SEGMENT TIMING IN TWELVE SOUTHEAST ASIAN LANGUAGES

Yanin Sawanakunanon¹

Abstract

In several studies the duration of segments (i.e. consonants and vowels) is measured to classify languages according to their speech rhythm. This research investigates whether Principal Component Analysis (PCA), a new method analyzing segment-timing of parameters for language classification, can be used to classify twelve Southeast Asian languages according to their timing The twelve Southeast Asian patterns. languages examined are Malay, Cebuano, Standard Thai, Southern Thai, Tai Yuan, Vietnamese, Hmong, Mien, Burmese, Sgaw Karen. Mon and Khmer.

Spontaneous speech from three speakers from each language was recorded. Vocalic, consonantal, voiced, and unvoiced intervals of 30 seconds of speech, not including pauses and hesitations, from each speaker were measured and analyzed using the three language typological classification models of Ramus et al. (1999), Grabe and Low (2002), and Dellwo et al. (2007). Eight parameters calculated from the duration of all intervals were then examined. In addition, Principal Component Analysis (PCA) was used to explore the relations among the parameters.

The results from the PCA show that the twelve languages can be classified into four groups: 1) Mon - Khmer; 2) Burmese -Hmong; 3) Vietnamese - Southern Thai - Tai Yuan; and 4) Malay - Cebuano. Standard Thai, Sgaw Karen, and Mien are not explicitly clustered with the other languages. The phonetic and phonological characteristics which seem to influence the twelve-language classification are the number of syllables in a word, the existence or non-existence of tone, and phonation type.

1. Introduction

Segments, as opposed to suprasegmental features, refer to consonants and vowels. In studies investigating segment timing or segment duration, it has been found that many factors may affect segment timing. Some of those factors are syllable structure, segment position in the syllable, syllable position in a word, phrase, or utterance, stress level, focus, sound environment, speech tempo, articulation process as well as intrinsic duration of the segment itself.

There are also studies of segment timing which aim to classify languages according to the temporal organization of segments in connected speech. Such studies can be said to have been developed from the study of speech rhythm. Rhythmic units in speech

¹ Ph.D. Candidate. Department of Linguistics, Faculty of Arts, Chulalongkorn University

can be determined by the recurrence of stressed syllables, all syllables, or moras. The recurrence of such units can be perceived as approximately equal in duration; that causes rhythm. It is widely accepted that there are three types of speech rhythm, namely, stress-timed, syllabletimed, and mora-timed, which are classified according to the units determining rhythm. Classic examples of stress-timed languages are English and German; syllable-timed languages are French, Spanish, and Italian (Pike 1945), and Japanese is a mora-timed language (Laver 1994).

However, acoustic studies fail to support the theory which claims that rhythmic units are equal in duration. Therefore, other phonetic and phonological explanations are considered. Dauer (1983) suggested that phonological, phonetic, lexical, and syntactic factors, rather than the speaker's equalize attempt to interstress or intersyllable intervals, may cause rhythmic She further suggested three differences. 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 languages. syllable-timed In addition. syllable weight plays some role in stress assignment. Heavy syllables tend to be stressed more than light syllables.

Vowel reduction, the centralization of unstressed vowels, is found in stress-timed While vowel reduction is languages. conditioned by phonetic factors in such languages, it seems to be conditioned by the phonological environment in languages with syllable-timed rhythm. Moreover, syllabletimed languages do not regularly have reduced variants of vowels in unstressed positions. 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 might be implied, from this statement, that all syllables tend to be syllable-timed prominent equally in languages; however, it might be just a tendency since there are some languages that have a mixture of characteristics from both rhythmic classes. Catalan, with syllable structures similar to those of Spanish, should be classified as a syllable-timed language, and yet has vowel reduction. On the other hand, Polish, which has a great variety of syllable structures, has no vowel reduction (Nespor 1990). Although Dauer's proposals may not hold true with every language, they give alternative ways of explaining factors determining speech rhythm.

Psycholinguistic studies about 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 and Segui 1981). Moreover, infants' ability to discriminate languages which belong to different types of rhythm and the ability to group languages which belong to the same type of rhythm suggest that there must be some characteristics in common between languages in the same group which make them different from another group (Mehler and Christophe 1995; Nazzi, Bertoncini and Mehler 1998; Ramus and Mehler 1999). This has led to the question of 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 with /a/ and all consonants with /s/ was used in a language discrimination experiment (Ramus and Mehler 1999). The results support the findings of the experiment with natural speech.

Ramus et al. (1999) then developed an acoustic model of rhythmic classification. This model creates three parameters from the duration of vocalic and intervocalic intervals, which are intervals of successive vowels and consonants respectively, and shows those parameters by plotting them on a graph. 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 (Δ C). They found that %V and Δ C show a 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 Polish has complex syllable tested. 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 results show that Polish is grouped with English and Dutch while Catalan is grouped with Spanish, Italian, and French. It might imply that languages in the study are grouped by the variation and complexity of syllable structure, not the existence of vowel reduction.

As good as it seems, the use of vocalic and intervocalic intervals raises some questions. How can infants or adult listeners distinguish between a nasal consonant, which is part of an intervocalic interval, and a nasal vowel, which is 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 that infants might roughly perceive speech signals on the basis of sonority and obstruency. The criterion used to determine sonorant and

obstruent sounds in this study was neither articulatory nor phonological but based purely on the acoustic properties of speech. On the other hand, Steiner (2003), with the concept of 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 later four are included in the intervocalic intervals. However, this study 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 assumed that listeners may be able to distinguish languages from 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. The normalization of unvoiced intervals is done 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. High varcoUV value can be linked to complex syllable structures, as found in English and German, whereas low value, as in French and Italian, seems to suggest simple syllable structures. It is interesting to see whether languages with significant vowel length distinction will have a high percentage of voiced interval, as VO is comprised of not only voiced consonants but also vowels.

