Instrumental Phonetic Study of the Rhythm of Malay

Wan Aslynn Salwani Wan Ahmad

Submitted in accordance with the requirements for the degree of

Doctor of Philosophy.

School of Education, Communication and Language Sciences
Newcastle University
October 2012

The candidate confirms that the work submitted is her own and that appropriate credit has been given where reference has been made to the work of others

Acknowledgement

First and foremost I would like to express my gratitude to my supervisors: Gerry Docherty and Katya Samoylova for their guidance. Without them I'd be lost. Their wisdom and professionalism are forever admired.

My appreciation extends to my employers: the Kulliyyah of Allied Health Sciences, International Islamic University and the Ministry of Higher Education Malaysia for making it financially possible for me to complete this thesis.

Special thanks to Stefanie Pillai and Jennifer Lee of the Faculty of Languages and Linguistics, Malaya University, Kuala Lumpur for their endless assistance and support during the data collection. I am also grateful to the faculty for letting me use their recording facilities during the period of my study. I hope that this is the beginning of productive research collaboration between us.

I appreciate the contributions from all the twenty-three subjects for both the experiments, for their 'voices' give so much insights into Malay speech.

This dedication is for the members of the Newcastle University phonetic research group. Their passion and enthusiasm strongly sealed my penchant for phonetics. They are friends for life.

To all my lovely friends thank you so much for the emotional support and encouragement especially when the goings were tough. Your friendships are eternally treasured.

The last but not the very least I'd like to convey my undying gratitude to my family: Mama, Abah, Adik and Luqman. Their love is felt through and through. The confidence that they have in me is greatly uplifting.

This is the steepest learning curve and a humbling experience but it is definitely worth it! Thank you Allah.

Pisang emas dibawa belayar;
Masak sebiji di atas peti;
Hutang emas boleh dibayar,
Hutang budi dibawa mati

Abstract

This thesis investigates the phonetic bases of the rhythmic properties of Malay speech. The aims of this research are twofold. First, it seeks to shed light on the rhythmic properties of a language that has not been analysed extensively for its rhythmic characteristics. In doing so, results are reported from the quantitative analyses based on a range of rhythm metrics (ΔC, %V, ΔV, *r*PVI and *n*PVI) derived from the duration consonantal and vocalic intervals within the speech signal. The second objective of the thesis is to test the validity of these rhythm metrics as a means of capturing the rhythmic properties of a language.

As an initial stage in the investigation of the rhythmic characteristics of Malay, short study is reported of the phonetic correlates of stress in Malay. Stress is a concept closely related to rhythm, and the way in which stressed syllables are located and realised can contribute significantly the rhythmic characteristics of the language. A list of 12 words was produced in isolation and in sentences by three native speakers of the standard variety of Malay. The words were divided into three categories: disyllabic monomorphemic words, polysyllabic monomorphemic words and polysyllabic morphologically complex words. The results reveal that in words produced in isolation, duration was found to correlate with stress; word-final vowels were significantly longer than vowels in other positions for both disyllabic and polysyllabic monomorphemic word types, whilst, the penultimate vowel was significantly longer than the other vowels in morphologically complex words. However, in

connected speech, it was found that these durational differences were not present.

In the subsequent acoustic phonetic investigation of the rhythmic properties of Malay, the material for analysis comprised recordings of 10 sentences produced by 20 speakers of standard Malay (10 males and 10 females). The recordings were first analysed using the rhythm metrics proposed by Ramus et. al (1999) and Grabe & Low (2002). The results indicated that Malay clustered with other so-called syllable-timed languages like French and Spanish on the basis of all metrics. However, notwithstanding the overall findings for these metrics, the results revealed a large degree of variability in values across speakers and sentences, with some speakers having values in the range typical for stressed-timed languages like English. In light of this variability, further analysis was carried in order to identify any other features of the signal which might be more robust to cross-speaker and cross-sentence variability, and which might therefore offer an alternative basis for capturing the rhythmic properties of Malay. Spectrographic analysis revealed that the duration of single vowels displayed a high degree of consistency and regularity for all speakers and sentences. In order to test this finding further, the same analysis was carried out on the data from the first experiment on the properties of stress, producing a result which was consistent with that of the second experiment. The results are discussed in light of recent studies critiquing the current state of our understanding of the rhythmic properties of speech

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

One of the central objectives of phonetic research is to understand the ways in which languages differ in respect of their sound patterning and phonetic realisation. The principal focus of this research is the rhythmic properties of speech production, with a particular focus on those which lend malay its particular rhythmic characteristics. However, rhythmic behaviour is of course not restricted to speech, since it is part of the underlying structure of most actions that involve motion. Evidence of rhythmicity can be clearly seen in both the natural environment, such as the sound of a wave on the beach, and in human activities, like the sound of the ball being bounced.

Desdain and Windsor (2000) describe rhythm as a succession of strong and weak elements that occur in a repetitive structure. Rhythm is formed by a combination of two or more different types of interval. These combinations can consist of any sequence; for example, finger tapping is usually made up of a tapping interval followed by a silence interval, and these two intervals will be repeated over a period of time. Rhythm has a strong temporal dimension, since the repetition of intervals can be perceived at a constant rate (Ouelette and Tardiff, 1996). However, timing and rhythm are not one and the same thing. Desdain and Windsor (2000:7) describe timing as a concept that covers ‘all types of succession and duration; rhythm usually implies a particular kind of succession of events, one where the events occur with some repetitive structure'.

1.1. Statement of the problem

Many factors are taken into consideration when characterising the rhythmic properties of languages. Among them are the interactions of phonetic features, such as the duration of stressed and unstressed syllables, and phonological features, such as the syllable structure of a language. Studies like those of Pike (1947), Dauer (1983) and Grabe and Low (2002) were dedicated to describing the production of the rhythmic characteristics of languages and to identifying the rhythmic differences between languages.

Two terms widely used to describe the rhythmic property of languages are 'syllable-timed' and 'stress-timed.' Originally, languages were thought to be in either one or the other of these categories. The basis of the regularity or isochrony underpinning the rhythmic quality of a language is related to either the regulation of the interval between syllables or feet, thus syllable-timed language was defined as having isochronous syllable intervals, while a stress-timed language was identified by isochronous feet (Pike, 1947). Over time, the original conceptualisation of these terms became obsolete, because the isochrony of these respective features was shown not to exist, for example in studies by Roach (1982) and Dauer (1983); instead, investigators found that there is no simple dichotomy into which separate languages can be grouped based on rhythm. Nonetheless, the terms 'syllable-timed' and 'stress-timed' are still regularly deployed when describing speech rhythm.

Ramus, Nespor, and Mehler (1999) and Grabe and Low (2002) devised the approach to describing speech rhythm which has greatest contemporary currency. They found that the rhythmic organisation of languages can be arranged in a continuum based on how they align to set of acoustically-defined rhythm metrics. These rhythm metrics involve a set of calculations that directly measure the consonantal and vocalic intervals of the speech signal. Ramus et al. analysed the rhythmic features of languages by calculating the standard deviations of the consonantal and vocalic intervals (ΔC and ΔV respectively) and the percentage of the vocalic proportion (%V). In contrast, Grabe and Low (2002) formulated a 'Pairwise Variability Index' (PVI) to analyse the successive consonantal (*r*PVI) and vocalic (*n*PVI) intervals. This work has suggested that the rhythm metric values seem to correspond broadly with the previous rhythmic descriptions for a wide range of languages. For example, English (Roach, 1982) and German (Benton, Dockendorf, Jin, Liu & Edmonsdson, 2007), which are generally considered to be stress-timed languages are differentiated from Spanish, which is considered to have the characteristics of a syllable-timed language (Bradlow, 1995). At the same time, the scores for the rhythm metrics are useful in the way they reflect the syllable and vocalic variability of the languages and how this underpins their varying rhythmic characteristics.

Nevertheless, some concern exists regarding the the application of these measures to some languages. For example, according to the rhythm metrics proposed by Ramus et al. (1999), Thai was considered in the same zone as syllable-timed languages, while Grabe and Low (2002) placed Thai on the opposite end of the continuum closer to languages generally thought of as stress-timed. Subsequently, the validity of using these metrics to capture the rhythmic features of languages has been questioned by investigators such as Arvaniti (2009) and Fletcher (2010) who argue that the scores for these rhythm metrics are influenced by factors like cross speaker-variability or the nature of the materials from which the metrics are calculated. Ross, Ferjan, and Arvaniti (2008) found variability of rhythm metrics among different speakers of the same language. As a result of the findings, Arvaniti (2009) concluded that rhythm metrics only draw on one aspect of timing, namely segmental duration, and that other timing features must be considered when characterising rhythm.

This thesis explores the notion of rhythm metrics with close reference to the standard variety of the Malay language. Malay provides good ground to test the validity of metrics, because it has not been extensively studied. Grabe and Low (2002) found Malay to be on the stress-timed end of the continuum, but their finding was based on the speech of only one Malay speaker. This result is rather debatable, especially since previous literature based on the auditory observations (Maris, 1980:21) described Malay as 'a syllable-timed language.'

One of the issues addressed in this thesis is the conflict between the instrumental and auditory descriptions of the rhythmic characteristics of Malay. By analysing a new set of Malay data, which includes more speakers than the Grabe and Low (2002) study, using rhythm metrics, the aim of this study is to provide a characterisation of Malay rhythm. At the same time, the Malay data set also can be used to investigate the questions raised by Arvaniti (2009) about the validity of rhythm metrics.

The aims of this research are twofold: First, to shed light on the rhythmic property of Malay—a language that has not been extensively quantitatively analysed in the past. Although some studies, such as Maris (1980), attempt to describe Malay rhythm based on auditory observation, it is still not clear to what extent this is reflected in the speech signal. Second, this research will further test the validity of the measurement techniques of describing speech rhythm available in the literature.

1.2. Research Questions

The following research questions are designed to fulfil the aims of the study. The first research question is: *is there any evidence for an acoustic correlate of stress in Malay connected speech?* Before exploring the rhythmic features of Malay, this issue related to stress must be addressed first. Stress is a feature related to rhythm—the greater the effects of stress on the phonetic realisation of the language, the more likely it is to be rhythmically categorised as stress-timed (Dauer, 1983). Malay has been classified as 'stressless' language (Maris, 1980; and Zuraidah, Knowles and Yong, 2008), although there are very few instrumental studies of the phonetic properties of stress in Malay. Thus, a better understanding of this is necessary in order to draw a complete picture of the rhythmic structure of Malay.

Subsequently, the second research question is: *How does Malay align to the rhythm metrics proposed by Ramus et al. (1999); and Grabe and Low (2002)?* As the rhythmic structure of Malay has not been studied extensively in the literature, this question seeks to discover the position of Malay with respect to other languages that have been studied quantitatively. This is carried out by applying a range of existing rhythm metrics to malay, including ΔC, %V, ΔV, *r*PVI and *n*PVI (these metrics will be thoroughly discussed in chapter 2). As mentioned in the previous section, Malay is considered a syllable-timed language; however, the only documented study of Malay against these rhythm metrics, which was based on only one speaker (Grabe & Low, 2002), showed Malay to be positioned close to English in respect of PVI scores. Thus, Malay, in that study, is positioned in the same zone as languages that are considered to be stress-timed. This finding is rather contradictory from an auditory observation of Malay rhythm. Moreover, Malay also is considered by some previous investigators to be stressless, which, from the point of view of rhythmic organisation, means it should be different from stress-timed languages.

The findings relating to the second research question will be analysed in light of the current views on speech rhythm brought forward by researchers such as Arvaniti (2009) and Fletcher (2010). This leads to the third and final research question: *what aspects of the speech signal underpin the regularity associated with the speech rhythm of Malay?* Among the criticisms of rhythm metrics is the variability of the scores found across sentences and subjects within the same language. That is, the means of measuring the successive consonantal and vocalic interval durations are not able to withstand confounds arising from a range of independent factors. Thus, the third research question is designed to discover whether there is any aspect of Malay speech that can be considered as a strong candidate for acting as the foundation for the rhythmic characteristics of Malay.

1.3. Overview of the thesis

This thesis is organised into seven chapters. The statement of the problem and the aims of this thesis are included within this chapter. Chapter 2 discusses basic concepts related to speech rhythm, such as syllable, stress and foot, and provides a review of the literature on speech rhythm as it has developed from the point at which the notion of syllable-timed and stress-timed dichotomy were coined by Pike (1947) to the present day.

In Chapter 3, the discussion focuses on some background features of Malay related to the thesis, including historical information regarding Malay and some of its morphophonological features such as its phonological structure, phonotactic rules and syllable structure. Previous studies on prosodic features of Malay are also included in this chapter.

The first experiment investigating the properties of stress in Malay words is reported in Chapter 4. This chapter provides a detailed description of the methodology used to conduct this instrumental experiment. The findings are presented and discussed.The subsequent chapters: 5 and 6 centre on the implementation of the second experiment to identify the acoustic measures to characterise Malay rhythm. Chapter 5 outlines the methodology of the experiment, describing the materials and background of the subjects as well as the procedure during the recording session.

Chapter 6 then presents the results of the experiment evaluating how Malay aligns to a range of rhythm metrics. Next, the relationships between rhythm metrics and speech rate are explored. The findings of a further acoustic analysis of some of the subjects, which appeared to be the outliers in some of the rhythm metrics, are also presented in this chapter. This chapter addresses the issues of grouping and prominence of rhythm by presenting the results of an investigation of three different acoustic measures that may go some way capture Malay speech rhythm.

Chapter 7 summarises the findings of the study and discusses how these shed light on our understanding of Malay rhythm. The limitations of the present study, the generalisations that could be made from this study thesis, and recommendations for future lines of research are also presented in this chapter.

In summary, rhythm is a phenomenon that can be observed in most movements in the environment including speech. The purpose of this thesis is to provide an instrumental description of Malay speech rhythm. The report presented is undertaken with close reference to the rhythm metrics available in the current literature. The next chapter provides a comprehensive description of speech rhythm.

Chapter 2

Literature Review: Rhythm

2.1. Introduction

This chapter provides an overview of the framework and theoretical underpinnings of features related to speech rhythm focussing on recent keynote studies on rhythm as well as some issues surrounding the understanding of speech rhythm.

Speech rhythm is described as the regularity associated with the rhythmic characteristics of languages as perceived by listeners. Speech rhythm can be studied from two viewpoints. One is the production element of rhythm which involves articulatory and acoustic dimensions; while the other element is the perceptual side of rhythm. This chapter focuses on the production aspect of rhythm in particular the acoustic component as it is the dimension most closely related to the general theme of this thesis.

To capture different aspects of rhythm requires appeal to aspects of the organisation of speech which are quite complex and subject to different interpretations. Some of the key aspects of speech which are linked to rhythm are discussed in the next section.

2.2. Concepts related to rhythm

In this section, we explore some of the basic concepts, which are brought to bear in the discussion of rhythm. These features are syllable, stress and foot and they are not straightforward themselves, therefore there is a need to unpack them methodically so that a better overview of speech rhythm can be provided.

2.2.1. Syllable

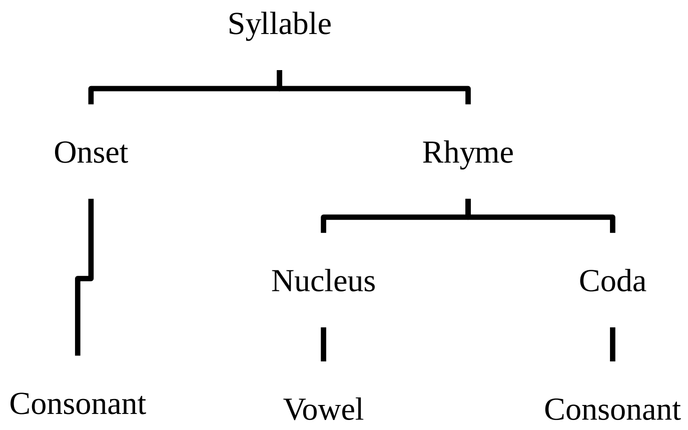
At the heart of the account of speech rhythm is this notion of recurrence or regularity of speech. A feature which is strongly connected to this is the syllable. The notion of syllable is regularly used in phonological analysis but the definition of syllable is less straightforward than what it might appear to be. According to Crystal (2003), there are two approaches to defining a syllable. Phonetically, a syllable is described as a string of sounds in which some are intrinsically more sonorous than others. Each peak of sonority corresponds to the centre of a syllable. The sonority peaks usually correspond to the vowels since they are produced with more energy and with no obstacles in the vocal tract, whilst the less sonorous part of the syllables are generated by the closure and the narrowing of vocal tract typically associated with the production of consonantal sounds.

Relative degrees of sonority are generally based on relative intensity as a starting point for estimating the perceptual saliency or loudness of a particular sound. Based on these criteria, Ladefoged (1982:25) outlined a partial sonority ranking for English, as follows: a＞æ＞ɛ＞ɪ＞u＞ i＞ l＞ n＞ m＞ z＞ v＞ s＞ ʃ＞ d＞ t＞ k. Blevins (1995) suggests that this particular scale agrees to a significant extent with most universal and language-specific phonological sonority scales.

Meanwhile, from a phonological viewpoint, syllable is regarded as the ways sounds combine in a particular language to produce typical sequences (Crystal, 2003). Typically, a syllable is analysed as a combination of an optional onset and a rhyme. The rhyme is further subcategorised into a nucleus and an optional coda. The nucleus, which is the most sonorous element, can occur on its own and usually is positioned at the centre of the syllables and it is occupied by vowels. A nucleus is obligatory for most of the syllables across most languages. Meanwhile, the optional onset and coda usually comprise of sounds that do not typically occur on their own; usually this class of sounds occurs at the edges of a sequence of sounds, consonants usually occupy these positions.

Sonority is an important defining characteristic of a syllable.
Ladd (1996) observed that in any of the syllable structure, only sounds of higher sonority are permitted to be the syllable peak, the recognition of sonority as a part of syllables could be traced back to Sievers (1881). To illustrate, for the word *dog* /dɒg/, where /d/ is the onset, /ɒ/, which is the most sonorous of these sounds is the nucleus and /g/ is the coda. The typical tree structure of a syllable is illustrated in Figure 2.1.

Figure 2.1 Tree representation of a CVC structure



Syllables are categorised as open or closed depending on whether they have a coda present or not. Open syllable is defined as a syllable with no coda; these are types of open syllable: V, CV and CCV. Examples of English words with these types of syllables are *a*, *to* and *plea* respectively. Closed syllable, on the other hand, is defined as a syllable that has a coda such as VC, CVC and CVCC, instances of words with these types of syllables in English are *at*, *cat* and *send* respectively.

In analysing the syllable structure of speech, one of the challenges is to demarcate between the boundaries of the syllables. In some cases, this is straightforward but in another it is more problematic especially when so-called ambisyllabicity occurs; by which a single sound could possibly belongs to two syllables. An example of ambisyllabicity in English is the word *selfish* /sɛlfɪʃ/, there are two plausible syllabification for it: /sɛl. fɪʃ/ and /sɛlf. ɪʃ/. Two approaches which can be used to resolve this issue; the first is to apply the maximum onset principle. The principle stipulates that as many consonants as possible will be syllabified into the onset. Thus, /sɛ. lfɪʃ/ seems to conform to the principle; however this type of syllabification is not phonotactically acceptable in English since the sequence /lf/ is not allowed as an onset of a syllable. Alternatively, the second method which is to rely on native speakers' intuition is potentially useful to demarcate syllable boundaries. Schütz (1985) observes that native speakers usually are able to identify how many syllables are there to a word as well as to which syllable a sound could be assigned. Thus, the two plausible and acceptable syllabifications for the word *selfish* according to the intuition of native speakers are /sɛl. fɪʃ/ and /sɛlf. ɪʃ/; both these alternatives are in alignment to the phonetic definition of the syllables, in which the sounds are demarcated according to the sonority. The former syllabification is demarcated by separating the consonant cluster /lf/ and assigned to different syllables. Although, the latter is also phonotactically acceptable in English; but it is rarer as a sequence.

In the present study of rhythm, the syllable is conceived largely in terms of its phonetic properties. In Malay, the syllabification process which results from the application of the phonetic definition is well aligned to native speakers' intuition as to where the syllable boundaries are drawn. Furthermore, the syllabification process is rather straightforward as, apart from borrowed words, the language does not have consonant clusters; section 3.3.3 outlines the typical syllable structures available in Malay.

2.2.2. Stress

Generally, stress can be analysed in two domains: the first is lexical stress; the second domain is the stress in connected speech. Since this thesis deals with rhythm, we are interested in stress realised in phrases or sentences but both types of stress will be described in this section. To begin, the phonetic correlates of lexical stress will be looked at first, followed by the phonetic correlates of sentence stress.

Stress is defined as ‘a degree of force used in producing a syllable’ (Crystal, 2003:435). Unlike the concept of syllable, the realisation of stress is rather straightforward to capture acoustically. A number of studies including Fry (1955), Bolinger (1958), Ladd (1996) and Dogil (1999) find that stress commonly correlates with longer duration, increased loudness and higher F_0.

In terms of lexical stress, Ladefoged (2005) outlines that the position of stress is fixed in relation to the word in some languages, for example stress falls on the first syllable of Czech word regardless the number of the syllables. while in Polish and Swahili, stress is always on the penultimate syllable. On the other hand, Ladefoged (2005) notes that other languages (including English and German) have no fixed position for stress within a word.

In order to understand further the acoustic properties of stress in individual words, Fry (1955) analysed recordings of twelve English speakers producing a list of English words for which a change of function from noun (n) to verb (v) is associated with a shift of stress from the first to the second syllable. The spectrograms of these recordings were analysed for their duration and intensity. He found that the consonantal segments were not ‘materially affected (1955:765) by the shifting stress...’, however the vowel segments showed a difference in duration ratio between the noun and the verb form. The example given in Fry (1955) is the word *permit*, the results for the mean durations for the vowel segments for the two syllables in this word (in ms) are given below

Table 2.1The mean durations for the two vocalic segments of the word *permit* which functioned as both noun and verb (in ms) (Fry, 1955)

	E	*i*
Noun	120	90
Verb	60	120

The table shows that corresponding to the word class where the word *permit* belongs; the vowel durations tend to be longer compared to the unstressed vowels. The stress in the word *permit* (n) falls on the first syllable, the mean duration for the first syllable is longer (120 ms) compared to the second syllable (90 ms), whereas the word *permit* (v) which is known to have stress on the second syllable can be observed to have a longer vowel segment on the second syllable (120 ms) compared to the first vowel segment (60 ms). At the same time, Fry (1955) also finds that the maximum intensity recorded is aligned with the corresponding longer vowel durations. He also finds a steady relationship between these two acoustic elements across the words studied. Fry (1955) concludes this experiment by stating that 'duration and intensity ratios are both cues for judgments of stress' (1955:768). Studies of lexical stress in other languages also show that lengthening of syllables is a cue for prominence in Swedish and Dutch (Nakatani et. al, 1981 and Nooteboom, 1972). A stressed vowel can be lengthened from 30 ms to 70 ms depending on the degree of stress.

Another acoustic property which has been found to be a correlate of stress in some languages is pitch or F_0[1]. Physically, F_0 is acoustic manifestation of the vibration of the vocal folds (Kent & Reid, 2003). F_0 prominence is associated with a pitch level or movement that makes a syllable stand out from its context (Roach, 2000). This feature, together with greater length and loudness discussed earlier, is considered as the correlates of stress in languages like English (Fry, 1955). Typically, F_0 can be found in both lexical and sentence domains.

Within the domain of lexical stress, there are two levels identified; they are primary and secondary stress. According to Ladd (1996) the differences in F_0 are the most important cues for listeners to rely on when making judgements about stress in isolated words. Roach (1998) defines primary stress as the most prominent syllable in a particular word, marked as "ˈ". Whereas, secondary stress is a stress weaker than the primary stress but stronger than a stressless syllable, it is marked as "ˌ". Some of the examples of these in English words are 'introduction' /ˌɪntɾə ˈdʌkʃən/ and 'organization' /ˌɔɾgənə ˈzeɪʃən/.

Another domain of stress that can be analysed is at the sentence level. The variation of the stress position in connected speech may partially contribute to the different perception of rhythm which speakers may hear in different languages (Couper- Kuhlen, 1986).

---

[1] Sometimes Pitch and F_0 are used interchangeably, but strictly speaking 'pitch' is a perceptual phenomenon and F_0 is a production phenomenon (Roach, 1998). For this study, F_0 will be used.

In a phrase or a sentence level, a stressed word is identified as the accented word. It is a word, which that stands out in a stream of speech because of its prominence (Crystal, 2003:436). Within the identified accented word, there is a nuclear accent or a tonic syllable which is the most prominent syllable within the accented word.

Ladd (1996) found that in American English, within an utterance of connected speech the tonic syllable in the accented word was recorded to be 37 to 41 ms longer than the other syllables within the same accented word.

At the same time, the nuclear accent also functions as lexical stress. It means that a certain word is emphasised because it conveys the most important message to the listeners or it provides new information to the listeners. For example, in German and Swedish, the accented words that carry new information are reported to be at least 30% longer than words which contain known information to the listeners (Ladd, 1996 and Heldner and Strangert, 2001). F_0 movement, on the other hand, is reported to be the most prominent feature of stress in Polish (Dogil, 1999) but the difference in duration between accented syllable and non-accented syllable is quite small (around 15 ms).

Fletcher (2010) states that there are different degrees of stress in the connected speech. Based on the literature, there are three levels of stress; unstressed reduced syllable, unaccented stressed syllables and accented stressed syllables. Heldner and Strangert (2001) find that the unstressed reduced syllable is 50% shorter than stressed unaccented syllables, for example in Swedish the unstressed reduced syllable is less than half of the unaccented syllable by 70 to 100 ms.. De Jong (2004) also finds similar results for Dutch. Of the three levels, stressed accented syllable are longer than the rest, Turk and

Sawusch (1997) find that the accented stressed vowel in American English is 21% longer than unaccented stressed vowel. Greenberg, Carvey and Chang (2003) observe that accented stressed vowel can be 60 to 100% longer than the unstressed reduced syllable

However the signalling of stress is different cross-linguistically. Hoequist (1983) reports that the durational marking of stressed syllables in Spanish is less marked than that of English. A similar finding is also reported by den Os (1988) for Italian and Dutch respectively. Fant, Krukenberg and Nord (1991) find that the duration of nuclear syllables for Swedish and English is between 100 ms and 150 ms longer than unaccented syllable, however French accented syllables are only 50 ms longer than unaccented ones.

Previously, the studies on the acoustic correlates of stress concentrated on durational, intensity and F_0 differences only; however there are other acoustic characteristics that have been found to differentiate between accented and unaccented syllables. These acoustic features of stress also can differ across languages. For example, Fletcher (2010) reports that vowel quantity contrasts also accentuate the differences between the nuclear syllable and unstressed syllables, this characteristic is evident in languages like Finnish (Suomi, 2007) and Tamil (Keane, 2006).

Nuclear syllables also usually show a relatively small coarticulatory influence from adjacent syllables. Lindblom, Aguelee,
Sussman and Eir Cortes (2007) illustrate that in American English, a nuclear vowel in an intervocalic position between velar and alveolar consonants is produced fully with no reduction. The nuclear vowel is less affected by the

neighbouring sounds compared to unstressed vowel in a similar intervocalic position.

Stress plays an important role in languages like English; according to Sabater (1991) an appropriate stress placement is essential for intelligibility purpose. For example if ‘profitable’ [‘prɒfitəbl] is realised as [prɒfi’təbl], misunderstanding can occur if the wrong syllable is accented. Pronunciation of a word with inappropriate stress pattern also can distort the shape of the word. Psychoacoustic studies shows that listeners expect certain variation of prominence to rely on when trying to process an utterance (Roach, 1998). At the same time, stress also is important for the perception of speech rhythm (Arvaniti, 2009), this will be discussed further in section 2.3.

These stressed and unstressed syllables in connected speech are usually combined in a foot. The notion of foot is going to be discussed next as it is also a concept that is closely related to speech rhythm.

2.2.3. Foot

According to Crystal (2003:183), the foot is a ‘unit of rhythm in languages displaying isochrony, i.e. where the stressed syllables fall at approximately regular intervals throughout an utterance’. The use of the term ‘foot’ in the studies of speech rhythm is drawn from the use of the same concept in describing metrical phonology particularly prose and verse. Abercrombie defines the foot in relation to prose and verse, as ‘the space in time from the incidence of one stress-pulse up to, but not including, the next stress pulse’ (1973:11). Couper-Kuhlen (1986:55) illustrates ‘the space’ with an example from a nursery rhyme; there are four feet in this phrase as denoted by the four stresses:

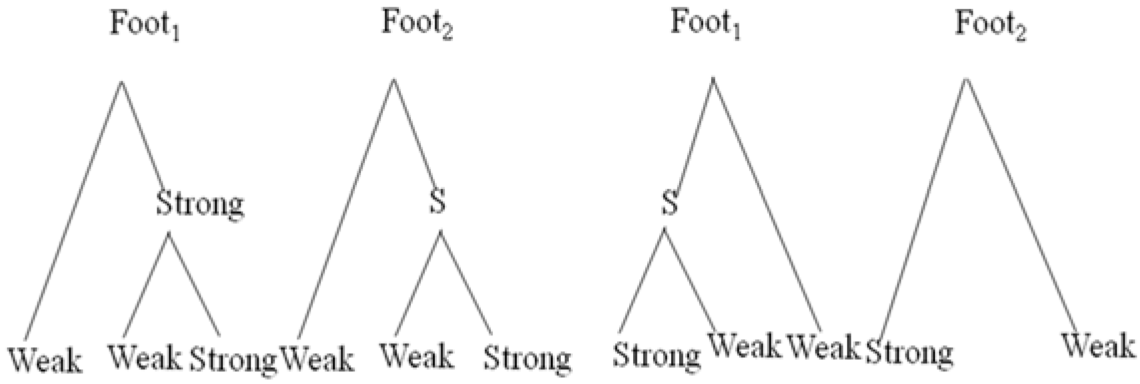
(a) ◌̍◌̍◌̍This is the ◌̍house that ◌̍Jack ◌̍built

Typically it is implied that in a stress-timed language, the feet would be of equal length (Pike, 1947 and Abercrombie, 1967), (the definitions of stress-timed and syllable-timed are given in 2.4). According to Couper-Kuhlen (1986) there are two kinds of feet across languages. For languages like French, right-headed feet are more dominant in which they have strong final syllable preceded by a succession of weak syllables. Whereas languages like English have left-dominant feet. This means that a strong syllable is followed by weak syllables. To illustate further, two sentences from French and English taken from Couper-Kulhen (1986:18) are given in Figure 2.2, together with their metrical foot tree.

Figure 2.2 Examples of right-dominant feet and left-dominant feet for the languages of French and English.

English:

sympho'nie fantas'tique[2] 'Buckingham 'Palace



As the concepts related to rhythm are discussed in this section; the next part of this chapter will address both the perception and production phenomena of speech rhythm across languages. The perception of rhythm will be looked at at a superficial level since this research concentrates on the production aspect of rhythm. Thus, the discussion on rhythmic production will be thoroughly reviewed.

[2] The translation for the French phrase is *Fantasy Symphony*

2.3. Speech rhythm: perception

The perception of the rhythmic quality of connected speech involves listeners tuning in to certain regulated features within the speech signal. Thus, Arvaniti (2009:59) defined rhythm as ‘the perception of series of stimuli as series of groups and repetitive patterns’. However, the link between perception and production is complex and not as straightforward as it seems because this regularity is not necessarily entirely a product of evenly-timed events within the signal (Terhardt and Schütte, 1976).

The fact that the alignment of perception and production is complex is illustrated by work on Perceptual-Centre or P-Centre. Morton, Markus and Frankish (1976:405) refer to P-Centres as ‘the psychological moment of occurrence of a word’.

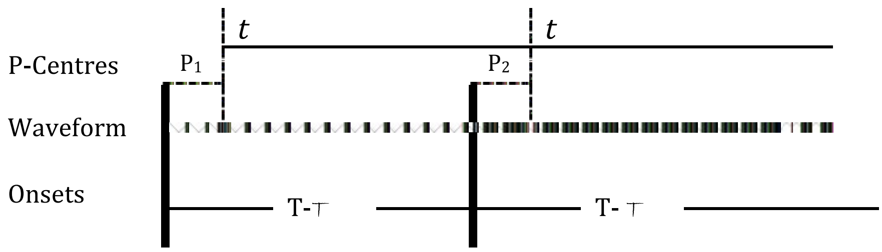
The term P-centre was coined following an experiment conducted by Morton et al. (1976) where they presented a pair of sounds in alternations to two subjects; these sounds were controlled by a computer. The first sound has a fixed interval; the interval of the subsequent sound would be determined by the subjects. They would adjust the knob until they perceived that the pair of sounds occurred at equal intervals. From this experiment, Morton et al. (1976) found that the alignment of the moment of occurrence was different from the acoustic isochrony.

The schematic diagram in Figure 2.3 illustrates the marking of p-centre by the listeners, where the wavy line is the representation of a waveform; the thick vertical lines are the acoustic onset of the intervals; the vertical serrated lines are the point on which the subjects perceived as the onset of the intervals; the horizontal serrated lines, which are labeled as P_x are the interval differences between the acoustic and perceived onset; *t* is the duration of the intervals marked by the subjects to be perceptually isochronous; and, ⊤ is the relative P-centre distance between the two intervals (which is obtained by calculating the difference between P_1 and P_2 durations). The diagram shows that the onset of the dark bar is not aligned to the onset of *t*.

This led Morton et al. (1976) to propose that what the listeners judged as isochronous and regular was some form of perceptual correlate which does not map easily onto the acoustic signal; instead, subjects were found to diverge systematically from acoustic regularity. The pattern which can be observed, in this experiment, is for sequence interval /ba~ma/, the duration of P_1 for the initial bilabial plosive was shorter compared to the P_2 of the following interval, the bilabial nasal. Meanwhile, for sequence interval /ma~ba/ the P_1 duration of the bilabial nasal was longer than that P_2 duration of the following bilabial plosive.

