups: vowel, approximant, syllabic lateral, syllabic nasal, lateral, nasal, fricative, and affricate. The first four groups are included in the vocalic intervals, and the latter four are included in the intervocalic intervals. However, Steiner (2003) suggested that lateral and nasal intervals can classify languages well, and that some classes of consonants might play a more important role than others in language grouping.

Dellwo, Fourcin, and Abberton (2007) took a different approach. They gave an example of the problem in classifying nasal consonants and nasal vowels. They also hypothesized that listeners may be able to distinguish languages based on the difference between voiced and voiceless sounds. Voiced interval (VO), instead of vocalic interval, is used in the parameter %VO, the proportion of voiced interval in the sentence. Voiceless or unvoiced interval (UV), instead of intervocalic interval, is used in the parameter varcoUV, which is the variation coefficient of the standard deviation of unvoiced intervals. Unvoiced intervals are normalized to reduce any effect of speech rate. The results seemed to support the traditional classification. English and German, which are stress-timed languages, are grouped together with high varcoUV values and low %VO. French and Italian, with low varcoUV and high %VO values, are separated from the other two languages. A high varcoUV value can be linked to complex syllable structures, as found in English and German, whereas a low value, as in French and Italian, seems to suggest simple syllable structures.

Not only have there been debates regarding segmentation of vocalic and intervocalic intervals, but also alternative parameters and calculations have been introduced. Low, Grabe, and Nolan (2000) proposed a different calculation of vocalic and intervocalic intervals. In their previous studies (Grabe, Post, & Watson, 1999), English had more vocalic variability than French. They related this finding to vowel quality and explained that English has high variability in vowel durations because it has both full and reduced vowels. French does not have reduced vowels, and that makes the level of vocalic variability lower than that of English. Therefore, they focused on the difference in the variability of vowel duration and computed a Pairwise Variability Index (PVI) which expressed the level of variability in successive measurements. Two versions of PVI were proposed (Grabe & Low, 2002): normalized PVI (nPVI) was used with vocalic intervals, and raw PVI (rPVI) was used with intervocalic intervals. They argued that their PVIs would capture the characteristics of rhythm better than Ramus, et al.'s ΔV and ΔC. Two sets of data of which one had three successive long vowels that followed three successive short vowels, and another which had long and short vowels alternating, would have the same standard deviation of vocalic interval durations although the patterns differed. The results suggested that the vocalic nPVI provided a better separation of languages than the intervocalic rPVI.

The vocalic nPVI values of six languages were also compared by Grabe and Low (2002) with Ramus, et al.'s %V values. English and German, which represent stress-timed languages, have high vocalic nPVI values and low %V values. French and Spanish, representing syllable-timed languages, have low vocalic nPVI values but high %V values. Thus, it seems that these two parameters can reflect a rhythmic characteristic which, in this case, is vowel duration. The conclusion by Grabe and Low (2002) that Thai and Tamil, which have vowel length distinctions, are stress-timed languages because of their high nPVI values are questionable since they also have high %V values which are a characteristic of syllable-timed languages. Therefore, languages which have vowel length distinctions should be carefully examined.

In spite of the varieties of methods used in segmentation and statistical analysis, it can be seen that these studies use timing of segmental intervals to classify languages. They also discuss phonetic and phonological factors shared by groups of languages which make them different from the others. Moreover, this kind of language classification is always compared with the classic rhythm class hypothesis. Whenever unclassified or mixed-rhythm languages are tested, they will be compared with the reference languages, such as English, French, and Japanese, to determine their rhythm class.

3. Method

3.1 Languages, Speakers, and Speech Materials

The three languages analyzed in this study are Burmese Mon, Surin Khmer, and Hanoi Vietnamese. Vietnamese is tonal and rich in monosyllabic words. Mon and Khmer are non-tonal languages and have a great deal of sesquisyllabic words[2]. Moreover, phonation type is contrastive in Mon and Vietnamese. A phonation contrast is found between modal and breathy vowels in Mon and phonation co-occurs with tones in Vietnamese (i.e., a creaky tone and a glottalized tone). Short and long vowels are phonologically different in Khmer. Vietnamese has one pair of short and long vowels. The aforementioned phonetic and phonological characteristics of the three languages can be summarized in Table 1.

Table 1 Phonetic and phonological characteristics of the three languages

Languages	Vowel length	Number of syllables in a word	Phonation contrast	Tonal/ non-tonal
Vietnamese	phonemic (1 pair)	monosyllabic	glottalized and creaky tones	tonal
Mon	non-phonemic	sesquisyllabic	modal and breathy vowels	non-tonal
Khmer	phonemic	sesquisyllabic	-	non-tonal

These three languages will be investigated with the three models of Ramus et al. (1999), Grabe and Low (2002), and Dellwo et al. (2007). The characteristics of the three languages shown in Table 1 have never been the focus of attention before as factors which might contribute to segment timing patterns. Therefore, it is interesting to see whether these characteristics will have some effect on segment timing patterns by using the three language classification models.

The three native speakers of each language ranged in age from 25 to 35 years old. The Vietnamese speakers were from Hanoi and Hai Duong and spoke Hanoi dialect. The three Mon speakers are from Mudon, Myanmar. The Khmer speakers spoke Surin dialect. They are all from Surin Province, Thailand.