Not only have debates regarding segmentation of vocalic and intervocalic intervals arisen, but alternative parameters and calculations have also been introduced. In 2000, Low, Grabe, and Nolan (2000) proposed a different calculation of vocalic and intervocalic intervals. In their previous studies (Grabe, Post, and 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 computed duration and a Pairwise Variability Index (PVI) which expresses the level of variability in successive measurements. Two versions of PVI are proposed (Grabe and Low 2002): normalized PVI (nPVI) is used with vocalic

intervals, and raw PVI (rPVI) is 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 has three successive long vowels that follow three successive short vowels, and another has long and short vowels alternating, would have the same standard deviation of vocalic interval durations although the patterns differ. 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 stresstimed 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. Thai and Tamil were compared with these languages. Thai has high %V value and the highest vocalic nPVI value among the six languages. Tamil has the highest %V value and about the same level of vocalic nPVI values as English and German. Grabe and Low, then, concluded that Thai, based on its high nPVI value, is a stress-timed language. But Tamil is not classifiable, which might be because of its high values of both parameters. It would seem, however, that these authors have not

considered that high nPVI value in Thai and Tamil is probably the result of the difference between durations of long and short vowels. The claim that a high vocalic nPVI value reflects high variability in vocalic interval durations might be true. Nevertheless, the existence of full and reduced vowels might not be the only case. Languages which have vowel length distinctions might have high vocalic nPVI values, as can be seen in Thai and Tamil. Therefore, the claim that Thai has stress-timed rhythm 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 segmental intervals to classify languages. They also discuss phonetic and phonological factors shared by groups of languages which make them different from others. Moreover, this kind of language classification is always compared with the classic rhythm class hypothesis. Models which support this hypothesis are, therefore, considered good models. Whenever unclassified or mixedrhythm languages are tested, they will be compared with the reference languages, such as English, French, and Japanese, to determine their rhythm class. Most languages studied are European languages and some major eastern languages. In this study, twelve Southeast Asian languages will be examined.

2. Method

2.1 Materials

The twelve languages analyzed in this study were selected from five language families spoken in Southeast Asia. Mon, Khmer, and Vietnamese are members of the Austroasiatic family. Malay and Cebuano are Austronesian languages. Southwestern Tai is the only major branch that is substantially represented in Southeast Asia. Standard Thai, Tai Yuan, and Southern Thai are, therefore, representatives of Tai-Kadai languages. Burmese and Sgaw Karen are members of the Tibeto-Burman family. Lastly, Green Hmong and Mien are representatives of the Hmong-Mien family. Some phonetic and phonological features are shared among certain languages. Malay and Cebuano are distinct from the other ten languages as they are polysyllabic and nontonal. In contrast, Standard Thai, Southern Thai, Tai Yuan, Vietnamese, Green Hmong, and Mien are tonal and rich in monosyllabic words. Burmese and Sgaw Karen are also tonal languages but have a large number of sesquisyllabic words. Mon and Khmer, despite the fact that they are non-tonal languages, also have a great deal of sesquisyllabic words. Therefore, they could be considered to occupy an intermediate position between Austronesian languages and the other languages in terms of tonality and number of syllables in a word. Moreover, phonation type is contrastive in certain languages. For example, phonation

contrast is found between modal, breathy, and creaky vowels in Mon. It can also cooccur with tones as found in glottalized tones in Burmese, Sgaw Karen, Vietnamese, and breathy and creaky tones in Hmong and Short Burmese. and long vowels phonologically different in Standard Thai, Tai Yuan, Southern Thai, Vietnamese (one pair), Khmer, and Mien. The other languages, nevertheless, do not have such a distinction. Lexical stress can be considered as either fixed or free. The lexical stress of the twelve languages falls on the ultimate syllable. Only Malay and Cebuano have variable lexical stress. The aforementioned phonetic and phonological characteristics of the twelve languages can be summarized in Table 1.

With some similarities and differences, these twelve 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 twelve languages shown in Table 1 have never been brought into attention before as factors which might contribute to segment timing patterns. Therefore, it is interesting to see whether these characteristics will have some effects on segment timing patterns by using the three language classification models.

Languages	Vowel length	Number of syllables in a word	Phonation contrast	Tonal/ non- tonal	Lexical stress
Standard Thai	phonemic	monosyllabic	-	tonal	fixed
Tai Yuan	phonemic	monosyllabic	-	tonal	fixed
Southern Thai	phonemic	monosyllabic	-	tonal	fixed
Vietnamese	phonemic (1 pair)	monosyllabic	glottalized and creaky tones	tonal	fixed
Mon	non-phonemic	sesquisyllabic	modal, breathy, creaky vowels	non-tonal	fixed
Khmer	phonemic	sesquisyllabic	-	non-tonal	fixed
Sgaw Karen	non-phonemic	sesquisyllabic	glottalized tones	tonal	fixed
Burmese	non-phonemic	sesquisyllabic	glottalized and creaky tones	tonal	fixed
Green Hmong	non-phonemic	monosyllabic	breathy and creaky tones	tonal	fixed
Mien	phonemic	monosyllabic	-	tonal	fixed
Malay	non-phonemic	polysyllabic	-	non-tonal	free
Cebuano	non-phonemic	polysyllabic	-	non-tonal	free

Table 1 Phonetic and	phonological	characteristics of	of the twelve	languages
	phonological	characteristics		languages

Vowel length distinction directly affects segment timing. The number of syllables in a word could play some role in timing Sesquisyllable is a syllable patterns. composed of a minor (reduced) syllable followed by a major (full form) syllable. Thus, the duration of vowels in a language with a great number of sesquisyllabic words might vary more than that of a language which has mostly monosyllabic words. Most polysyllabic languages have simple syllable structures, which mean a vowel in each syllable might be of approximately equal duration. The vowel duration of monosyllabic, sesquisyllabic, and polysyllabic languages, hence, is expected to behave differently.

All the monosyllabic and sesquisyllabic languages in this study have fixed stress falling on the last syllable. For monosyllabic words, that means all syllables should receive stress. For sesquisyllabic and disyllabic words, that means the difference between the durations of unstressed and stressed syllables in the first and second syllables must be obvious. As for polysyllabic languages which have a great number of polysyllabic words, the size of such difference might be smaller as, in a word, there is only one stressed syllable but the others are unstressed. There are a number of studies showing relations between phonation and vowel duration (Blankenship 2002; Gordon and Ladefoged 2001; Kirk, Ladefoged and Ladefoged 1993; Kirk, Ladefoged and Ladefoged 1984; Luangthongkum 1990; Samely 1991; Silverman, Blankenship, Kirk and Ladefoged 1995; Wayland, Gargash and Jongman 1994; Wayland and Jongman 2003). Therefore, phonation will surely affect segment timing. As for tones, it is hard to say whether they could affect segment timing. However, as they co-occur with different phonation types in some languages, it is interesting to see whether those languages will have similar or different timing patterns from languages which do not have phonologically contrastive phonation.