Figure 2.3 The schematic diagram of P-Centre

⊤ $=P_2-P_1$



Following this, Fowler (1979) conducted two experiments to study the production and perception of P-centres. The first experiment focused on the production of evenly-timed syllable intervals. A male adult was instructed to produce sequences of nonsense sentences. Each of the sentences consisted of

six monosyllables. The monosyllables rhymed with /ad/ but each had different initial consonants. These sentences were either homogeneous in composition like '*mad mad mad mad mad mad*' or alternating in composition like '*mad sad mad sad mad sad*'. The subject was instructed to read these sentences in a slow rhythmic rate and by stressing each syllable. 'Rhythmic' in this experiment was referred to as 'equality of intervals between syllable onsets'
(Fowler, 1979:378). Measurements were taken of the duration of the interstress intervals between acoustic onsets of the syllable as a function of their position in an utterance. The interstress interval was marked between the onset of the stressed vowel and the onset of the following stressed vowel. Fowler (1979) did not include the initial and final intervals in order 'to avoid the contamination of the rhythmicity effects' (1979:377) due to the effects of initial and final lengthening. Thus, only the durational measurement, for both homogeneous and alternating utterances, of the interstress intervals between syllables 2 and 3, 3 and 4 and 4 and 5 were taken (in ms). The absolute mean differences for each of the successive interstress intervals were also calculated.

The results for the homogeneous utterances showed that the interstress intervals were produced in near isochrony as the mean deviation from isochrony ranged between 19.7ms and 27.5ms. Subsequently, a one-way ANOVA also showed there is no significant effect of the interstress interval position on the vowel durations for the homogeneous utterances. This was because the durational differences between these intervals were rather small.

On the other hand, for the alternating utterances, the results indicated that there are big deviations from the isochrony, the mean deviations ranged from 18.4ms to 80.2ms. In fact, a one-way ANOVA showed a significant effect of the interstress interval positions on the durations, which suggested durational variability across the productions of the successive interstress intervals. A post-hoc analysis showed that interstress intervals with /s/ were significantly longer than interstress intervals with /m/.

Although the data were rather limited, Fowler (1979) concluded that when the subject in this experiment was asked to produce stress-timed utterances for the alternating sentences, the subject deviated from acoustic isochrony in systematic ways. According to Fowler (1979), The deviations were apparent in intervals beginning with acoustically long-duration consonants such as /s/ and ending with short-duration consonants /m/ were longer than those of the intervals with the reverse type of consonants. This finding supports the observation made by Morton et al. (1976) earlier in which they suggest that the sound properties of the intervals somewhat contribute to the way in which the listeners aligned their P-centres.

The deviation seemed to support the notion brought forward by Morton et al. (1976) in which to hear an utterance as stress-timed, the listeners require the deviation from the acoustic isochrony just like the speaker in this experiment produced.

The results of the first experiment were further substantiated by the second experiment. The second experiment focused on the perception by subjects of stress-timed rhythm. There were 12 alternating stimuli used for this experiment (similar to that of the first experiment). Each of the stimuli consisted of six monosyllables. These stimuli then were divided into two groups. The interstress interval duration within one group was manipulated and generated with very little standard deviation (12ms). While, the interstress interval duration of the stimuli in the other group were not controlled for and they have larger standard deviation (125ms). The unaltered stimuli were paired with the altered stimuli. Ten subjects were asked to identify which one of the utterances was ‘rhythmic’. The findings revealed that nine out of ten subjects chose the unaltered version of the six pairs to be rhythmic.

With regard to the first experiment, the speaker regulated something other than the acoustic components of the interstress intervals and, further to this, the second experiment showed that the listeners did not depend on the isochrony of the durations of the interstress intervals as a cue for rhythmicity. These experiments did not establish what was being regulated by the speakers or what was being perceived as regular by the listeners. Moreover, Morton et. al (1976) and Fowler (1979) suggested that P-centre markers were an abstract feature.

By using a similar stimuli described in Fowler (1979), studies like Patel, Lofqvist and Naito (1999) suggested cues such as the interval between syllable onsets, intervals between first formant amplitude slope maxima and between fundamental frequency amplitude slope maxima and intervals between articulator velocity maxima might be considered as P-centre markers. However, none of these features were found by Patel et al. (1999) to be acoustically isochronous. Perception of rhythm is clearly a complex matter. And it is nevertheless clear that languages can be differentiated by the rhythmic characteristics. Notwithstanding that we do not understand the perceptual basis of these differences; researchers have over the decades try to capture them in a typology of rhythmic properties.

2.4. Speech rhythm: Production

This section reviews the way in which people have looked at different aspects of speech production to try to work out what the basis is for the differences that people hear in the rhythmic properties across languages. It looks at the types of typology that people have applied to speech rhythm in order to characterise how it is differentiated across languages by looking from the articulatory and acoustic perspectives.

2.4.4. Speech rhythm typology

Amongst the earliest attempt to describe speech rhythm and to categorise languages according to their rhythmic properties was Pike (1947). He believed that there were two categories of languages which were characterised by having distinctive temporal organisation where elements like the syllable and stress would recur at regular intervals. These two categories of language were described as syllable-timed and stress-timed respectively.

Pike's earlier work on speech rhythm, which was quite influential, has a primarily articulatory basis to a lot of extent. The centre of Pike's typology of rhythm was the notion of a chest-pulse system during speech. Chest-pulses were described as pulse-like puffs of air from the lungs resulting from the alternate contractions and relaxations of the breathing muscles (Grabe & Low, 2002).

According to Pike (1947) languages could be categorised into two groups in terms of their rhythmic properties: first: the syllable-timed languages were claimed to have similar length of the successive syllables. Whereby in stress-timed syllable, the rhythmic feet or the intervals between stresses were argued to be equal.

Abercrombie (1967) also considered rhythm as purely an articulatory phenomenon, and like Pike (1947), he claimed that the respiratory muscles provide the instruments to produce speech rhythm. His views on rhythm were largely influenced by Stetson's 'motor' theory of syllable production (1928). According to Laver (1994:67), Stetson (1928) claimed that each syllable is produced on an individual pulse of egressive airflow created by a specific construction of the muscles of the respiratory system. Speech rhythm was defined as a periodic recurrence of movement produced by pulmonic air-stream mechanism. He considered chest-pulses and stress-pulses as foundations for speech rhythm. As he wrote 'syllable-timed rhythm, the periodic recurrence of movement is supplied by the syllable producing process: the chest-pulse, and hence the syllable, recur at equal intervals of time they are isochronous… stress-timed rhythm, the periodic recurrence of movement is supplied by the stress producing process: the stress pulse and hence the stressed syllables are isochronous.' (1967:97).

Abercrombie (1967) also supported the notion brought forward by Pike (1947) that languages could only belong to either a syllable-timed or stress-timed groups. Pike (1947) and Abercrombie (1967) proposed that languages like Spanish and Italian are syllable-timed and that English and Dutch belong to the stress-timed category.

Draper, Ladefoged and Whitteridge (1967) performed an experiment to test Stetson's theory of syllable production by using electromyography (EMG). In this experiment they used EMG to analyse the muscle activities that might indicate the existence of chest pulses and stress pulses when a syllable is produced by the speaker. The muscles or group of muscles investigated were the external intercostals, the internal intercostals, the latissimus dorsi, rectus abdominis, internal and external obliques and the diaphragm. They expected that each syllable produced would be accompanied by a chest pulse produced by the action of the internal intercoastal muscles; however the experiment failed to detect the observation Stetson (1928) suggested in which the muscle contraction could be consistent with the production of syllable. These results brought into questions the articulatory basis of Abercrombie's (1967) binary dichotomy for describing the rhythm of languages.

Despite that, the syllable timed and the stressed timed dichotomy has been a powerful factor motivating subsequent studies, and it can be observed that studies on speech rhythm still use these two labels as a convenience (Roach, 1982 and Grabe and Low 2002); although, in doing so, they do not advocate any longer the articulatory or physiological basis of these terms.

2.4.5. Acoustic evidence for the lack of isochrony

After the stress pulse and syllable pulse theory was challenged by Ladefoged (1967), phoneticians like Roach (1982) and Dauer (1983) attempted to search for isochrony of either inter-feet or inter-syllable intervals depending on the rhythmic category of a language. This section discusses the findings of the studies by Roach (1982) and Dauer (1983) as well as the phonological explanation by Dauer (1983) of cross linguistic differences in speech rhythm.

Roach (1982) was interested in defining an instrumental technique to categorise languages as either stressed-timed or syllable-timed. In his study, English, Russian and Arabic (conventionally classified as stressed-timed languages) as well as French, Telugu and Yoruba (conventionally classified as syllable-timed languages) were studied. A speaker of each language was given pictures to describe. Roach (1982) elicited spontaneous speech from them. Intensity metre traces of each of the recordings (which were approximately two minute's long) were made and then segmented manually. The durations of the interstress intervals were measured and these measurements were then divided by the number of interstress intervals the recordings contained. The measurement of inter-stress intervals was considered from the offset of the stressed syllable to the onset of the next stressed syllable.

Roach (1982) advanced two hypotheses for the foundation of his study. The first was that syllable-timed languages should exhibit a greater percentage of deviation in inter-stress interval than stressed-timed languages, because in stressed-timed languages, the inter-stress intervals would have the same duration regardless of the number of syllables in an interval. However, the results did not support this hypothesis as the standard deviation of the inter-stress syllable duration for both groups of languages were approximately of the same length. For example, the standard deviations of both Arabic (stress-timed language) and Telugu (a syllable-timed language) were 870 ms and 874 ms respectively.

The second hypothesis was that the inter-stress interval durations would be longer in proportion to the number of syllables in syllable-timed languages in which the duration of inter-stress intervals would increase depending on the number of syllables in it. To test this hypothesis a Pearson's correlation coefficient was calculated to study the association between percentage deviation and the number of syllables per inter-stress intervals for each of the languages. However, the results indicated that both categories of languages exhibited similar properties. For example, Russian (stress-timed) and

Telugu (syllable-timed), had identical correlation coefficients of 0.61 indicating that the duration of inter-stress intervals tends to be longer depending on the number of syllables regardless of the category of the language.

At the same time, Roach (1982) outlined some methodological problems encountered in designing an acoustic based approach. The first was identification of stress, especially since this was a spontaneous recording of speech; it was rather difficult to place stress. The researcher had to depend on subjective information of stress features given by native speakers of the languages studied.

The second problem was the issue of syllabification (described in 2.2.1) as it was essential to identify the exact syllable boundaries for the purpose of the analysis of the recordings. Problems may arise for words with polysyllabic structures where syllable boundaries are difficult to determine especially when analysing spontaneous speech in which the complex phonetic properties of connected speech can make syllabification segmentation even harder than it is for speech produced in isolation. Deterding (2001) encountered the exact same problem when he carried out the syllabic segmentation of connected speech to compare the rhythm of British English and Singapore English. He employed the maximum onset principle where the intervocalic consonants are assigned to the following syllable as long as that did not violate phonotactical constraints of the syllable onset, for example ‘sister’ would be syllabified as /sɪ. stə/ but ‘elder’ is /el.də/ because /ld/ was phonotactically impossible in English. In order to solve the syllabification issue, Deterding (2001) took the midpoint of the ambiguous segment as the boundary of the two syllables.

The third methodological problem highlighted by Roach (1982) was to identify the point at which the beginning of the inter-stress intervals should be drawn. There were two choices available; either to draw the inter-stress interval from the peak of the stressed vowel to the corresponding following peak or to draw the line from the beginning of the phonological stressed syllable which included the onset as well. Roach (1982) decided to choose the latter as a method to segment the inter-stress intervals in his study. A similar problem was also faced by Deterding (2001) where he found that it was difficult to identify the exact end of a syllable in phrase final position due to the fact that the speakers of Singapore English have a tendency to lengthen their final syllables. To remedy this problem, Deterding (2001) decided to exclude the final syllable from the analysis.

Roach (1982) concluded that the categorisation of rhythm in languages could not depend strictly on measuring time intervals based on the conventional dichotomy of stress-timed and syllable-timed languages since the results of his study indicated similarities in these languages.

Roach (1982) also pointed out that the results of the analysis might be influenced by speaker differences as only one subject was analysed for each of the languages studied.

Dauer (1983) reanalysed the notion of the rhythm categories coined by Pike (1947) by looking closely at what constituted properties of syllable-timed and stressed-timed languages. She studied five languages: English, conversational Thai (stressed-timed languages), Spanish (a syllable-timed language) and Greek and Italian (languages for which the rhythmic characteristics were not yet classified). Two subjects for each language were chosen and based on demographic information these subjects spoke different varieties of four of the languages, as for conversational Thai, she used the data from Luangthongkum, (1977). The subjects were asked to read a passage from a modern novel or play where the dialogue required the character speaking in normal everyday language. A native speaker and a phonetician listened and marked which syllable was stressed in the recordings. Inter-stress intervals were measured from the onset of the first stressed vowel to the onset of the next stressed vowel. The stressed and unstressed syllable structures for all the languages and the types of vowels that constituted these syllables were analysed.

Based on her observations, the length of inter-stress intervals merely indicated slow and fast speakers with no connection to rhythmic patterns. She found two differences between syllable-timed languages (English and conversational Thai) and other languages. Firstly, the average duration of monosyllabic inter-stress intervals in English and Thai was about the same as disyllabic inter-stress intervals (300ms). Secondly, the maximum number of syllables in continuous natural speech in inter-stress intervals for English was five, whereas it reached up to nine or ten in Spanish, Greek and Italian.

The syllable structures of the languages in Dauer (1983) were examined and it was found that syllable length variation correlated with the syllable structures that were permissible in the languages. Based on the analysis of the syllable structure in her English and Spanish materials, Dauer (1983) found that the most occurring syllable types in English were CV (34%) and CVC (30%). Based on the transcription, 82% of CV syllables were stressed and 66% of CVC syllables were stressed. While in Spanish 58% of the transcriptions analysed showed that CV type was the most occurring syllable. Unlike English, there were no particular syllables that were more salient in Spanish.

According to Dauer (1983) stressed and unstressed syllables were differentiated by centralisation of vowels in unstressed syllables and this was evident in stress-timed languages. In the transcribed text of English, she found that 92% of unstressed syllables have centralised vowels like [i] (it is assumed that the actual transcription was /ɪ/ since it was a more appropriate transcription to indicate a centralised quality), /ə/, and /ɚ/ while the stressed syllables were made up of vowels like /ɔ/, /ou/, /ɛ/ and /ei/. On the other hand, in Spanish, there was no correlation between syllable stress and the types of vowels. Dauer (1983) also observed that lexical or word stress level was often realised by changes in length, pitch contour, loudness and vowel quality in languages like English and Spanish, thus the ratio for the duration of the non-final open stressed syllable for English was 1:6 times longer than those of the unstressed syllable while 1:1 was the corresponding ratio for Spanish.

The conclusion drawn from the results of Dauer's (1983) study was that there were many factors, such as syllable structure, vowel quality and stress, which contributed to the rhythm of a language and there was no clear line to separate these languages into two different categories. Dauer suggested that it was more feasible to describe languages in the dimension of 'having a more or less stress-based rhythm' (1983:59). Dauer (1983) also noted the influence of speech rate on inter-stress interval durations in this study.

Whilst there are limitations of the work carried out by Roach (1982) and Dauer (1983), nevertheless, these investigations were quite influential on subsequent work on speech rhythm and laid the foundation for the next wave of work on rhythm. Recent work, which is going to be discussed in the next section, focusses on trying to discover the timing patterns within continuous speech which might differentiate languages by virtue of their rhythmic characteristics.

2.4.6. Rhythm metrics

Over the last 10-20 years, recent research into speech rhythm has shifted its focus to search for new forms of analysis to differentiate languages which are perceptually different from the point of view of their rhythm. As a result, another perspective of analysing speech rhythm has been developed. The two highlights from these works are: a) to develop new methodologies in order to better capture the production of rhythm; and b) to shed light on the typology of speech rhythm. Two major studies related to durational correlates of rhythm are discussed in this section.

Ramus et al (1999) were interested in obtaining an instrumental measurement which would differentiate languages that have been conventionally classified as either stressed-timed or syllable-timed and to clarify how rhythm might be extracted from the speech signal.

The starting point for Ramus et al's (1999) study was infant speech perception. They believed that infants are able to discriminate different speech rhythms at birth although they do not have pre-existing phonological knowledge about the languages since rhythm does not rely on 'complex and language- dependent phonological concepts' (1999:270) but more on phonetic cues from the speakers around them, therefore, Ramus et. al (1999) argued that a purely phonetic definition of language rhythm should be attempted.

Eight languages were studied by Ramus et. al (1999): English, Dutch, Polish, French, Spanish, Italian, Catalan and Japanese. Four female native speakers of each language were recruited. The subjects were asked to produce short news-like declarative statements. The text was originally in French and was loosely translated to the respective languages matching the number of syllables which was between 15-19 and the duration of each text/ utterance was approximately 3s. 160 utterances were collected from these languages.

Phonemes for each sentence were segmented using auditory and visual cues. The segments were located as precisely as possible using a phoneme inventory of each language. The phonemes were classified as consonants and vowels except for glides. In cases where glides were prevocalic, they were treated as consonants whilst postvocalic glides were identified as vowels. Measurements for consonantal intervals were located between the onset and offset of the consonant or clusters of consonants and vocalic intervals were identified as being located between onset and offset of a vowel or cluster of vowels. For example, the phrase ‘next Tuesday on’ which can be transcribed as /nɛkstjuzdeiɔn/ has the following consonantal and vocalic intervals:

/n/,/ɛ/, /kstj/, /u/, /zd/, /eiɔ/ and /n/ (Ramus et al, 1999:272).

Three variables were calculated from these intervals. The first is the proportion of vocalic intervals within the sentence, which was calculated by adding the duration of all the vocalic intervals, dividing it by the total duration of the sentence and multiplying the resultant number by 100 (%V). The second variable was the standard deviation of the duration of vocalic intervals within each sentence (ΔV) and the third variable was the standard deviation of the duration of consonantal intervals within each sentence (ΔC).

The distribution of these languages in respect of their %V measure allowed descriptions of speech rhythm which were somewhat similar to the traditional classification made by Pike (1947) and Abercrombie (1967). In Table 2.2, languages which are traditionally described as syllable-timed tend to cluster at one end of the continuum, while languages described as stress-timed tend to cluster at another end of the continuum. Ramus et. al (1999) believed that this finding was significant as it was the first indication of rhythmic structure that supported the conventional description of speech rhythm. ΔC and %V values of these languages appeared to be directly related to the syllable structures of these languages. The more syllable types the languages permitted; the greater the variability of the consonantal interval durations resulting in higher values of ΔC. However it was rather difficult to interpret ΔV since it was affected by vowel reduction, contrastive vowel length, vowel length in specific contexts and long vowels. Ramus et. al (1999) concluded that languages could be classified into a few categories based on the values of %V and ΔC.

Table 2.2 Results of the rhythm metrics across the languages (Ramus et.al, 1999:272)

Total number of measurements, proportion of vocalic intervals (%V), standard deviation of vocalic intervals over a sentence (ΔV), standard deviation of consonantal intervals over a sentence (ΔC), averaged by language, and their respective standard deviations[a]

Languages	Vocalic intervals	Consonantal intervals	%V (SD)	ΔV (SD) (× 100)	ΔC (SD) (*100)
English	307	320	40.1 (5.4)	4.64 (1.25)	5.35 (1.63)
Polish	334	333	41.0 (3.4)	2.51 (0.67)	5.14 (1.18)
Dutch	320	329	42.3 (4.2)	4.23 (0.93)	5.33 (1.5)
French	328	330	43.6 (4.5)	3.78 (1.21)	4.39 (0.74)
Spanish	320	317	43.8 (4.0)	3.32 (1.0)	4.74 (0.85)
Italian	326	317	45.2 (3.9)	4.00 (1.05)	4.81 (0.89)
Catalan	332	329	45.6 (5.4)	3.68 (1.44)	4.52 (0.86)
Japanese	336	334	53.1 (3.4)	4.02 (0.58)	3.56 (0.74)

[a] ΔV, ΔC and their respective SDs are shown multiplied by 100 for ease of reading.

Another major study related to speech rhythm is Grabe and Low (2002). Similar to Ramus, et al. (1999), they also took a direct approach of measuring the speech timing from the acoustic signal. However, Grabe and Low (2002) were different from Ramus et al (1999) study in the way in which they used an index to express the variability level in successive interval measurement known as Pairwise Variability Index (PVI) to provide evidence for rhythmic classifications of speech from durational measurements.

Eighteen languages were studied; each of these languages was presented by one speaker. The subjects read ‘The North Wind and The Sun’ in their respective languages. Subjects were asked to read the text once at their own pace.

Some of these languages were pre-classified according to their traditional rhythmic description: syllable-timed such as Spanish and French, stress-timed such as British English and Dutch and mora-timed[3] such as Japanese. While other languages such as Malay which did not have a pre-existing classification was categorised as an 'unclassified' language. There were also other languages that Grabe and Low (2002) considered as 'mixed' languages. These languages displayed both stress-timed and syllable-timed characteristics, for example, Polish was classified as stress-timed but at the same time, it did not exhibit vowel reduction. Both unclassified and mixed languages were grouped into the same category.

Similar to Ramus et. al (1999), the spectrogram measurements of these recordings were obtained and segmented as either vocalic intervals or intervocalic intervals. The vocalic interval was defined as the signal between vowel onset and vowel offset characterised by vowel formants. The vocalic intervals included monophthongs, diphthongs, or two or more vowels in a sequence. Intervocalic interval was described as the stretch of signal between vowel offset and vowel onset regardless of the number of consonants included. Table 2.3 sums up the description of these intervals.

Table 2.3 Summary of the description of the vocalic and intervocalic intervals segmented in Grabe and Low (2002)

Intervals	Description
Intervocalic	The consonantal intervals or the stretch of sounds between vowel offset and vowel onset
Vocalic	The vowel intervals which included monophthong, diphthongs and sequences of vowel.

They used several guidelines to segment vowels in the spectrograms. In fricative vowel sequences, the onset of the vowel was taken to be the onset

[3] Morae was considered as a sub unit of a syllable in which consisted of only a vowel and (if any) a preceding coda (Crystal, 2003:299). In mora-timing, the successive mora intervals were considered isochronous (Grabe and Low, 2002:521)

of the second formant, and in vowel-voiceless fricative sequences, the vowel was considered terminated where the friction begins. In vowel- voiced fricative sequences, the vowel was terminated at the onset of the high frequency energy for the fricative. Nasal vowel sequences were segmented by observing the fault transitions (Grabe and Low, 2002:524). For glides in initial position, Grabe and Low (2002) excluded it from the vocalic portions if their presence was indicated by clearly observable changes in formant structure or in the amplitude of the signal, otherwise glides were included in the vocalic portions. Pauses and hesitation were excluded in the analysis.

They measured the vocalic and intervocalic intervals; and these measurements were computed into PVI scores. The index score indicated the level of variability in successive measurements. The raw PVI (intervocalic *r*PVI) was compiled by calculating the difference between successive durations of intervocalic interval. As shown by the following formula:

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

Where *m* equalled number of intervals, vocalic or intervocalic, and *d* was the duration of the kth interval.

In order to control for speech rate across the speakers
Grabe and Low (2002) normalised the PVI values of the vocalic intervals (*n*PVI) by calculating the difference between successive durations of the vocalic interval and dividing it to the mean duration of these intervals:

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

The researchers chose to normalised the vocalic intervals because they claimed that vowels in English (at least) indicated ‘lowest level of prosodic hierarchy’ (Grabe and Low, 2002:526). At the same time, the duration of the vocalic intervals provided the best basis of comparison across languages in order to determine rhythmic differences. The supporting evidence for this assumption came from the results of their previous study (Low and Grabe, 1995) which showed that listeners perceived Singapore English as being syllable-time compared to British English from the variability of the vocalic intervals. Low and Grabe, (1995) found that British English has higher vowel variability that that of Singapore English.

Grabe and Low (2002) predicted that stressed-timed languages would show higher vocalic *n*PVI and higher intervocalic *r*PVI. Syllable-timed languages, on the other hand were predicted to have lower vocalic *n*PVI and lower intervocalic *r*PVI. Mora-timed languages were also predicted to have similar results as those of syllable-timed languages. Grabe and Low (2002) also predicted that mixed languages would have combination of both depending on whether the languages allowed vowel reduction as well as complex syllable structure in their system.

PVIs of both vocalic and intervocalic measurements were plotted against each other. The results indicated that the PVI scores of these languages fit the profiles of the conventional rhythmic classification of these languages. Languages like English, Dutch and German (stressed-timed) exhibited higher vocalic *n*PVI, and low *n*PVI values were found for French and Japanese. However, there was no support for strict categorical distinction between languages. The scatter plot (Figure 2.4) sums up the results:

Figure 2.4 PVI profiles for data from eighteen languages (Grabe & Low, 2002:530)

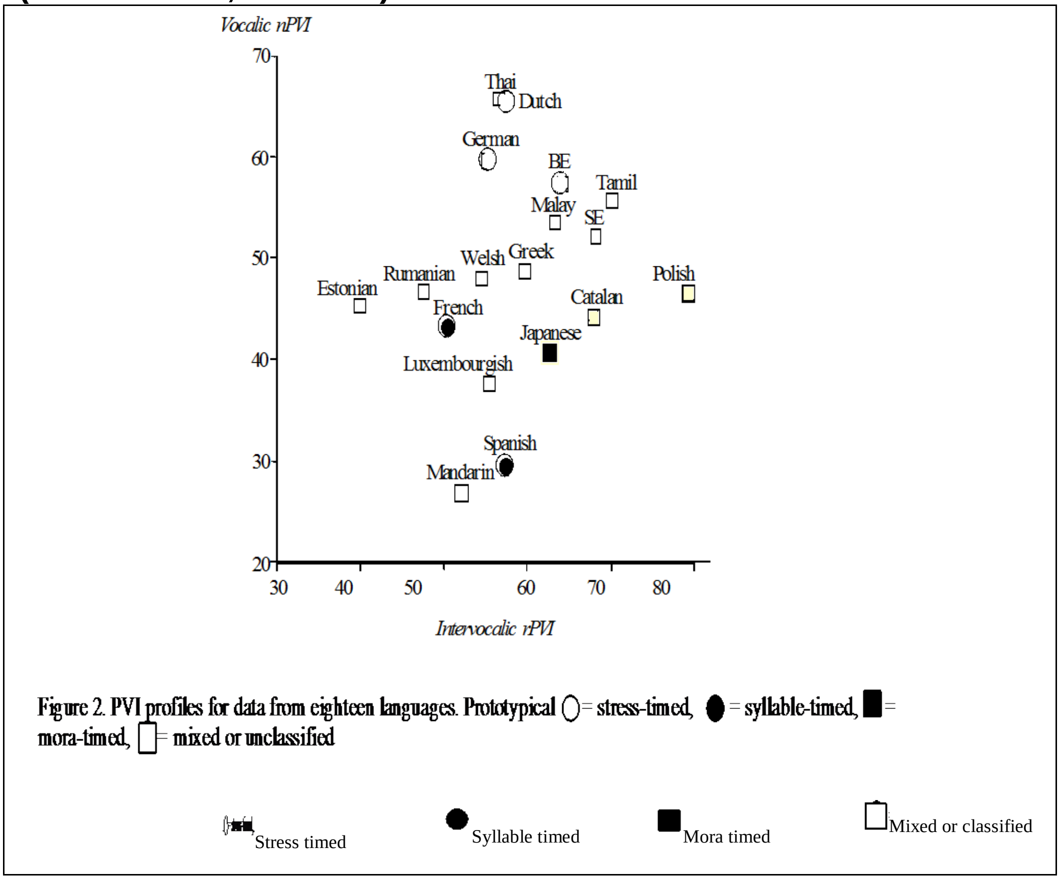


Figure 2. PVI profiles for data from eighteen languages. Prototypical = stress-timed, = syllable-timed, = mora-timed, = mixed or unclassified

The conclusions made by Grabe and Low (2002) were that the acoustic analysis of their data supports a weak distinction between syllable-timed and stressed-time languages and that there was a considerable overlap between these two categories.

In retrospect, these two metrics graded the languages differently by being based on different calculations; Ramus et al. (1999) looked at the durations of intervals, whereas Grabe and Low (2002) looked at the relationship between one interval to another interval. However both Ramus et. al (1999) and Grabe and Low (2002) did not explain other potentially relevant factors, such as how the texts chosen may also influence the results, and there is no discussion of the way in which speaker differences could affect the production of these utterances.

Unlike Dauer (1983), Ramus et. al (1999) and Grabe and Low (2002) focused on differentiating the languages according to rhythmic characteristics by analysing their durational measurements. However, they did not provide an explanation of what caused the rhythm metrics scores for these languages to be different from each other.

Following the use of durational measurements to calculate rhythm metrics; Dellwo and Wagner (2003) challenged the approach taken by Ramus et al (1999) because Dellwo and Wagner (2003) found that ΔC is dependent on speech rate (this study will be discussed in Section 2.3.4), thus, making the rhythmic continuum proposed by Ramus, et al. (1999) less than robust because it could not withstand the influence of speech rate. At the same time, factors such as the differences across different materials and across subjects were not addressed in both the Ramus et al. (1999) and Grabe and Low (2002) studies.

2.4.7. Rhythm as a function of speech rate

Dellwo and Wagner (2003) found that there is an interaction between speech rate and the rhythm metrics proposed by Ramus et al. (1999) in particular ΔC and %V. In an experiment to study the proportion of variance for %V and ΔC in relation to speech rate; they found that speakers producing different speech rates could change the values of the rhythm metrics of their respective languages.

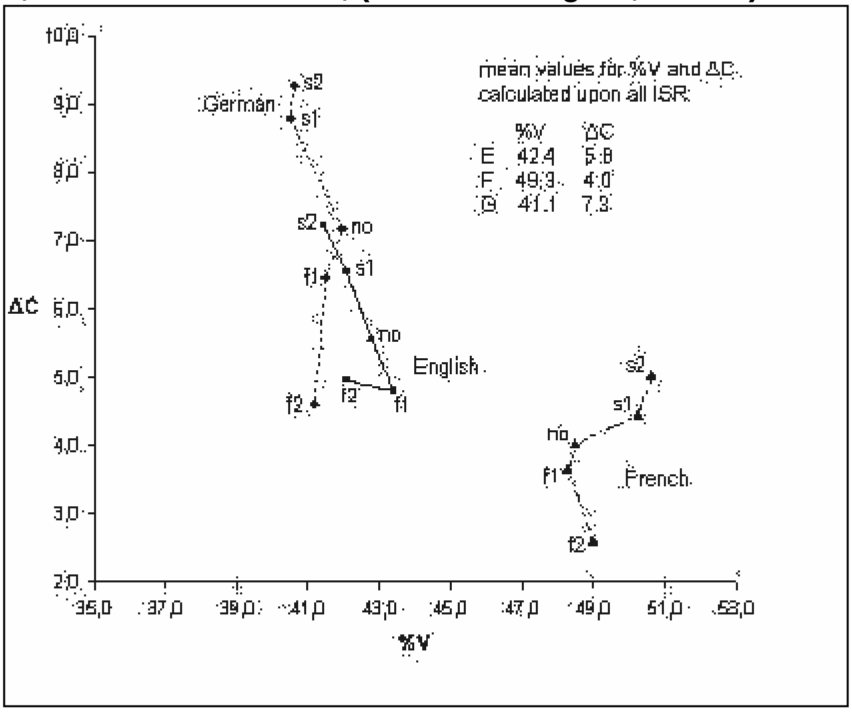
Three languages: English, German and French were studied in Dellwo and Wagner (2003). There were 5, 4 and 7 native speakers respectively. The original material was taken from a German text containing 76 syllables. This material was then translated into English (76 syllables) and French (93 syllables). After the subjects were familiar with the texts, the ‘normal rate’ reading was recorded. Then, these subjects had to undergo two recording phases, in each phase; the subjects had to read the texts twice. In the first phase, the subjects were instructed to read the text more slowly than their normal rate once, subsequently, they had to repeat the reading in an even slower pace than their first reading. In the second phase, the subjects were instructed to produce the text faster than their normal speech rate once and followed by an even faster speech rate than the first reading second. The recordings of all the subjects were then labelled as *no* for the normal rate, the first slow reading was labelled as *s1* and the next was *s2*, while first and second fast readings from the second phase were labelled as *f1* and *f2* respectively.

The %V and ΔC for all these utterances were also calculated following the criteria outlined in Ramus et al. (1999). The results showed that there was a positive correlation between speech rate and the rhythm metrics; when the speakers speak faster, the rhythm metrics tend to lean towards stress-timed whereas when they speak slowly, the results of the rhythm metrics tend to be on the syllable-timed end of the continuum.

A graph plotted for %V against ΔC for all speech rates recorded (Figure 2.5), Dellwo and Wagner (2003) found that there were two clusters of languages. French was positioned in one cluster while German and English were relatively in the same cluster. This result agreed with the grouping of languages based on the rhythm metrics proposed by Ramus et al (1999).

French, a language which was categorised in the syllable-timed end of the continuum was recorded with the fastest speech rate with a mean of 11.10 syllable/second for *f2*; while, the means *f2* for German (8.18 syllable/second) and English (7.95 syllable/second) were lower. German and English were positioned in the stress-timed end of the continuum according to Ramus et al. (1999).

Figure 2.5 %V against ΔC for all the five different speech rates of French, German and French, (Dellwo & Wagner, 2003:3)



When comparisons were made for the results of %V obtained by Ramus et.al (1999) and their study, Dellwo and Wagner (2003) acknowledged that variability within languages exists in terms of their rhythm metrics; for example the mean values of %V for French in this study was 49.38% while the former's result for the same language was 43.6%. For ΔC, a negative correlation was recorded to the ratio of speech rate.