Spontaneous speech in stories told by speakers with moderate tempo was sampled at 16 kHz and recorded with a unidirectional microphone directly on a laptop computer hard drive. Approximately 30 seconds of clear speech, not including pauses and hesitations, was selected from each speaker for acoustic analysis.

3.2 Acoustic analysis

The data were segmented and labeled, using the Praat software system (Boersma & Weenink, 2010), into vocalic and consonantal intervals, and voiced and unvoiced intervals. These intervals were identified regardless of syllable and word boundaries. In addition, consonant-vowel and syllable boundaries were also marked for reference. Pauses, as well as syllables preceding and following pauses, were excluded from the analysis. Utterance-final syllables were excluded to avoid lengthening effects. It was also impossible to identify the point where a stop sound ended or began when it occured before and after pauses. Utterance-initial syllables were consequently excluded for consistency. Segmentation was made as accurate as possible despite the fact that there was co-production or coarticulation – that is, overlap in articulatory movements. Particular measurement issues that required careful consideration are discussed below.

Vowels were marked between the points where clear patterns of vowel formants appeared whether the acoustic excitation was voiced or voiceless or both. Other acoustic properties were also used to help identify such points. A vocalic interval was marked between the two points. A consonantal interval was then marked between two vocalic intervals.

For glides, Ramus et al. (1999) included pre-vocalic glides in consonantal intervals and post-vocalic glides in vocalic intervals. Grabe and Low (2002) used formant frequency and amplitude movements to classify glides. They included glides in vocalic intervals unless there were

2 Sesquisyllabic structure is composed of a minor syllable followed by a major syllable.

observable changes in formant and amplitude of speech signals. In this study, pre-vocalic glides were included in consonantal intervals because constriction in initial position is quite audible. Post-vocalic glides were included in vocalic intervals because there is not enough constriction at the end when the vocal tract is coming to shape ‘u’ or ‘i’. These acoustic criteria for glides then agreed with the measurements of Ramus et al. (1999).

As for voiced and unvoiced intervals, Dellwo et al. (2007) used acoustic cues to locate them. A voiced interval of successive voiced segments, beginning from the onset of a voiced segment to the offset of the next one, was marked across syllable and word boundaries. Similarly, an unvoiced interval was marked from the onset to the offset of an unvoiced segment, or successive unvoiced segments were marked.

A glottal stop closure at the end of a glottalized tone and in the middle of a creaky tone was treated as a consonant. So, it was treated as a part of consonantal or unvoiced intervals. Figure 1 illustrates the segmentation of the four types of intervals in Praat.

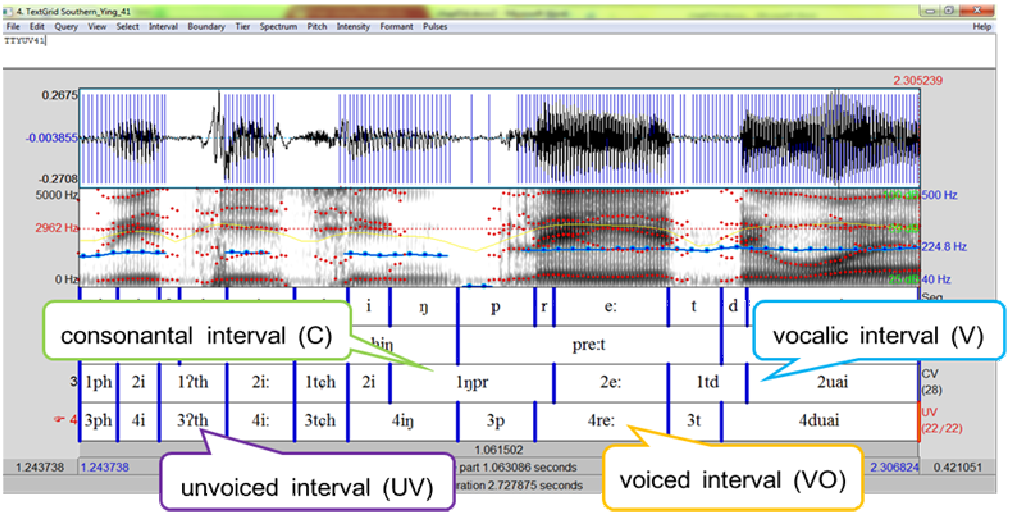


Figure 1: Segmentation of vocalic and consonantal intervals, and voiced and unvoiced intervals

3.3 Language classification parameters

Durations of the four intervals were obtained and expressed as eight parameters in the three models. Duration measurements of vocalic and consonantal intervals were used in the two models proposed by Ramus et al. (1999) and Grabe and Low (2002). The three parameters, which are the proportion of vocalic intervals (%V), the standard deviation of the duration of vocalic intervals (ΔV), and the standard deviation of the duration of consonantal intervals (ΔC), were used in Ramus et al. (1999)’s model. The values of these three parameters were extracted from each utterance. The proportion of vocalic intervals (%V) is the sum of the duration of vocalic intervals divided by the total duration of the utterance. Therefore, %V will show whether the utterance has a proportion of vowels or consonants. The standard deviation of the duration of vocalic intervals (ΔV) and the standard deviation of the duration of consonantal intervals (ΔC) are also calculated per utterance and indicate how the duration of either vocalic or consonantal intervals in each utterance varied. Ramus et al. (1999) found that languages with reduced vowels are likely to have low value of %V and high value of ΔC. Therefore, Mon and Khmer were expected to have such patterns. Although ΔV does not classify languages well compared with %V and ΔC in their paper, it shows the difference between a language with vowel length distinction and languages with no such distinction in my preliminary study. In the current study, Khmer was expected to have high value of ΔV as its short and long vowels are phonemically contrastive. We will also see whether ΔV could capture the difference of phonation types in modal and breathy vowels in Mon.