Three native speakers of each language ranging in age from 25 to 35 years old participated in the present study. The Standard Thai speakers speak Bangkok dialect as their native language. The Tai Yuan speakers speak Chiang Mai dialect. The Southern Thai speakers are all from Nakhon Sri Thammarat. The Vietnamese speakers are from Hanoi and Hai Duong and speak Hanoi dialect. The three Mon speakers are from Mudon, Myanmar. The Khmer speakers who speak Surin dialect are all from Surin Province, Thailand. The Sgaw Karen and the Burmese speakers are all from Rangoon, Myanmar. The speakers of Green Hmong are from Tak and Mae Hong Son, Thailand. The Mien speakers are from

Chiang Rai, Thailand. The Malay speakers are from Putarajaya, Kuala Lumpur, and Terengganu and speak Standard Malay in daily life. The Cebuano speakers are from Iligan City, Agusan del Sur, and Marbel, the Philippines. Cebuano is their mother tongue.

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 thirty seconds of clear speech, not including pauses and hesitations, was selected from each speaker for acoustic analysis.

2.2 Acoustic analysis

The data were segmented and labeled, using Praat software, 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 is also impossible to identify the point where a stop sound ends or begins when it occurs before and after pauses. Utterance-initial syllables were thus excluded for consistency. Segmentation was made as accurate as possible despite the fact that there is coproduction or coarticulation - overlapping in Particular articulatory movements.

measurement issues that deserved to be viewed carefully are discussed here.

Vowels were marked between the points where clear patterns of vowel formants appear, whether the acoustic excitation is 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 intervals unless vocalic there were observable changes in formant and amplitude of speech signals. In this study, pre-vocalic glides were in consonantal intervals because constriction in initial position is quite audible. Post-vocalic glides were 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 beginning from the onset to the offset of a voiced segment, or successive voiced segments were 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. 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 interval

2.3 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 (Δ V), and the standard deviation of the duration of consonantal intervals (Δ C), were used in Ramus, et al. (1999)'s model. Grabe and Low (2002) used normalized PVI (nPVI_V) with vocalic intervals, and raw PVI (rPVI C) with consonantal intervals. Durations of voiced and unvoiced intervals were used in Dellwo, et al. (2007)'s model. The three parameters were derived as the proportion of voiced intervals (%VO) and the variation coefficient of the standard deviation of unvoiced intervals (varcoUV). As Low et al. (2000) found no effect of speech rate on consonantal intervals, the same result should be found in the case of unvoiced intervals as well. Therefore, the standard deviation of the duration of unvoiced intervals (ΔUV), where the duration of unvoiced intervals is not normalized, is added in order to compare its

result with that of varcoUV where the duration of unvoiced intervals is normalized.

However, some parameters seem to share similarities. Take, for example, parameters using consonantal and unvoiced intervals. Voiceless consonants are members of these two intervals. Therefore, it might be said that ΔC , rPVI_C, ΔUV , and varcoUV represent variations of consonantal interval duration. Similarly, vowels belong to both vocalic and voiced intervals. Thus, it can be said that ΔV and nPVI V represent variations of vocalic interval duration. As mentioned earlier in §1, as these parameters believed to reflect phonological are characteristics of a language, the explanation for similar values of some parameters could be the same. For instance, high values of ΔC and rPVI_C could be a result of complex syllable structures in a language. For this reason, Sawanakunanon (2012) proposed to look at the eight parameters by using a statistical model, Principal Component Analysis (PCA), to reduce redundancy.

PCA is a statistical method of converting a set of correlated parameters into a set of values called principal components. The number of principal components will be less than or equal to the number of original parameters. The analysis was done in R and the results were statistically tested by ANOVA and followed by Tukey's HSD post-hoc test in the case of a statistically significant difference between at least one pair of languages. The values of the eight parameters of each language were taken as raw data for the analysis in this step.

3. Results

In this section, the values of the eight parameters as well as language classification by these parameters are presented in §3.1. The PCA results can be found in §3.2.

3.1 The eight parameters

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).

3.1.1 %V, Δ C, and Δ V

Table 2 presents the number of vocalic and consonantal intervals, total duration, average proportion of the duration of vocalic intervals (%V), average standard deviation of the duration of vocalic intervals (Δ V), and the average standard deviation of the duration of consonantal intervals (Δ C) across all utterances of each language.

Languages	Vocalic Consonantal Languages intervals intervals		Total Duration (Sec)	%V	ΔV	ΔC
BM	519	519	93.10	61.47	54.18	27.93
СВ	545	548	93.03	52.52	44.32	41.51
HM	491	500	96.40	54.75	56.17	35.33
KM	460	460 479		51.27	66.60	47.23
MI	481	487	93.14	57.27	59.44	37.73
ML	503	511	93.59	53.27	18.28	38.18
MN	453	462	94.55	55.43	62.02	42.00
SG	463	465	92.11	57.34	62.54	36.27
TH	486	509	93.28	49.97	56.73	42.00
TT	532	553	92.41	54.15	16.23	35.97
TY	520	533	94.22	48.23	40.76	43.64
VN	560	585	103.13	49.31	43.91	41.05

Table 2: Total number of vocalic and consonantal intervals, total duration, 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).

BM = Burmese; CB = Cebuano; HM = Hmong; KM = Khmer; MI = Mien; ML = Malay; MN = Mon; SG = Sgaw Karen; TH = Standard Thai; TT = Southern Thai; TY = Tai Yuan; VN = Vietnamese

It can be seen from Table 2 that Burmese has the highest value of %V but has the lowest value of ΔC compared to other languages. According to the background of Ramus et al. (1999), a language with a high %V value and a low ΔC value has simpler syllable structure. For Burmese in this case, however, this explanation is doubtful since some languages, like Malay or Cebuano, have simple syllable structures. also Another interesting thing is that Khmer has the highest values for both ΔC and ΔV . The two parameters capture the variability of consonantal and vocalic intervals which could be the results of complex syllable structures. Using ANOVA in statistical tests, it is found that the values of all three parameters are significantly different with p < 0.05.

2 Figure demonstrates language classification by plotting the graph with the values of %V, ΔV , and ΔC . The graph plotted on the %V and ΔC plane shows language classification the best according to Ramus et al. (1999). In this study, it shows roughly three groups of languages. Tai Yuan, Vietnamese, and Standard Thai are clustered together with high ΔC values but low %V values. Malay, Southern Thai, and Hmong have lower values of ΔC but higher values of %V. Mon and Sgaw Karen have ΔC values in the same range as Malay, Southern Thai, and Hmong, but have higher %V values. Khmer, Mon, and Burmese, however, are not clearly grouped with other languages. The other two graphs illustrate more scattered patterns and, thus, do not show language groupings as clearly as the $\%V-\Delta C$ graph does. Nevertheless, it should be noted that Khmer and Burmese are not grouped with other languages on all three graphs.