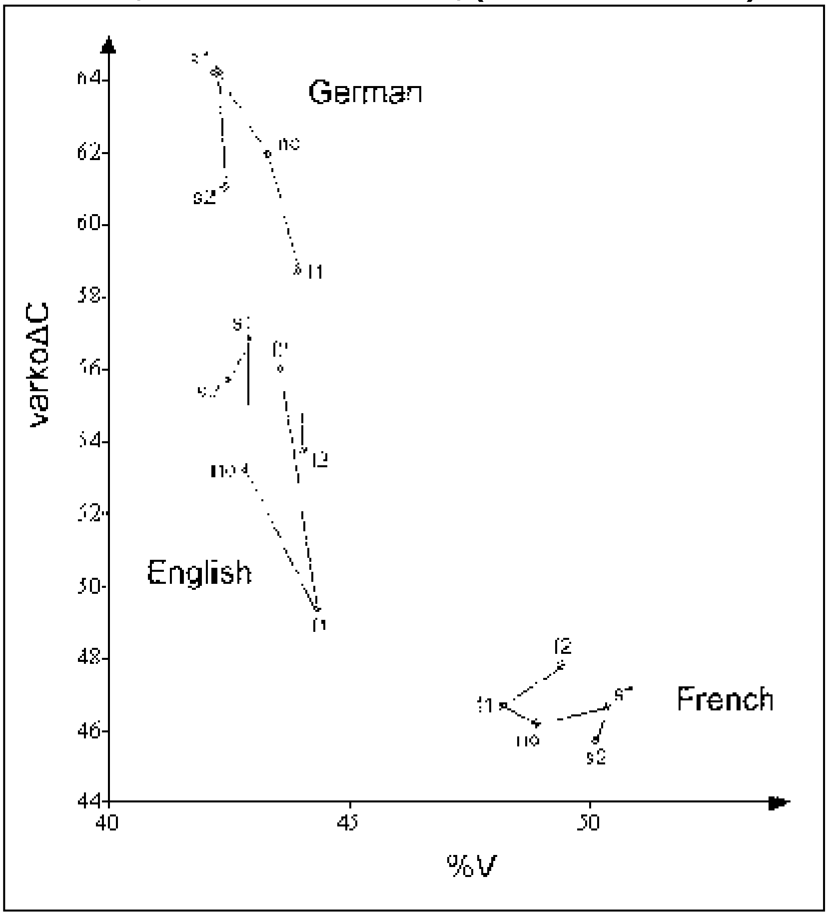
Dellwo and Wagner (2003) found that the results of this study revealed characteristics of the language, which is the number of syllables that the speakers are able to produce in a second. This was highly influenced by the language's phonetic, phonology and phonotactic structures, as per French compared to English and German. They claimed that the metrics advocated by Ramus et al (1999) were not able to distinguish rhythmic clusters of these languages as speech rate can influence the values of the rhythm metrics.

Dellwo (2005) argued that there was a need to address the high variation of ΔC in order to better capture different rhythmic classes. This was due to the fact that the results of ΔC in Dellwo & Wagner (2003) are influenced by speech rates. When speech rate was slower, it permitted the speakers to have more variability in the durations of consonantal intervals, while the opposite observation was noted for faster speech rate. Thus he proposed varcoΔC, a proportional variation of the ΔC divided by the mean of consonantal duration:

$$Varco\Delta C = \frac{\Delta C * 100}{meanC}; \mathrm{C = durational\ of\ consonantal\ measurement}$$

In this study, Dellwo (2005) examined German, English and French. There were 12 speakers of German and 7 each for English and French. The same materials and techniques from Dellwo and Wagner (2003) described earlier was used to obtain the data. He calculated %V, ΔC and VarcoΔC for all speech rate. The results for %V and ΔC showed similar patterns to that of Dellwo and Wagner (2003); although more subjects were included than the previous study. Meanwhile, VarcoΔC reveals clearer clustering patterns compared to ΔC in the previous study when the values were plotted against %V (Figure 2.6):

Figure 2.6 %V against VarcoΔC for all the five different speech rates of French, German and French, (from Dellwo 2005)



Initially, Dellwo (2005) assumed that the relative changes of the values of VarcoΔC according to speech rate were minor because the complexity of consonants in a language did not really change due to speech rate. The results for French agreed with this assumption; however German and English showed that speech rate had a great effect on ΔC. From this finding, Dellwo (2005) suggested that the values of VarcoΔC for these languages did not necessarily determine the absolute syllable complexity instead the values could be an effect of normal utterances.

Based on the use of durational measurements of consonantal and vocalic intervals to capture the rhythmic properties of the languages studied, Ramus et al. (1999) and Grabe and Low (2002) have made an assumption that these languages could be straightforwardly compared, however, according to Dauer (1983) (section 2.4.2), what might appear to be the differences in the rhythmic characteristics of languages as defined by some of these measures might actually be attributed to other factors. Therefore, this methodology should take into consideration the morphophonological features of each language when attempting to compare across the data.

From a methodological point of view, the contribution of these studies has been to provide a systematic ways of analysing the durational interval of speech in order to unpack more information on the production of speech rhythm. What we have learned is that there is no simple typology to describe rhythmic properties of languages as other factors, such as speech rate as demonstrated earlier seems to determine the rhythmic characteristics of the languages studied. This leads back to the notion brought forward by Dauer (1983) in which the differences in the phonological structures within these languages must be looked at in order to capture their rhythmic properties.

Investigations subsequent to Ramus et al. (1999) and Grabe & Low (2002) have acknowledged that there are many factors which could contribute to the properties of speech rhythm; and have increasingly attempt to look at this from a broader perspective. In relation to the current work on characterising speech rhythm of Malay, durational values will be looked at not from a narrow contexts but from a wider perspective and by considering morphophonological factors of Malay, which might influence the duration of the intervals. This particular notion of rhythm will be looked at next.

2.5. Rhythm and Prominence

Arvaniti (2009) argued that the rhythm metrics proposed by Ramus et al. (1999) and Grabe & Low (2002) for categorising speech rhythm were less than reliable because their focus was largely on timing. In her opinion, timing and rhythm are two separate features and should not be equated; this is because of the typological issue in which some of the languages could not be positioned straightforwardly within the traditional rhythmic categories. The example given by Arvaniti (2009) was Thai in which, by using the rhythm metrics proposed by Ramus et al. (1999) was positioned in the syllable-timed end of the continuum. However, the very same language could also be positioned at the opposite end of the continuum on the basis of a PVI analysis (Grabe & Low, 2002).

The second reason Arvaniti (2009) is worried about these metrics is that the materials used might affect the metrics. In her own experiment on English, German, Italian, Spanish, Greek and Korean, it was initially assumed that if the metrical values of a language was constant, spontaneous speech should yield the same result as the read speech; however she found that this was not the case for the languages she studied. She argued that some materials within the same language may be inclined to be stressed-timed and some may be inclined to be syllable-timed, an example in English (as taken from Arvaniti, 2009:49) is given below:

Type	Example of the sentences
Stressed-timed	*The production increased by three fifth in the last quarter of 2007.*
Syllable-timed	*Lara saw Bobby when she was on the way to the photocopy room.*

At the same time, variability in the way in which subjects produced these utterances could influence the rhythm metrics. For example, Roach (1982) acknowledged the fact that individual differences could influence results, and this was substantiated further by Arvaniti (2010) and Wiget, White, Grenon, Rauch, and Mattys (2010) in which they found that speakers from the same variety producing the same materials also could have variable results on the same rhythm metrics.

In retrospect, the foundations for the rhythm metrics proposed by Ramus et. al (1999) and Grabe and Low (2002) were derived from the assumption that there was a simple and straight-forward relationship between duration and abstract phonological categories, such as syllable structure and vowel reduction patterns (Arvaniti, 2009).

Fletcher (2010) also outlines several problems with the fundamental assumption of the findings in Ramus, et. al (1999) and Grabe and Low (2002). This is because there are many factors influencing segmental duration (as listed by Klatt, 1976) for example the presence of geminate consonants or of vocalic length distinction. Other language specific factors also may affect duration and they include: vowel duration, phrase final lengthening, allophonic changes of both consonants and vowels and stressed and unstressed syllables.

Thus, there is a need to have some possibility for developing a fuller understanding of rhythmic properties of speech, which takes into consideration a wider range of factors, which might be contributing to the rhythmicity. This was essential in order to avoid circularity and move away from relying on a single aspect of speech timing when describing rhythm (Fletcher, 2010); methodologically it might not be easy to find a single phonetic parameter which correlates straightforwardly with rhythmic differences across languages because there are other factors that are clearly involved. However, arguably, it is worth testing the measures like PVIs and ΔC, ΔV and %V as they might be helpful in bringing to light factors that might be relevant to the study of rhythm.

Arvaniti (2009) also suggests that it is worthy to go back to the basic definition of rhythm. Psychologically, Woodrow (1951), Fraisse (1963, 1982) define rhythm as ‘the perception of series of stimuli as series of groups of similar and repetitive patterns’. The ‘grouping of stimuli’ is not only limited to duration, it involves a lot of categories such as relative intensity and F_0 peaks. And most of the time these groupings are strongly connected to the phonological structure of the particular language and the listeners’ perception of the variation of prominence within a phrase.

Arvaniti (2009) went further to stress that a new typology of rhythmic types should rely on grouping and prominence so that a psychologically valid understanding of rhythm can be achieved and at the same time can be applied to all existing languages. In view of this, while the thesis carries out analysis based on the pre-existing rhythm metrics, it also goes beyond those to look at what other factors might be associated with a particular rhythmic characteristic of Malay.

In this section, stress and rhythm were discussed at length. The current metrics which are heavily based on durational measurement were reviewed thoroughly. The counter arguments for these metrics were also established. Arvaniti (2009) and Fletcher (2010) advance the need to move beyond metrics based simply on interval measurements as a means of describing speech rhythm. The next chapter of the literature review looks at Malay and its morphophonological properties.

Chapter 3

Literature review: Malay and Its Morphophonology

3.1. Introduction

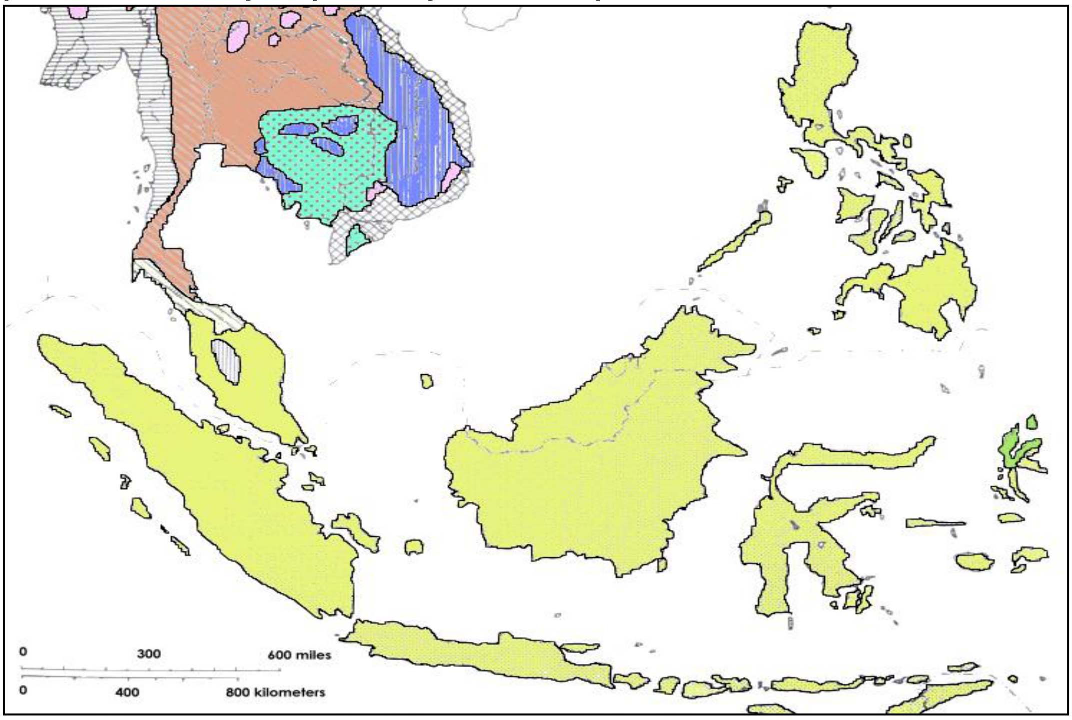
This chapter presents the relevant aspects of Malay, which are important to the present research. First by giving some general background of the language and second to highlight those which are relevant to the current study in particular the morphophonological features of Malay.

3.2. Malay: Some general characteristics

Malay belongs to the nuclear Malayo-Polynesian branch of Austronesian languages. According to Nik Safiah (1993:25), speakers of these languages occupied the coastal and lowland areas of Southeast Asia; some travelled further eastward to the Pacific Ocean while other groups went westwards as far as Madagascar and southwards to New Zealand.

Malay is spoken by the people (mostly of Malay ethnicity) who live in the Malay Peninsula, Southern Thailand, The Philippines, Singapore, Central Eastern Sumatra, The Riau Islands, and part of the coast of Borneo (see Figure 3.1). It is an official language in Malaysia, Brunei, Indonesia and one of the national languages spoken in Singapore. There are about 780-790 million speakers of this language worldwide.

Figure 3.1 The map of Austronesian language in Southeast Asia
(area in which Malay is spoken is yellow-coded)



In Malaysia, Malay is known as *Bahasa Melayu* with 16.5 million speakers. In Indonesia, it is known as *Bahasa Indonesia* (170 million speakers), in Brunei and Singapore, the variants are known as *Bahasa Melayu Brunei* (0.25 million speakers) and *Bahasa Melayu Singapura* (3.25 millions) respectively (Zaharani, 1998)

3.2.8. Malay in Malaysia

Malaysia (Figure 3.2) is a multiethnic country in which 50.4% of the population is of Malay ethnicity, 23.7% is of Chinese ethnicity, 11.0% is of the Indigenous group and 7.1% is of Indian ethnicity (Population and Housing Census, 2000). The diversity of the society makes Malaysia multilingual with languages like Malay, Tamil, Iban, Dusun, Chinese dialects such as Mandarin, Hokkien, Hakka and Cantonese as well as English spoken across the country. Following a National Language Act in 1967, Malay is established as the national language due to its role as a medium of interaction between different ethnic groups in Malaysia. This language is used as for official purposes in government administration, media communication and in the public schools.

Malay is the first language of 45% of the Malaysian population while the remaining 55% of the Malaysians considered Malay as their second language, learning it formally in schools and universities (Zaharani, 1998). Malay is characterised by a variety of dialects which include Perak, Kedah Negeri Sembilan and Kelantan dialects. The geographic origin of a speaker can be identified by the dialect he or she speaks. The standard Malay used in government administration, in broadcasting and in national schools can be historically traced back to the Johore-Riau dialect. This dialect also is widely used in the west-coast of Peninsula Malaysia in states such as Selangor and Kuala Lumpur (Tajul, 2000). Some of the key morphophonological features of the standard variety of Malay are presented in the next section.

Figure 3.2 Map of Malaysia

3.3. Phonological features of Standard Malay

There are 21 consonantal phonemes in standard Malay (Table 3.1). Glottal stop [ʔ] is regarded as a realisational variant and is not phonemic (Teoh, 1994 and Tajul, 2000) in this variety (see section 3.3.1). Subsequently there are six vowel phonemes in Malay (Figure 3.3). Vowel length is not phonemically contrastive within the Malay vowel system. There are three diphthongs found in Malay, they are /oɪ/, /ɐɪ/ and /ɐu/ (Yunus Maris, 1980:41) (Figure 3.3).

The phonemic system presented here captures the lexicon which is not loan words. Words in Malay may form different categories depending on their origins. Some of the loan words are borrowed from languages like English, Arabic and Tagalog. From speakers of Malay point of view, the origins of the words are irrelevant. Thus, the phonology is rather more complex than is reflected in the typical account given of Malay; for example some words like *kompleks* 'complex' and *kritik* 'critique', which were borrowed from English are allowed in Malay despite the consonantal clusters in the initial and final positions of these words respectively. The phonotactic features of the language are discussed in section 3.3.3.

Table 3.1 The consonant phonemes of standard Malay

	Bilabial	Labio-dental	Dental	Alveolar	Post-alveolar	Palatal	Velar	Glottal
Nasal	m			n		ɲ	ŋ	
Plosive	p b			t d			k g	
Affricate					ʧ ʤ			
Fricative		f v		s z	ʃ			h
Approximant				ɹ		j		
Lateral				l				
Other consonant: w								

Figure 3.3 The vowel phonemes of standard Malay

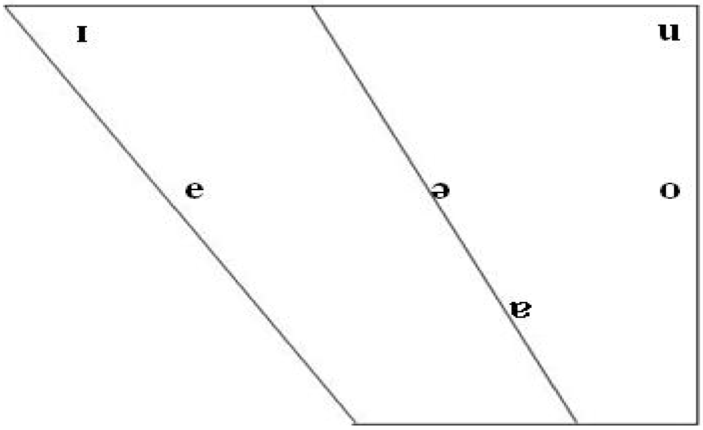
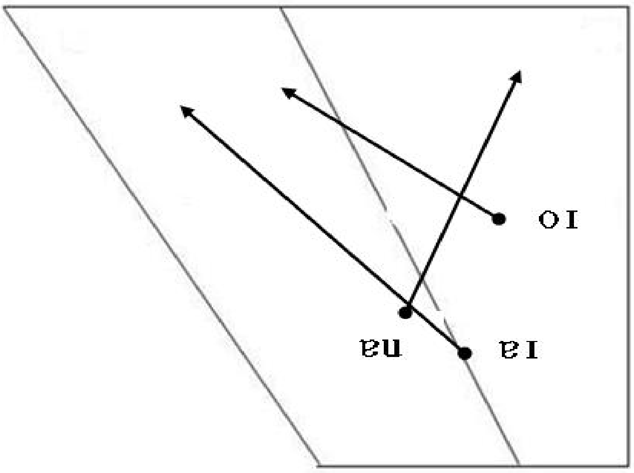


Figure 3.4 The diphthong phonemes of standard Malay



3.3.1. Word structure

Malay is classified as an agglutinative language in which words are formed by merging morphemes together. Generally there are two types of words in Malay; they are morphologically simple words and morphologically complex words. There are three types of word formations found in morphologically complex lexical items namely affixation, compound words and reduplication.

According to Teoh (1994), there are three types of affixes permissible in Malay, namely prefixes, suffixes and circumfixes. A circumfix is a type of affix which requires both prefix and suffix to be attached simultaneously to the stem word. Some of the prefixes and suffixes can change the word class of a stem. For example, the prefix '*pel*-' attached to a verb stem will produce a noun, as in the stem word *ajar* 'to teach'(verb) will change to a noun; *pelajar* 'student'. Similarly the suffix '-*an*' if attached to the end of the stem alters the word class, for example the word. *ajaran* 'teachings'(noun). Other affixes such as prefix '*bel*-' and suffix '-*kan*' will not change the word class of the stem, for instance, *belajar* 'to learn' (verb) and *ajarkan* 'to teach' (verb).

The derivation of new words by circumfixation will always change the word class. Malay allows attachment of more than one prefix in circumfixation, for example the word *pembelajaran* 'the process of learning' (noun) is derived from prefixes '*pem*-' and '*bel*-' and suffix '-*an*', they are attached to the verb stem *ajar* 'to teach'.

New words can also be formed by the combination of two root words. The process of compounding does not change the word class of the newly combined words, for example the word *kereta* 'car' (noun) when combined with the word *api* ‘fire’ (noun) will become *keretapi* ‘train’ (noun). Affixation also is permissible on compound words, and this sometimes can alter the word class, for example the prefix ‘*peng-’* and suffix ‘*-an*’ can be added to the word *ambil alih* which means ‘to take over’ (verb) – seperately *ambil* means ‘to take’ (verb) and *alih* ‘to move’ (verb) – to become *pengambilalihan* ‘the process of taking over’ (noun).

Another way to form a new morpheme in Malay is reduplication. According to Zaharani (1994) reduplication usually is used to indicate plurality for example the word *kotak* ‘one box’ if duplicated becomes *kotak-kotak* ‘more than one box’. If reduplication process happens on a verb, it alters the word class to noun, for example. the word *main* ‘to play’ (verb) if duplicated becomes *main-main* ‘banter’ (noun).

In respect to the word order, Subject Verb Object (SVO) is a predominant type in Malay (Asmah, 1985). It is as follows:

Dia	*Ambil*	*bola*
3sing	take-Pres	ball
He/She	Takes	the ball

3.3.2. Phonetic realisation of Malay

This section lists some of the significant characteristics of the phonetic realisation of the phonological system described previously. Some of these are features of connected speech but others are equally found in words in isolation.

One of the features of the phonetic realisation of Malay is glottal stop insertion (Tajul, 2000). When there is a vowel hiatus- in which two vowels sounds occurring in adjacent syllables across word boundaries- a glottal stop

[ʔ] is inserted between the two vowel sounds. An instance of glottal stop insertion in connected speech is in the phrase *hari ini ada kuliah* 'there is a lecture today' is realised as [hɐɹɪʔɪnɪʔɐdɐkulɪjɐh] (Tajul, 2000:63). At the same time, glottal insertion also occurs in phrase-initial position when the phrase begins with a vowel; for example, speakers of standard Malay typically realise the phrase initial word *apa* in the phrase *apa kamu mahu?* 'what do you want' as [ʔɐpɐ kɐmu mɐhu](Tajul, 2000).

According to Tajul (2000), for Malay words ending with /k/, the /k/ in the final position is usually realised as glottal stop [ʔ]. For example, the word *ajak* /ɐʤɐk/'to invite' is usually pronounced as [ɐʤɐʔ]. Moreover, in morphologically complex words, when a stem ending with a voiceless velar stop is suffixed with vowel initial suffixes '*-an*' and '*-i*', the velar stop manifests itself as glottal stop, for example the word *masuk* 'enter' (verb) when attached to the suffix '*–an'* becomes *masukkan* 'to enter (an object)' (verb) and it is realised as [mɐsuʔkɐn] in standard Malay (Teoh, 1994:59) and the word *memasuki* 'to enter (an object)', which is derived by attaching the suffix '*-i*', is realised as [məmɐsuʔki]. For both the words where stem-finally /k/ is followed by a suffix which begins with a vowel, *'-an*' or *'-i'*, the /k/ in the stem is realised by glottal stop but the velar closure is retained after the glottal stop.

Another feature of the phonetic realisation of Malay is that voiceless stops /p/ and /t/ are unreleased when word final or followed by another consonant, words like *tetap* 'permanent' is realised as [tətɐp̚] (Tajul, 2000).

In general, when there is a vowel hiatus in both simple and morphologically complex words a glottal stop will be inserted between these vowels (Teoh 1994); for example, the words like *peel* ‘attitude’, *ejaaan* ‘spelling’, *merasai* ‘to love’ and *menemui* ‘to find’ will be realised as [peʔel], [eʤɐʔɐn], [mərɐsɐʔɪ] and [mənəmuʔɪ] respectively. However a different pattern is observed when a stem ends with a high vowel and is attached to the suffix ‘-an’ (section 3.3.2), in this case, the vowel hiatus is resolved by the insertion of a glide, either a /j/ or a /w/. When the stem ends with a high front vowel /ɪ/, /j/ becomes inserted; like the word *undian* ‘vote’ is realised as [undɪjɐn]. On the other hand, when the stem ends with a high back vowel /u/, /w/ becomes inserted, an example of this is in the word *perbatuan* ‘mileage’ which is realised as [pəbɐtuwɐn].

One more common phonological feature of Malay is /ɹ/ in stem-final position is only realised when there is a following vowel (Teoh, 1994:43), for example the word *kisar* ‘blend’ (noun) is realised as [kɪsɐ]. The same occurs when the suffix ‘*–kan’* is attached to the stem *kisar,* it becomes *kisarkan* ‘blend’ (verb) and it is realised as [kɪsɐkɐn] in which [ɹ] is dropped. However, when it is attached to the suffix *‘-an’*, it becomes *kisaran* ‘blended’ it is realised as [kɪsɐɹɐn]. These examples indicate that Malay is a non-rhotic language.

3.3.3. Phonotactics and Malay syllable structure

This section describes the phonotactic constraints of Malay and its syllable structures. According to Teoh (1994) the phonotactics of Malay does not allow for consonant clusters in either the onset or the coda position of the syllable. For example the borrowed word ‘film’ becomes ‘*filem’* in Malay, a schwa is inserted between the consonant sequences in the coda position, this word is realised as [fɪləm]. However, not every borrowed word will adapt to the phonotactic constraints as some of these words may be permitted in Malay, the examples of the words can be seen in section 3.3.

Another phonotactic rule available in Malay is that the phonemes /ʧ/, /ɲ/and /ʤ/ are not allowed in the coda position. However some exception applies in the case of /ʤ/ in which it is permitted in the loan words such as in the words *kolej* ‘college’ and *imej* ‘image’ (Teoh, 1994:55).

Phonotactically, according to Zaharani (1998) nasals that occur before plosives across a syllable boundary should be homorganic. For example the word *sumbang* ‘to contribute’ which is realised as [sum. bɐŋ], both [m] and [b] are bilabials. Another example of homorganic occurrence between plosives and nasals in Malay word is *bantal* ‘pillow’. Next, the syllable structure of Malay will be looked at.

Most of the monosyllabic stems in Malay have a CV structure, e.g. *ru* ‘casuarinas’, or CVC structure, e.g. *dan* ‘and’. According to Zaharani (2004) it is not possible to have a monosyllabic stem without an onset in Malay.

For polysyllabic stems in SM, Teoh suggested that there are four permissible syllable types: V, CV, VC and CVC. Zaharani (1998) observed that in polysyllabic words an independent syllable which consists of the nucleus only is permitted, for example the word *epal* [e.pal] ‘apple’ has a V.CVC structure where the initial syllable is made up of a nucleus. Compared to some other languages, Malay has a relatively simple set of possible syllable structures.

3.4. The prosodic features of Malay

The prosodic characteristic of Malay are somewhat understudied instrumentally. Most of the studies reported in the literature are based on auditory and perceptual observations. In this section, related studies on the prosodic feature of Malay are presented. These include the rhythm, and stress of Malay, two key aspects that are related to the current study.

Previous studies describing Malay rhythm are very limited and based only on auditory observation. Farid (1980), Baskaran (1982) and Teoh (1994) believe that Malay is a syllable-timed language; i.e they agree in claiming that the basis of rhythm in Malay is the isochronous productions of syllables. It is worth noting that they have not carried out any instrumental study in order to substantiate their claim and it is based purely on auditory judgement.

One of the studies of acoustic metrics for differentiating the rhythmic properties of languages, Grabe and Low (2002), included Malay as one of the 18 languages they investigated. Based on the calculation taken from consonantal and vocalic intervals from a single speaker of Malay, the results suggested that Malay is positioned in the stress-timed end of the PVI based continuum (see section 2.4.3).

Clearly this finding is completely the opposite of the auditory observations described above. However, the drawback of

Grabe and Low's (2002) study is that, it is based only on one speaker; and there is a possibility that the results could be influenced by individual differences in the way in which the materials were produced. Hence, it is not possible to say whether that study is able to reflect the rhythmic properties of speech in the general population of Malay speakers.

Based on the discussion in 2.2.2, stress is a prosodic feature related to the study of speech rhythm. As with the study of rhythm, the available literature does not provide sufficient background to the phonetic properties of Malay stress. Earlier studies like Farid (1980), Baskaran (1982), Suhaila (1994) and Teoh (1994) claim that, based on auditory observation, there is no systematic variation of syllable prominence in Malay connected speech. Suhaila (1994) further claims that stress has no particular function in Malay connected speech. However, the first systematic instrumental study of the phonetic properties of Malay stress was reported by Zuraidah, Knowles and Yong (2008).

Zuraidah et. al (2008) conducted an experiment to identify the stress characteristics of Malay words. This was a preliminary study prior to the compilation of a larger corpus for the analysis of the prosodic characteristic of Malay. Their investigation focussed on two features which they hypothesized might be correlates of the lexical stress in Malay: syllable duration and the variation of F_0 in the vowel in the final two syllables.

Zuraidah et. al (2008) studied 111 words produced in isolation by two female Malay native speakers. The material included simple and complex words as well as a few loan words from English; however it was not clear how many words were listed in each of the categories. The simple words had a maximum of four syllables, whilst the complex words were formed through four processes which were a) prefix + root, b) prefix + root +root, c) prefix + root + suffix, and d) prefix + prefix + root + suffix; the examples given were *mem* + *baca* ‘reading’, *ber* + *bagai* + *bagai* ‘a variety of something’, *mem* + *per* + *bagai* + *kan* ‘to have a variety’ respectively. Zuraidah et al. (2008) gave no account of how the words were distributed across different conditions, which made up the design of the experiment. The rationale for choosing these words appear to be to investigate the role of different morphological structures on the realisation of stress in Malay but what was noticeable about this study was that in the presentation of the findings of the study there was very little discussion of this.

The production of these words were recorded and digitized using MAC Speech Lab 2, they were segmented into 422 syllables. Each syllable was annotated with its duration and a phonological representation. Native speaker judgments on where the initial and final boundaries of the syllable were employed. For instance, when a glottal stop was inserted before the words with an initial vowel and between two vowels sequence, the inserted glottal stops are treated as the onset consonant in the CV syllable; as the word *istiadat* ‘ceremony’ is segmented as /ʔɪs. tɪ. ʔa. dat/.

Likewise when glides are inserted between two adjacent vowels in some of the words like *sosial* ‘social’ is segmented as /so. sɪ. jɐl/; the inserted glide is treated as the onset to the final syllable. These phonological words were then divided into CV(C) sequences. Nevertheless, an issue which can be raised here instance relates to segmentation; Zuraidah et al. (2008) did not provide a detailed description of how this was carried out. This is particularly problematic in the case of vowel-glide sequences where it is well known that drawing a boundary between a vowel and a glide is far from straightforward in spectrographic analysis.

Zuraidah et al (2008) focused their research particularly on the final two syllables of those words as they observed that noticeable prosodic changes in Malay words in isolation usually occur in these two syllables. Hence, the final syllable was labeled 0, the penultimate syllable 1, and the preceding syllables as 2, 3 and 4. For the F_0 plot, they observed that typically in 92 of the 111 words, the F_0 rises to a peak in the penultimate syllable followed by complete fall to the bottom of the speakers' range on the last syllable.

The measurements for syllable durations showed that final syllable is the longest with the mean of 411.41 ms. It has to be noted that this study measured the whole syllable, which included both the consonantal and the vocalic segments. The ANOVA test also revealed that syllable position had a significant effect on the syllable duration. Post hoc analysis revealed that the duration of the final syllable was significantly longer than the other syllables. The researchers also observed that there was no distinct variation in the realisation of vowels and consonants of Malay that could be attributed to the position of the syllable within the word.

The shortcoming of this experiment is that the morphological properties of these words were overlooked; there were imbalances in the selection of the words according to their morphological types, and when analysing the data, Zuraidah et al. (2008) conflated the different word types. By doing so, they missed the possibility of investigating the role of morphological complexity on the way in which the kind of prosodic structure of these words is realised. Therefore, different morphological word types should to be analysed separately in order to ascertain a stress related pattern corresponding to the types of words.

Consequently, there are also some reservations regarding the use of syllable duration to determine the citation patterns of Malay since naturally syllables with CVC structure have longer duration than the CV syllable structures, a predominant structure in Malay. Furthermore, it is also difficult to control and compare across different syllable structures.

Subsequently, using the same methodology, Zuraidah et. al (2008) analysed the connected speech data taken from a different study; Zuraidah and Knowles, (2006) which comprised a seven-hour recording of audio broadcast interviews. The participants consisted of nine educated Malay speakers aged between 45 and 53 years old. These participants spoke standard spoken Malay. However, similar to the data of the isolated words, there were also no description on how the words were distributed across different conditions and the types of sentences available in this recording.

Zuraidah et. al, (2008) found that final lengthening could only be observed at phrase final position; words produced in non-phrase final position did not show lengthening of final syllable as they were observed to have when spoken in isolation. Statistically, Zuraidah et al. (2008) also found that, in this continuous speech data, when they exclude the phrase final syllable in their calculation, the syllable positions within the phrase has no effects on the syllable duration and F_0 unlike the findings of the words produced in isolation.

At the same time, they also found that there there was variability in syllable duration, a finding which is contrary to the auditory-based claim made previously by Farid (1980), Baskaran (1982), Suhaila (1994) and Teoh (1994) that Malay is syllable-timed. Perhaps it is not suprising to find high variability of syllable duration across the data because different types of syllable structures would yield different durational values; since syllables with more sounds like CVC would have longer duration that those syllables with CV structures. Furthermore, in their analysis, Zuraidah et al. (2008) did not include comparison of syllable durations across the same syllable structures to further support their claim for non-isochrony.

Zuraidah et al. (2008) concluded that 'there are no phenomena in spoken Malay corresponding to what phonologists call stress' (2008:10). They noted that no prominent patterns of the pitch and duration were established in the wider context of connected speech. Ultimately, this raises a question as to how Malay speakers achieve the contrast between new and old information. This is something which we will return to following the results of the experiment presented in Chapter 4; which seeks to investigate this issue further.

To the best of our knowledge, Zuraidah et al (2008) is the only study which looks at Malay stress instrumentally, and as pointed out above the study does have some important limitations. In view of the fact that understanding lexical stress is probably important in relation to comprehending the rhythmic properties of Malay. It was decided to dig deeper into the area which Zuraidah et al. (2008) studied and to undertake a further study of lexical stress in Malay focussing on morphological structures and different numbers of syllables.

In conclusion, this chapter presents general background of the language and some of the aspects of Malay morphophonology relevant to this research. The next chapter aims to delve further into the area investigated by Zuraidah et al (2008) by taking into consideration the different types of morphological structures of Malay words and different number of syllables of the words which might influence the acoustic correlate of stress in Malay.

Chapter 4

Identifying Acoustic Correlates of Stress in Malay Words

4.1. Introduction

In order to understand rhythm, it is really important to know how stress operates in a particular language. As it is demonstrated in section 2.2, the features known to be the acoustic correlates of stress are duration, intensity and pitch (Fry, 1955; Lindblom et al., 2007; Jones, 2006). It is inherently important to know how these correlates work in a particular language since the same parameters may be well be implicated in determining the rhythmic properties of a language. The current situation for Malay is that the only instrumental study (Zuraidah et al., 2008) suggested that there is no stress in Malay. Hence there is a need to validate this claim. The concern about previous study is that they did not consider factors that might be important such as the morphological features of Malay. Therefore there is a need to carry out further work in order to validate what Zuraidah et al. (2008) found in a slightly thorough manner.

The discussion in the previous chapter (section 3.4) shows that Zuraidah et al. (2008) has shed some light on the stress patterns of Malay in both isolation and continuous speech contexts. However, the study was not without some limitation (section 3.4) and there is more to be investigated in relation to the acoustic correlates of Malay stress. This chapter aims to further explore the stress pattern of Malay by looking at other possible acoustic property that might be essential to Malay. At the same time, this experiment also served as a preliminary investigation to the second experiment in which to identify the acoustic correlates of Malay rhythm.