Grabe and Low (2002) used PVI measurement aiming to show variability of interval duration. However, while the model of Ramus et al. (1999) aims to show variation in each utterance, Grabe and Low (2002) focuses on the difference between duration of two successive intervals. Accordingly, the PVI value represents variability of duration of adjacent intervals, not variability in an utterance. The raw pairwise variability index (rPVI) is used with consonantal intervals. It shows durational differences between two successive intervals on average. In this paper, it is referred to as rPVI_C for readability and can be computed by using the following formula:

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

In equation (1), 'd' is the duration of the consonantal interval, 'k' is consonantal interval k, $|d_k - d_{k+1}|$ is the absolute value of the durational difference between the preceding and the following consonantal intervals, and 'm' is the number of consonantal intervals in the utterance. The value of rPVI_C is the sum of the durational differences between two successive intervals in each utterance divided by '$m-1$'. In this study, Khmer was expected to have high rPVI_C as its initial consonant cluster is more complex than the clusters of Mon and Vietnamese.

The normalized pairwise variability index (nPVI) which is used with vocalic intervals is referred to nPVI_V in this paper and is calculated by the following formula:

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

In equation 2, 'd' is the duration of the vocalic interval, 'k' is vocalic interval k, $|d_k - d_{k+1}|$ is the absolute value of the durational difference between the preceding and the following vocalic intervals, and 'm' is the number of vocalic intervals in the utterance. $(d_k + d_{k+1})/2$ is the average duration of the preceding and the following vocalic intervals and is used to normalized the durational difference between the two vocalic intervals. The value of nPVI_V is the sum of the normalized durational differences between two successive intervals in each utterance divided by '$m-1$' and multiplied by 100.

A high value of PVI shows that there is a great variation between two successive intervals. According to Grabe and Low (2002)'s study, stress-timed languages are likely to have higher value of nPVI_V than that of syllable-timed languages as a result of the durational difference between vowels in stressed and unstressed syllables. In the current study, Mon and Khmer were expected to have higher value of nPVI_V because they have a great number of sesquisyllabic words compared with Vietnamese, which is a monosyllabic language.

Dellwo et al. (2007) proposed sound segmentation into voiced (VO) and unvoiced (UV) intervals instead of vocalic and consonantal intervals as in Ramus et al. (1999) and Grabe and Low (2002). As voiced intervals consist of vowels and voiced consonants, the characteristics of vowels and consonants are responsible for duration and proportion of voiced intervals (%VO). The proportion of voiced interval (%VO) is calculated by the duration of voiced interval in an utterance divided by the duration of the utterance and multiplied by 100.

Unvoiced intervals are only composed of voiceless consonants. The more unvoiced segments occur sequentially, the longer the unvoiced intervals. The variation coefficient of the standard deviation of unvoiced intervals (varcoUV) is computed from equation (3):

(3) $$varcoUV = \frac{\Delta UV}{\overline{UV}} \times 100$$

In equation (3), 'ΔUV' refers to the standard deviation of the duration of unvoiced intervals in the utterance and '$\overline{UV}$' is the average duration of unvoiced intervals in the utterance. The varcoUV value is calculated by dividing the standard deviation of the duration of unvoiced intervals (ΔUV) by the average duration of unvoiced intervals ($\overline{UV}$) and multiplying by 100. The division by $\overline{UV}$ is an attempt to reduce the effect of different speech rate.

As Low et al. (2000) found that there was no effect of speech rate on consonantal intervals, the same result should also be found in the case of unvoiced intervals. Therefore, the standard deviation of the duration of unvoiced intervals (ΔUV), where the duration of unvoiced intervals was not normalized, was analyzed in this study to compare its result with that of varcoUV where the duration of unvoiced intervals is normalized. It is also found in a preliminary study that ΔUV yielded a similar result to ΔC and provided clearer picture of language classification according to statistical analyses.

The values of the three parameters were extracted from each utterance similar to the model of Ramus et al. (1999). Dellwo et al. (2007) found that stress-timed languages were likely to have lower value of %VO and higher value of varcoUV because of their complex syllable structures. The same pattern was expected to be found in Mon and Khmer which have more complex initial clusters than Vietnamese.

The calculation of the eight parameters was done in Microsoft Excel. The results were statistically tested by ANOVA and followed by Tukey's HSD post-hoc test (Tukey's Honestly Significant Difference Test) to ascertain if there was a statistically significant difference.

4. Results

The results of the eight parameters are illustrated below, beginning with the three parameters from Ramus et al. (1999), followed by the two parameters from Grabe and Low (2002) and the last three ones from Dellwo et al. (2007).