Figure 2: Distribution of languages over the %V and ΔC plane (top left), %V and ΔV plane (top right), and ΔV and ΔC (bottom left). Error bar represents ±1 S.D.

3.1.2 PVI Results

PVI measurement is another method intended to show the variability of interval duration. (Grabe and Low 2002) However, it focuses on the difference between the duration of two successive intervals. Accordingly, PVI value represents the variability of duration of adjacent intervals not the variability in an utterance. The raw pairwise variability index (rPVI) is used with consonantal intervals as stated in §1. In this paper, it is referred to as rPVI_C for readability and can be computed by using the following formula:

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

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

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

The values of nPVI_V and rPVI_C of all speakers of the twelve languages are presented in Table 3.

rPVI C Languages nPVI V BM 32.09 52.37 CB 40.32 41.91 HM 42.58 50.17 KM 53.01 70.72 MI 39.72 57.74 40.80 48.59 ML MN 46.42 62.14 SG 40.72 56.78 TH 47.00 54.63 TT 37.45 50.21 TY 45.57 48.71 VN 44.27 53.71

Table 3: rPVI_C and nPVI_V values

BM = Burmese; CB = Cebuano; HM = Hmong; KM = Khmer; MI = Mien; ML = Malay; MN = Mon; SG = Sgaw Karen; TH = Thai; TT = Southern Thai; TY = Tai Yuan; VN = Vietnamese

The value of rPVI C in Khmer is the highest and is much higher than for the other 11 languages. High rPVI_C value represents more variability in two successive consonantal intervals. Burmese has the lowest rPVI_C value while those of the other ten languages do not differ much. The value of nPVI_V can be interpreted in the same way. The higher nPVI_V value suggests that two adjacent vocalic intervals in Khmer and Mon have greater variability than in the other ten languages. The language with the lowest value of nPVI V is Cebuano. This could be due to the fact that its syllable structure is simple most of the time. The ANOVA tests show significant differences on both parameters in this model (p < 0.05).

Figure 3 shows the twelve languages plotted on the rPVI_C and nPVI_V planes. Most languages are clustered in the middle of the graph with not so high or so low values of rPVI C but more on the lower values of nPVI V. The language distribution in Figure 3 agrees with all three graphs in §3.1.1 following Ramus et al. (1999) in several ways. For example, Khmer and Burmese are not grouped with other languages; Mien and Sgaw Karen are always grouped together, and Southern Thai and Malay are clustered together in all four graphs. This agreement could be because the five parameters used in both the Ramus et al. (1999) and Grabe and Low (2002), models are computed from the durations of consonantal and vocalic intervals.

3.1.3 %VO, varcoUV, and ΔUV

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 the duration and proportion of voiced intervals (%VO). Unvoiced intervals are only composed of voiceless consonants. The larger the number of sequential unvoiced segments, the longer the unvoiced intervals. Languages which have a wider range of number of onsets and codas could have higher variability in unvoiced interval

durations (varcoUV). VarcoUV is also normalized to reduce the effect of different speech rate. As Grabe and Low (2002) found, speech rate does not affect consonantal duration to the extent that it could affect the values of rPVI_C. Therefore, they propose that consonantal interval duration is not necessary for normalization. Consequently, Δ UV, the standard deviation of consonantal interval duration which is the unnormalized version of varcoUV, is included in this section.

The total number of voiced and unvoiced intervals, the proportion of the duration of voiced intervals (%VO), the average duration of voiced intervals (meanVO), the variation coefficient of the standard deviation of unvoiced intervals (varcoUV), the standard variation of unvoiced intervals (Δ UV), and the average duration of unvoiced intervals (meanUV) are presented in Table 4.



rPVI_C = raw pairwise variability index in consonantal interval duration nPVI_V = normalized pairwise variability index in vocalic interval duration

BM = Burmese; CB = Cebuano; HM = Hmong; KM = Khmer; MI = Mien; ML = Malay; MN = Mon; SG = Sgaw Karen; TH = Thai; TT = Southern Thai; TY = Tai Yuan; VN = Vietnamese

Figure 3: rPVI	C and nPVI V	values across all languages.	Error bar represents ± 1 S.D.
		<i>() ()</i>	

Languages Voiced intervals		Voiced intervals	Unvoiced intervals	%VO	varcoUV	ΔUV
-	BM	268	253	78.87	34.63	26.87
	CB	311	294	75.68	42.17	32.93
	HM	333	319	71.20	33.09	28.95
	KM	312	305	72.20	34.75	32.64
	MI	283	275	77.97	29.93	22.39
	ML	238	221	79.73	34.75	29.01
	MN	324	308	74.79	41.47	33.79
	SG	281	264	72.34	35.97	34.05
	TH	336	323	73.03	44.59	35.16
	TT	314	300	78.76	39.99	25.94
	TY	320	312	73.20	42.21	34.43
	VN	324	311	75.44	37.34	30.41

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

BM = Burmese; CB = Cebuano; HM = Hmong; KM = Khmer; MI = Mien; ML = Malay; MN = Mon; SG = Sgaw Karen; TH = Standard Thai; TT = Southern Thai; TY = Tai Yuan; VN = Vietnamese



%VO = proportion of voiced interval; varcoUV = variation coefficient of the standard deviation of unvoiced interval; ΔUV = S.D. of unvoiced interval; BM = Burmese; CB = Cebuano; HM = Hmong; KM = Khmer; MI = Mien; ML = Malay; MN = Mon; SG = Sgaw Karen; TH = Thai; TT = Southern Thai; TY = Tai Yuan; VN = Vietnamese

Figure 4: Distribution of languages over the %VO and varcoUV plane (left) and the %VO and Δ UV plane (right). Error bar represents ±1

From Table 4, it can be seen that Malay has the highest values of %VO followed by Burmese, Southern Thai, and Mien. Burmese and Mien are among the top four languages which have high %V values as seen in §3.1.1. The value of varcoUV is based on the standard deviation of unvoiced intervals (Δ UV) and the average duration of unvoiced intervals (meanUV). High Δ UV and low meanUV will result in high varcoUV, and vice versa. VarcoUV values which show durational variation of unvoiced intervals are highest in Standard Thai and lowest in Mien. The results of varcoUV and Δ UV are quite similar and confirm that it might not be necessary to normalize consonantal duration since the effect is so small. The ANOVA test found that there were significant differences in the means of the values of %VO, varcoUV, and Δ UV among the twelve languages (p < .05). Although the values of varcoUV and Δ UV do not differ much, the graph plotted on the %VO and Δ UV plane seems to exhibit clearer language distribution, i.e. the twelve languages are shown in two groups. Malay, Burmese, Southern Thai, and Mien are grouped together with higher values of %VO and lower values of Δ UV while the other languages are clustered with lower values of %VO and higher values of Δ UV.