For the purpose of this study and following Fry (1955) this experiment will only explore the duration and intensity of Malay vowels as the acoustic correlates of stress. Although F_0 is another acoustic correlate associated with stress as demonstrated in languages like English and Polish (Dogil, 1999) (section 2.2.2); it was excluded from this experiment because it was rather clear that F_0 was found not to be significant in Malay connected speech; Even though there seemed to be a pattern in isolated words where F_0 reaches the peak on the penultimate syllable (Zuraidah et. al., 2008). The decision to include only duration and F_0 was supported further by Maris (1980), in which based on his auditory observation; the length and the loudness of the vowels – rather than pitch – are more prominent as a differentiator of the relative stress between the syllables in Malay speech (see section 3.6) furthermore, it is useful to study the role of intensity on Malay as it was not looked at instrumentally in Zuraidah et. al. (2008).

Thus, this experiment focusses on the duration and intensity of the vowels in the target words from three morphological environments as possible correlates of stress in Malay. The three categories of words are:

a) disyllabic monomorphemic words;

b) polysyllabic monomorphemic words;

c) polysyllabic morphologically complex words.

4.2. Methodology

This section describes the design of this experiment and provides detailed information about the process of data collection, participants, recording procedures and segmentation criteria employed in this study.

4.2.4. Materials

In this study, the participants were required to read out the list of target words in isolation and the list of sentences in which the target words were embedded. According to Ladefoged (2003) connected speech is more complex from the point of view of the measurements because of the variability of the data; therefore, prepared reading material was chosen over spontaneous speech. Although a read sentence approach might compromise the naturalness of the production, it was more structured and provided more consistency for data comparison across speakers.

4.2.4.1 The target words in isolation

There were four target words for each of the categories mentioned in 4.1. In total, there were twelve target words (see appendix A for the word list). The target words of the first two categories had an open final syllable, like the word *pedati* ‘cartwheel’. The last category had a closed final syllable, like the word *membidakan* ‘to bid’. Some of the words in the last category were from the root words available in the first two categories. These words were derived by the addition of suffixes, the example of the words in this category are *kemukakan* /kəmukəkɐn/ ‘to present’ from the stem word ‘*muka*’ and *membidakan* /məmbidɐkɐn/ ‘to nail’ from the stem word ‘*bida*’

The target words were chosen in such a way that the process of segmenting the recording later would be relatively straightforward. These words were constrained to contain a limited set of consonants, namely plosives and nasals. The rationale for this was that these manners of articulation have acoustic features that make them relatively easy to distinguish from the surrounding vowels (see section 4.2.4 for details on segmentation criteria). Words containing diphthongs and sequence of vowels were excluded from the selection of the target words because it was important to extract the acoustic measurements from stable single vowels for the purpose of consistency across all the target words.

4.2.4.2 The target words in sentence context

In order to compare between the target words in isolation and in connected speech, the reading material also included a list of twelve sentences. Each of these sentences contained one of the target words (see appendix B for the list of sentences).

All the sentences were constructed so that the words, preceding the target word, always ended in a vowel since all twelve target words began with a consonant. Restrictions were also placed on the first sound of the word following the target word. If the target word ended in an open syllable the following sound of the next word was a consonant; however, if it ended in a closed syllable, the next sound of the following word was a vowel (Table 4.1). These restrictions were necessary so that the target words could be easily distinguished from other words on the spectrogram.

Table 4.1.The examples of the selection of words preceding and following target words.

An example of a final vowel sound before the target word[*] and an initial consonant sound following the target word with an open final syllable[**]	Jangan biar[4] **paku** berselerak (Sentence 20) /dʒɐŋɐn bɪjɐ[*] **pɐku**[**] **b**ərsəlerɐk/ *Don't let the nail scatter around*
An example of an initial consonant sound following the target word with a close final syllable[♯].	Dia **kemukakan** agenda baru (Sentence 6) /dɪə **kəmukəkɐn**[♯] ɐdʒenda bɐru/ *He/She presents a new agenda*

As this experiment focussed only on the target words, some of the other words in these sentences contained semi vowels and vowel sequences.

4.2.5. The Participants

Three subjects participated in this pilot experiment. They were female students at the Newcastle University with ages ranging from 26 to 29 years old. None of them reported any kind of hearing or speech problems. They are native speakers of standard Malay from the same region i.e. the west coast of peninsular Malaysia. These subjects have been in the United Kingdom for a period of time ranging between two to four years.

Participants from the same demographic area were chosen in order to eliminate possible accentual variability. Furthermore, as the main focus of this thesis is to describe the rhythmic patterns of standard Malay, it was deemed necessary to use the same variety in this pilot investigation of Malay stress so that associations between these experiments can be made constructively.

[4] In standard Malay speech, 'r' in the final syllable of this word is silent and only appears orthographically (Yunos, 1980 and Teoh, 1994) (See section 3.3.1)

4.2.6. Recording procedure

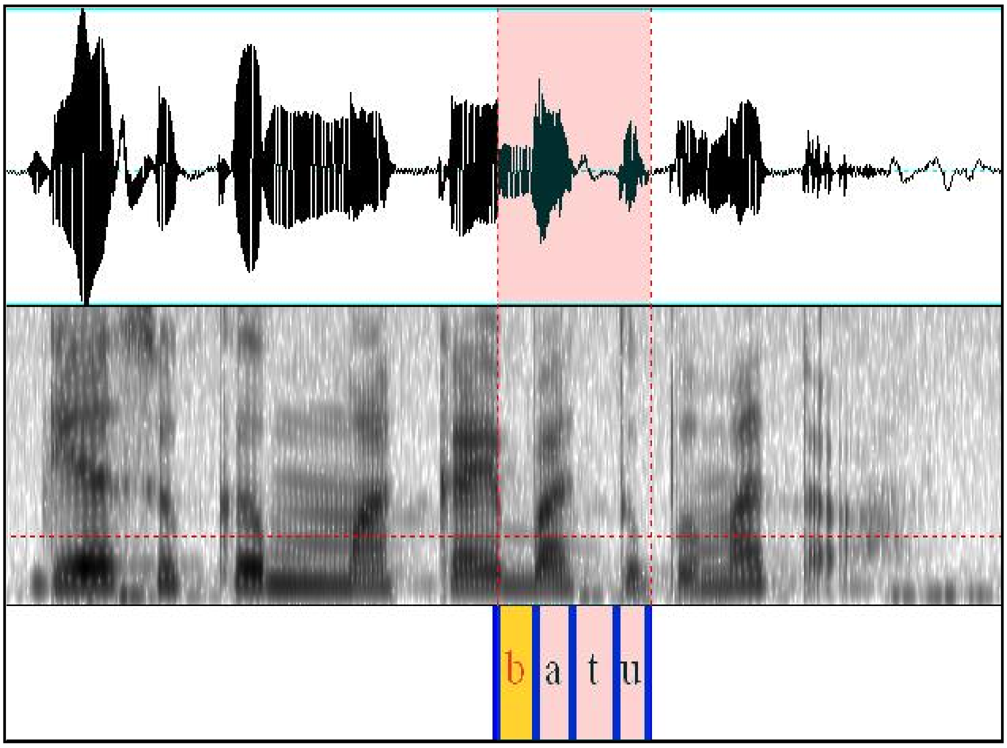
The recording session took place in the speech laboratory at King George VI Building, Newcastle University. A desk computer and an inline Logitech 320 microphone were used for the recordings. The microphone was positioned approximately 10 cm from the subjects' mouths.

Before the actual recording took place, the subjects practiced using the recording apparatus by producing random words and phrases so that they are familiar with the recording procedure. The subjects were given the materials (see appendix A for the copy of reading list given to the participants). They were instructed to produce speech at their normal speaking rate. These utterances were digitised at the sampling rate of 22 Hz, 16 bit Mono.

4.2.7. Segmentation criteria

The spectrograms of the recorded sound files were produced using PRAAT version 5.0.25 (Boersma & Weenik, 2006) with the window length set to 5 ms and the dynamic range to50 dB. The spectrograms are analysed using the Fourier method with Gaussian window shape. The TextGrid function was used to create a segmentation tier where the boundaries between consonantal and vocalic intervals could be labelled (Figure 4.1).

Figure 4.1 An example of the segmented target word in the sentence



As the target words comprised the combination of plosive or nasal consonants and vowels, this section describes how the vowels were acoustically distinguished from the consonants on the spectrograms.

4.2.7.3 The demarcation of plosive-vowel boundary

A plosive is produced when a complete closure is made in the vocal tract for a period of time and then the air is rapidly released. According to

Ladefoged (2005), there are typically three physical stages in the production of plosives and they are:

1. The approach stage when the active articulator approaches and makes a complete closure with the passive articulator;
2. The hold stage when the closure is maintained for a period of time;
3. The release stage when the active articulator moves away from the passive articulator.

These stages of the plosives could be identified clearly in the spectrogram. The examples of the voiceless and voiced plosives in CVC contexts are given in Figure 4.2 and Figure 4.3 respectively.

The onset of the voiceless plosive closure in CVC syllable type can be identified by decreasing the overall amplitude and accompanying it with a cessation of voicing. The spectrogram of the closure stage appears to have no energy in comparison to those of the neighbouring vowels. It is also important to note that according to Turk et. al (2006), some voiceless plosives have a voicing that can continue through the initial part of the closure phase which can be observed in the waveform (Figure 4.2).

Meanwhile, the onset of the vowel following the plosive is drawn at the release stage (Figure 4.2 and Figure 4.3). The acoustic correlate of the release stage was consistent with the appearance of bursts on the spectrogram which signified the end of the closure phase. This is following the criterion set by Turk et. al (2006) in which they identify the release phase of the plosive as a part of the adjacent vowel. This criterion is constantly applied throughout the data.

Figure 4.2 An example of a voiceless plosive in an intervocalic position

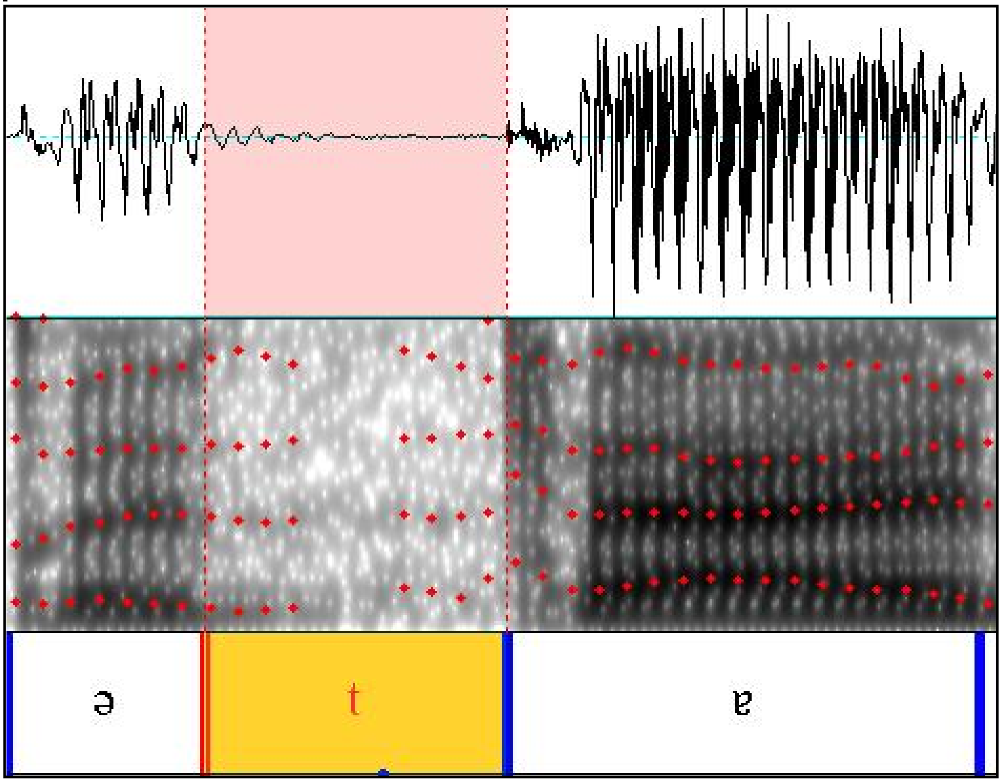
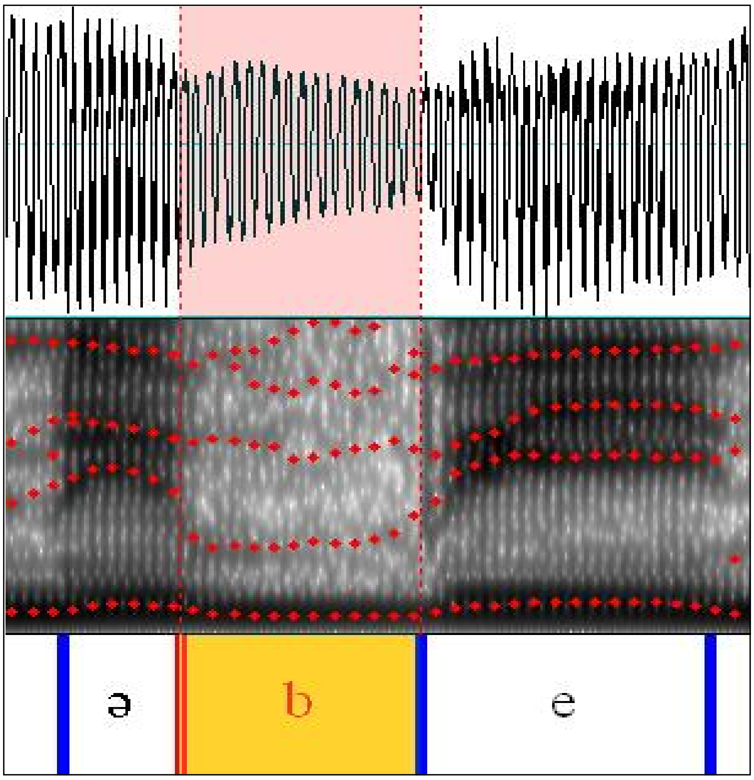


Figure 4.3 is an example of a voiced plosive in an intervocalic sequence. It can be observed that voicing continues throughout the intervals. To demarcate the onset of the closure, the F_2 is observed to the point where it starts to decrease. According to Turk et al. (2006) there is a dip in F_2 energy during the closure phase due to the closure of the vocal tract. The onset of the following vowel was drawn at the beginning of the release phase of the preceding plosive sound.

Figure 4.3 An example of a voiced plosive in an intervocalic position



It is also important to note that during the production of velar plosives: [k] and [g] in a pre-vocalic sequence, there is a tendency for several bursts to be produced in the release stage. According to Turk et. al (2006) these bursts are caused by an uncontrolled Bernoulli Effect. Bernoulli Effect is observed when a constant flow of air is passed through the oral tract and a section of the tract is constricted. At the point of constriction, the flow will speed up and there will be a drop pressure against the wall of the tract, this creates a small vibration, which was consistent with the bursts that appear in the spectrogram. In this situation, the offset of the plosive is drawn at the end of the first burst, which also marks the beginning of the following vowel.

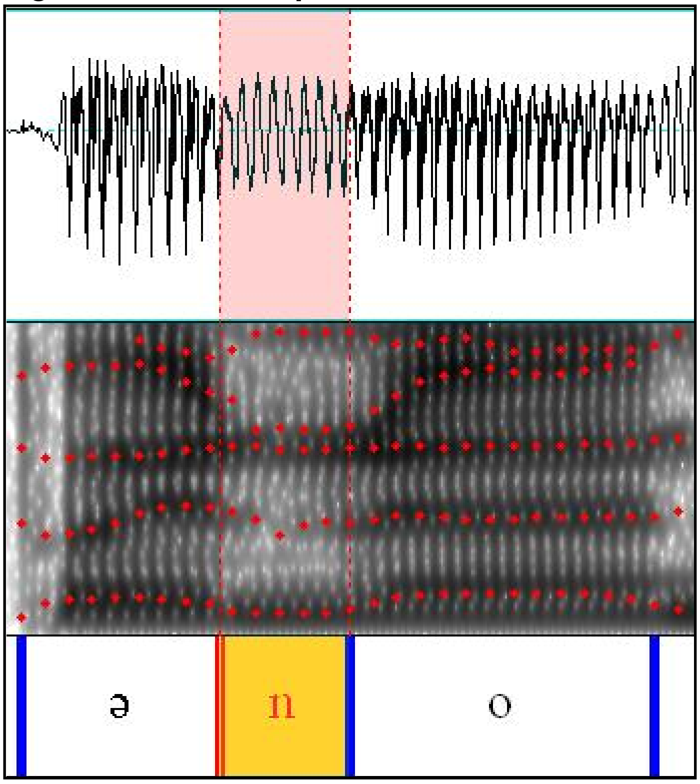
4.2.7.4 The demarcation of nasal-vowel boundary

A nasal is produced when the velum is lowered with a complete closure in the oral cavity and the air is let out through the nasal cavity (Ladefoged, 2005). Nasals are voiced; therefore, the vocal folds vibrate throughout their production. These vibrations generate sound waves and parts of these waves are absorbed by the cavity wall. This absorption is marked on the spectrogram by the appearance of antiformants. Figure 4.4 shows an example of antiformants for [n]. It is illustrated by the white space along the formants.

Similar to the production of plosives, the production of the nasal consonants also involves a complete closure, at some point, along the length of the oral region of the vocal tract. However according to Fant (1973), the principal difference between nasals and plosives is that for nasal consonants the veropharyngeal port is open during the time of the supraglottal closure. Thus, there is no increase in the pressure behind the constriction, and hence there is no decrease in overall amplitude (as can be seen on the highlighted nasal segment in Figure 4.4).

Figure 4.4 is an example of a nasal in an intervocalic position taken from the data. The onset of the nasal is marked at a point in the spectrogram where the appearance of antiformants could be observed. The offset of the nasal is fixed at a point where the antiformants disappear, and are replaced by the formant structure of the following vowel.

Figure 4.4 An example of a nasal in an intervocalic position



The next section discusses the measurement and the calculation applied to the segmented intervals in this experiment.

4.3. The measurements

After the spectrograms were segmented and labelled, these vowels were labelled V_1, V_2, V_3 or V_4 corresponding to their positions in the target words.

As for the measurement of vowel duration, the interval of the vowel was taken at the time between the onset and the offset of the segmented vowel. As a measure of intensity reading, the intensity contour of the spectrogram was used as measured by PRAAT. The default intensity range setting within PRAAT was applied in this analysis; the range was 50-100 dB. The maximum point of intensity within each word was identified as the intensity value of that vowel, then, the identified duration and the intensity of the segmented vowels were recorded in an Excel spreadsheet.

As described earlier in section 4.2.6, the distance of the microphone from the speakers' mouth was controlled; however, the researcher could not control the background noise and could not control the fact that these speakers have intrinsically quieter or louder voices. Therefore the pre-normalised intensity readings pooled across speakers were not really valid until they were statistically normalised by utilising the z-score results for comparison across speakers.

4.4. The statistical analysis

A non-parametric test was carried out to determine if the data was normally distributed to establish homogeneity of variance. Shapiro-Wilk test was used for a sample size smaller than 50 and Kolmogorov-Smirnov test was used for a sample size greater than 50. The results showed that each of the cases was significant ($p< 0.05$) which revealed the samples were normally distributed. The sample sizes for both the durations and the intensity of disyllabic monomorphemic words, the polysyllabic monomorphemic words and polysyllabic morphologically complex words were 48, 72 and 96 respectively.

A two-way analysis of variance (ANOVA) was also carried out for each of the word types in order to look at the interaction between the two independent variables: the vowel positions (V_1-V_4) and the types of production (isolated vs. Sentence). Post hoc (Tukey) analysis was also run to compare each mean against all the others in the data sets.

4.5. The Results

This section presents the findings of the two acoustic measurements: the vowel duration and the vowel intensity of the three types of the target words.

4.5.8. Disyllabic monomorphemic words

The means and the standard deviations of the vowel durations of the target words produced in isolation and in connected speech are given in

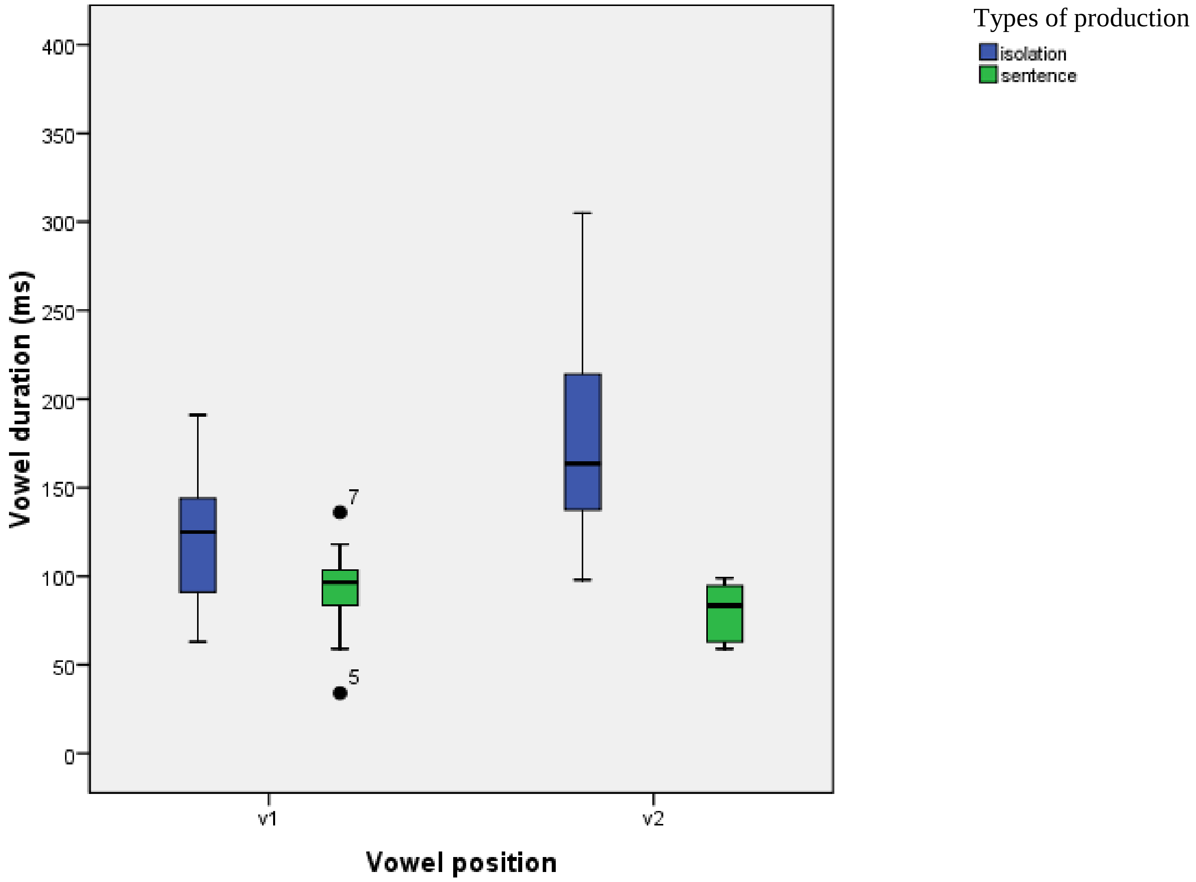
Table 4.2. The overall results are also shown in this table. The mean of the vowel durations in isolation was 151.38 ms (SD 61.30), whereas the mean of these durations in sentences was lower: 85.71 ms (SD 22.34). It was also observed that the standard deviation was greater for the vowel durations in isolation than that of the connected speech.

Table 4.2 The means and standard deviations of the vowel durations of the disyllabic monomorphemic target words

	Mean (ms)	Standard deviation
In isolation	151.38	61.30
In sentences	85.71	22.34
Overall	118.54	56.43

Figure 4.5 illustrates the boxplot of these durations in relation to the different vowel positions and types of production. It can be observed that, generally, the vowel durations produced in isolation are longer than those in embedded context. The vowel duration of V_2 produced in isolation was the longest, whereas the same vowel position in the connected speech was the shortest.

Figure 4.5 The boxplot of the duration of the vowels (ms) in isolation and in sentence across the syllables of the disyllabic monomorphemic target words



A two-way ANOVA was conducted to determine if the vowel position and the types of production affect vowel duration. The results showed a significant effect of vowel position on the duration, ($F_{(1,44)}$= 3.41, p< 0.05); likewise a significant effect of the types of production was found on the vowel duration ($F_{(1,44)}$= 29.30, p< 0.05), indicating that the vowels produced in isolation were longer than those produced in connected speech. Finally, there was no significant interaction between the vowel positions and the types of production, ($F_{(1,44)}$= 8.01, p> 0.05). The evident on the boxplot (Figure 4.5) suggested there was a tendency of interactions between the two independent variables however; the interaction was not registered as statistically significant.

The means and standard deviations of the pre-normalised intensity of target vowels in different production contexts as well as the overall results are given in

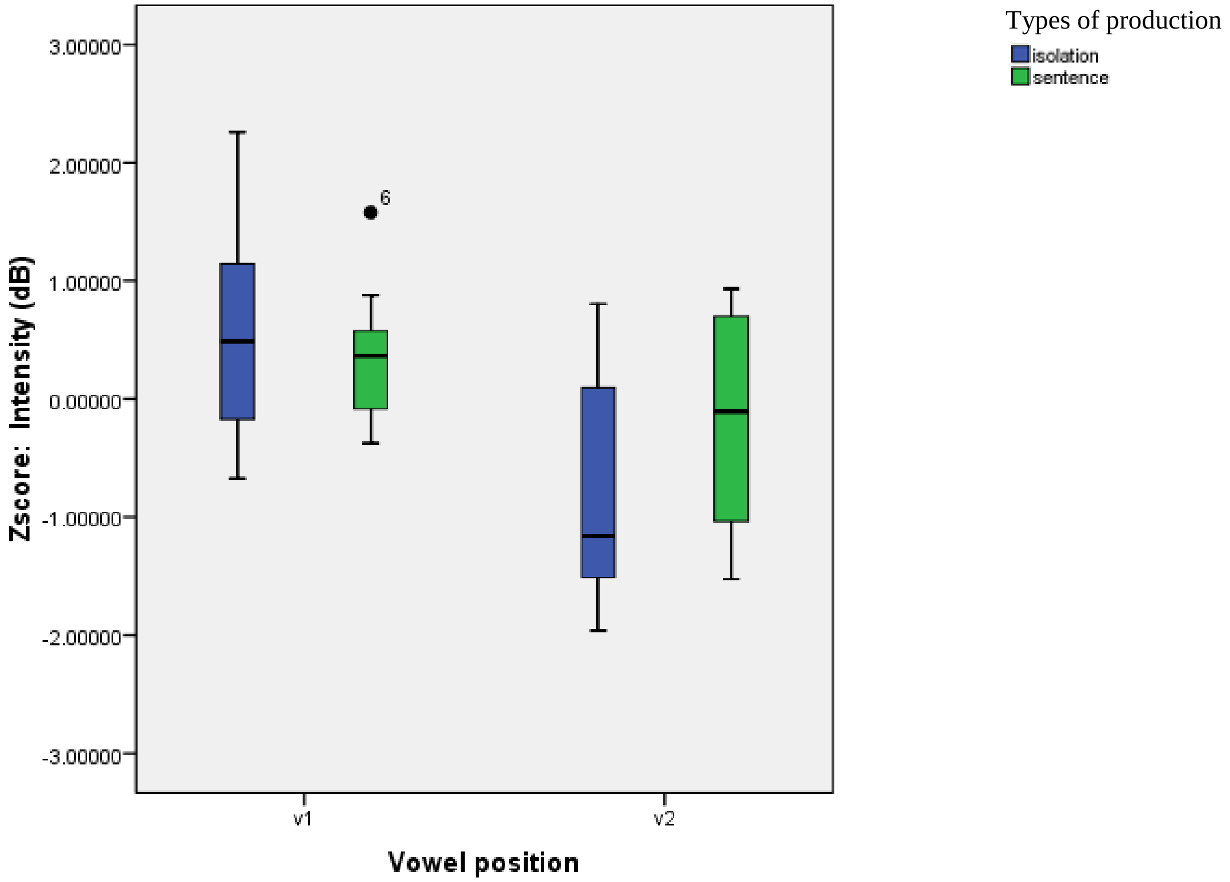
Table 4.3. Generally, it can be observed that the means of the pre-normalised intensity of the vowels are relatively consistent between the two types of production, where in isolation the mean was 66.75 dB (SD 5.03), and in connected speech it was 67.37 dB (SD 4.43). For the purpose of comparison across speakers, these intensity values were normalised using z-scores.

Table 4.3 The means and standard deviations of the pre-normalised vowel intensity of the disyllabic monomorphemic target words

	Mean (dB)	Standard deviation
In isolation	66.75	5.03
In sentences	67.37	4.43
Overall	67.06	4.70

The boxplot in Figure 4.6 illustrates the normalised intensity of the different vowel positions and different types of productions. In general, the intensities across the vowels in isolation were consistent, whereas the intensity of V_2 was longer than that of V_1 in connected speech.

Figure 4.6 The boxplot of the normalized intensity of the vowels in isolation and in sentence across the syllables of the disyllabic monomorphemic words



A two-way ANOVA was conducted to study the effect of vowel position and the types of production on the normalised intensity values. There was a significant effect of the vowel position on the normalised intensity, ($F_{(1,44)}$= 15.80, p< 0.01) where V_1 was louder than V_2. However, there was no significant effect of the types of production on the normalised intensity, ($F_{(1,44)}$= 0.62, p> 0.05). There was also no significant interaction between these two independent variables on the normalised intensity values, ($F_{(1,44)}$= 3.32, p> 0.05).

4.5.9. Polysyllabic monomorphemic words

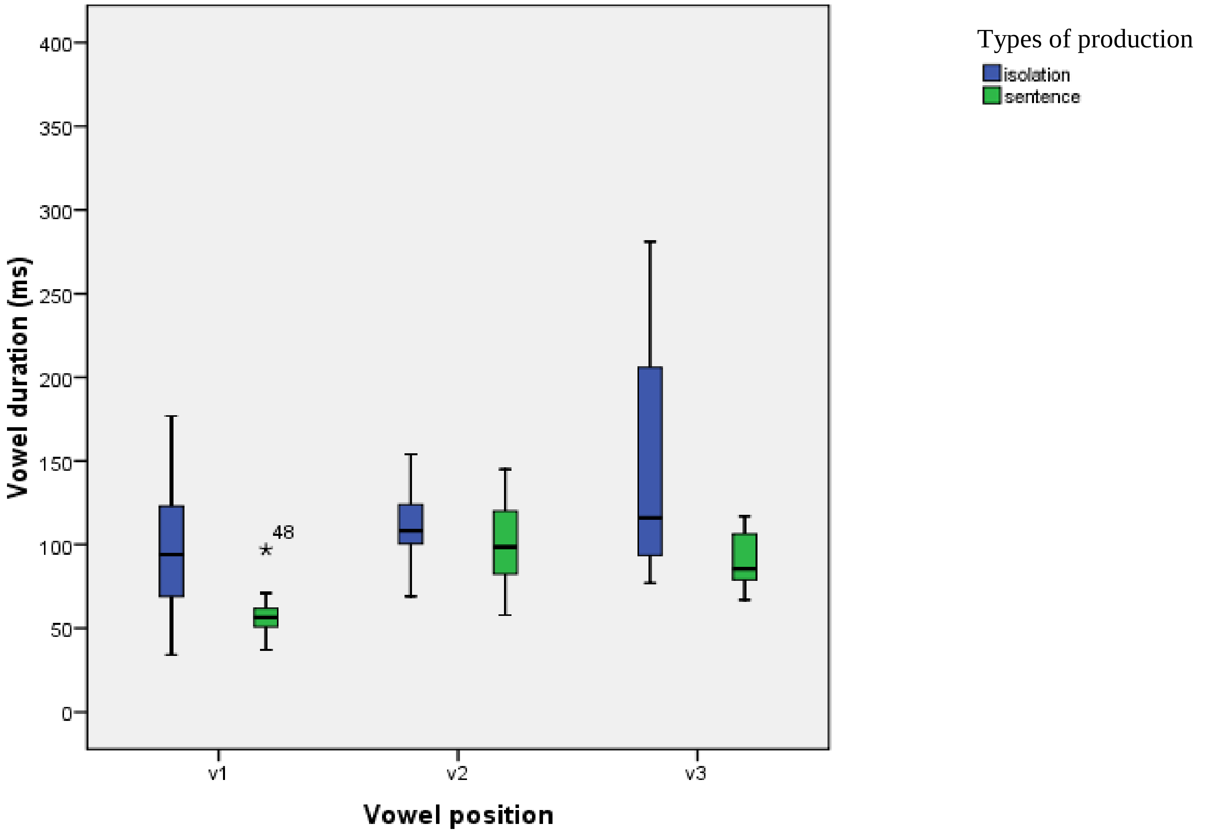
Table 4.4 sums up the mean durations and standard deviations of the vowel duration in the target words of the polysyllabic monomorphemic words. It can be seen that the mean vowel durations across the syllables were longer in isolated context with 119.7 ms (SD 53.06) than the mean vowel durations in sentence context with 82.67 ms (SD 26.45).

Table 4.4 The means and standard deviations of the vowel durations of the polysyllabic monomorphemic target words

	Mean (ms)	Standard deviation
In isolation	119.25	53.06
In sentence	82.67	26.45
Overall	100.96	45.52

The boxplot below (Figure 4.7) illustrates the vowel durations across syllables in different production contexts. Similar to Figure 4.5, the durations of the vowels across the syllables were longer in isolated condition than those in the embedded condition. It can be seen that the shortest duration was for the V_1 produced in the embedded condition and the longest duration was for the final vowel (V_3) produced in the isolated condition.

Figure 4.7 The boxplot of the duration of the vowels (ms) in isolation and in sentence across the syllables of the polysyllabic monomorphemic target words



A two-way ANOVA was conducted to determine if the vowel positions and the types of production in these target words affect the vowel durations. There was a significant effect of the vowel position on the vowel durations ($F_{(2,66)} = 8.05$, $p < 0.05$). Post hoc analysis revealed that V_1 was significantly shorter than V_2 and V_3. At the same time, a significant effect of different types of production on the vowel duration was also found, ($F_{(1,66)} = 17.06$, $p < 0.05$) in which the vowel durations of the target words produced in isolation were longer than those produced in sentences. Meanwhile, no significant interaction was found between the vowel positions and the types of production on the durational values ($F_{(2,66)} = 2.51$, $p > 0.05$).