4.1 %V, ΔC, and ΔV

The parameters %V, ΔC and ΔV were analyzed following Ramus et al. (1999). The number of vocalic and consonantal intervals, total duration, the average proportion of the duration of vocalic intervals (%V), the average standard deviation of the duration of vocalic intervals (ΔV), and the average standard deviation of the duration of consonantal intervals (ΔC) across all utterances of each language are presented in Table 2.

Table 2 shows that Mon has the highest value for the proportion of vocalic intervals (%V) and Vietnamese has the lowest. This result refutes the hypothesis that Mon and Khmer would have low values of %V because they have reduced vowels in minor syllables of sesquisyllabic words. The ANOVA test shows a significant difference ($p < .05$) and Tukey's HSD test shows that Mon is significantly different from Vietnamese ($p < .05$). The high value of %V, which means a high proportion of vowel duration, in Mon could be because of its breathy vowels, as duration of breathy vowels is found higher than that of modal vowels in some studies (Blankenship, 2002; Kirk, Ladefoged, & Ladefoged, 1984; Luangthongkum, 1990; Wayland & Jongman, 2003). For Vietnamese, a glottal closure in creaky and glottalized tones was treated as a consonant, as mentioned in §3.2. Vowels occurring with such tones are shorter in duration than vowels occurring with non-phonation tones. These tones, hence, contribute to a lower value of vocalic duration in Vietnamese.

Table 2: Total number of vocalic and consonantal intervals, total duration, the proportion of the duration of vocalic intervals (%V), the standard deviation of the duration of vocalic intervals (ΔV), and the standard deviation of the duration of consonantal intervals (ΔC)

Languages	Vocalic intervals	Consonantal intervals	Total duration (Sec)	%V	ΔC	ΔV
KM1	168	175	36.06	50.97	48.88	63.52
KM2	146	155	32.26	49.21	45.57	69.92
KM3	146	149	34.25	54.03	44.86	68.33
KM	460	479	102.57	51.27	46.38	67.42
MN1	142	146	31.11	57.05	42.94	56.10
MN2	134	135	30.22	59.10	41.64	70.69
MN3	177	181	33.22	46.98	44.70	54.25
MN	453	462	94.55	55.43	42.81	62.02
VN1	203	214	33.53	46.68	35.35	39.35
VN2	189	195	36.12	55.81	42.22	46.05
VN3	168	176	33.08	46.18	47.21	47.63
VN	560	585	103.13	49.31	41.05	43.91

KM = Khmer; MN = Mon; VN = Vietnamese; 1 = Speaker 1; 2 = Speaker 2; 3 = Speaker 3

The standard deviation of the duration of consonantal intervals (ΔC) was highest in Khmer followed by Mon and Vietnamese. This result supports the hypothesis that Mon and Khmer have high values of ΔC because their syllable structures are more complex than those of Vietnamese. However, although the numeric pattern of the result supports the hypothesis, the differences were not statistically significant.

Figure 2: Distribution of languages over the %V and ΔC plane (top) and the %V and ΔV plane (bottom)

%V = proportion of vocalic intervals; ΔV = S.D. of vocalic interval duration;
ΔC = S.D. of consonantal interval duration
KM = Khmer; MN = Mon; VN = Vietnamese; 1 = Speaker 1; 2 = Speaker 2; 3 = Speaker 3

As for the standard deviation of the duration of vocalic intervals (ΔV), it is found that Khmer has the greatest variation in vocalic interval duration, as expected with its highest value of ΔV among the three languages. Mon comes in second and Vietnamese has the lowest ΔV value. The high ΔV value in Khmer could be a result of the durational difference between short and long vowels. Moreover, the difference between reduced vowels in minor syllables and full vowels in major syllables in sesquisyllabic words might play some role. Although Mon also has a great number of sesquisyllabic words, its ΔV value is lower than that of Khmer. The durational difference between normal and breathy vowels in Mon also resulted in the lower ΔV value than that of Khmer. The result for ΔV, therefore, suggests that the durational difference between short and long vowels is larger than that between reduced and full vowels, as well as normal and breathy vowels. The ANOVA test found a significant difference ($p < .05$) and the Tukey's HSD test found that Mon and Khmer were significantly different from Vietnamese ($p < .05$).

Ramus et al. (1999) suggested that a graph plotted on the %V and ΔC plane gives the best language classification. However, since the ΔC values in Mon, Khmer and Vietnamese are not significantly different, a graph plotted on %V and ΔV plane consequently provides better classification (see Figure 2).

The average values of each language are shown with dotted circles. From Figure 2, the %V and ΔV graph shown in the bottom displays Mon and Khmer together with higher values of %V and ΔV than those of Vietnamese. Mon and Khmer share phonetic characteristics, resulting in similar vowel timing patterns. They are sesquisyllablic languages. The factor that may play the most important role in this part, thus, could be the durational difference between reduced and full vowels in sesquisyllabic words. Vowel length distinction and phonation type distinction in vowels are also important factors causing high values of %V and ΔV.

4.2 PVI Results

PVI measurement is used by Grabe and Low (2002) to show variability of interval duration. The durations of vocalic and consonantal intervals were used to compute the raw pairwise variability index of consonantal intervals (rPVI_C) and the normalized pairwise variability index of vocalic intervals (nPVI_V). The values of both parameters of all speakers of Mon, Khmer, and Vietnamese are presented in Table 3.