Compared to language distributions in §3.1.1 and 3.1.2, it is noticeable that Burmese and Khmer are not separated from other languages as found in graphs following the Ramus et al. (1999) and Grabe and Low (2002) models. It could be the result of a different method in sound segmentation.

It might be said that the results of the eight parameters in §3.1.1-3.1.3 as shown above are not consistent and it is not easy to judge which model or parameters are best in classifying languages according to their segment timing behavior. Also, some phonetic and phonological characteristics might distribute to several parameters to different degrees. Hence, considering the parameters and the phonetic and phonological properties of the languages as a whole might give clearer picture of how the parameters and such properties are relevant. In the next section, a statistical technique called Principal Component Analysis or PCA will be applied to reduce redundancy in interpreting the results of each parameter.

3.2 Principal Component Analysis (PCA)

The eight parameters in this study can be classified into two groups: 1) consonantal parameters and 2) vocalic parameters. The four parameters which fall into consonantal parameters are ΔC , rPVI C, varcoUV, and ΔUV . All four parameters show a variability of consonantal interval duration. The other four parameters, %V, ΔV , nPVI V, and could be considered vocalic %VO, parameters. They can be further divided into parameters which show the proportion of vocalic intervals and those which show a variability of vocalic interval duration. The similarities among parameters as seen here can be captured by the Principal Component Analysis (PCA). The analysis in the section was done using R, a statistical computing software. The results will then be discussed in relation to phonetic and phonological properties of the twelve languages.

The PCA result shows that the values of the eight parameters can be converted into three principal components as shown in the PCA matrix in Table 5.

PC	PC1	PC2	PC3
	ΔC	nPVI_V	%V
Parameters	ΔUV	ΔV	%VO
	rPVI_C		
	varcoUV		

Table 5: Members of each principal component

PC1 = Principal Component 1; PC2 = Principal Component 2; PC3 = Principal Component 3; ΔC = S.D. of consonantal interval duration; ΔUV = S.D. of unvoiced interval duration; rPVI_C = raw pairwise variability index of consonantal interval duration; varcoUV = variation coefficient of unvoiced interval duration; nPVI_V = normalized pairwise variability index of vocalic interval duration; ΔV = S.D. of vocalic interval duration; % V = proportion of vocalic interval; % VO = proportion of voiced interval

It can be seen, in Table 5, that PC1 is composed of four parameters, i.e. ΔC , ΔUV , rPVI_C, and varcoUV. In other words, PC1 represents the durational variation of consonantal and unvoiced intervals. PC2 represents the variation of vocalic interval duration as it is comprised of nPVI_V and ΔV . PC3, which consists of %V and VO, shows the proportion of the duration of the vocalic and voiced intervals.

The factor score for the PC of each language was then computed and statistically tested with an ANOVA test. It is noted, however, that factor score values could be in different directions from the raw values of the eight parameters. That is, for example, low values of ΔV and nPVI_V do not guarantee a low PC2 value.

Statistically significant differences are found among languages in all three principal components (p < .05). Tukey's HSD posthoc test was then analyzed to find which language pairs are different at the .05 level of significance. It is found that, for PC1, Mien, Malay, and Vietnamese are significantly different from Mon and Sgaw Karen.

As for PC2, 19 out of 62 pairs of languages are significantly different. The value of PC2 in Khmer is different from that of most languages. It is significantly different from Sgaw Karen, Standard Thai, Burmese, Hmong, Malay, Vietnamese, Southern Thai, Tai Yuan, and Cebuano. Mon is found different from Malay, Vietnamese, Standard Thai, Tai Yuan, and Cebuano. Mien is significantly different from Standard Thai, Tai Yuan, and Cebuano. Lastly, Sgaw Karen is found to be significantly different from Tai Yuan and Cebuano. For PC3, Cebuano, Tai Yuan, and Standard Thai are shown to be significantly different from Mien and Hmong. To sum up, the results of the statistical tests on principal components show that PC2 seems to be the most different component while PC1 and PC3 are more similar among the twelve languages. Since there were significant differences in all three principal components, the PCs1-3values for all languages were plotted onto a 3D-plane in order to show language classification as displayed in Figure 5.



PC1 = Principal Component 1; PC2 = Principal Component 2; PC3 = Principal Component 3; BM = Burmese; CB = Cebuano; HM = Hmong; KM = Khmer; MI = Mien; ML = Malay; MN = Mon; SG = Sgaw Karen; TH = Thai; TT = Southern Thai; TY = Tai Yuan; VN = Vietnamese

Figure 5: Distribution of languages on PC1, PC2, and PC3 plane

Considering the values of all three PCs in Figure 5, it can be seen that the twelve languages could be classified into four groups. Starting from the leftmost of the graph, Mon and Khmer are grouped together with high values of PC1 and the lowest values of PC2. PC3 values of both languages are neither high nor low. Burmese and Hmong are grouped together in the back of the graph with relatively high values of PC1 and PC3 while PC2 values are in the middle range. At the bottom right of the graph, Vietnamese, Tai Yuan, and Southern Thai are clustered together with low values of PC1, high values of PC2, and PC3 values in the middle range. Right at the corner where the PC2 and PC3 axes cross, Malay and Cebuano are together with low values of PC1 and PC3 and a high value of PC2.

Figure 5 also shows that Standard Thai is not grouped with other languages with mid value of PC1, higher value of PC2, and low value of PC3. Sgaw Karen and Mien are close in the back of the graph with higher values of PC2 and PC3. However, their PC1 values are very different as Sgaw Karen has the highest value of PC1 while that of Mien is the lowest. Therefore, they are not really clustered together.

To sum up, the PCA results clarify language into four groups, which are 1) Mon-Khmer 2) Burmese-Hmong 3) Malay-Cebuano and 4) Vietnamese, Tai Yuan, and Southern Thai. Standard Thai, Sgaw Karen, and Mien, however, are not clearly clustered with other languages.