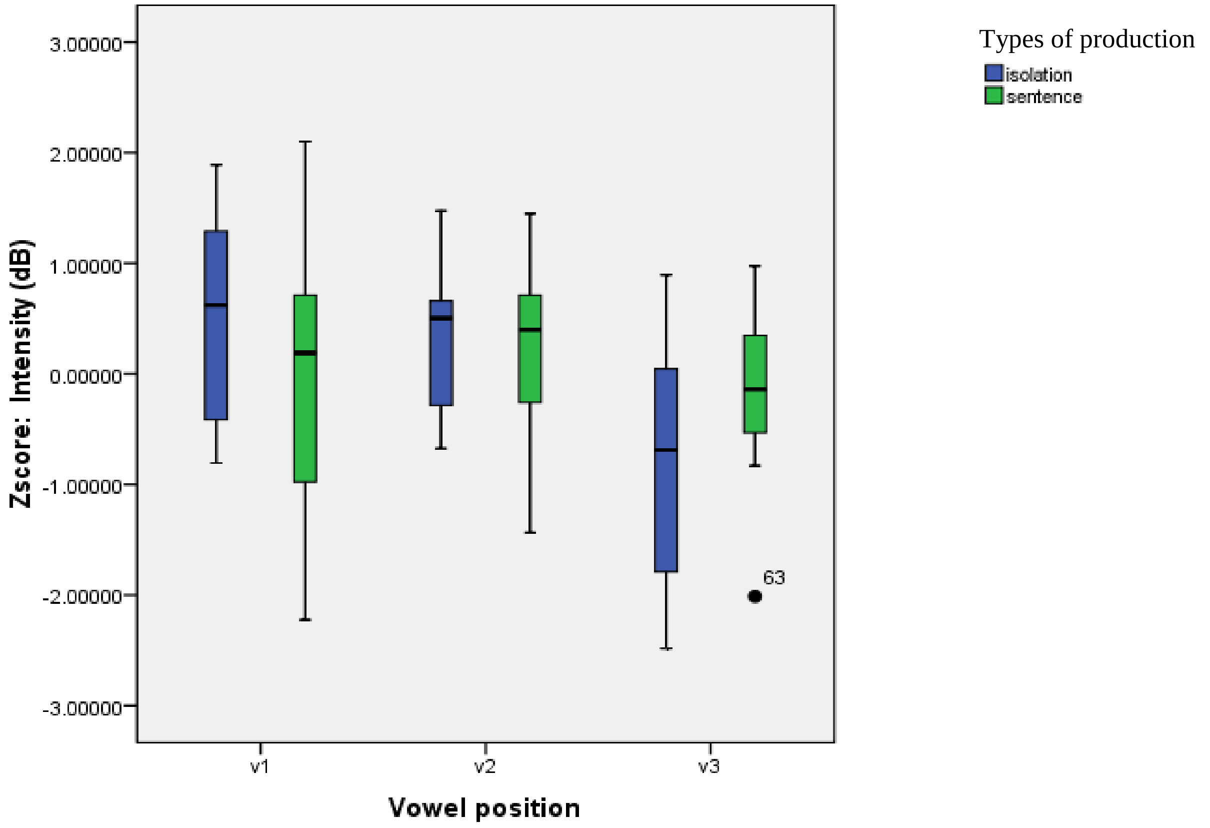
The means and standard deviations of the pre-normalised intensity of target vowels in the different production contexts, as well as the overall results are demonstrated in Table 4.5. Generally, it can be observed that the means of the pre-normalised intensity of the vowels were consistent between the two types of production in which, in isolation, the mean was 66.28 dB (SD, 3.82) and in connected speech it was 66.37 dB (SD 4.04). For the purpose of comparison, these pre-normalised intensity values were also normalised using z-scores.

Table 4.5 The means and standard deviations of the pre-normalised vowel intensity of the polysyllabic monomorphemic target words

	Mean (dB)	Standard deviation
In isolation	66.28	3.82
In sentences	66.37	4.04
Overall	66.28	3.63

The boxplot in Figure 4.8 illustrates the normalised intensity of the different vowel positions across both vowel contexts. In isolated context, the normalised intensity of the V_3 has the longest range across the vowels and that of the V_2 has the shortest range. Meanwhile, in the sentence context, the normalised intensity of the V_1 has the longest range and that of the V_3 has the shortest range. When the comparison of these two vowel contexts was made, it could be observed that the V_1 of both contexts have the highest values in the normalised intensity whereas the V_3 of those contexts have the lowest values in the normalised intensity. It is also noted that the normalised intensity of the V_3 in the isolated contexts has a longer range, in contrast to its counterpart in the sentence condition.

Figure 4.8 The boxplot of the normalized intensity of the vowels in isolation and in sentence across the syllables of the polysyllabic monomorphemic target words



A two-way ANOVA was conducted to study the effect of vowel position and the types of production on the normalised intensity values. There was no significant effect of the vowel position on the normalised intensity, ($F_{(2,66)}$= 5.15, p> 0.05). In addition, there was no significant differences between the types of production on the normalised intensity, ($F_{(1,66)}$= 0.48, p> 0.05). There was also no significant interaction between these two independent variables on the normalised intensity values, ($F_{(2,66)}$= 2.59, p> 0.05).

4.5.10. Polysyllabic morphologically complex words

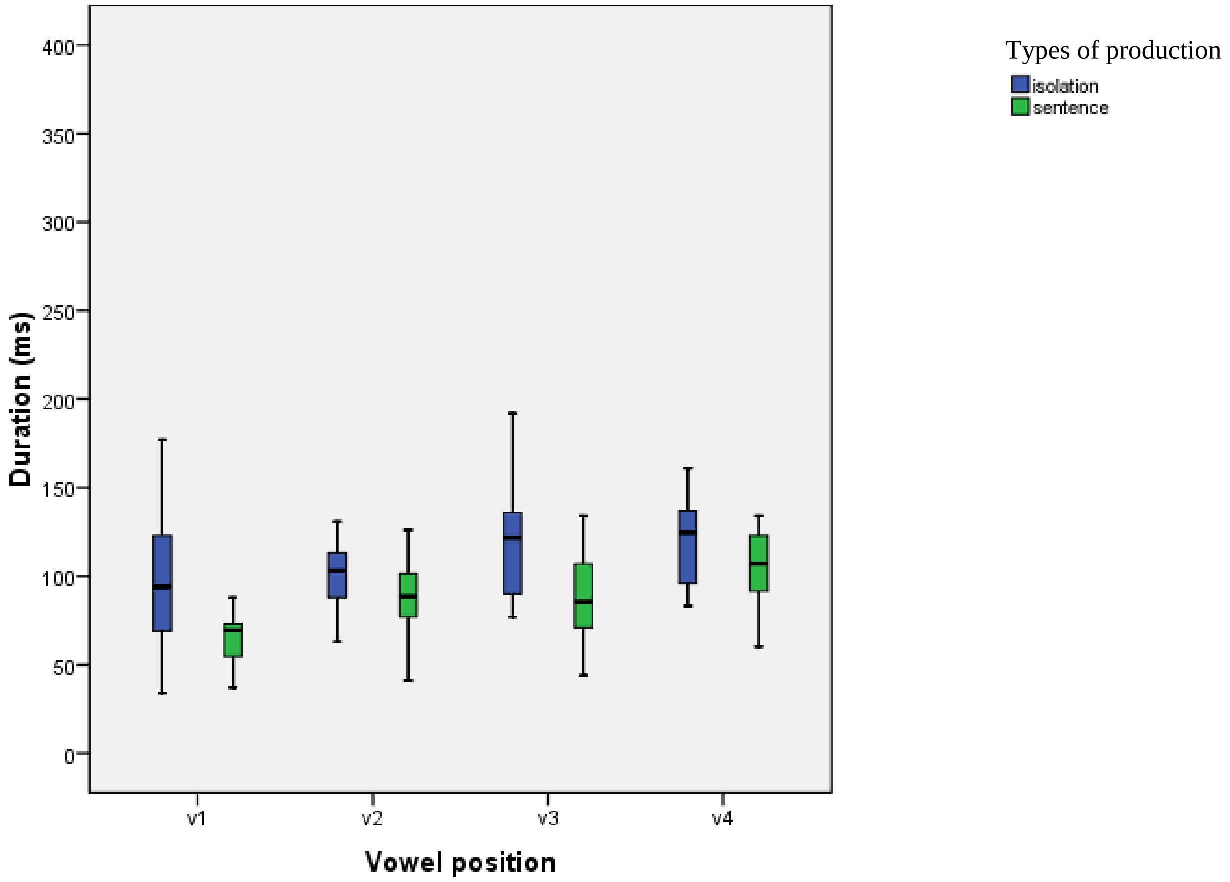
Table 4.6 shows the mean durations and standard deviations of the vowel duration in the polysyllabic morphologically complex target words. It can be seen that the mean vowel durations across the syllables are longer in the isolated context with 108.92ms (SD 31.96) than the mean vowel durations in the sentence context with 85.79 ms (SD 26.32).

Table 4.6 The means and standard deviations of the vowel durations of the polysyllabic morphologically complex target words

	Mean (ms)	Standard deviation
In isolation	108.92	31.96
In sentence	85.79	26.32
Overall	97.35	31.37

Figure 4.9 illustrates the vowel durations across syllables in different production contexts. Comparison of the types of production on the durational variability showed that there was more variability of the duration of V_1 in which the vowel duration in isolation was longer than that in connected speech. Meanwhile less durational variability was found across the vowels of the target words produced in connected speech.

Figure 4.9 The boxplot of the duration of the vowels (ms) in isolation and in sentence across the syllables of the polysyllabic morphologically complex target words



A two-way ANOVA was conducted to determine if the vowel positions and the types of production in these target words affect the vowel durations. There was a significant effect of the vowel position on the vowel durations ($F_{(3,88)} = 5.82$, $p < 0.01$). Post hoc analysis revealed that V_1 was significantly shorter than V3 and V4. At the same time, a significant effect of different types of production on the vowel durations was also found, ($F_{(1,88)} = 17.19$, $p < 0.01$) where the vowel durations of the target words produced in isolation were longer than those produced in sentences. Subsequently, no significant interaction was found between the vowel positions and the types of production on the durational values of the vowel, ($F_{(3,88)}$= 0.87, p> 0.05).

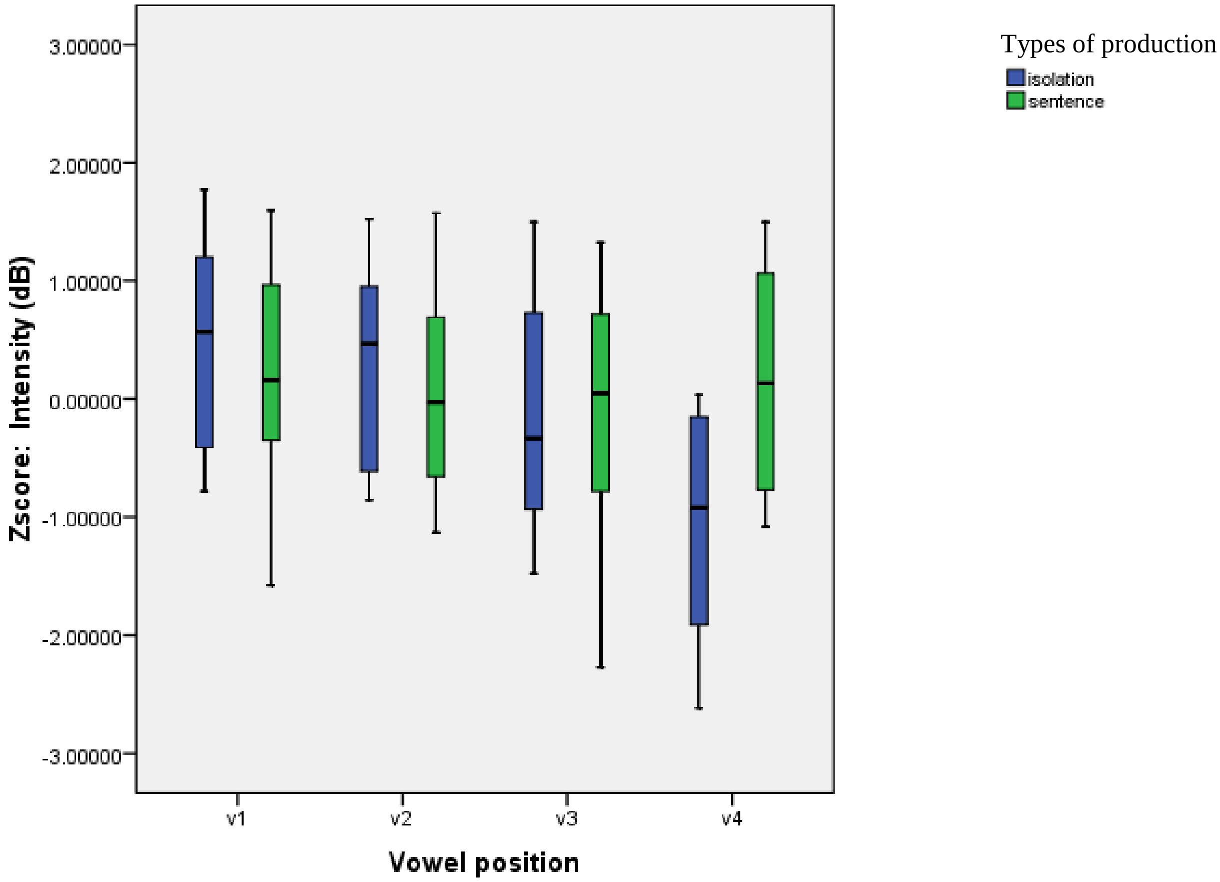
The means and standard deviations of the pre-normalised intensity of the target vowels in the different production contexts as well as the overall results are demonstrated in Table 4.7. Generally it can be observed that the mean of the pre-normalised intensity of the vowels were consistent between the two types of production, where, in isolation, the mean was 65.00 dB (SD 4.30) and in connected speech it was 66.71 dB (SD 3.76). For the purpose of comparison, these intensity values were also normalised using z-scores.

Table 4.7 The means and standard deviations of the pre-normalised vowel intensity of the polysyllabic morphologically complex target words

	Mean (dB)	Standard deviation
In isolation	65.00	4.30
In sentences	66.71	3.76
Overall	66.36	4.03

The boxplot in Figure 4.10 illustrates the normalised intensity of the different vowel positions across both types of production. The intensity of the syllables across the two production contexts was relatively consistent except for the intensity of V_4 in isolated context. This particular vowel has the lowest intensity from the rest of the syllables.

Figure 4.10 The boxplot of the normalised intensity of the vowels in isolation and in sentence across the syllables of the polysyllabic morphologically complex target words



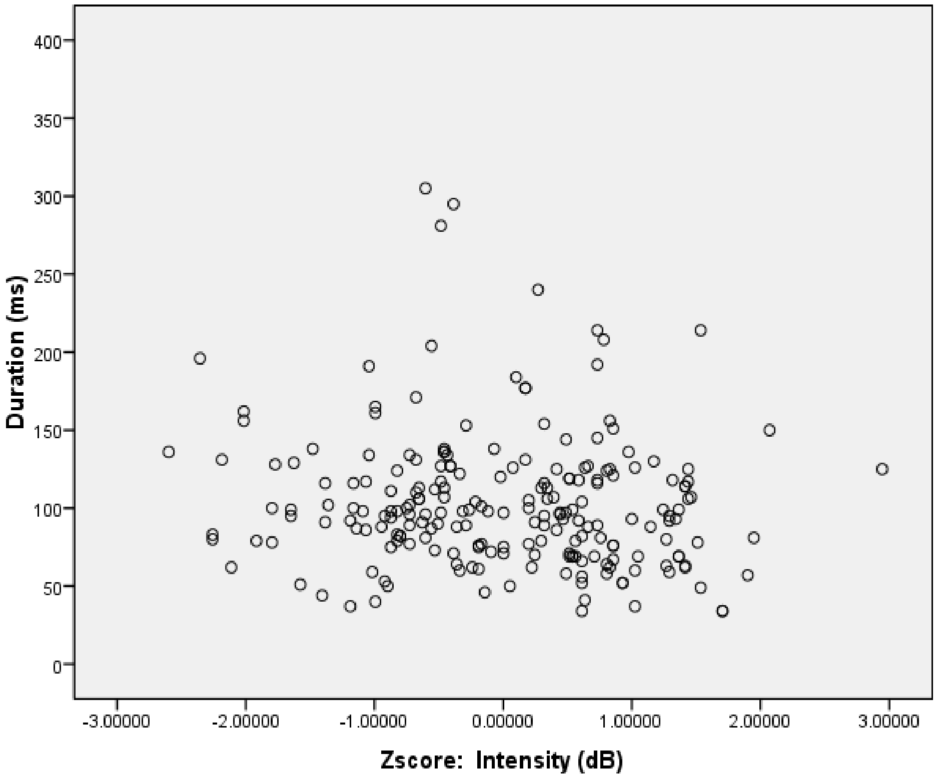
A two-way ANOVA was conducted to study the effect of vowel position and the types of production on the normalised intensity values. There was no significant effect of the vowel position on the normalised intensity, ($F_{(3,88)}$= 3.50, p> 0.05). No significant differences between the types of production in the normalised intensity was found ($F_{(1,88)}$ =0.87, p> 0.05). There was also no significant interaction between these two independent variables on the normalised intensity values, ($F_{(3,88)}$= 2.59, p> 0.05).

4.5.11. The relationship between the vowel duration and the normalised intensity

A correlation analysis was also carried out to see if there was any relationship between vowel duration and vowel intensity of Malay words. This step is conducted in accordance with Fry (1955) who found that both these acoustic features in English were correlated, where stressed vowels were longer and louder compared to other vowels in the same words. This was done in order to identify whether there was statistical evidence to suggest that these two acoustic characteristics could be associated to lexical stress.

The vowel duration and the normalised intensity across all the vowel positions and across all the three word types were computed and were analysed using the Pearson's correlation test. The result revealed a non significant, weak, negative correlation between these two variables ($r= -0.12$, $p= 0.428$). The scatter plot in Figure 4.11 sums up this correlation result. Based on this analysis, both the acoustic features in Malay seemed to be independent of each other.

Figure 4.11 Distribution across the three types of words over normalized intensity (Z-score: Intensity (dB)- Duration plane



4.6. Summary

For the vowel duration in both the disyllabic and polysyllabic monomorphemic words in isolation, what seemed to be consistent is the lengthening of the final vowel. This specific lengthening pattern might be considered initially as a cue for stress. The boxplot of the disyllabic monomorphemic words in isolation suggested a considerable difference between V_2 and V_1; meanwhile the post hoc analysis of the polysyllabic monomorphemic words produced in isolation revealed that the duration of V_3 was longer than V_1. However, it should be noted that final vowel lengthening is a common speech pattern across languages irrespective of whether the vowel is stressed or not (Klatt, 1979 and Crystal & House, 1988) this characteristic was discussed further in section 2.2.3), so it is possible that the observation made in this study is simply a reflection of this more general property of phonetic realisation.

Moreover, the other pattern that emerged in the polysyllabic morphologically complex words produced in isolation is the penultimate vowel duration (V_3), which was found to be longer than the first two vowels of the target words. The post hoc analysis revealed that V_3 was significantly longer than the preceding vowels, but no significant difference was detected between V_3 and the following (word-final) vowel. In terms of the morphological structure, the penultimate syllable of the polysyllabic morphologically complex words corresponded to the final syllable of the disyllabic monomorphemic words. This lengthening characteristic also showed that the prosodic characteristics of Malay words in isolation were strongly influenced by the morphological structure of the words.

This experiment suggested that there was stress in the target words produced in isolation. This might suggest a difference in the analysis of accent or stress within those morphologically complex words. However, both of these lengthening effects disappeared when the words were produced in connected speech; thus, the lengthening of the final vowel presented in the results could be a result of word final lengthening and not stress.

The results of this experiment show that the final vowels of both the monomorphemic word types produced in isolation were lengthened. However, these durational effects disappeared in connected speech and these were consistent with Zuraidah et al. (2008) who also found that word-final vowels were longer than other vowels in different positions of the words produced in isolated context. Similarly, Zuraidah et al. (2008) also found that this lengthening effect ceased to exist when the words were embedded in a phrasal context. Hence, the lengthening of final syllable appeared in both Zuraidah et al. (2008) and the current experiments is a consequent of finality marking and not related to lexical accent of the words produced in isolation. Further analysis also suggests that the word final lengthening feature is only found in both disyllabic and polysyllabic monomorphemic isolated words, whereas stem-finally lengthening feature was found for polysyllabic monomorphemic complex words in isolation.

For the normalised intensity, only the results for disyllabic monomorphemic target words were found to be significant, in which V_1 were louder than V_2. However, no other significant effects of vowel position, production type, subjects and morphological contexts were found on the normalised intensity. The consistency of the normalised intensity across the data proposed that intensity was not a robust indicator of stress in Malay.

A correlation test run on these two phonetic variables also revealed no relationship between the vowel duration and the normalised intensity. This outcome suggested that unlike in English (Fry, 1955) these two phonetic cues did not have any association in Malay that could lead to acoustic correlate of stress.

With regard to the acoustic correlates of stress in connected speech, the findings in this experiment were consistent with the auditory observations of Farid (1980), Suhaila (1994) and Teoh (1994) which led to the claim that there was no durational and intensity prominence within Malay phrases (see section 3.4). This production experiment revealed that connected speech of Malay does not have stress as conventionally defined in previous literature such as in Fletcher (2010) but it does have significant phrase-final lengthenining line with what has been found in a number of other languages such as Finnish (Suomi, 2007). However, this finding does raise an interesting question about how Malay speakers accomplish the pragmatic function of distinguishing between old and new information within speech which in many other languages is tied to the phrase-level prominence contour. Zuraidah (1996) provides account about how the pragmatic difference between old vs. new information is conveyed within continuous speech in Malay.

According to Zuraidah (1996), there is a tendency for the speakers to increase their tempo before reaching to the new information that they want to disseminate to the listeners. She claimed that there is a fall-rising pitch and increased loudness on these demonstrative pronouns rather than on the major lexical items such as nouns, verbs or adjectives, which carry new information to the conversation The practice of inserting demonstrative pronouns such *ni* ‘this’ and *tu* ‘that’ after the production of the new information is also a useful cue for the listeners. It is also important to note this observation was based on auditory information; it would be useful in future studies to test these claims instrumentally.

To illustrate, consider the excerpt taken from Zuraidah (1996:256) below:

A	*Jalan*	*tak*	*jam*	*ya*	*Doktor*
	Road	not	jammed	yes	Doctor
B	*Nampak*	*pagi*	***NI***	*bagus*	
	Look	morning	**THIS**	good	

A The road is not congested, is it Doctor
B It looks good this morning

In this conversation between two people above, A and B, in which B is a doctor about whether if the road was congested where B was in the morning. To introduce new information to A, B inserts the demonstrative pronoun *ni* 'this' before the adjective which carries the new information.

Another pragmatic characteristic used by Malay speakers to differentiate new and old information according to Zuraidah (1996) is the insertion of grammatical words such as *eh* and *lah* after the intended new information. Although there is no systematic phonetic patterns in the realisation of Malay stress at phrasal level, the speakers differentiate between new and old information pragmatically by assigning demonstrative nouns and grammatical items after the intended new information.

This experiment only looked at one facet of the acoustic correlate of stress which was the production aspect of the speech however in order to get an overall picture of Malay stress, there is a need to look at the production side in more natural, unscripted materials, and to look at the perceptual dimension, in order to test the extent to which Malay listeners perceive differences in prominence across the syllables within a word. It is also interesting to see how

Malay listeners respond to material from languages, such as English or Spanish, which do have a very clear correlates of lexical and phrase-level stress.

As stress is a concept related to rhythm; the results of this experiment will be taken into account when describing the rhythmic characteristic of Malay in the next chapters.

Chapter 5

Acoustic Measures of Malay Rhythm: The Methodology

5.1. Introduction

This chapter explores the methodology used in the collection of data for the experiment conducted to identify the acoustic measures of Malay rhythm. This includes the materials used, the background of the participants, the recording procedure, the segmentation criteria to demarcate the vowels and the consonants, and the calculations and statistical analysis applied to the vocalic and consonantal intervals.

5.2. The Materials

Similar to the first experiment, prepared reading material was chosen over spontaneous speech. Although a read sentence approach may compromise the naturalness of the production, it was rather structured and provided more consistent data for comparison across speakers.

Participants were recorded producing the reading list consisting of ten sentences. The selection of the consonants used in the materials was controlled throughout. Glides and semi vowels were excluded from this list. The decision to exclude these sounds was due to foreseeable difficulty in which could be encountered when segmenting the acoustic signal of these sounds particularly if they were in an intervocalic positions. However there were some exceptions, for example in sentence 4, orthographically *mengukur* ‘to measure’ has a final alveolar approximant [ɹ]; however, phonologically, this particular sound was not realised in this variety of Malay, thus the phonological transcription was /məŋuko/. These precautions were taken in order to eliminate any ambiguous and confusing spectrogram so that a straightforward analysis could be carried out (see section 3.3.1).

The shortest of the ten sentences comprised 20 consonantal and vocalic intervals (sentence 6) and the longest sentence comprised 52 consonantal and vocalic intervals (sentence 10). Table 5.1 indicates the distribution of both the intervals across the sentences.

Table 5.1 The number of consonantal and vocalic intervals distributed across the sentences

Sentences	1	2	3	4	5	6	7	8	9	10
Intervals										
Consonantal	20	14	15	18	14	10	17	19	17	26
Vocalic	22	15	15	17	15	10	16	19	16	26

The sentences included all the possible syllable structures which were permitted in Malay: CVCV, V.CVC, CVC, CVCC (Ismail, 1994) (see section C of the appendix for the full list of the materials).

5.3. The Participants

Twenty participants were selected to participate in this study. They were ten male and ten female subjects. Their ages were between 22 and 34. These subjects spoke the standard variety of Malay as their L1. They were students of the University of Malaya from various schools. All of them reported normal hearing and no history of language problem in their L1.

These subjects came from the same region of the west coast of Malaysia. This was consistent with the demographic background of the participants from the preliminary study so that a legitimate relationship could be established between these two experiments in the discussion chapter. Another reason for carefully selecting the subjects was to eliminate discrepancy of different features of speech production which might arise if the subjects were not the native speakers of the standard variety of Malay. Inconsistency of the subjects' selection between the first and the second experiments could lead to invalid observation of the investigation.

5.4. Recording procedure

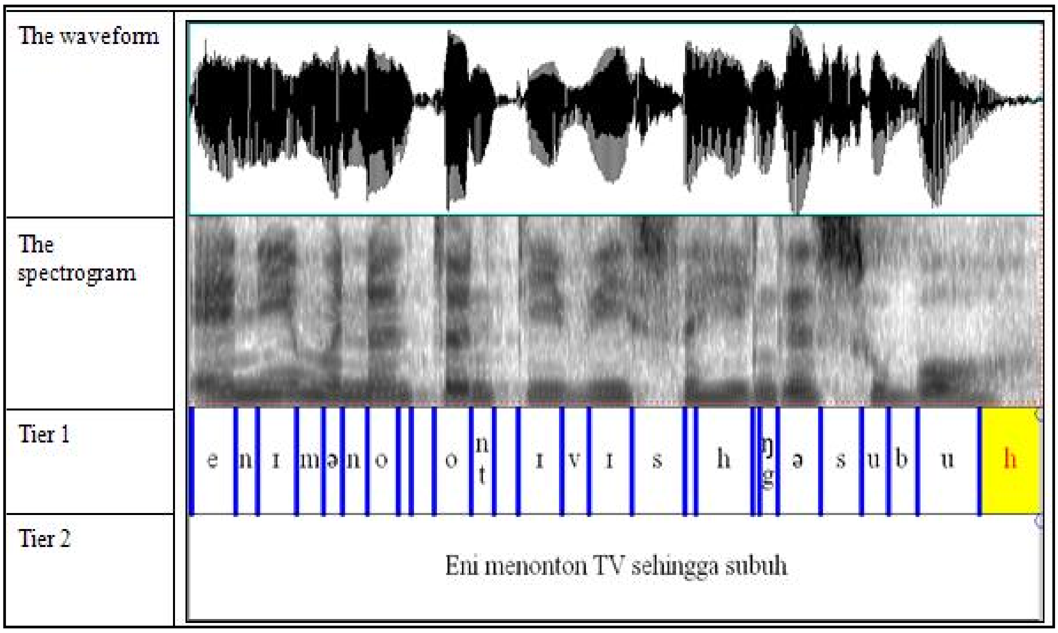
This section looks at the details of the recording sessions of the main study. The sessions took place in a speech lab in the faculty of Languages and Linguistics in University Malaya. A computer equipped with computer language recording (CLR) software, model 4500 KAY, together with a CD 2688M microphone and speakers were utilised in this experiment. The speech software PRAAT 5.0.25 (Boersma and Weenik, 2006) was used later to process the recordings.

During the recordings, the microphone was positioned approximately 10 cm from the subjects' mouths. A controlled length between the subjects' mouths and the microphone was to reduce any noise and irregularity due to voice quality of the subjects. Before the actual recording took place the participants produced random utterances in order to familiarise themselves with the session. The subjects were asked to read the sentences naturally at their normal speaking rate as suggested by Ladefoged (2005:94) in order to ascertain data close to the natural speech. The recorded speech materials were digitised at 22 kHz, 16 bit-Mono.

The spectrograms of the recorded sound files were produced using PRAAT version 5.0.25 (Boersma & Weenik, 2006) with the window length set to 5 ms and the dynamic range to 50 dB. The spectrograms were analysed using the Fourier method with Gaussian window shape.

The spectrograms were annotated with two text grids: the first text grid was for the segmented lines together with their broad phonetic transcription, and the second text grid was for the orthographic transcription of the sentence (Figure 5.1)

Figure 5.1- An example of how the spectrogram is annotated to two tiers in Praat



In order to carry out the calculations of the rhythm metrics, the onset and the offset of each of the consonantal and vocalic intervals were identified by the inspection of the wide-band spectrogram and the waveform following the criteria used in the preliminary study (see 4.24). The next section explores other criteria of sound segmentation not discussed previously.

5.5. Segmentation criteria

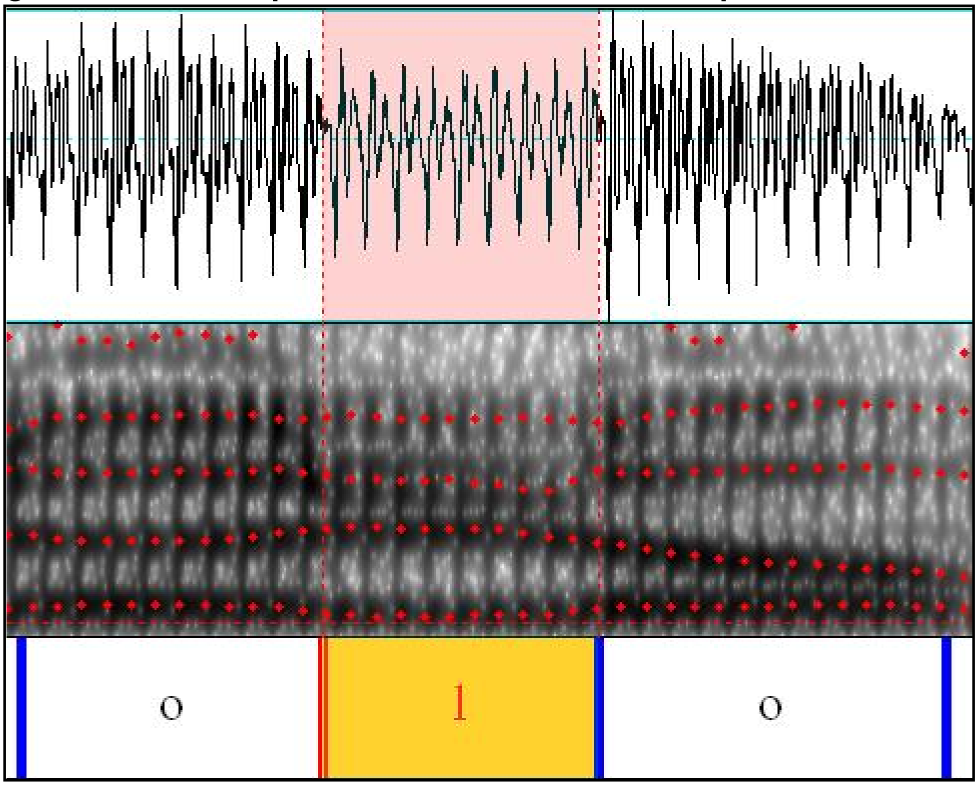
This section presents the criteria applied to segment the consonantal and vocalic intervals of the recordings. The consonantal sounds included in the materials of this second experiment were plosives, nasals, laterals, fricatives and affricates. The same criteria for segmenting plosives and nasals in the first experiment were applied in this study. The characteristics of the demarcation of the two consonantal sounds were discussed in sections 4.2.4.1 and 4.2.4.2 respectively. This section discusses the demarcation characteristics of the rest of the consonant-vowel sounds.

5.5.1. The demarcation of the lateral-vowel boundary

Ladefoged (2005) defined lateral as a sound produced by the closure of the central body of the tongue and air was let out sideways. This constriction was not sufficiently narrow to affect the glottal airflow through the vocal cavity or to cause any turbulence. Due to the closure along the middle of the vocal tract, some of the air was not allowed to escape; thus, it was absorbed by the cavity wall. As a result of this absorption, antiformants appeared acoustically. Antiformants were characterised by the white lines which appeared on the spectrogram. A detailed description of antiformants was given in 4.2.4.2. Apart from the antiformants, the steady vertical striation was another acoustic property of the laterals that could be observed on the spectrogram because they were voiced. The laterals were also identified by the existence of spectral discontinuity caused by the blockage of air down the central of the oral cavity (Stevens 1997:25).

In intervocalic position (Figure 5.2), the existence of the lateral could be observed when the spectrogram energy was somewhat reduced compared to the energy manifested by the two corresponding vowels. An onset of the lateral could be drawn by observing the spectral discontinuity following a vowel-like formant pattern, whereas the onset line of the following vowel was drawn when a more distinctive vowel-like formant pattern began after the spectral discontinuity.