Table 3: rPVI_C and nPVI_V values

Languages	**rPVI_C**	**nPVI_V**
KM1	56.46	64.97
KM2	52.84	78.40
KM3	49.76	67.26
KM	53.01	70.72
MN1	46.83	48.11
MN2	45.65	66.54
MN3	47.26	71.97
MN	46.42	62.14
VN1	41.27	54.59
VN2	45.18	48.25
VN3	47.22	58.03
VN	44.27	53.71

KM = Khmer; MN = Mon; VN = Vietnamese; 1 = Speaker 1; 2 = Speaker 2; 3 = Speaker 3

From Table 3, the value of rPVI_C in Khmer is higher than Mon and Vietnamese, as expected. Its more complex initial cluster resulted in a high value of rPVI_C, which represents more variability in two successive consonantal intervals. However, the ANOVA test does not show a statistically significant difference, in contrast to the result for ΔC in the Ramus et al. (1999) model.

The value of nPVI_V can be interpreted in the same way. Khmer, again, has the highest value which suggests that adjacent vocalic intervals in Khmer have greater variability than in the other two languages. The result in this part is as expected and similar to the result of ΔV in the model of Ramus et al. (1999), as Mon has the second highest nPVI_V value and Vietnamese comes last. The ANOVA test shows a significant difference ($p < .05$) and the Tukey's HSD test shows that Khmer and Vietnamese are significantly different. Figure 3 shows the three languages plotted on rPVI_C and nPVI_V plane.

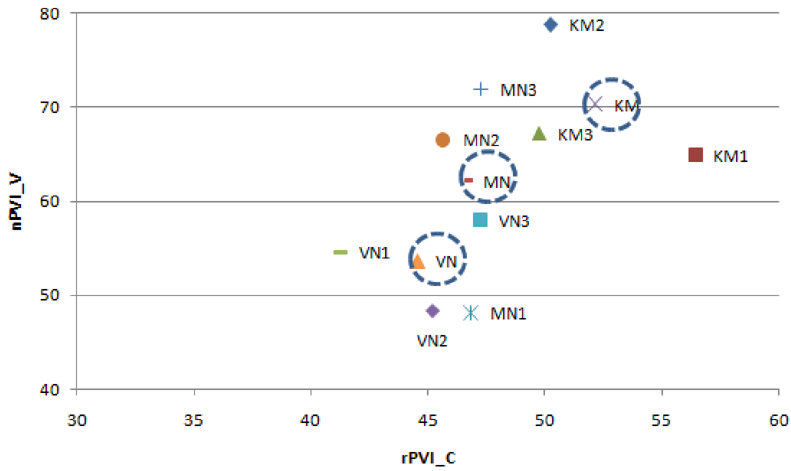


Figure 3: Distribution of languages over the rPVI_C and nPVI_V plane

rPVI_C = raw pairwise variability index in consonantal interval duration;
nPVI_V = normalized pairwise variability index in vocalic interval duration;
KM = Khmer; MN = Mon; VN = Vietnamese; 1 = Speaker 1; 2 = Speaker 2; 3 = Speaker 3

The language distribution in Figure 3 shows Khmer with high values for both rPVI_C and nPVI_V. Mon is between Khmer and Vietnamese. The result in this part, thus, supports Grabe and Low's (2002) claim that both parameters reflect the difference between reduced vowels in unstressed syllables and full vowels in stressed syllables as found in Khmer and Mon. Moreover, the values of rPVI_C and nPIV_V of Vietnamese in the current study are similar to those of Romano et al. (2011).

4.3 %VO, varcoUV, and ΔUV

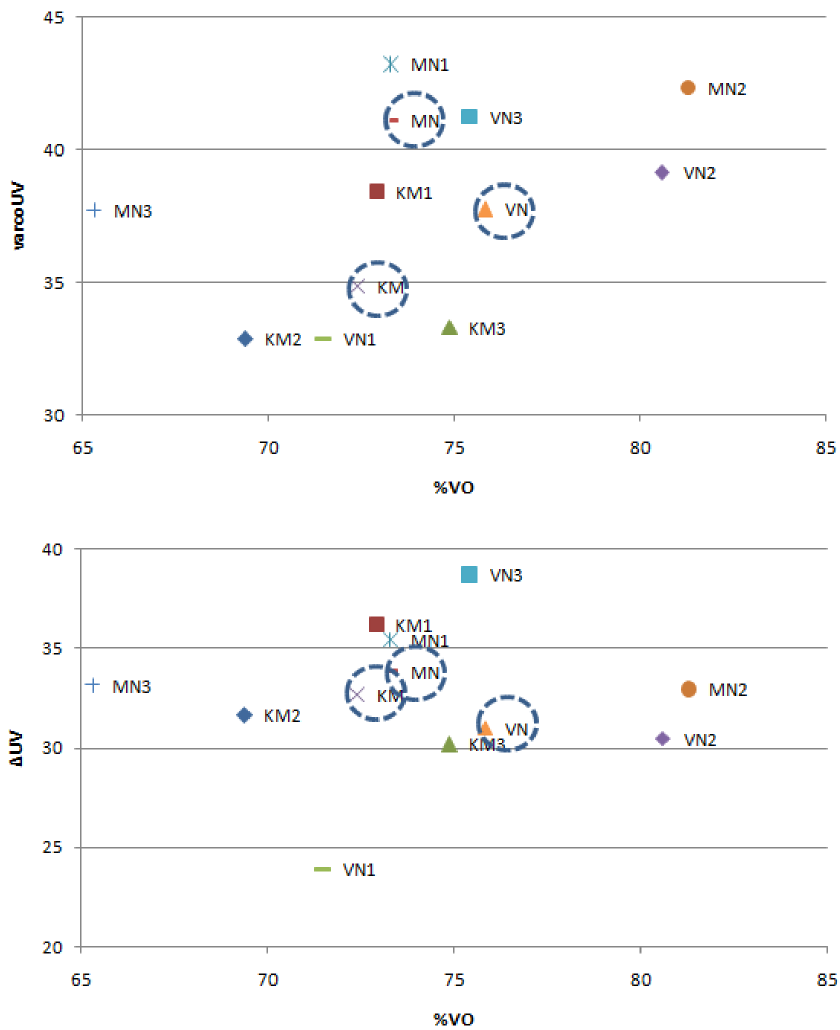
This section describes the analysis of durations of voiced (VO) and unvoiced (UV) intervals following the model of Dellwo et al. (2007). The total number of voiced and unvoiced intervals, the proportion of the duration of voiced intervals (%VO), the variation coefficient of the standard deviation of unvoiced intervals (varcoUV), and the standard variation of unvoiced intervals (ΔUV) are presented in Table 4.