4. Discussion

The results of the three models in the analysis in this study as shown in §3 do not reflect language classification as clearly as in Ramus et al. (1999), Grabe and Low (2002), or Dellwo et al. (2007). It is possible that some languages in the three models have very similar phonetic and phonological properties while some are very different. For example, English has a greater variety of syllable structures and its most frequently occurring syllable type is more complex than that of French and Spanish (Dauer 1983). On the contrary, the differences in syllable structures of the twelve languages in this study might not be that large. Some languages, however, are always grouped together as seen in §3. Therefore, it can be implied that those languages have some phonetic and phonological characteristics in common. Likewise, some languages which are not clustered with other languages might have different phonetic and phonological properties. Considering these properties of the languages would help understand the distribution of languages better. In this section, the relations between segment timing and phonetic and phonological properties as well as common features shared among languages in the same group will be discussed.

4.1 Segment timing and phonetic and phonological characteristics

With Principal Component Analysis, we have seen that correlated parameters can be considered as a group of parameters called principal components. The three principal components were examined to see which phonetic and phonological characteristics might affect the values of relevant parameters. The relations between possible characteristics and each parameter are shown in Table 5.

From Table 5, PC1, which is composed of ΔC , ΔUV , rPVI C, and varcoUV, represents durational variation of consonantal and unvoiced intervals. The values of PC1 might be affected by syllable structure complexity, phonations of consonants, and manners of articulation. Complex syllable structures which allow numbers of consonants to occur consecutively can result in high variation in consonantal interval durations. On the other hand, languages which have simpler syllable structures such as Malay and Cebuano show lower variation as their consonantal intervals generally contain only one consonant. Thus, it might be concluded that languages which allow more types of syllable structures are likely to have higher values of PC1 than languages with simpler syllable structures. Manner of articulation and phonation of consonants could also influence the parameters in PC1. A language with many stop or fricative consonants could have higher durational variation of consonantal or unvoiced intervals than a language with less various types of consonants.

Although varcoUV and Δ UV seem to be similar, they differ in that Δ UV is normalized to reduce the effect of speech rate. As the values of both parameters do not agree in all twelve languages, it is possible that speech rate influences certain types of sounds more than others. Since all consonants in unvoiced intervals are voiceless, the differences in phonation of consonants could be a reason behind this phenomenon. However, the durations of consonantal and unvoiced intervals in all languages in this study might not be very different since there are only six pairs of languages which are significantly different by the values of PC1 as mentioned in §3.2. This might be explained by the fact that the twelve languages are similar in terms of consonantal sounds in their phonological inventories or the data analyzed from each language contain many consonantal sounds of the same types.

PC2 exhibits durational variation of vocalic intervals through nPVI_V and ΔV . As both parameters are based on vowel duration, the existence or absence of vowel length distinctions is a factor that will definitely influence their values. A language whose short and long vowels are contrastive phonemically is expected to have a higher durational variation of vowels and, of course, vocalic intervals. On the other hand, vowel duration in a language with no vowel distinctions might not vary much. Thus, lower variation of vowel duration is expected. Syllable structure complexity and variation could affect the durations of vocalic intervals as well. A language which has more complex and various types of syllable structures is likely to have greater

durational variation of vocalic intervals than a language which has simpler and less various syllable structures. This study supports the findings of Ramus et al. (1999), Grabe and Low (2002), and Dellwo et al. (2007) as illustrated in Figure 5, which Malay and Cebuano, which have simple syllable structures, are grouped together with similar values of PC2.

Parameters	PC1			PC2		PC3		
Phonetic and phonological characteristics	ΔC	ΔUV	rPVI_C	varcoUV	Λ [−] ΙΛdu	۸V	Λ%	%V0
syllable structure complexity								
lexical stress								
number of syllables in a word								
vowel length distinction								
phonation of vowels								
phonation of consonants								
segment's sonority								
manner of articulation								
tone								
tone with phonation								

Table 6 Parameter effects of phonetic and phonological characteristics

According to the ANOVA test on PC2 stated in §3.2, Khmer, Mon, and Mien are significantly different from Standard Thai, Burmese, Hmong, Malay, Vietnamese, Southern Thai, Tai Yuan, and Cebuano. Let's consider the members of PC2 to find which phonetic and phonological characteristics could explain the high or low values of the two parameters.

As shown in Table 2, it can be seen that Khmer, Sgaw Karen, and Mon are the three languages having the highest values of ΔV . Sesquisyllabic words, which are in large numbers in the three languages, might be a reason behind this phenomenon since the greater variation of the durations of the vocalic intervals could be due to the differences between unstressed and stressed vowels in the first and second syllable of a sesquisyllabic word. For Khmer, vowel length distinction could be another factor for the high value of ΔV . The difference between breathy and modal vowels could contribute to the high value of ΔV in Mon. Phonation types affect the ΔV value in Sgaw Karen in the same way as there are breathy, modal, creaky, and glottalized tones. There are quite a number of studies, as mentioned previously, which found that phonation types affect vocalic durations. Hence, they could lead to greater durational variation of vocalic intervals. nPIV_V, another member of PC2, is found lowest in Malay and Cebuano. The fact that these two languages have many polysyllabic words might be used to explain the low values of nPIV_V in

Malay and Cebuano. The differences between vowel durations in successive syllables are smaller than those of sesquisyllabic languages like Mon and Khmer. Another common characteristic that Malay and Cebuano share is that they have free lexical stress while the other languages have fixed lexical stress. Thus, there might not be a clear pattern of the durations of stressed and unstressed syllables in languages with free lexical stress.

For PC2, it can be concluded that the values of ΔV and nPVI_V are affected by vowel distinctions, syllable structure length complexity, lexical stress, and phonation types. The ANOVA test in §3.2, which found that 19 out of 62 language pairs are significantly different, shows that there are more differences among languages in vocalic durations as seen in PC2 than in consonantal and unvoiced interval durations as shown in PC1. That the twelve languages in this study are more different in vocalic sounds than in consonantal sounds may explain this.

The last principal component, PC3, represents proportions of vocalic and voiced intervals in utterances. The two parameters that fall into PC3 are %V and %VO. The phonetic and phonological characteristics which could influence the values of the two parameters are syllable structure complexity, phonation types, vowel length distinctions, and a segment's sonority.

Syllable structure has a direct influence on %V and %VO as the occurrences of vowels and consonants will surely be different. For example, a language with simpler syllable structure would have about the same number of vowels and consonants. As a result, the values of %V and %VO are likely to be higher than those of a language with more complex syllable structures in which vowels occur less frequently than consonants. However, as voiced consonants are also included in voiced intervals (VO), a language which allows sonorants, i.e. semivowels (/w, j/) or liquids (/l, r/), as the second element in initial clusters could result in high %VO value.