Figure 5.2 An example of a lateral in intervocalic position



5.5.2. The demarcation of fricative-vowel boundary

According to Ladefoged (2005) fricative consonants were produced by forming a narrow constriction along the length of the vocal tract, generating turbulence noise in the vicinity of this constriction. At the same time some adjustment of the glottal opening was made to maximise the amplitude of the turbulence noise that is generated near the supraglottal constriction (Stevens 1997). The constriction, along with the high turbulence velocity of the air that passed through, was typically signalled as an aperiodic waveform and a noisy spectrogram acoustically. The highlighted segments in Figures 5.3 and 5.4 are examples of these acoustic properties of voiceless and voiced fricatives respectively. According to Turk et. al (2006), noise frication on the spectrogram was the most identifiable acoustic criteria of both the voiceless and voiced fricative.

The spectrogram (Figure 5.3) is an example of a voiceless alveolar fricative in intervocalic position. The onset of the fricative was marked on the spectrogram at a point in which the appearance of the high frequency noise could be observed. In addition, the offset of the fricative was fixed at a point at which the noise has declined in amplitude, and it was replaced by the formant structure of the following vowel.

Figure 5.3 An example of a voiceless alveolar fricative in intervocalic position

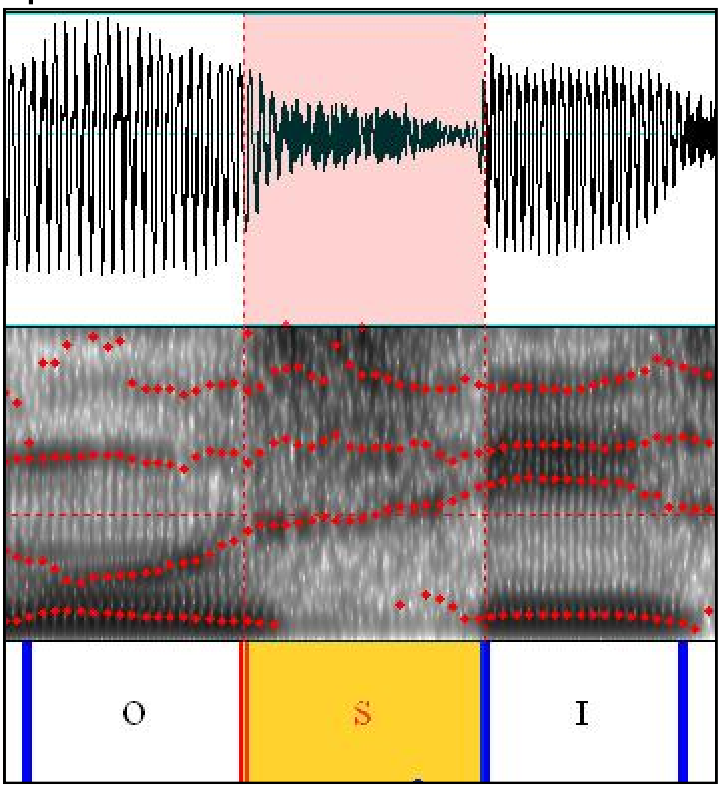
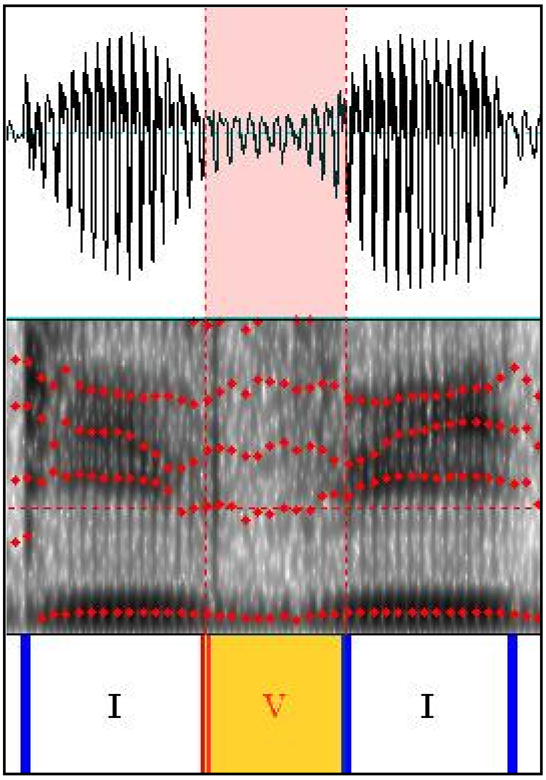


Figure 5.4 is a spectrogram of a voiced fricative. Due to the vocal folds vibrating during a typical production of a voiced consonant (Ladefoged 2005),

it was quite clear that during the production of the fricative, there was a decrease in the amplitude of the formants of the vowel which corresponds to the interval. The onset of the fricative was marked on the spectrogram at a point in which the appearance of the frication noise began. Meanwhile the offset of the fricative was drawn when the regular striation of the waveform began.

Figure 5.4 An example of a voiced labiodentals fricative in intervocalic position

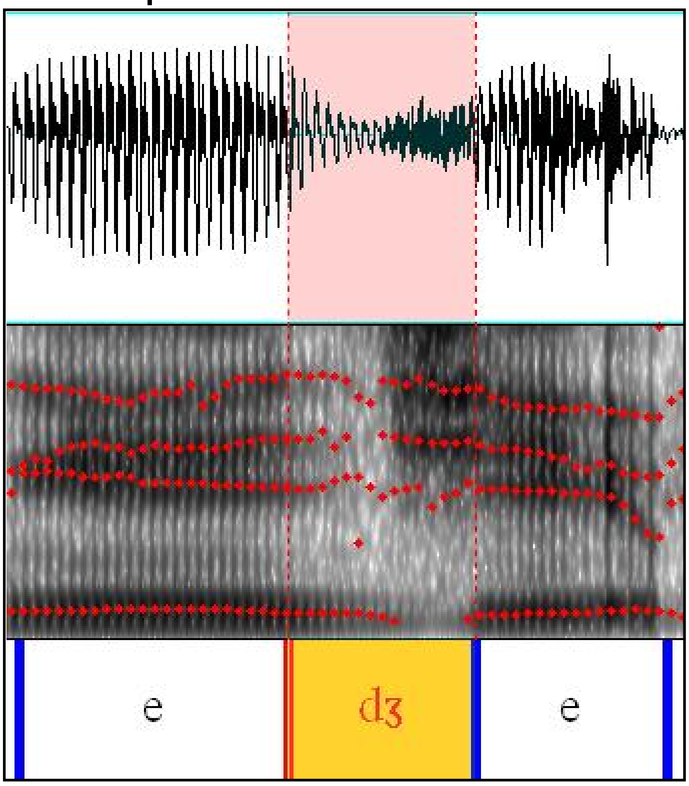


5.5.3. The demarcation of affricates and vowel boundary

An affricate was generally produced when there was a built up of air pressure behind a complete closure in the vocal tract and it was gradually released. This release produced a plosive but the separation which followed was sufficiently slow to produce audible frication. This frication was similar to that of the fricative consonant, however duration of this particular frication was somewhat shorter that the individual fricative sound (Crystal, 2003:16).

According to Turk et. al (2006), the criteria to identify the onset for the affricates were similar to those for the plosive, whilst the criteria for identifying its offset were similar to those for fricatives. The highlighted segment in Figure 5.5 is an example of a voiced palatoalveolar affricate. This sound was produced when a closure was made in the front part of the tongue blade at a point immediately posterior to the alveolar ridge. Following the initial release of this closure, the tongue blade assumed a shape and position similar to that of a palatoalveolar fricative. The offset of the preceding vowel was marked on the spectrogram at a point in which the vowel-like formant spectrum disappears, and it was replaced by a closure like spectrum corresponding to the air turbulence behind the constriction. Since the vocal folds vibrated throughout the production of a voiced fricative, the continuous friction appeared on the spectrogram. The onset of the following vowel was fixed at a point in which the noise frication declined and a vowel-like formant structure appears.

Figure 5.5 An example of an affricate in an intervocalic position



5.6. The rhythm metric measurements

Replicating the methodology used in Ramus et. al (1999) and Grabe and Low (2002), A consonantal interval comprised a single consonant sound or a sequence of consonant sounds. The same characterisation was also applied for a vocalic interval. In the cases where there was a glottal stop insertion on vowel hiatus, the vowels were considered as two vocalic intervals while the glottal stop was considered as a consonantal interval (see section 3.3.1).

The durations of the final syllable of the last words in these sentences were excluded. This was done to avoid a final lengthening effect. Deterding, (2002) pointed out that final vowel lengthening is one of the marked features of the languages in South East Asia. Plosives in initial position also were not included since it was difficult to determine the beginning of the closure when it is in the initial position.

SPSS software was used to generate the data needed for the rhythm metrics: ΔC, ΔV and %V calculations. A spreadsheet was created with four defined columns. The columns were the sentence numbers (coded 1- 10), subject numbers (coded 1- 20), and the interval types (coded 1- consonantal intervals and 2 for vocalic intervals)

5.6.1. ΔC, ΔV and %V.

The values of ΔC were derived by calculating the standard deviation of the consonantal interval durations. Likewise, the values of ΔV were derived by calculating the standard deviation of the vocalic interval durations. Meanwhile, the %V is the sum of vocalic intervals divided by the overall duration of the intervals and multiplied by 100.

The calculations of these rhythm metrics were carried out for each of the subjects producing the sentences.

5.6.2. *r*PVI and nPVI.

For this research, the term 'intervocalic *r*PVI' and 'vocalic *n*PVI' used in Grabe and Low (2002) were referred to as *r*PVI and *n*PVI respectively. The rhythm calculation of the *r*PVI and *n*PVI was derived by replicating the formula used in Grabe and Low (2002). The values of *r*PVI were derived by calculating the differences in duration between successive consonantal intervals. The formula is as follows:

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

Where *m* is the number of intervals, in the case of above equation it is referred to as the consonantal intervals, *d* is the duration of the *k*th intervals. The calculation of *r*PVI was not normalised.

Meanwhile, *n*PVI was calculated by mean of difference between successive vocalic intervals divided by the sum of the same interval in order to control speech rate. For this formula, the duration of successive vocalic intervals were entered using the following equation, for the *n*PVI calculation, *m* was referred to as the consonantal intervals;

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

This formula was programmed into an Excel spreadsheet. In order to obtain the values of *r*PVI for this study, the readings of the consonantal interval duration were entered into the Excel spreadsheet.

5.6.3. Speech rate

To calculate the speech rate, the total duration of the sentence in seconds was divided by the number of the syllables. The total durations did not include the final syllable and the first syllable beginning with initial-plosive syllable in some sentences. This exclusion was consistent with the criteria described in section 5.6.

5.7. The statistical analysis

A non-parametric test was carried out to determine if the data was normally distributed for each of the variables. This was conducted in order to verify the homogeneity of variance. Kolmogorov-Smirnov test was used for the sample sizes in this dataset. The sample sizes for the consonantal and vocalic intervals were 3400 and 3420 respectively. The results showed that each of the cases was significant ($p < 0.05$) which revealed that the samples were normally distributed.

A two-way analysis of variance (ANOVA) was also carried out for each of the word types in order to look at the interaction between the two independent variables: the vowel positions and the types of production. A post hoc (Tukey) analysis was also run to compare the means in which the degree of variability, if any, could be determined.

Chapter 6

Acoustic Measures of Malay Rhythm: The Results

6.1. Introduction

This chapter is divided into two main sections. The first half of the chapter presents the findings of the experiment described in chapter 5 and discusses the results for each of the rhythm metrics (ΔC, %V, ΔV, *r*PVI and *n*PVI) investigated. This section also includes analysis of the relationship of these rhythm metrics with speech rate and considers other subject variation.

The second half of this chapter looks at the evaluation of the rhythm metrics and subsequently proposes three alternative acoustic measures investigated as a possible basis for capturing Malay rhythm.

6.2. The results of the rhythm metrics: ΔC, ΔV and %V

According to Ramus et. al (1999) the values of the standard deviation of consonantal intervals or ΔC reflected the categorization of the languages based on their traditional descriptions of rhythm. High ΔC value was associated with the traditional category of stress-timed while; low ΔC value was associated with syllable-timed. Ramus et. al (1999) also suggested that the values of ΔC were related to the variability of syllable structure which were available within a language.

To recap, the overall value ΔC was determined by calculating the mean standard deviations of the consonantal interval durations of every subject producing the ten sentences. For this dataset, the overall ΔC was 34.5. Comparing this value to the list of languages in Ramus et. al (1999), Malay appeared to be in the syllable-timed end of the continuum together with Catalan (45.2) and Japanese (35.6). Whilst languages with higher ΔC than Malay from this dataset on the other end were English (53.5) and Polish (51.4)

Meanwhile, %V was the calculation of the percentage of proportional duration of the vocalic intervals from the total duration of the utterances . %V score for Malay from this data was 51.11. This result was consistent with languages in the syllable-timed end of the continuum such as Catalan (45.6) and Japanese (53.1) in Ramus et. al (1999).

ΔV was the standard deviation of vocalic interval and its value was obtained by using similar procedure to that of ΔC. The ΔV score for Malay was 36.6 and this outcome was also close to the values of ΔVs found for Catalan (36.8) and French (37.8). Once again the score for ΔV of this dataset was consistent with those of other languages in syllable-timed end of the continuum. According to Ramus et. al (1999), the values of ΔV was not associated to the description of rhythm. Despite this, the current analysis included all three rhythm metrics as described in section 2.3 since there was a possibility ΔV also could indirectly be related to the rhythmic description of Malay. Table 6.1 displays the overall results of these rhythm metrics (ΔC, %V and ΔC) for Malay in the current study.

Table 6.1 The overall scores for ΔC, %V and ΔC of Malay

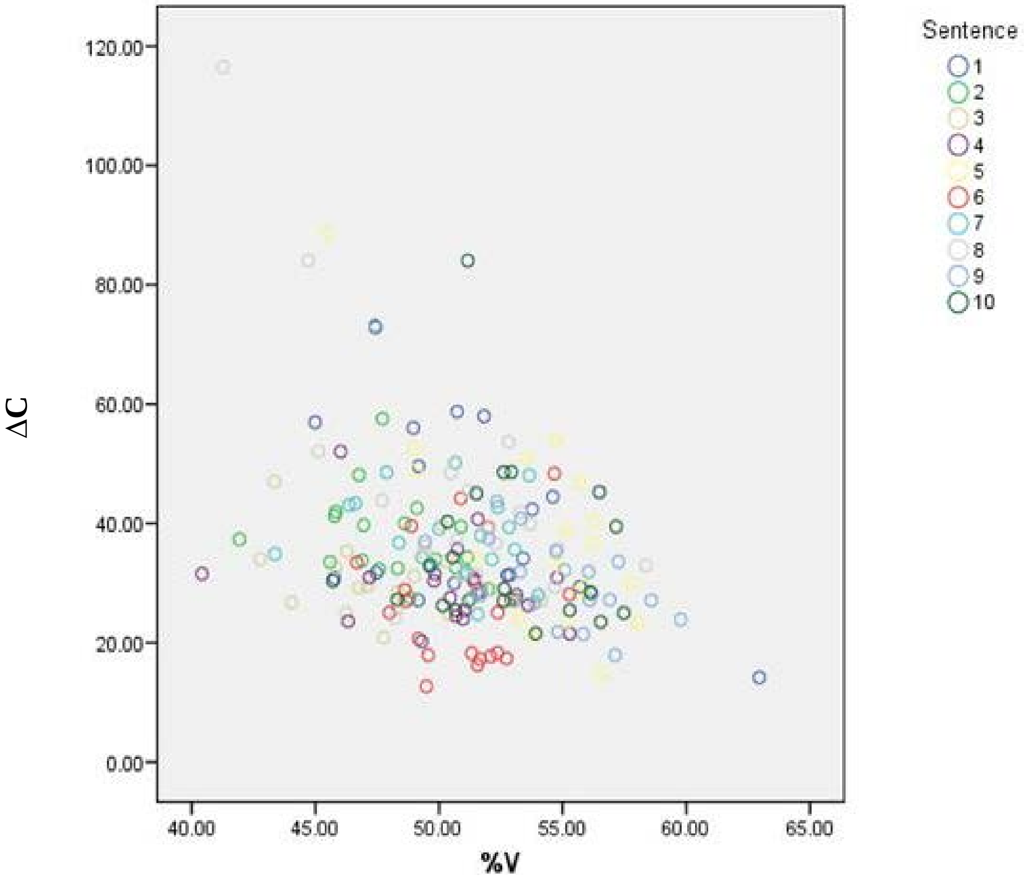
	ΔC	%V	ΔV
Overall scores	34.6	51.1	36.6

It must be noted, however that these results did not reflect the durational variability of individual sentences and speakers as exhibited in Figure 6.1. There was variability in the values of ΔC and %V across sentences and speakers. Different sentences and individual differences in the way in which the sentences were produced probably affected the variability of the rhythm metrics, these two elements are discussed further later in this chapter (see section 6.5)

The majority of ΔC scores in Figure 6.1 were spread between 20 and 40; these values were consistent with the values of ΔC for languages in the syllable-timed end of the continuum. But interestingly, there were also some ΔC cases which had consistent values with those languages from the stress-timed end of the continuum. Higher ΔC suggested that some of these subjects' consonantal interval durations were more variable than the rest of the subjects.

Meanwhile, for %V, the cases in the scatter plot centred mostly between 45 and 55. The lowest %V value was 40 and the highest was 62.96. %V represented the proportion of vocalic intervals and it could be observed in Figure 6.1 and Figure 6.2, most subjects still maintain a relatively balanced proportion of consonantal and vocalic interval durations. The %V pattern across the data was consistent with the features of languages in the syllable-timed end of the continuum.

Figure 6.1 The distribution of the sentences spoken by twenty subjects over the %V-ΔC plane



The variability across the data was apparent and the next section investigated further the variability of each of the rhythm metrics proposed by Ramus et. al (1999).

Table 6.2 and Table 6.3 present the scores for ΔC, ΔV and %V across sentences and subjects respectively.

Table 6.2 The scores for ΔC, ΔV and %V across the sentences

Sentence	ΔC	%V	ΔV
1	36.63	51.42	31.71
2	36.44	53.12	43.45
3	30.46	48.10	29.22
4	31.57	50.08	29.42
5	37.33	54.03	46.97
6	26.29	50.87	33.81
7	39.52	50.16	48.48
8	44.43	50.73	33.81
9	30.01	54.87	47.17
10	35.39	52.93	40.65

Table 6.3 The scores for ΔC, ΔV and %V across the subjects

Subject	ΔC	%V	ΔV
S1	28.99	51.41	34.93
S2	32.62	50.08	31.97
S3	37.24	50.12	38.57
S4	29.61	51.35	34.73
S5	31.34	51.64	38.39
S6	31.98	52.06	33.22
S7	33.37	52.09	37.29
S8	32.88	52.02	32.56
S9	24.13	49.39	30.33
S10	41.20	50.04	36.53
S11	36.88	53.45	40.24
S12	54.75	48.29	43.44
S13	36.37	45.66	37.52
S14	54.75	48.29	43.44
S15	47.57	50.98	45.62
S16	30.35	53.43	38.39
S17	35.55	51.01	33.82
S18	32.33	52.54	32.49
S19	35.74	52.34	35.22
S20	27.25	54.10	33.68

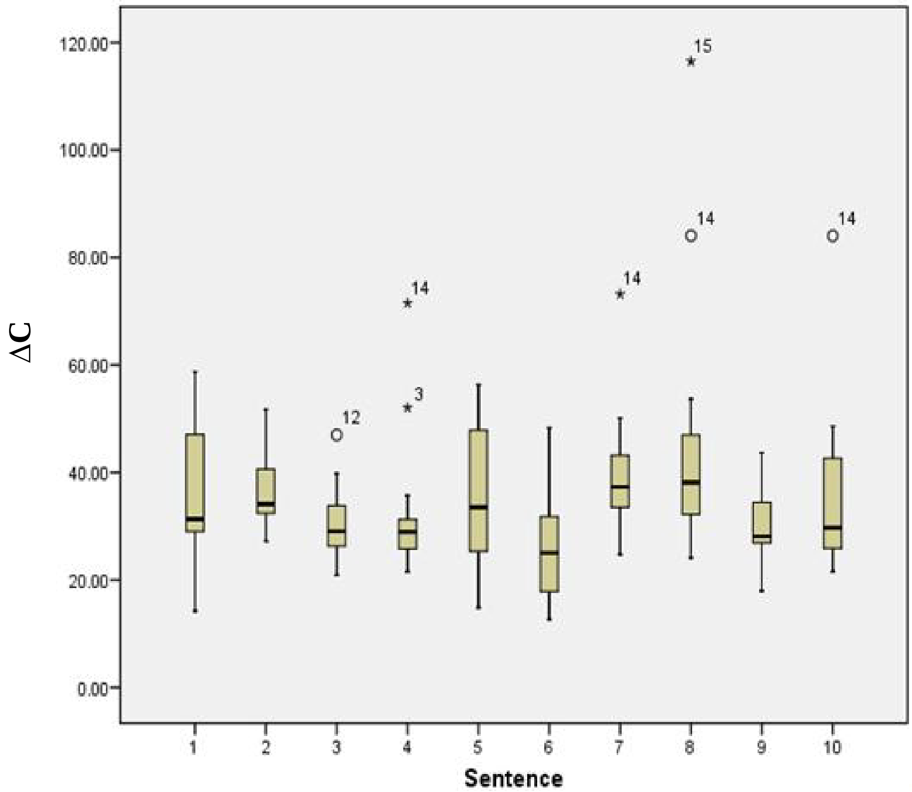
6.2.1. ΔC variability

Generally, according to Ramus et. al (1999) ΔC was closely linked to the syllable structures available in a language, however, the results of ΔC for the ten sentences shown in Table 6.2 demonstrated variability across the sentences. A one-way ANOVA was conducted to compare the effect of these sentences on the scores for ΔC. And the result showed a significant effect of sentences on ΔC values ($F_{9, 190}$= 3.48, $p < 0.01$). Subsequent post hoc comparison (Tukey) revealed that sentence 6 was significantly lower than sentence 7 and sentence 8 was significantly higher than sentences 3, 6 and 9.

The reason for these differences might be the way in which the sentences were structured which allowed durational variability. This was evident in Figure 6.2 in which the ranges of ΔC across the sentences were different from each other; it was observed that sentence 4 has the shortest ΔC range and sentences 4 and 5 had relatively equal longest range of ΔC than the rest of the sentences.

Another one-way ANOVA test was calculated to test the effect of different subjects on the values of ΔC, the result showed a significant effect of subjects on the values of ΔC ($F_{19,180} = 3.97$ $p < 0.01$) and post hoc comparison revealed that subject 15 was significantly higher than subject 9 and 20. At the same time, subject 14 was significantly higher than subjects 2, 4, 5, 6, 7, 8, 9, 11, 12, 16 and 18.

Figure 6.2 Boxplot of ΔC across the sentences



A further analysis in section 6.5 revealed that individual differences in the way in which these sentences were produced contributed to the variability of the scores for ΔC across the subjects.

Another reason for the variability across ΔC was the influence of speech rate. According to Dellwo (2004) ΔC was correlated negatively to speech rate and it was plausible that this feature also could influence the scores for ΔC of this dataset. The relationship of speech rate and the rhythm metrics is discussed in section 6.4

The overall ΔC value of this Malay data suggested that it is positioned in the syllable-timed end of the continuum. A low ΔC of 34.6 also indicated a limited variety of syllable structures, yet this overall value was not able to capture the observation made in respect of the variability across sentences and subjects. Thus, the single overall value of ΔC did not provide a good representation for the description of rhythm for Malay. The next section looks at the variability of the next rhythm metrics: %V.

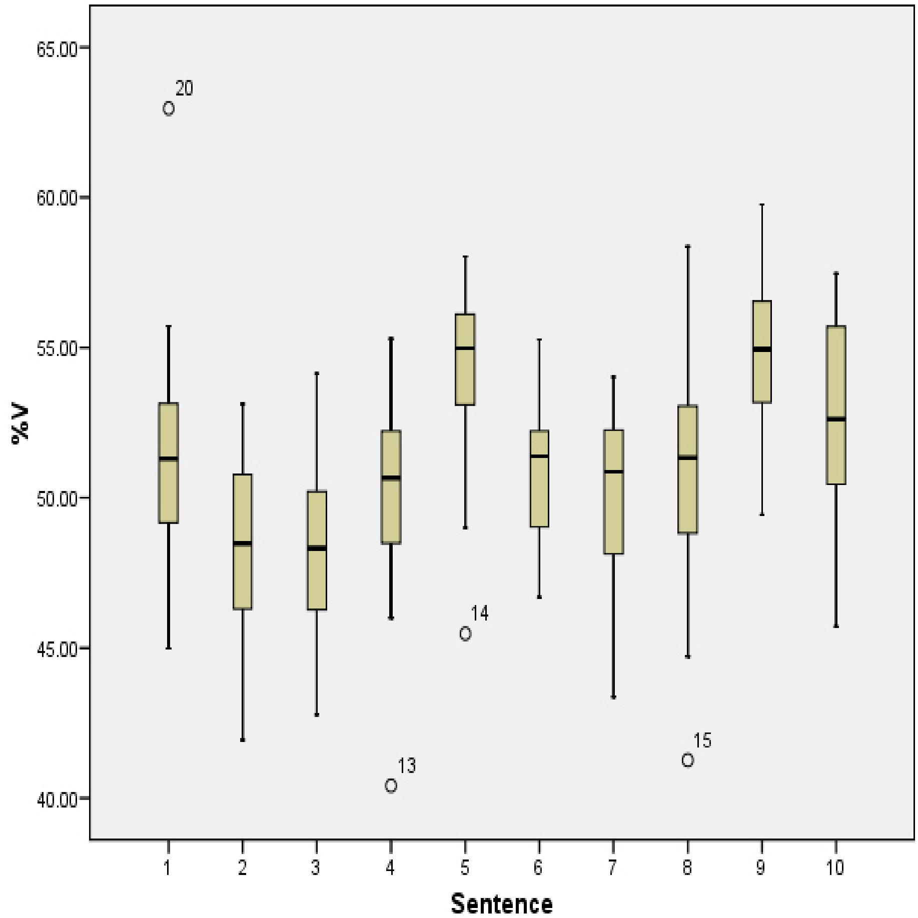
6.2.2. %V variability

The percentages of vocalic proportion (%V) of all the ten sentences and the twenty subjects are presented in Table 6.2 and Table 6.3. It can be observed that there was a wide range of differences of the values of the %V with the biggest vocalic interval proportion of 54.87 recorded in sentence 9 and the smallest percentage, 48.10, in sentence 3 (Table 6.2). A one-way ANOVA was carried out to compare the effect of the sentences on the values of %V in this data set. The result revealed a significant effect of the sentences on %V ($F_{9, 190} = 9.82$ $p < 0.001$). Post hoc comparisons showed that sentence 5 was significantly higher than sentences 2, 3, 4, 6, 7 and 8. Sentence 9 was also significantly higher than sentences 1, 2, 4, 6 7 and 8. Meanwhile sentence 3 was significantly lower than sentences 1, 9 and 10 and sentence 2 was significantly lower than sentence 10.

Based on the boxplot (Figure 6.3) some of the subjects had lower scores of %V than those of the typical range within the sentence. An example of this was the %V score for subject 20 in sentence 1, however there were also instances in which the subjects had lower %V scores than the lower interquartile range within the sentence, they were subject 13 (sentence 4), subject 14 (sentence 5) and subject 15 (sentence 8).

A one-way ANOVA was carried out to compare the effect of subjects on the %V scores. The results revealed significant effect of subjects on %V scores ($F_{19,180} = 3.17$, $p < 0.001$). Post hoc comparisons indicated that subject 13 had a significantly lower %V than subjects 1, 4, 5, 6, 7, 8, 11, 16, 18, 19 and 20.

Figure 6.3 Boxplot of %V across the sentences



6.2.3. ΔV variability

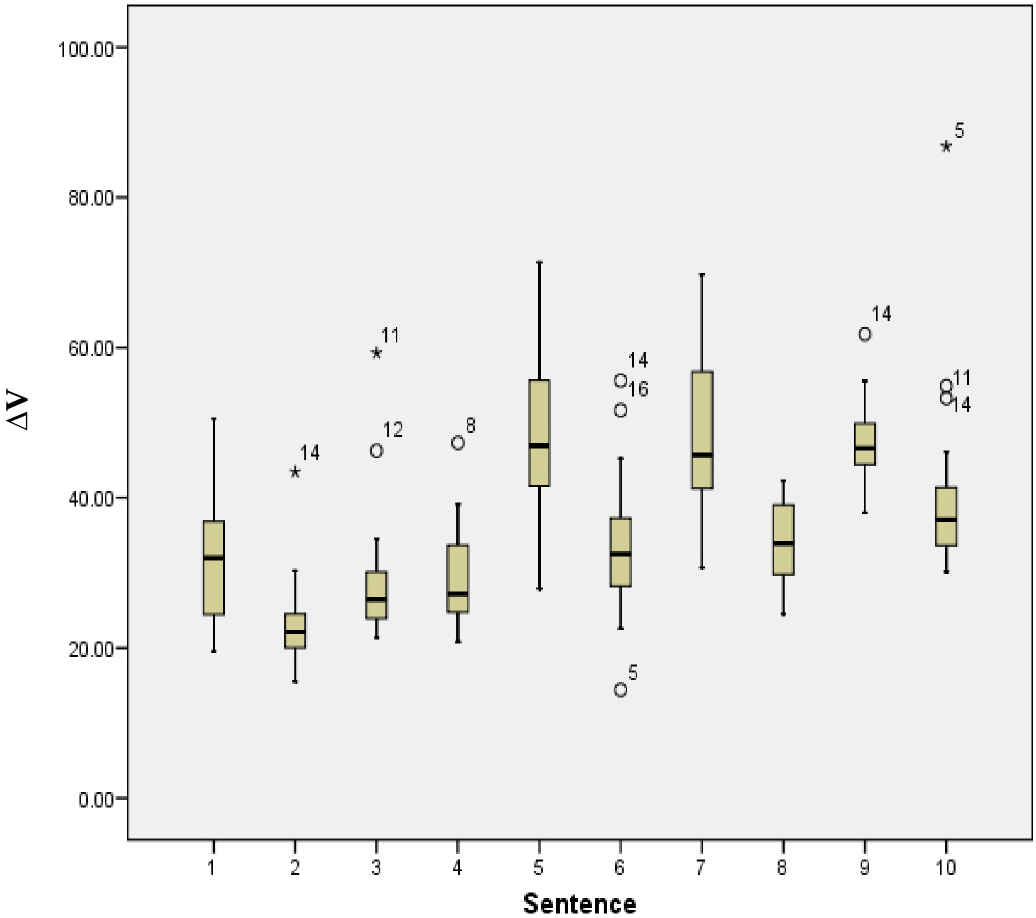
Based on the ΔV scores across the sentences and the subjects in Table 6.2 and Table 6.2 respectively, it could be observed that there was also variability for this rhythm metrics.

To investigate further, a one-way ANOVA was conducted to study the effect of the sentences on ΔV values. The result indicated that there was a significant difference between the sentences on ΔV ($F_{9, 190}$= 19.06, $p < 0.001$). Post hoc comparison indicated that sentence 1 was significantly lower than sentences 5, 7 and 9, while sentence 2 was significantly lower than sentences 6, 7 and 8. Sentence 4 was also significantly lower than sentences 5 and 7.

Meanwhile, sentence 5 was significantly higher than sentences 6 and 8. Sentence 6 was significantly lower than sentences 9 and 7 while sentences 7 and 10 were significantly lower than sentences 3 and 8 and sentences 2, 3 and 4. The variability of the sentence was illustrated in Figure 6.4 in which it could be seen that across the sentences, the interquartile ranges were different from each other. The way in which the sentences were structured contributed to different values of ΔV; it could be observed that the number of vocalic intervals across the sentences varied (Table 5.1).

A one-way ANOVA analysis to study the effect of subjects on the ΔV scores showed no significant differences between the subjects ($F_{19, 180} = 1.05$, $p > 0.001$). The finding showed there was relatively little variability of the vocalic interval duration across the subjects.

Figure 6.4 Boxplot of ∆V across the sentences



6.3. The results of the Pairwise Variability Index (PVI)

PVI, as suggested by Grabe & Low (2002), is able to differentiate languages in a way which reflects the traditional rhythmic description of the languages. The overall calculation of *n*PVI (normalised PVI of the vocalic duration) for Malay in this data was 44.8. To recap, the value was computed by calculating the mean difference between successive vocalic intervals divided by the sum of the same intervals. The data was normalised in order to control for speech rate. Based on the findings and the comparison with other languages in Grabe and Low (2002), Malay stood between Romanian and Estonian, two languages which were claimed to be located at the syllable-timed end of the continuum. The value of *n*PVI for this data suggested that Malay in this dataset was consistent with the features of syllable-timed languages in which a low *n*PVI indicated regular vocalic interval duration between the successive syllables. A higher *n*PVI, on the other hand, would be consistent with languages in the stress-timed end of the continuum since it represented irregular intervals which were associated with vowel reductions.

Subsequently, *r*PVI (the raw dataset of PVI of the consonantal duration) was obtained by calculating the differences of successive consonantal intervals of each of the subjects producing the sentences. According to Grabe and Low (2002), as well as differentiating languages with different rhythmic properties, the values of *r*PVI reflected the syllable structure of the language in which high scores of *r*PVI indicated more complex syllable structures. Meanwhile a low *r*PVI suggested there was a relatively simple syllable structure in the language. The *r*PVI score for this current study was 29.5. In comparison to other languages studied in Grabe & Low (2002), this *r*PVI value was considered low, which to some extent might reflect the relatively simple syllable structure of Malay as discussed in section 3.3.2. Table 6.4 sums up the overall scores for both the PVIs of this present data.

Table 6.4 The scores for *n*PVI and *r*PVI of Malay

	*n*PVI	*r*PVI
Overall results	44.8	29.5

A comparison of the PVI scores from this data and those from Grabe and Low (2002) indicated a considerable difference between these scores, the latter put Malay at the stress-timed end of the continuum. The *n*PVI and *r*PVI scores for Malay in Grabe and Low (2002) were 53.6 and 63.3 respectively.