Table 4: Total number of voiced and unvoiced intervals, the proportion of the duration of voiced intervals (%VO), the variation coefficient of the standard deviation of unvoiced intervals (varcoUV), and the standard variation of unvoiced intervals (ΔUV)

Languages	Voiced intervals	Unvoiced intervals	%VO	varcoUV	ΔUV
KM1	113	112	72.91	38.43	36.24
KM2	100	98	69.38	32.88	31.66
KM3	99	95	74.87	33.29	30.21
KM	312	305	72.20	34.75	32.64
MN1	104	102	73.28	43.24	35.44
MN2	81	72	81.29	42.36	32.96
MN3	139	134	65.30	37.72	33.16
MN	324	308	74.79	41.47	33.79
VN1	133	129	71.47	32.89	23.92
VN2	101	97	80.60	39.16	30.45
VN3	90	85	75.40	41.26	38.72
VN	324	311	75.44	37.34	30.41

KM = Khmer; MN = Mon; VN = Vietnamese; 1 = Speaker 1; 2 = Speaker 2; 3 = Speaker 3

From Table 4, the %VO values of the three languages can be seen to be not much different. However, it is noticeable that Vietnamese, which has the lowest value of %V as shown in §4.1, has the highest value of %VO. As the difference between the two parameters is that voiced consonants are included in %VO, the higher value of %VO suggests that the Vietnamese data have more voiced consonants than Mon and Khmer. Mon has the highest values of varcoUV and ΔUV. The results of the three parameters for Mon and Khmer, which have more complex syllable structures, show lower values of %VO and higher values of varcoUV than those of Vietnamese, as expected. The graph plotted on %VO and ΔUV plane in Figure 4 shows Mon and Khmer are grouped closer. Nevertheless, the ANOVA tests of the three parameters do not show statistically significant differences across languages.



%VO = proportion of voiced intervals; varcoUV = variation coefficient of the standard deviation of unvoiced interval duration; ΔUV = S.D. of unvoiced interval duration; KM = Khmer; MN = Mon; VN = Vietnamese; 1 = Speaker 1; 2 = Speaker 2; 3 = Speaker 3

Figure 4: Distribution of languages over the %VO and varcoUV plane (top) and the %VO and ΔUV plane (bottom)

5. Discussion

The eight parameters analyzed in this study were derived from the durations of vocalic-consonantal intervals and voiced-unvoiced intervals. Vocalic and consonantal intervals consist of vowels and consonants respectively. Voiced intervals include not only vowels but also voiced consonants, while unvoiced intervals consist of only voiceless consonants. Although there are some differences between these two groups of intervals, the analyses of the eight parameters were mostly based on the durations of consonantal and vocalic intervals.

According to the results in §4, the parameters acquired from consonantal intervals (ΔC, rPVI_C, varcoUV and ΔUV) do not show statistically significant differences between Mon, Khmer, and Vietnamese. This means that even though Mon and Khmer have more complex initial clusters than Vietnamese, the durational differences are not that large.

On the other hand, there are significant differences between Mon, Khmer, and Vietnamese in the vocalic interval parameters (i.e. %V, ΔV and nPVI_V). Therefore, this finding suggests that the three languages are more different in terms of vowels, especially in their durational variation. The largest differences among the three parameters are found in ΔV and nPVI_V, which show the durational variation of vocalic intervals. The parameter ΔV represents overall variation of vocalic intervals and nPVI_V measures variation of successive vocalic intervals. According to the results of the ANOVA and the Tukey HSD tests shown in §4.1 and §4.2, the value of ΔV in Vietnamese is significantly different from those of Mon and Khmer ($p < .05$) and the value of of nPVI_V in Vietnamese is significantly different from that of Khmer ($p < .05$). Figure 5, with the values of these two parameters plotted, shows that Mon and Khmer are grouped closer with higher values of ΔV and nPVI_V and leave Vietnamese in another corner of the graph with lower values of both parameters.