Phonation types also play some role in the value of %V. Languages with distinctions in phonation types of vowels, like Mon, or tones, like Burmese, Sgaw Karen, and Hmong, are found to have higher values of %V than languages with no phonation type distinctions like Malay and Cebuano. It is interesting to note, however, that Vietnamese has a lower value of %V even though it has phonation tones.

Vowel length distinctions might be another factor that affects the value of %V. Grabe and Low (2002) found that Thai and Tamil, languages with vowel length distinctions, have higher %V values than languages which do not have vowel length distinctions. The results in this study both support and oppose Grabe and Low (2002)'s findings. It is found that the %V value of Southern Thai, a language which has vowel length distinctions, is higher than that of Malay and Cebuano which do not have vowel length distinctions. However, Khmer, Standard Thai, Tai Yuan, and Vietnamese, unexpectedly have lower values of %V than Malay and Cebuano.

The results of %V and %VO are sometimes in the same direction. For example, Burmese has the highest %V value and the second highest %VO value. On the contrary, they could be in the opposite direction as can be seen from the case of Sgaw Karen. It has the second highest %V value but has a very low %VO value as it ranks tenth among the twelve languages. This finding suggests that voiced intervals play an important role as they can portray some information about consonants. A language which has a high occurrence of voiceless consonants, especially ones which are stops, affricates, and fricatives, or has many of them in its sound system, is likely to have a lower %VO value than a language which has high occurrence of voiced consonants, semivowels, or liquids. Thus, it might be suggested that sonority of consonants could also affect the value of %VO.

The ANOVA test on PC3, which shows that there are only six pairs of languages which are significantly different, implies that the durations of vocalic and voiced intervals in the twelve languages are not very different when considered in terms of proportion in an utterance. The different phonetic and phonological characteristics such as phonation types and vowel length distinctions, thus, seem to affect the durational variation of vocalic and voiced intervals as shown in PC2 more than the proportion of these intervals in utterances.

In conclusion, the results of PCA, which found that the twelve languages in this study are more different in the durational variation of vocalic intervals, suggests that the twelve languages are more different in terms of vowels, especially in their durational variation.

4.2 Common phonetic and phonological characteristics among languages

As mentioned in §1, the assumption behind the three models of rhythmic analysis is that the proposed parameters could reflect phonetic and phonological characteristics of different types of speech rhythm. Moreover, language groupings in a graph plotted with values of certain parameters would reflect language groupings according to their rhythm types, i.e. stress-timed, syllabletimed, or mora-timed languages. The model of Ramus et al. (1999) seems to show best a speech rhythm continuum with stress-timed languages on one end, a mora-timed language on the other end and stress-timed languages in between. This finding supports Dauer's (1987) suggestion that speech

rhythm should be considered as a continuum instead of separate groups.

The PCA result in this study, however, is shown on a 3D graph plotted with the values of PC1, PC2, and PC3, it cannot be presented as a straight line on a continuum as in Ramus et al. (1999). Nevertheless, it can be seen in Figure 5 that there are four groups of languages in three corners of the graph in a circular fashion. Group 1, which is composed of Mon and Khmer, is in the bottom left of the graph. Burmese and Hmong are in Group 2 and placed in the upper right. Malay and Cebuano are in the bottom right of the graph. This group will be referred to as Group 4 henceforth. Vietnamese, Tai Yuan, and Southern Thai belong to Group 3 and are located between Groups 2 and 4. Sgaw Karen and Mien at the top of the graph are not clearly grouped with other languages but can be found between Groups 1 and 2. Standard Thai is another language which is not obviously grouped with other languages but can be found between Groups 1 and 3-4. The properties of languages in each group will be discussed in this section.

Languages in the same group are expected to have some common phonetic and phonological characteristics. However, it is not always the case. Take Group 1 as an example. Although Mon and Khmer are both sesquisyllabic languages, they are different in that vowel length distinctions exist in Khmer but not in Mon. There is also a difference in phonation types in Mon vowels but not in Khmer. Even though all three characteristics affect the durations of vocalic intervals as explained in §4.1, the characteristic which seems to bring Mon and Khmer into the same group is that they have many sesquisyllabic words. Thus, it might be said that the main characteristic of languages in Group 1 is that they are sesquisyllabic languages.

Now let's turn to Group 2. Burmese is also a sesquisyllabic language; however, it is grouped with Hmong, which is а monosyllabic language. Stress pattern, hence, could not be a common characteristic shared by these two languages. However, both Burmese and Hmong have some common characteristics. They do not have length. phonemic vowel And more importantly, they are tonal languages. This could be the main characteristic which differentiates them from the non-tonal sesquisyllabic languages in Group 1. It is worth noting that some tones co-occur with phonation types. It is noticeable that Sgaw Karen, located between Groups 1 and 2, has dominant characteristics from both groups as it is a sesquisyllabic language with phonation-type tones.

It can be seen that Group 3 and 4 languages are close to each other on the graph. Nonetheless, they have different phonetic and phonological properties. The syllable structures of the languages in Group 3 are more complex than those of the languages in Group 4. Lexical stress patterns in the two groups are also different since Group 3 languages are monosyllabic while Group 4 languages are polysyllabic. Vowel length is phonologically different in the languages in Group 3 but not in Group 4. Lastly, the languages in Group 3 are tonal but those in Group 4 are non-tonal.

Although Vietnamese, one of the languages in Group 3, has phonation-type tones similar to the main characteristic of the languages in Group 2, its PC3 score, which shows the proportion of vocalic and voiced intervals, is different from that of Group 3 languages. However, since phonation-type tones do not exist in Southern Thai and Tai Yuan, the other two members in Group 3, it can be said that the languages in this group are tonal languages without phonation-type cooccurrence.

The languages in Group 4 are all polysyllabic languages. They are different from the other groups in several ways. They are non-tonal languages, have no distinction of phonation types on vowels, and have no vowel length distinctions.



BM = Burmese; CB = Cebuano; HM = Hmong; KM = Khmer; MI = Mien; ML = Malay; MN = Mon; SG = Sgaw Karen; TH = Thai; TT = Southern Thai; TY = Tai Yuan; VN = Vietnamese

Figure 6: Common characteristics of each group of languages

Standard Thai, which is placed between the languages in Group 1 and Group 3, share some characteristics with them. It is a tonal language with no phonation-type tones similar to Group 3, and has vowel length distinction just like Group 3 and Khmer, which is a member of Group 1. Numerous sesquisyllabic words are also found in Standard Thai. Thus, the stress pattern in those words will be the same as in Mon and Khmer.