There were two possible reasons for the discrepancies; firstly the differences in the number of subjects who took part in these studies. Twenty subjects took part in this study compared to only one in Grabe and Low (2002). Even within this present study, variability could be observed across the twenty subjects as shown in Figure 6.5. Thus it was not surprising that there were differences across the subjects between the two studies.

Secondly, the individual differences in the way which the subjects in both studies produced the sentences could also be linked to the differences of the PVIs scores. With regard to this analysis, in which the materials were controlled, individual differences were still apparent; hence it was not surprising for the differences to arise between the data from the current study and that from Grabe and Low (2002).

Based on the scatter plot in Figure 6.5, it can be observed that the nPVI values of the sentences produced by the subjects were distributed across the range between 20 and 80. The majority scores of the current study reflected features which was consistent with those of languages in the syllable-timed end of the scale. However, it was also noticeable that some of the *n*PVI values were similar to those of the languages in the stress-timed end of the continuum. The differences within the same study indicated variability of this rhythm metric to the factors of sentences and subjects.

As for the *r*PVI scores, the same scatter plot shows that most of the cases are distributed between the scores of 20 and 60. At the same time, some cases also scatter higher than the typical range. This illustrates the fact that variability of both the rhythm metrics existed similar to the ΔC, %V and ΔV discussed earlier. The next two sections discuss the variability issue of the PVIs further.

Figure 6.5 Distribution of sentences spoken by twenty subjects over the *r*PVI-*n*PVI plane

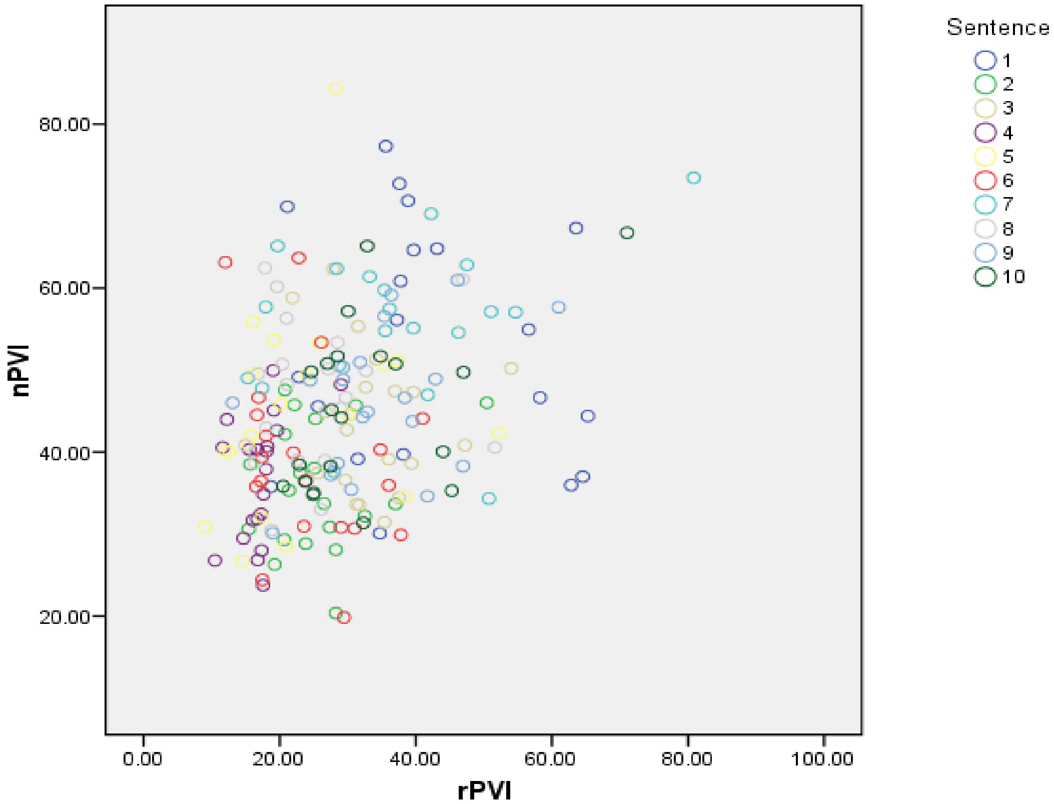


Table 6.5 and Table 6.6 present the scores for *n*PVI and *r*PVI of the sentences and subjects respectively.

Table 6.5 The scores for *n*PVI and *r*PVI across the sentences

Sentence	*n*PVI	*r*PVI
1	53.14	41.66
2	35.77	25.80
3	43.93	33.97
4	36.57	16.40
5	44.84	23.02
6	39.43	24.47
7	55.71	37.51
8	46.93	25.27
9	46.11	34.33
10	45.43	32.77

Table 6.6 The scores for *n*PVI and *r*PVI across the subjects

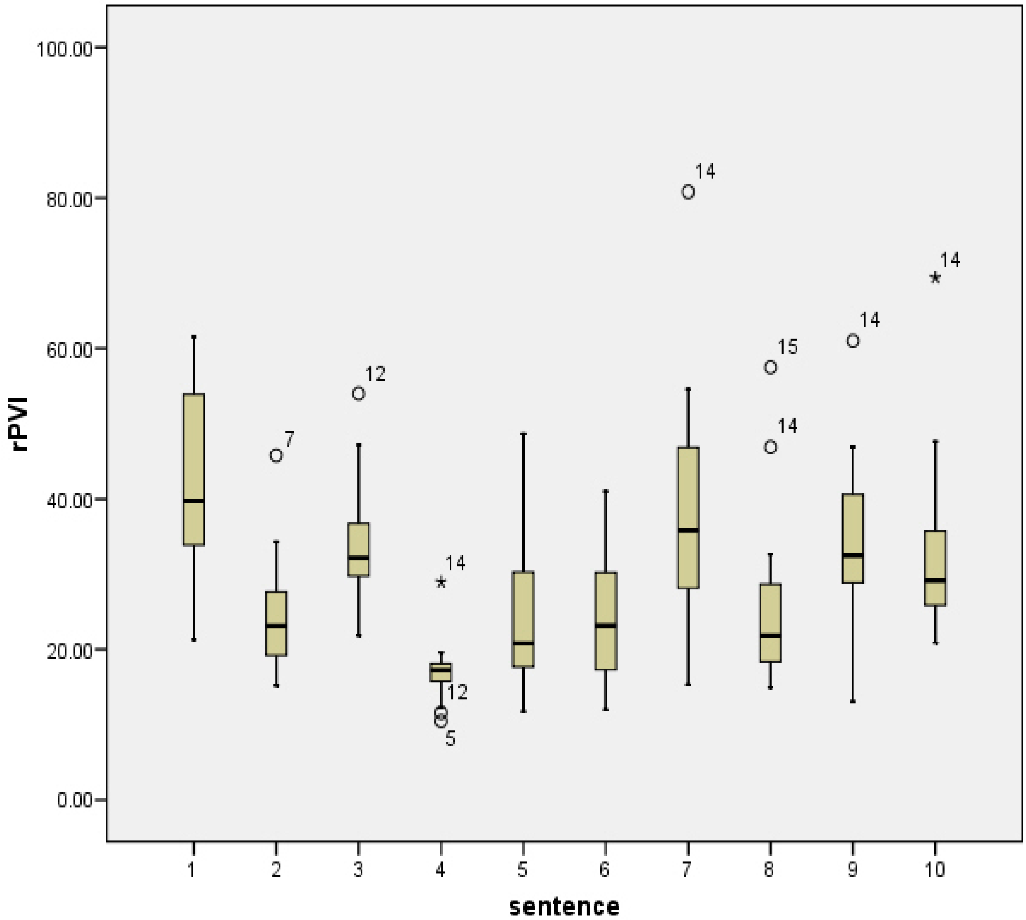
Subject	*n*PVI	*r*PVI
S1	44.83	20.46
S2	44.24	26.03
S3	50.88	29.56
S4	50.97	26.53
S5	50.70	27.24
S6	43.70	25.47
S7	48.47	32.40
S8	37.93	26.01
S9	47.80	21.74
S10	39.23	34.85
S11	38.85	28.82
S12	55.38	47.49
S13	44.39	30.57
S14	55.38	47.49
S15	42.43	38.10
S16	46.38	28.31
S17	34.95	32.67
S18	44.12	28.20
S19	44.96	32.63
S20	38.44	24.22

6.3.1. rPVI variability

Tables 6.5 and 6.6 for the *r*PVI values indicate varying scores of this rhythm metrics across sentences and subjects respectively. In order to investigate the variability of the sentences, a one- way ANOVA was run to analyse the effect of these sentences on *n*PVI. The results showed a significant difference across these sentences ($F_{9,\,190} = 10.17$, $p < 0.001$). Post hoc analysis revealed than sentences 1, 2, 3, 4, 5, 6 and 7 were lower than the other sentences. The variability of *r*PVI is illustrated in the boxplot (Figure 6.6), it could be observed that sentence 4 has the smallest range of *r*PVI than the rest of the sentences. The way in which the sentences were structured could contribute to such variability of the scores.

Subsequently, another one-way ANOVA was done to see the effect of the subjects on *r*PVI values and the result revealed that the subjects have significant effect on this rhythm metrics ($F_{19,180} = 2.55$, $p < 0.001$). Post hoc test showed that subject 14 was significantly higher than subjects 1, 2, 4, 5, 6, 8, 9, 16, 18 and 20. It is also worth noting that subject 14 also were found to be different in the other rhythm metrics: ΔC and %V.

Figure 6.6 Boxplot of *r*PVI across the sentences



Based on the mean values of the rhythm metrics across the sentences and subjects, the one-way ANOVA results and the boxplots; it was evident that *r*PVI exhibited high variability across the sentences and subjects. Thus the impression of the rhythm metrics results which positioned Malay in syllable-timed end of the continuum could not be taken at face value because the overall score of these rhythm metrics did not successfully explain the variability found across sentences and subjects.

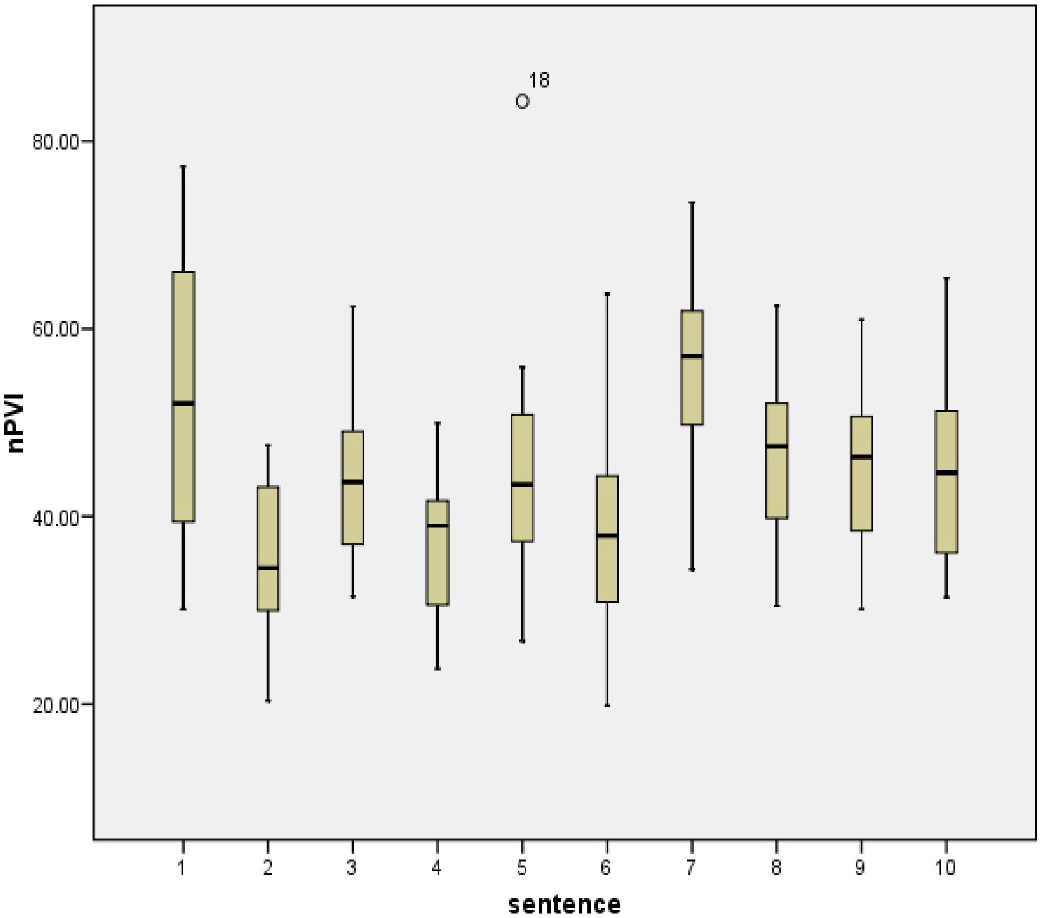
6.3.2. *n*PVI variability

The *n*PVI scores in both Table 6.5 and Table 6.6 show variability across sentences and subjects respectively. Sentence 2 had the lowest *n*PVI score while sentence 7 had the highest score. A one-way ANOVA conducted to study the effect of these sentences on the *n*PVI scores. The result revealed a significant effect of the sentences on the *n*PVI values ($F_{9, 190}$= 7.90, $p < 0.001$). Post hoc comparison showed that sentence 1 was significantly higher than sentences 2 and 4, sentence 2 was significantly lower than sentences 7, 8 and 9, the same was reported for sentence 3 (sentence 7), however sentence 7 was significantly higher than sentences 4, 5 and 6.

The significant differences in *n*PVI scores were predicted since the distribution of vocalic interval durations in these sentences varied from each other. A contributing factor to these differences might be due to different subjects producing these sentences. A one-way ANOVA was also conducted to study the effect of subjects on the *n*PVI values. There was a significant difference between the subjects on *n*PVI ($F_{19, 180}$ = 2.20, $p < 0.001$). Post hoc analysis revealed that subject 7 displayed significant lower result than subject 14.

Figure 6.7 is a boxplot of the distribution of *n*PVI across the sentences, it can be observed that there varying *n*PVI ranges between the sentences. Sentences 1 and 6 exhibited longer ranges of *n*PVIs which indicated considerable higher variability of vocalic interval durations than those of other sentences like sentences 2 and 4. Further to this, more evident of variability was observed with the outlier on sentence 5.

Figure 6.7 Boxplot of *n*PVI across the sentences



Another variable which was predicted to contribute to the variability of these rhythm metrics was speech rate (Dellwo and Wagner, 2004 and Dellwo, 2006), following this, the next section investigates the relationship between these rhythm metrics and speech rate.

6.4. Speech Rate and the rhythm metrics

The interaction of speech rate and the rhythm metrics could also contribute to the variability of the rhythm metric scores. According to Dellwo and Wagner (2004) ΔC was correlated with speech rate as demonstrated in the experiment they carried out in which ΔC and %V of five different speech rates were measured and they found that ΔC was rate-dependent whereas %V appeared to be more constant across different speech rates. Following the findings in Dellwo and Wagner (2004), an examination of the speech rate was carried out for the present data in order to investigate if similar patterns emerged with regard to the speech rate.

To recap, the speech rate was calculated by dividing the number of syllables by the overall duration of the utterance. The duration was converted from milliseconds (ms) to seconds for this calculation. This calculation was carried out for each sentence produced by each of the speakers. The calculation excluded any plosives in the initial position of a sentence and all the final sounds in all the sentences, so that this data would be consistent with the measurement of other parameters.

The mean speech rate across the dataset was 6.3 syllables/second. The standard deviation was 0.79. The speech rates of each of the sentences and subjects are in Table 6.7 and Table 6.8

Table 6.7 The speech rate across the sentences

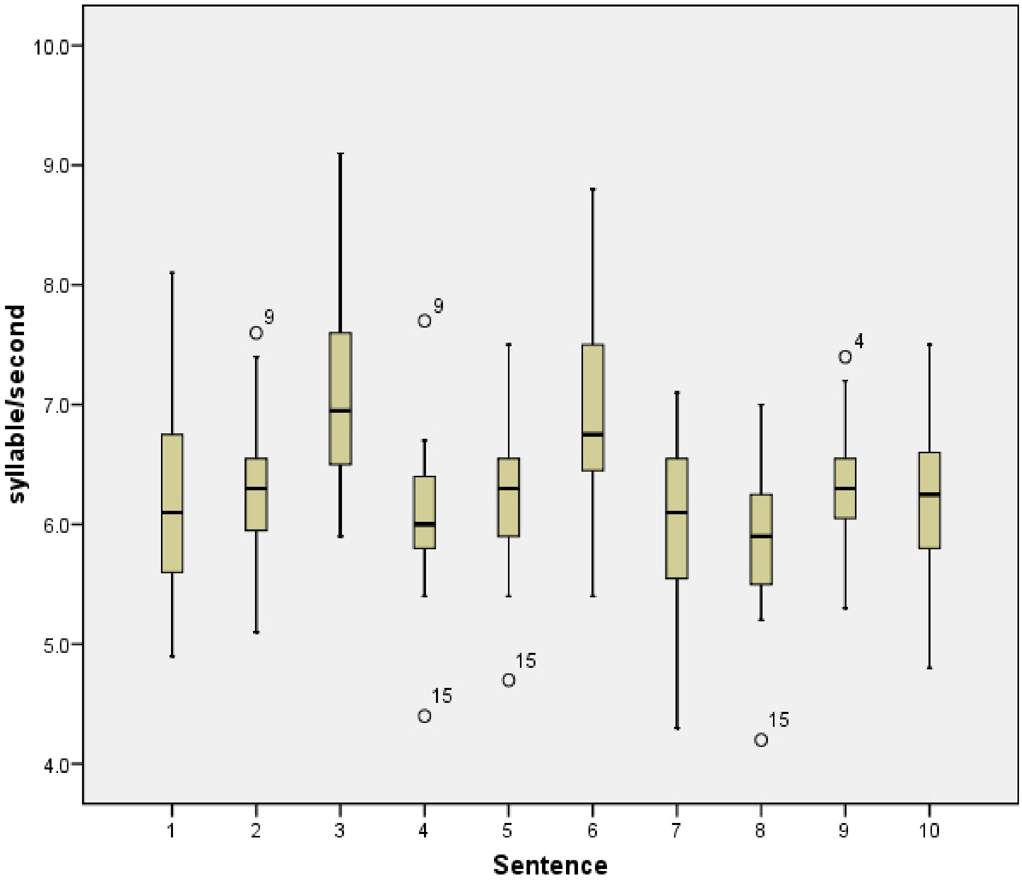
Sentence	Speech rate (Syllable/second)
1	6.2
2	6.3
3	7.1
4	6.1
5	6.3
6	7.0
7	6.0
8	5.9
9	6.3
10	6.3

Table 6.8 The speech rate across the subjects

Subject	Speech rate (Syllable/second)
1	6.4
2	6.5
3	6.3
4	7.2
5	7.2
6	6.4
7	6.3
8	6.2
9	7.6
10	6.1
11	5.7
12	6.1
13	6.0
14	6.1
15	4.9
16	6.3
17	5.8
18	6.6
19	6.3
20	6.5

One-way ANOVA was calculated to test the effect of these sentences on speech rate indicated that the sentences were significantly different with regard to the speech rate at ($F_{9, 190} = 6.432$, $p < 0.001$). Post hoc comparisons showed that the speech rate of sentence 3 was significantly higher than sentences 1, 2, 4, 5, 7, 8, 9 and 10, while sentence 6 was reported to be significantly higher than sentences 1, 4, 5, 7 and 8. This variability is illustrated in the boxplot (Figure 6.8) in which there are six outliers across the sentences. The variability of the speech rate could be attributed to the fact that the subjects were not told to control their rate during the recording session, thus they were free to produce the sentences at any rate they deemed comfortable.

Figure 6.8 Boxplot of speech rate across the sentences

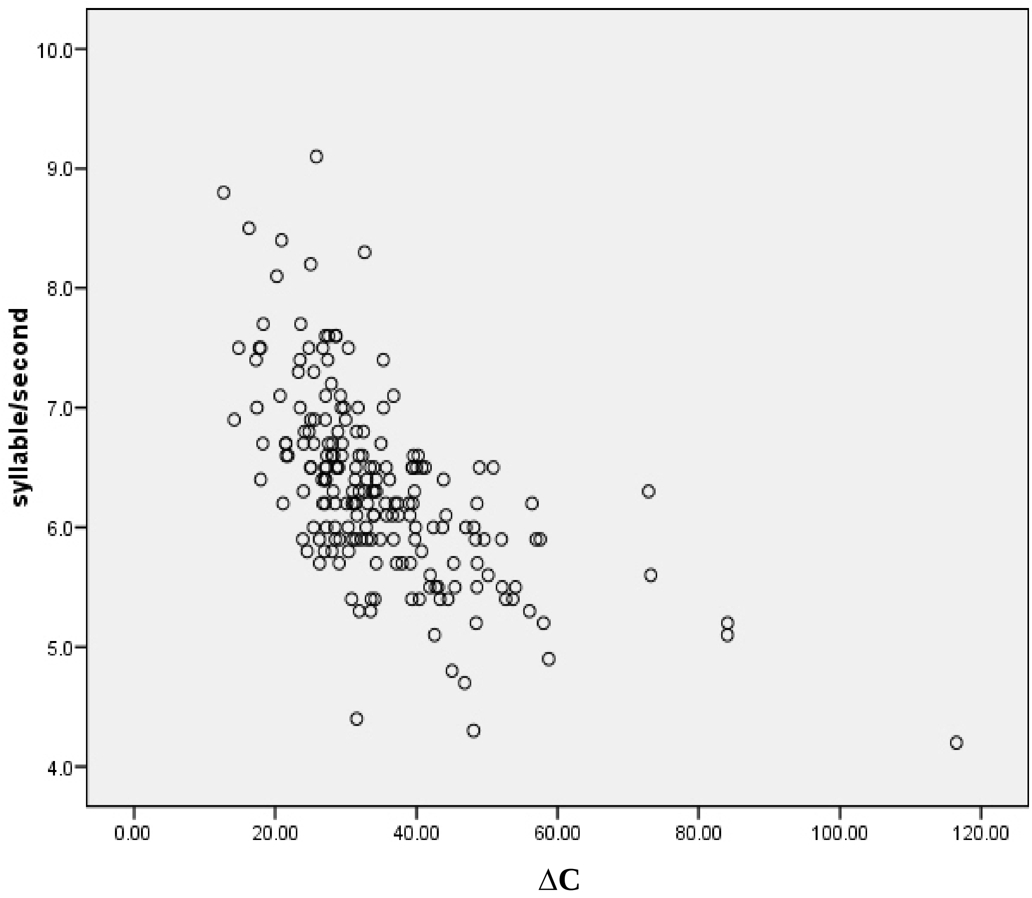


In order to investigate the relationship between speech rate and rhythm metrics correlation tests were used to analyse the speech rates against each of these rhythm metrics: ΔC, %V, ΔV, *n*PVI and *r*PVI of the current data. The purpose of the correlation tests was to find out whether these metrics, which were claimed to be closely related to the rhythm of the language, were related to speech rate. This was based on the assumption that the values of ΔC, %V, ΔV, *n*PVI and *r*PVI were influenced by how fast or slow the rate of the speech was. If these metrics were independent of speech rate, it could be claimed that to some extent they could be useful measures of rhythm.

6.4.1. Speech rate and ΔC

A Pearson's correlation was computed to assess the relationship between speech rate and ΔC. There was a strong negative significant correlation between these two variables ($r = -0.605$, $p < 0.05$). This finding is illustrated in the scatter plot (Figure 6.9). Overall the increases in speech rate were correlated with decreases in the values of ΔC. This further substantiated the claim by Dellwo and Wagner (2003) and Dellwo (2005) in which the variability of ΔC was significantly linked to speech rate.

Figure 6.9 Distribution of sentence-subjects over ΔC-speech rate (syllable/second) plane

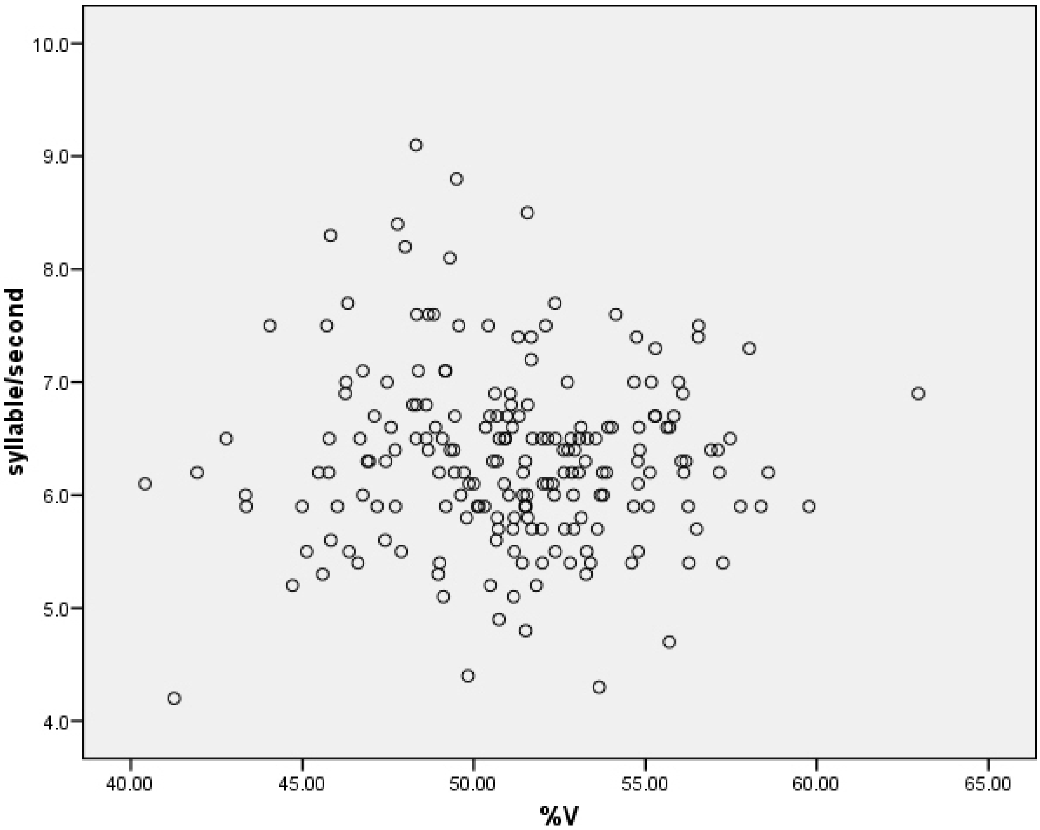


6.4.2. Speech rate and %V

A Pearson's correlation test conducted to study the relationship between speech rate and %V of Malay revealed that there was no significant correlation between the two variables ($r = -0.25$, $p > 0.05$). The scatter plot illustrates this finding (Figure 6.10); the values of %V were independent from speech rate. This finding was similar to that of Dellwo and Wagner (2004) in which it was found that the values of %V were more constant across different speech rates.

Dellwo and Wagner (2004) limited their speech rate investigation to ΔC and %V only. This present research, however, went further to look at relationship of the speech rate to other features of rhythm metrics: ΔV, *n*PVI and *r*PVI.

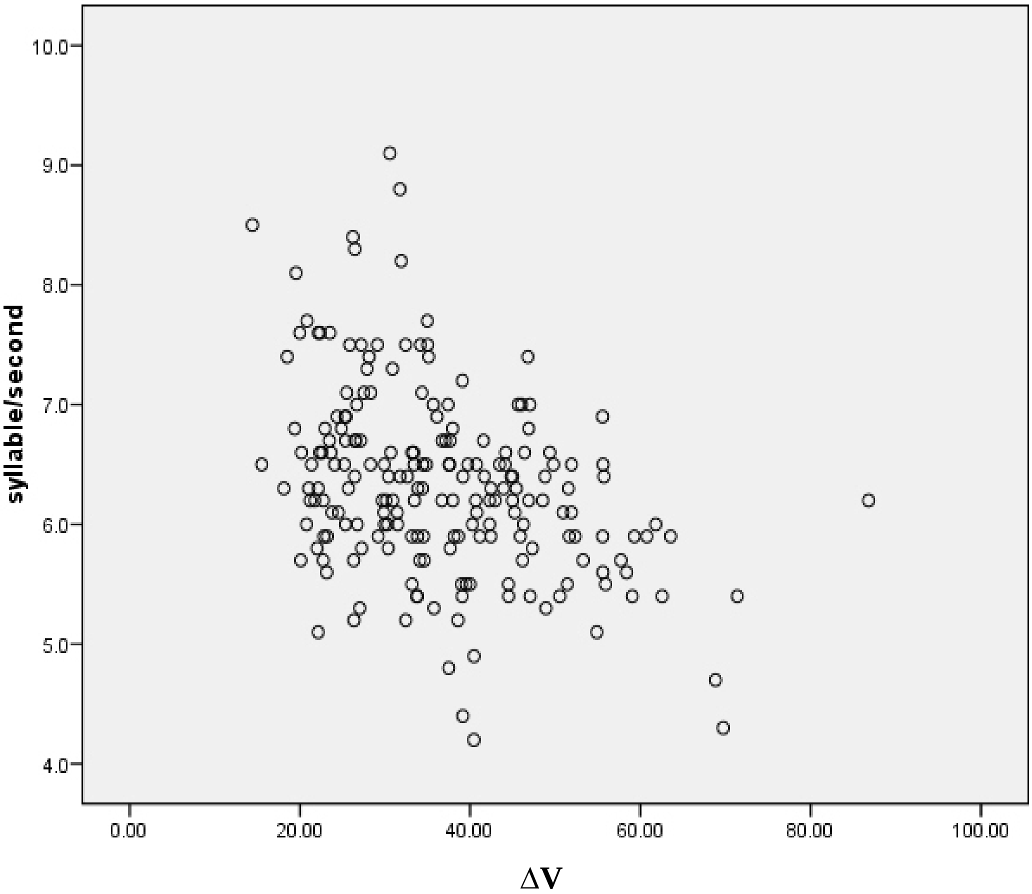
Figure 6.10 Distribution of sentence-subjects over %V-speech rate (syllable/second) plane



6.4.3. Speech rate and ΔV

A Pearson's correlation test was computed to assess the relationship between speech rate and ΔV. There was a negative significant correlation between these two variables ($r=-0.379$, $p< 0.05$). A scatter plot summarises the results (Figure 6.11). It can be observed that there was no relationship between ΔV and speech rate as the values of the former seemed to be constant across different speech rates.

Figure 6.11 Distribution of sentence-subjects over ΔV-speech rate (syllable/second) plane

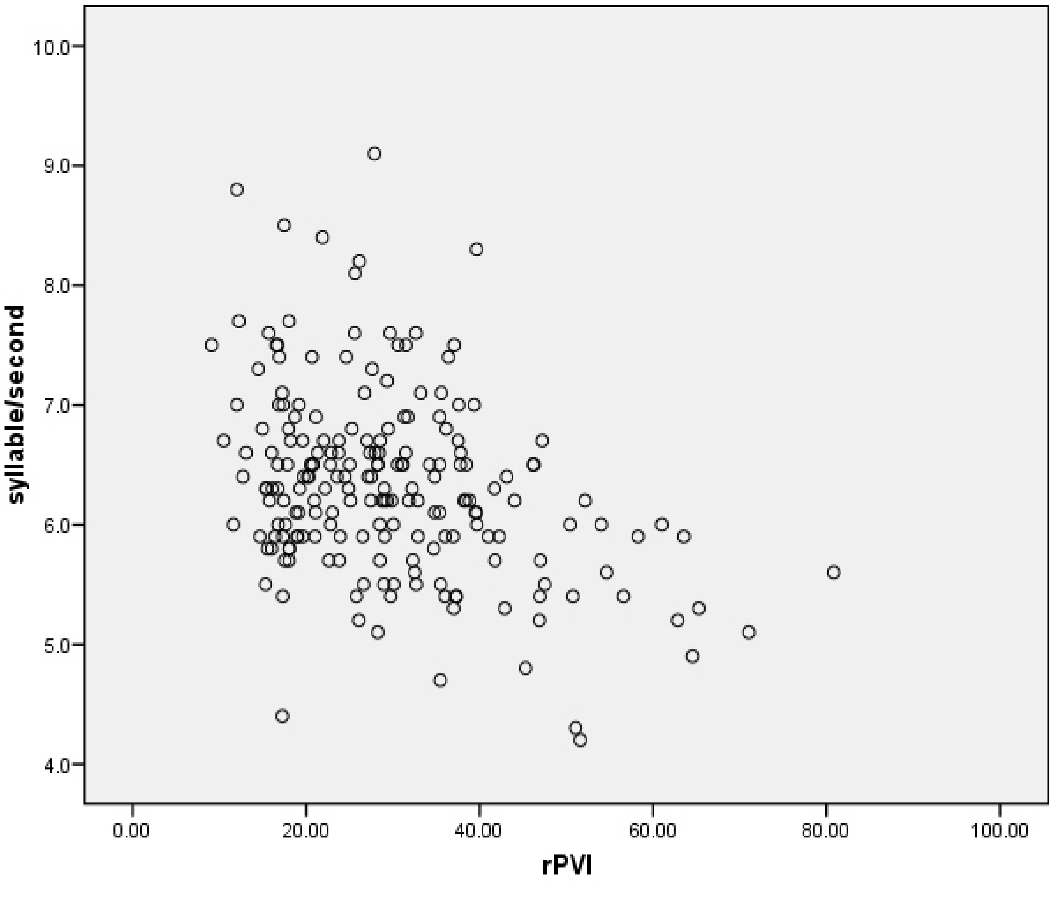


6.4.4. Speech rate and *r*PVI

The present study also analysed the relationship between speech rate and the Pairwise Variability Index. The investigation was carried out to look if there was any correlation between these two variables which contributed to the variability of the PVIs scores.

A Pearson's correlation analysis conducted to assess the relationship of speech rate and *r*PVI showed a negative significant correlation between these two variables (r=-0.368, p <0.005). A scatter plot illustrates this result (Figure 6.12). Increases in speech rate were correlated with decreases in the *r*PVI values. This finding demonstrated that the values of *r*PVI were dependent on the changes in speech rate. Thus far, the variability of the rhythm metrics cannot be interpreted without considering the influence of speech rate.