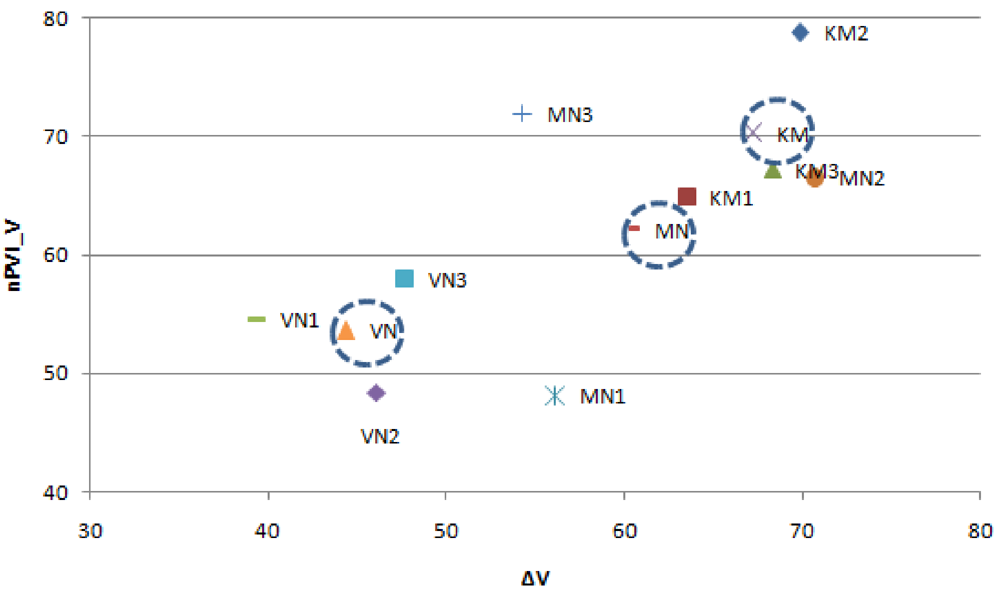


Figure 5: Distribution of languages over the ΔV and nPVI_V plane

As explained in §4, similar phonetic and phonological characteristics should result in similar segment timing patterns. Mon and Khmer are both sesquisyllabic languages and this characteristic differentiates them from Vietnamese, which is a monosyllabic language. The greater difference between the durations of reduced vowels in minor syllables and full vowels in major syllables leads to higher values of ΔV and nPVI_V. The characteristics of vowels themselves also matter. The durational difference between modal and breathy vowels in Mon is another factor that causes high values of both parameters. Moreover, vowel length distinction in Khmer enhances the durational variation of vocalic intervals.

According to the language classification models followed in the current study, values of parameters plotted on graphs can be considered as speech rhythm continuum. Greater variation in the durations of vocalic intervals is a characteristic of stress-timed languages, while less variation is an attribute of syllable-timed languages. Applying this concept to the two parameters plotted in Figure 5, the rhythm continuum would lie from syllable-timed rhythm at the bottom left of the graph to stress-timed rhythm at the top right. Although there are no exact reference points in the graph to determine the region of each type of rhythm, it can be said that Vietnamese, at one end of the continuum, has a characteristic of a syllable-timed language and Khmer, on another end of the continuum, seems to have stress-timed rhythm. As for Mon, it falls in the middle of the continuum but is closer to the stress-timed rhythm end.

6. Conclusion

This study analyzed segment timing in Mon, Khmer, and Vietnamese. The speech data were segmented into vocalic, consonantal, voiced, and unvoiced intervals. The interval durations were then measured and converted into eight parameters (%V, ΔV, ΔC, nPVI_V, rPVI_C, %VO, varcoUV, and ΔUV) following the previous works of Ramus et al. (1999), Grabe and Low (2002), and Dellwo et al. (2007).

The results showed that the variation of vocalic durations plays an important role in language classification. The different characteristics of vowels in each language led to different timing patterns. Khmer, a restructured language, has vowel length distinction resulting in the highest variation of vocalic durations. Mon, a register language, followed Khmer with the distinction in phonation types. Vietnamese, a tonal language in which some tones co-occur with phonation, has the lowest variation of vocalic durations. It was noted that suprasegmental features had various levels of influence on segment timing.

The ΔV and nPVI_V graph shows language classification echoing the statistical analyses. The distribution of languages in the graph resembles a speech rhythm continuum where Khmer is on the stress-timed end, Vietnamese is on the syllable-timed end, and Mon is in the middle close to Khmer.

Acknowledgments

This research was supported by the Royal Golden Jubilee program (RGJ), the Thailand Research Fund (TRF) (PHD/0186/2549). I am deeply grateful to my advisors, Professor Theraphan Luangthongkum and Professor Arthur S. Abramson for their guidance and support. I would also like to thank the reviewers for their invaluable comments. Thanks to all the nine speakers. Without their help, this research project would not have been possible.

References

Blankenship, B. (2002). The timing of nonmodal phonation in vowels. *Journal of Phonetics*, 30(2), 163-191.