As for Mien, although it is a tonal monosyllabic language and also has vowel length distinction, it is not grouped with the languages in Group 3 with its lower PC3 score. This could be due to the difference in consonantal sounds as Mien has a higher occurrence of fricatives and affricates, which could affect the durations of consonantal and unvoiced intervals.

The twelve languages as well as their dominant phonetic and phonological characteristics which cluster some languages together and make them different from other groups can be portrayed in Figure 6. The phonetic and phonological characteristics shown in Figure 6 had never been considered as factors that correlated with speech rhythm in past research. This could be because of the languages analyzed in previous works were mostly languages spoken in Europe which lack these characteristics. The results in this study, however, illustrated that these characteristics may have some contribution to speech rhythm.

5. Conclusion

This study aims to analyze segment timing in twelve Southeast Asian languages, namely, Standard Thai, Southern Thai, Tai Yuan, Mon, Khmer, Vietnamese, Burmese, Sgaw Karen, Malay, Cebuano, Green Hmong, and Mien. The speech data were segmented into vocalic, consonantal, voiced, unvoiced intervals. The interval and durations were then measured and converted into the eight parameters following the previous work of Ramus et al. (1999), Grabe and Low (2002), and Dellwo et al. (2007). The eight parameters are %V, ΔV , ΔC , nPVI_V, rPVI_C, %VO, varcoUV, and However, some parameters are $\Delta UV.$ similar in their concepts and yield similar results. Therefore, a Principal Component Analysis was done to group similar parameters together into three principal PC1 shows the components or PCs. durational variation of consonantal and unvoiced intervals. Likewise, PC2 exhibits the durational variation of vocalic intervals. PC3 represents the proportion of vocalic and voiced intervals. It is found that the twelve languages are more different on PC2, the durational variation of vocalic intervals.

The values of the three PCs were then plotted on a 3-axes graph.

The graph shows language classification according to the phonetic and phonological characteristics they share. The three dominant characteristics which seem to distinguish the twelve languages into groups are the number of syllables in a word, tones, and phonation types. First, the number of syllables in a word divides the twelve languages into three groups, i.e. monosyllabic, sesquisyllabic, and polysyllabic languages. The lexical stress patterns in these three language groups are different. As a result, the durations of vocalic intervals vary to different degrees. Second, tones divide the twelve languages into two groups, that is whether or not they are tonal or non-tonal languages. Last, phonation types also play a role. They occur with vowels in Mon and co-occur with tones in several languages. Some languages have all three characteristics, but some lack one two characteristics. The or different combinations can be put on a continuum as shown in Figure 6.

It is noted that even though the original results of the three rhythmic models link language classification to speech rhythm types as discussed in §1, such a conclusion cannot be claimed in this study. English, French, and Japanese are usually used as points of reference for each type of speech rhythm. However, they are not included in this research. So it is impossible to make a conclusion whether the continuum in this study could be considered as the rhythm continuum proposed by Dauer (1987). The language classification in this study, as a result, can only be claimed to the extent that languages are grouped together with similar behaviors in segment timing which are the of different phonetic results and phonological characteristics. Comparing the values of each parameter with other studies would not make either since the nature of the data is different. The data in previous works are in reading style while they are spontaneous speech in this study. Further study on the three reference languages using spontaneous speech is suggested to see what type of speech rhythm the languages in each group in this study represent.

Another important point to be made is that the discussion on the relations between each and phonetic parameter some and phonological characteristics in §4.1 could be better with statistical support as done in Easterday, Timm, and Maddieson (2011). They studied the influence of some phonological characteristics on certain rhythmic parameters as well as the relations between each parameter by doing correlation analysis. They found in their study that %V and ΔV are correlated in the same direction. The result in this study also supports this finding as the value of ΔV is high when the value of %V is high. Thus, the correlation analysis of phonetic and

phonological characteristics and rhythmic parameters is suggested for further study.

6. 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 36 speakers of the twelve languages. 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.
- Crystal, D. 2008. A dictionary of linguistics and phonetics (6th ed.). Oxford, UK: Blackwell Publishing.
- Dauer, R. M. 1983. Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics* 11:51-62.
- Dauer, R. M. (1987, August 1-7, 1987). Phonetic and Phonological Components of Language Rhythm. Paper presented at the 11th International Congress of Phonetic Sciences 5, Tallinn, Estonia, USSR.
- 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.

- Easterday, S., Timm, J., & Maddieson, I. (2011, August 17-21, 2011). The effects of phonological structure on the acoustic correlates of rhythm. Paper presented at the *17th International Congress of Phonetic Sciences (ICPhS 17)*, Hong Kong.
- 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.
- Gordon, M., & Ladefoged, P. 2001. Phonation types: A cross-linguistic overview. *Journal of Phonetics*, 29.4:383-406.
- Grabe, E., & Low, E. L. 2002.
 Durational variability in speech and the rhythm class hypothesis. In C.
 Gussenhoven & N. Warner (Eds.), *Laboratory Phonology* 7:515-546. 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.
- Kirk, P. L., Ladefoged, J., & Ladefoged,P. 1993. Quantifying acoustic properties of modal, breathy and creaky vowels in Jalapa Mazatec. In

American Indian linguistics and ethnography in honor of Laurence C. Thompson, edited by A. Mattina & T. Montler. Missoula, MT: University of Montana Press.

- 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. 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 The *Cognitive Neurosciences*, edited by M.S. Gazzaniga, 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 *Logical issues in language acquisition*, edited by I. M. Roca, 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.
- Samely, U. 1991. Kedang (Eastern Indonesia), some aspects of its grammar. Hamburg: Helmut Buske Verlag.
- Sawanakunanon, Y. 2012. Segment Timing in 12 Southeast Asian Languages: Implications for Typological Classification. Ph.D. dissertation, Chulalongkorn University, Bangkok, Thailand.

- Silverman, D., Blankenship, B., Kirk, P., & Ladefoged, P. 1995. Phonetic structures in Jalapa Mazatec. *Anthropological Linguistics* 37: 70 88.
- Steiner, I. 2003. A refined acoustic analysis of speech rhythm. Paper presented at the *LingColl 2003*, Budapest, Hungary.
- Trask, R. L. 1996. *Dictionary of Phonetics and Phonology*. Londong and New York: Taylor & Francis Routledge.
- Wayland, R., Gargash, S., & Jongman, A. 1994. Acoustic and perceptual investigation of breathy voice. *Journal of the Acoustical Society of America*, 97:33-64.
- Wayland, R., & Jongman, A. 2003. Acoustic correlates of breathy and clear vowels: the case of Khmer. *Journal of Phonetics* 31.2:181-201.