Boersma, P., & Weenink, D. (2010). Praat: doing phonetics by computer (Version 5.1.43). Retrieved from http://www.praat.org/

Botinis, A., Bannert, R., Fourakis, M., & Pagoni-Tetlow, S. (2002, April 11-13, 2002). *Crosslinguistic Segmental Durations and Prosodic Typology.* Paper presented at the Speech Prosody 2002, Aix-en-Provence, France.

Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics*, 11, 51-62.

de Jong, K. J. (1991). An Articulatory study of consonant-induced vowel duration changes in English. *Phonetica, 48*, 1-17.

de Jong, K. J. (2004). Stress, lexical focus, and segmental focus in English: patterns of variation in vowel duration. *Journal of Phonetics, 32*(4), 493-516.

Dellwo, V., Fourcin, A., & Abberton, E. (2007, August 6-10, 2007). *Rhythmical classification based on voice parameters*. Paper presented at the 16th ICPhs 2007, Saarbrücken, Germany.

Galves, A., Garcia, J., Duarte, D., & Galves, C. (2002, April 11-13, 2002). *Sonority as a basis for rhythmic class discrimination*. Paper presented at the Speech Prosody 2002, Aix-en-Provence, France.

Grabe, E., & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. In C. Gussenhoven & N. Warner (Eds.), *Laboratory Phonology 7* (pp. 515-546). Berlin: Mouton de Gruyter.

Grabe, E., Post, B., & Watson, I. (1999, August 1-7, 1999). *The acquisition of rhythm in English and French*. Paper presented at the 14th International Congress of Phonetic Sciences, San Francisco.

Greenberg, S. M., Carvey, H., Hitchcock, L., & Chang, S. Y. (2003). Temporal properties of spontaneous speech - a syllable-centric perspective. *Journal of Phonetics, 31*, 465-485.

Kirk, P. L., Ladefoged, P., & Ladefoged, J. (1984). *Using a spectrograph for measures of phonation types in a natural language*. UCLA Working Papers in Phonetics 59.

Laver, J. (1994). *Principles of phonetics*. Cambridge: Cambridge University Press.

Low, E. L., Grabe, E., & Nolan, F. (2000). Quantitative characterisations of speech rhythm: Syllable-timing in Singapore English. *Language and Speech*, 43(4), 377-401.

Luangthongkum, T. (1977). *Rhythm in standard Thai*. (Unpublished doctoral dissertation). University of Edinburgh, Edinburgh.

Luangthongkum, T. (1990). The interaction between pitch and phonation type in Mon: Phonetic implications for a theory of tonogenesis. *Mon-Khmer Studies*, 16-17, 11-24.

Mehler, J., & Christophe, A. (1995). Maturation and learning of language during the first year of life. In M. S. Gazzaniga (Ed.), *The Cognitive Neurosciences* (pp. 943-954). MIT Press: Bradford Books.

Mehler, J., Dommergues, J. Y., Fraunfelder, U., & Segui, J. (1981). The syllable's role in speech segmentation. *Journal of Verbal Learning and Verbal Behavior*, 20(3), 298-305.

Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language Discrimination by Newborns: Toward an Understanding of the Role of Rhythm. *Journal of Experimental Psychology: Human Perception and Performance*, 24(3), 756-766.

Nespor, M. (1990). On the rhythm parameter in phonology. In I. M. Roca (Ed.), *Logical issues in language acquisition* (pp. 157-175). Dordrecht: Foris.

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

Ramus, F., & Mehler, J. (1999). Language identification with suprasegmental cues: A study based on speech resynthesis. *Journal of the Acoustical Society of America*, 105(1), 512-521.

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

Roach, P. (1982). On the distinction between “stress-timed” and “syllable-timed” languages. In D. Crystal (Ed.), *Linguistic controversies* (pp. 73-79). London: Edward Arnold.

Romano, A., Mairano, P., & Calabro, L. (2011, August 17-21, 2011). *Measures of speech rhythm in East-Asian tonal languages*. Paper presented at the The 17th International Congress of Phonetic Sciences (ICPhS 17), Hong Kong.

Sawanakunanon, Y. (2003). A comparison of rhythm in Thai tracheoesophageal and normal speech. In T. L-Thongkum (Ed.), *The Accent of Alaryngeal Thai Speakers*. Bangkok: Chulalongkorn Printing House. (in Thai)

Steiner, I. (2003). *A refined acoustic analysis of speech rhythm*. Paper presented at the LingColl 2003, Budapest, Hungary.

Suomi, K. (2005). Temporal conspiracies for a tonal end: Segmental durations and accentual f0 movement in a quantity language. *Journal of Phonetics, 33*, 291-309.

Surinpiboon, S. (1985). *The accentual system of polysyllabic words in Thai*. Unpublished master's thesis. Chulalongkorn University, Bangkok. (in Thai)

Teeranon, P. (2000). *Rhythmic unit and tonal variation in Thai connected speech*. Unpublished master's thesis. Chulalongkorn University, Bangkok. (in Thai)

Warner, N., & Arai, T. (2001). Japanese mora-timing: A review. *Phonetica, 58*, 1-25.

Wayland, R., & Jongman, A. (2003). Acoustic correlates of breathy and clear vowels: the case of Khmer. *Journal of Phonetics*, 31(2), 181-201.