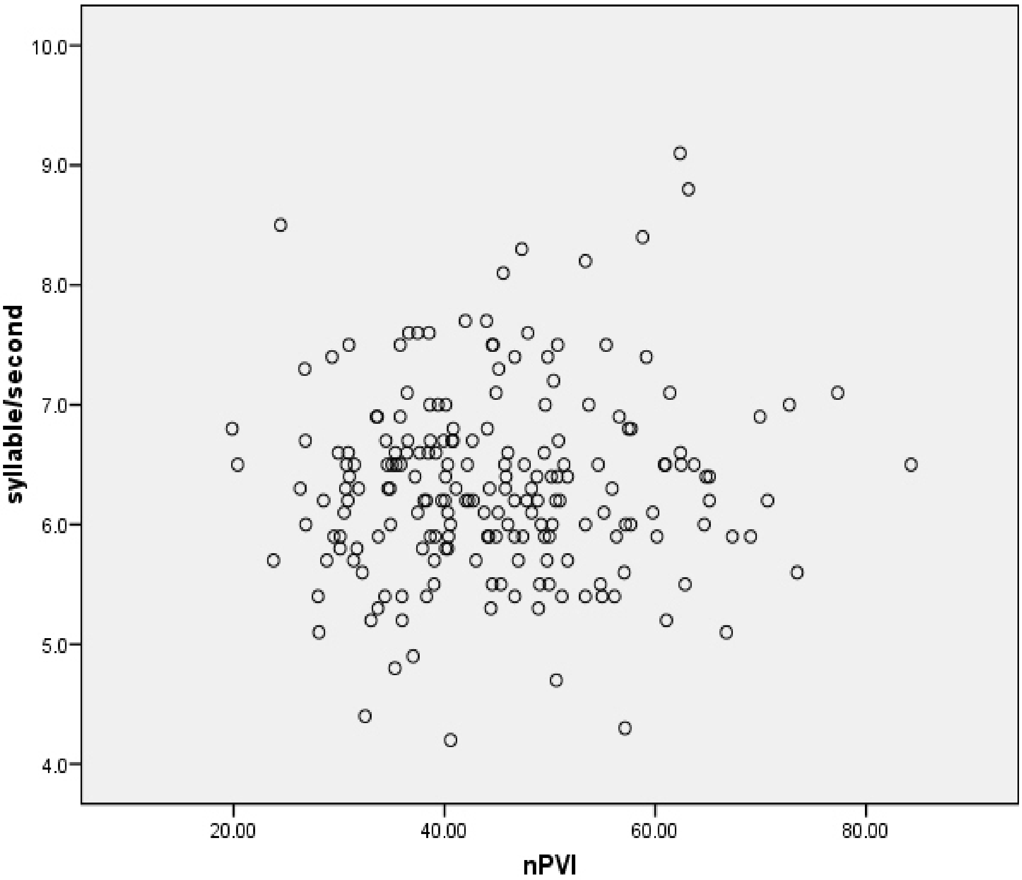
Figure 6.12 Distribution of sentence-subjects over *r*PVI-speech rate (syllable/second) plane



6.4.5. Speech Rate and *n*PVI

A Pearson's correlation test was carried out to analyse the relationship between speech rate and *n*PVI. There was no correlation between the two variables (r= 0.055, p=>0.005). A scatter plot illustrates this result (Figure 6.13). This result was somewhat similar to that of %V, therefore it can be explained that those *n*PVI values were constant despite the changes in speech rate.

Figure 6.13 Distribution of sentence-subjects over *n*PVI-speech rate (syllable/second) plane



Of all the rhythm metrics, three out of the five, ΔC, ΔV and *r*PVI, showed significant correlations with speech rate at p<0.05 level. The findings demonstrated further that apart from being sensitive to the functions of sentences and subjects; these rhythm metrics were not robust in the face of differences of speech rate.

The next section focuses on the durational analysis of some of the subjects and sentences in order to understand cross-speakers' variability that arises from the rhythm metrics calculations.

6.5. Further acoustic analysis

An acoustic analysis was carried out to investigate in more detail the variability of the rhythm metrics scores across the subjects and sentences. It was designed to compare the analysis of individual subjects to see why some stood out from the general patterns observed. The analysis was carried out on two productions of sentence 5 by subjects 7 and 14.

This analysis was carried out on subject 14 because his/her scores for the rhythm metrics: ΔC, %V and *r*PVI appeared to be the outliers for sentence 5. The results of this analysis were then compared to the same analysis carried out on subject 7. Subject 7 was chosen because his/her scores were the closest to the overall values of the rhythm metrics for this dataset. The comparison looked at the overall consonantal and vocalic durations in the sentence produced by these subjects.

This is the phonemic transcription of sentence 5: /tuʤuh eko kətupɐt sotoŋ enɐk dɪmɐsɐk munɐh/ which was translated as ‘Munah is cooking seven calamari with glutinous rice’. The first consonantal interval, which was a plosive, as well as the final interval was excluded from the analysis. The exclusion was consistent with the interval duration values used as inputs for the calculation of the rhythm metrics.

6.5.1. Subjects 14 and 7 producing sentence 5

There were 60 tokens of both the consonantal and vocalic segments across the two subjects. The overall duration for subject 14 was 2581 ms and for subject 7, it was 2493 ms. The duration of each of the segments within each of the utterances was compared in a Pairwise fashion

An independent-samples t-test was conducted to compare the overall duration of sentence 5 produced by subject 14 and 7. There was no significant difference between the sentence duration for subject 14 (M=89.0, SD=44.61) and subject 7 (M=85.97, SD=44.61), ($t_{(2)}$=4.35, p>0.05). The result suggested that the there was not much difference in terms of the overall durations between subject 14 and subject 7. The next sections will investigate both the durational components of the sentence which were the consonantal and vocalic intervals.

There were 14 consonantal intervals in sentence 5; the sample size for this analysis across the two subjects was 28. The consonantal intervals comprised four types of single consonantal sounds (plosive, nasal, fricative and affricate) and two types of consonantal combinations (fricative and plosive-nasal). The total consonantal interval durations for subjects 14 and 7 were 1332 ms and 1126 ms respectively. The duration of each of the consonantal intervals within each of the utterances was analysed.

An independent-samples t-test was conducted to compare the consonantal interval duration of sentence 5 produced by subject 14 and 7. There was a significant difference in the duration of subject 14 (M=95.14, SD=56.38) and subject 7 (M=80.43, SD=33.0) ($t_{(8)}$=12.72, p<0.05). Thus, from the results of the standard deviations, higher variability of consonantal intervals duration produced by subject 14 than those of subject 7 contributed to high ΔC and *r*PVI values to the former.

There were 15 vocalic intervals in sentence 5; the sample size across these two subjects was 30. The duration of overall vocalic intervals was 1249 ms for subject 14 and 1367 ms for subject 7. The vocalic interval duration of subject 14 was shorter than that of subject 7. Duration of each of the vocalic segment within each of the utterances was compared in a Pairwise fashion.

An independent-sample t-test was conducted to compare the vocalic interval duration of sentence 5 produced by subjects 14 and 7. There was a significant difference in the duration of subject 14 (M=83.27, SD=30.9) and subject 7 (M=91.13, SD=55.71) ($t_{(12)}$=29.42, $p<0.05$). This result suggested that individual differences have an effect on the durations of the vocalic intervals.

The proportion of the sentences occupied by consonantal interval was greater than the proportion of vocalic interval for subject 14 in the production of sentence 5. This explained why subject 14 had higher scores of rhythm metrics than the average when the calculation of were related to consonantal interval duration; namely ΔC and *r*PVI while he/she had lower scores of rhythm metrics when the calculation was related to the vocalic interval duration; namely %V and *n*PVI.

6.5.2. Sentence 4

In all the analysis of the rhythm metrics for this dataset, sentence 4 appeared to have seven outliers (subjects 6, 3 and 14 for ΔC, subject 13 for %V, subject 8 for ΔV and subjects 5, 12 and 14 for *r*PVI). Following this, an analysis of the sentence was also carried out in order to understand why this sentence attracted more outliers in the calculations of the rhythm metrics than the rest of the sentences.

The phonemic transcription of sentence 4 is /lolo məŋuko benteŋ dɪbɪnə tɐmɐn temɐ kulɪm ɪtu/– ‘Lolo is measuring the wall erected by Kulim Theme Park’. There were 18 consonantal intervals and 17 vocalic intervals. The total duration of the sample size for this analysis across the twenty subjects was 700. Z-scores were computed for the raw consonantal and vocalic interval durations. The decision to analyse the data with z-score was made in order to normalise the differences between the interval durations across the twenty subjects.

The overall mean duration for sentence 4 was 2986 ms. Meanwhile, the mean consonantal interval duration was 1484 ms and the mean vocalic interval duration was 1496 ms. The minimum raw interval duration of this sentence was 77 ms, z=-1.28 (subject 3) while the maximum raw interval duration was 246 ms, z=2.98 (subject 14). Interestingly, both of these values were traced back to the same consonantal interval consisting of two sounds, which were nasal and plosive [ŋd] across words boundary. From the analysis, it could be observed that this particular combination of sounds along with the fact that these consonants belonged to two adjacent syllables, there was some flexible in which its duration can be short or long.

The analysis in this section suggested that variability of durational values were influenced by the factors of sentences and subjects; even though controlled material was used in this experiment. This investigation revealed that the differences of the durational proportion of the consonantal and vocalic intervals could result the variability of the rhythm metrics scores.

6.6. An evaluation of the rhythm metrics

In a way, these rhythm metrics have been successful in a sense because they positioned Malay overall with the group of languages which has similarity in some respect with Malay. Thus, the overall results of the rhythm metrics were consistent with the auditory observation in which it can be described as a syllable-timed language.

On the other hand, these rhythm metrics have some weaknesses too. The data revealed a lot of variability across sentences and subjects. At the same time, if the scores of the rhythm metrics of individual subjects are examined; there is a possibility that we might come to a different conclusion. Given the knowledge we now have on the variability of sentences and subjects; this is not a surprising conclusion because it would be quite possible to find subjects with stress-timed values of rhythm metrics within the current dataset. At the same time, this might be the same reason Grabe and Low (2002) might misrepresent Malay as a stress-timed language since they based their study on only one speaker

These findings support the recent critique by Arvaniti (2009) and Fletcher (2010) on the validity of these metrics for describing rhythm. The main criticism of the rhythm metrics was surrounding the issue of variability of the scores. This was demonstrated by the findings in the previous sections in which the rhythm metrics were sensitive to the function of sentences, subject and speech rate. Arvaniti (2009) believed that speech rhythm was not entirely about timing; it was about the perceptual regularity from the acoustic signal. Timing was still relevant to describe rhythm but the focus attempting to identify a feature which was regulated as a basis of speech rhythm which could withstand the effects of variables such as sentences and subjects.

Thus, other alternative approaches were explored to identify whether there were other features which could be observed to be regular and therefore possibly associated with the rhythmic characteristics of Malay. The features explored were the syllable duration, inter-intensity minima duration as well as the duration of the single vowels.

6.6.1. Syllable duration

This feature was investigated following the prediction made by the traditional typology of rhythm (Pike 1947) in which the syllable durations of syllable-timed languages were hypothesised to be isochronous.

The measurements of the duration of the syllables were taken. However, the measurements of the syllables in final position and plosive-initial syllables (if any) in the beginning of the sentences were excluded from this analysis. The remaining syllables were divided into syllable types: CVC, CV and V. These syllables were analysed separately.

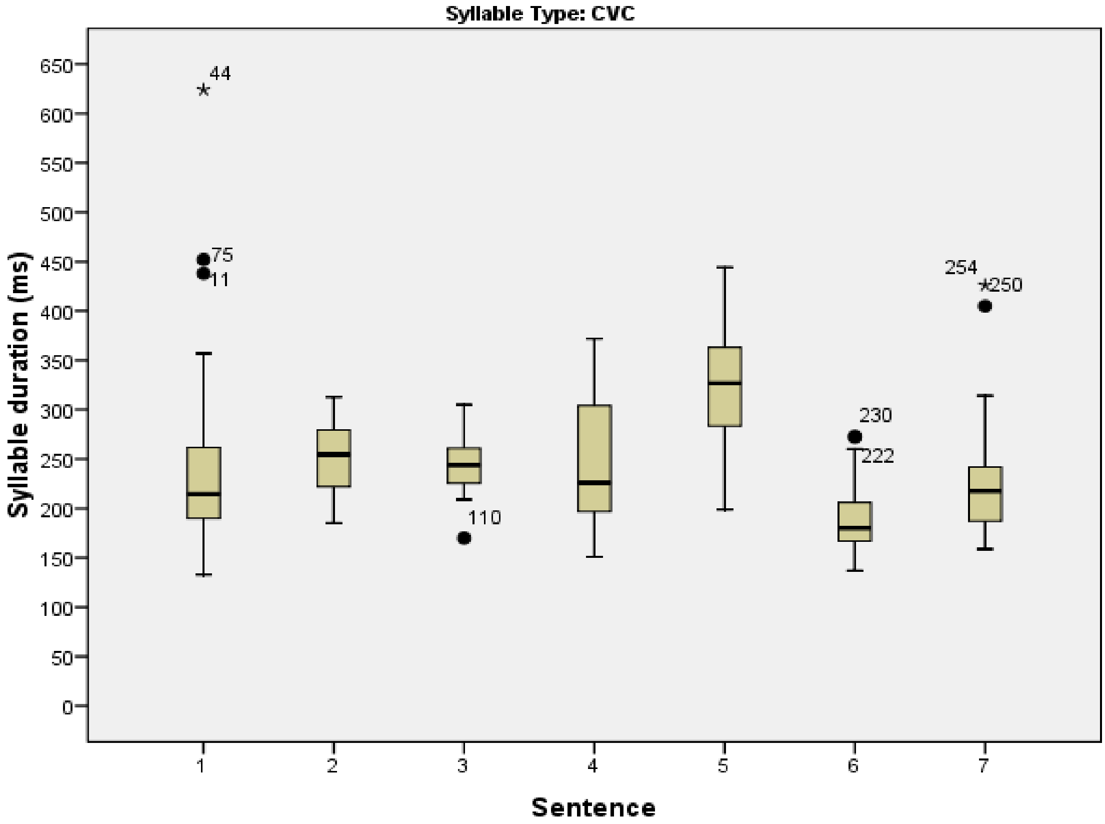
6.6.1.1 Syllable type: CVC.

CVC type syllables were identified in seven of the sentences in the dataset. There were 184 sample sizes for this syllable type across speakers. The mean and standard deviation of the full data set were 238.61 ms and 66.49 respectively (Table 6.9). The boxplot (Figure 6.14) comparing the CVC duration shows that there are considerable variations of the duration across the sentences. Sentence 1, with three outliers, has the most variable duration than the other sentences. While, sentence 3 is the least variable in which most of the duration are in a narrow range compared to that of other sentences. This result suggested that there was no durational isochrony for CVC syllable type as there was variability across sentences.

Table 6.9 Mean and standard deviation of the duration of CVC syllable type of the full dataset

Mean (ms)	Standard Deviation
238.61	66.49

Figure 6.14 Boxplot of the duration of CVC-type syllable of the full dataset across the sentences



As demonstrated in the boxplot (Figure 6.14), there was variability of the duration of CVC syllable type; a decision was made to exclude 5% of each of the minimum and maximum extreme outliers. This action was motivated by the fact that extreme outliers could distort the statistical test based on sample means, leading to central tendency of the data being difficult to capture. Furthermore, this exclusion minimised the influence of variables such as different sentences and subjects rate which could lead to faulty conclusions of the data. In general, the same decision was taken for all the features presented in this section.

The sample size across the sentences after the exclusion was 167. The following are the results for CVC syllable type after 10% minimum and maximum exclusion of the extreme outliers. The mean recorded was 219.9 ms and the standard deviation was 31.8 (Table 6.10)

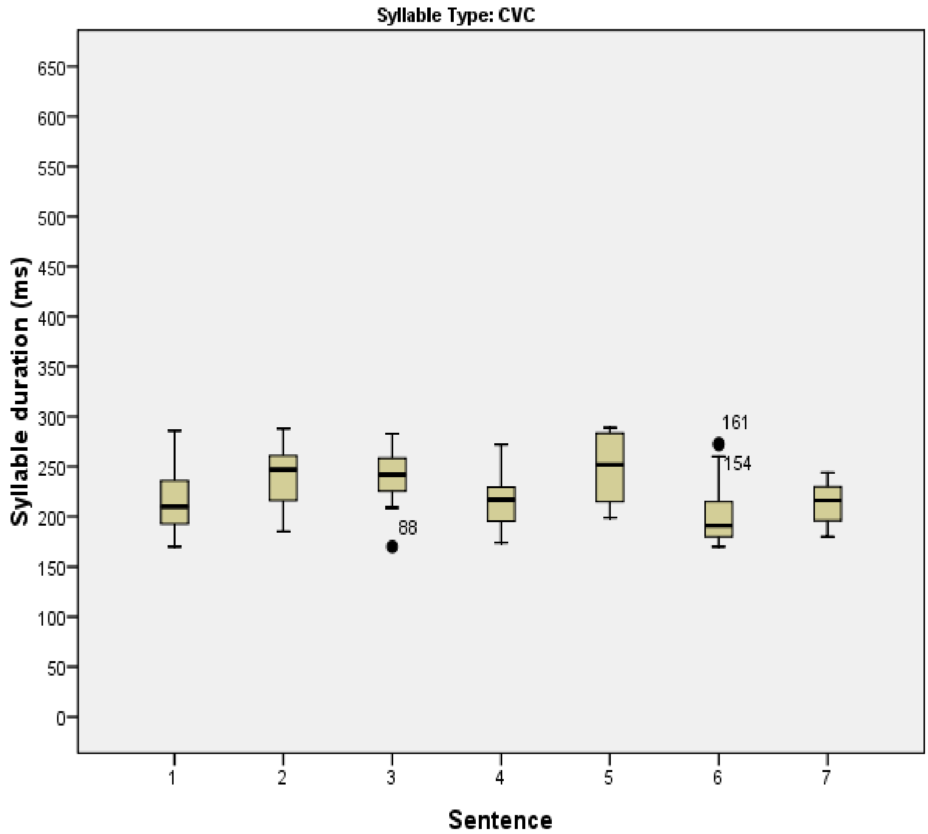
Table 6.10 Mean and standard deviation of the duration of CVC-type syllable of the dataset after the exclusion of 10% of the extreme values

Mean (ms)	Standard Deviation
219.9	31.8

The boxplot (Figure 6.15) illustrates the distribution of the CVC syllable duration after the exclusion of the extreme data on either side of the maximum and minimum durations across the sentences. Although the ranges of the CVC durations are narrow across the sentences; there are still variability observed. The medians of the CVC syllable durations across the sentences are not consistent with one another, for example the median for sentence 5 (252 ms) is the highest while that of sentence 6 (200 ms) is the lowest. The ranges of the CVC syllable durations are also not consistent across the sentences, for example the distribution of the syllable durations for sentence 6 are skewed to the low end of the interquartile range while the opposite is observed for sentence 5.

The evidence from the results of both the full dataset and the 5% exclusion of either side of the maximum and minimum extreme outliers dataset showed that there was variability across the sentences for the durations of CVC syllable type. Thus, it can be suggested that the durations of this type of syllable were not isochronous as claimed by Pike (1947). The next syllable type analysed is CV syllable type. The same methodology was used for pooling the syllable duration measurement across the data.

Figure 6.15 Boxplot of the duration of CVC-type syllable of the dataset after the exclusion of 10% of the extreme values across the sentences



6.6.1.2 Syllable type: CV.

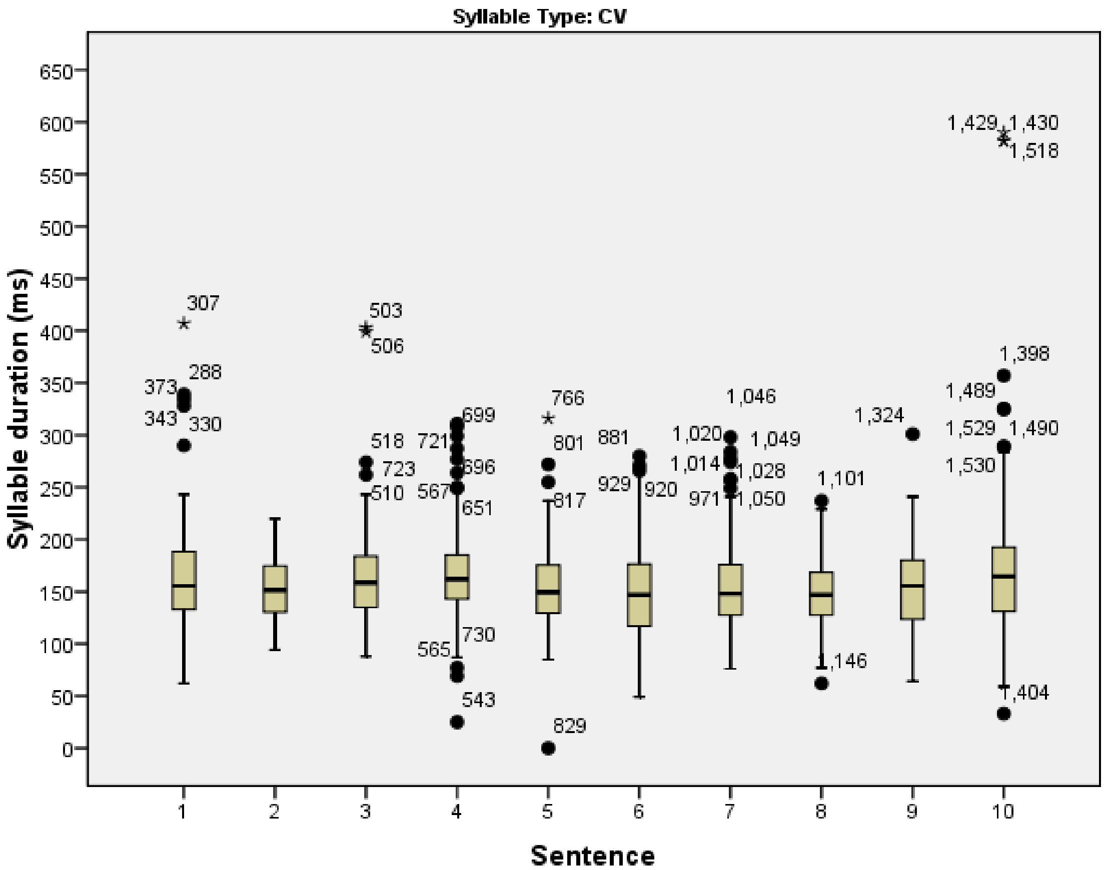
The sample size for this analysis was 1049 across the sentences. The table shows the mean and standard deviation of the duration of CV syllable type across the full dataset (Table 6.11).

Table 6.11 Mean and standard deviation of the duration of CV syllable type of the full dataset

Mean (ms)	Standard Deviation
160.68	48.48

Further observation of the boxplot (Figure 6.16) comparing the sentences for the CV durations of the full dataset shows the medians were relatively consistent. Despite this, there are also variations observed in the boxplot. Sentence 10, with 10 outliers, is the most variable than the other sentences. While, sentence 2, with no outlier, is the least variable. Many outliers are also present across the sentences. These outliers were consistent with the fact that there were influences from variables such as sentences and subjects.

Figure 6.16 Boxplot of the duration of CV-type syllable of the full dataset across the sentences



Following the variability and the outliers in the analysis of CV syllable type duration, 5% extreme outliers from each maximum and minimum value were identified and excluded from the analysis. Similar to the rationale presented in the previous section; this was done in order to identify the central tendency of the dataset and minimise the influence of the variables.

The sample size after the exclusion was 945 across the sentences. The results of the mean and standard deviation of the dataset after the exclusion are presented in Table 6.12.

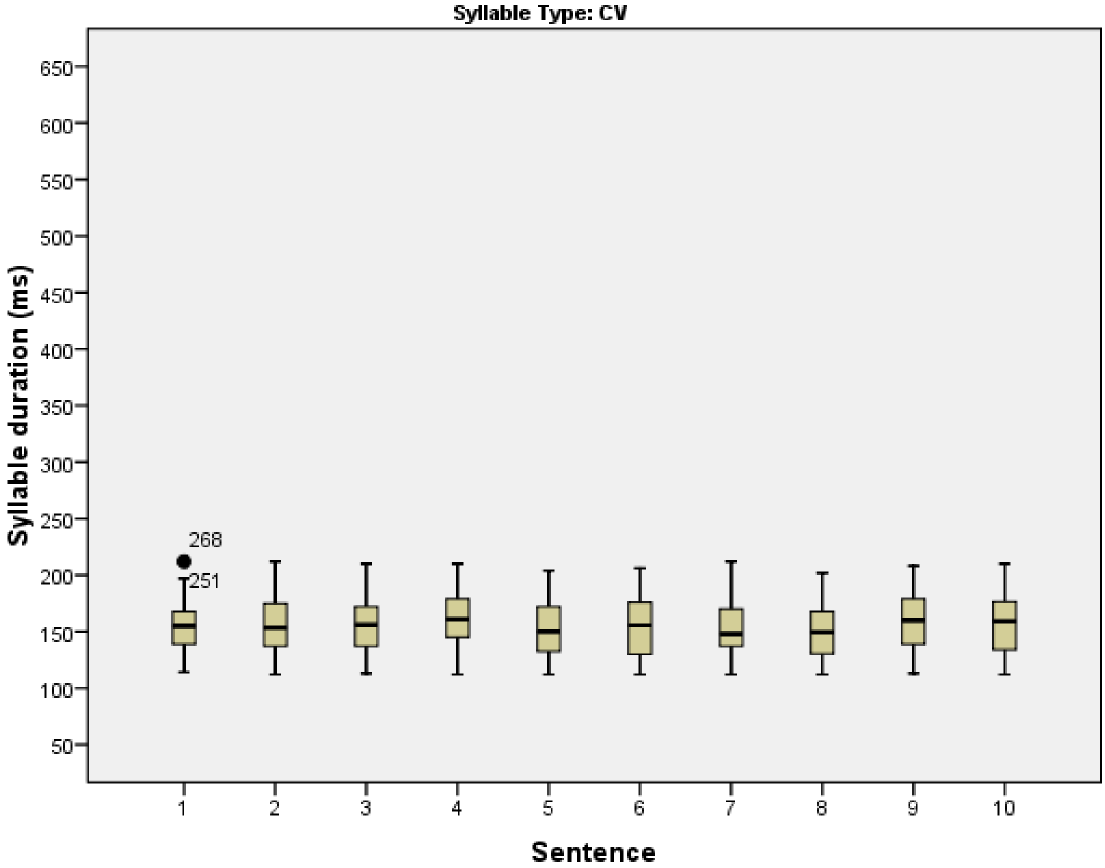
Table 6.12 Mean and standard deviation of the duration of CV-type syllable of the dataset after the exclusion of 10% of the extreme values

Mean (ms)	Standard Deviation
156.00	24.98

The boxplot (Figure 6.17) illustrates the duration of CV syllable type after the exclusion of the 5% outliers on either side of the data. Although it can be observed that the ranges of the CV syllable durations are somewhat uniform; there were still variability. Within the sentence, there are two outliers identified in sentence 1. The variability across the sentences were also identified, sentence 6 has different CV duration than that of sentence 9.

The findings suggested that the duration of CV type syllable was influenced by the sentences. Another syllable type available in Malay: V is discussed in the next section.

Figure 6.17 Boxplot of the duration of CV-type syllable of the dataset after the exclusion of 10% of the extreme values



6.6.1.3 Syllable type: V

This type of syllable was found in four out of the ten sentences The sample size across the sentences were 98. The scores for the mean and standard deviation of the full dataset are presented in Table 6.13.

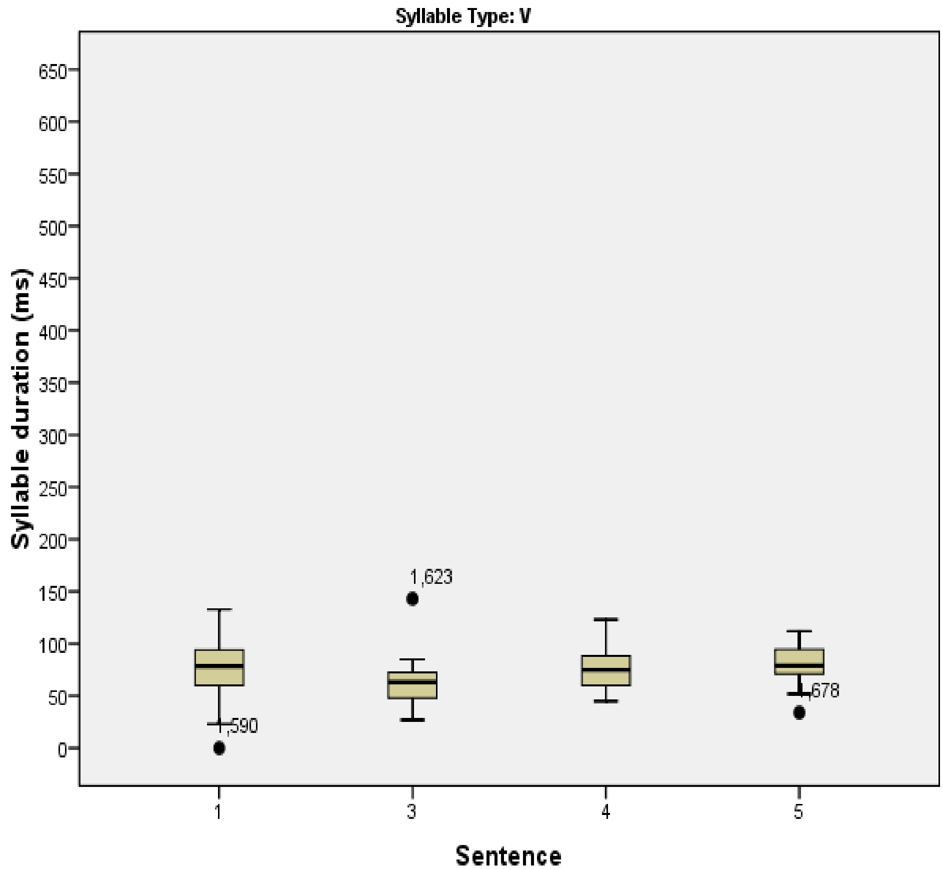
Table 6.13 Mean and standard deviation of the duration of V-syllable type of the full dataset

Mean (ms)	Standard Deviation
77.43	27.00

Further analysis of the boxplot (Figure 6.18) of the full dataset indicates relatively consistent duration of this type of syllable. However, some variability across the sentences was found; in which sentence 5 has the narrowest durational range while sentence 1 has the widest durational range across the sentences. Only two outliers are present in the boxplot.

Comparison made between the results of this section and CVC (Figure 6.14), CV (Figure 6.16) of the full datset implied that the duration of V syllable type was relatively consistent even without the exclusion of the extreme outliers.

Figure 6.18 Boxplot of the duration of V-type syllable of the full dataset across the sentences



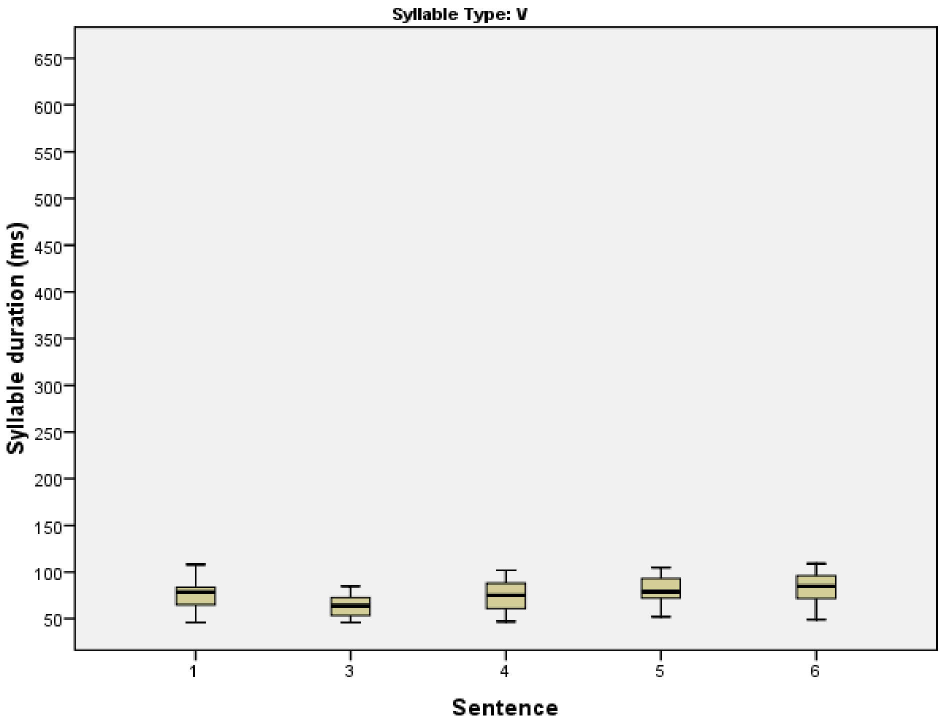
Similar to the previous analysis, the duration of V type syllable was also treated with the exclusion of 5% of both the maximum and minimum outliers. The sample size for this analysis was 89. The results of the mean and standard deviation are presented in Table 6.14.

Table 6.14 Mean and standard deviation of the duration of V-type syllable of the dataset after the exclusion of 10% of the extreme values

Mean (ms)	Standard Deviation
76.23	16.51

Further analysis on the boxplot (Figure 6.19) shows that the duration of V syllable was uniform across the sentences. Comparison of the results, within the V syllable type, between the full dataset and the exclusion indicated not much difference. It could be implied that the duration of V syllable was relatively consistent despite the influence of sentences and subjects.

Figure 6.19 Boxplot of the duration of V-type syllable of the dataset after the exclusion of 10% of the extreme values across the sentences



In this section, all the types of syllable available in Malay were analysed. The analysis on CVC syllable type revealed that its duration varied across sentences. A similar result was also found after the exclusion of the 5% of the maximum and minimum extreme outliers. Thus, no isonchrony was detected within this syllable type. Subsequently, a similar result was found for the CV duration, although after the exclusion of the extreme values, the dataset still showed some variability across the sentences.

However, a considerable isochrony was found within the V type syllable on both the full dataset as well as the exclusion of the extreme values dataset. Following this promising indication of regularity from the speech signal, this feature will be analysed further by taking into account other single non-syllabic vowels in section 6.7.3.

Another feature from the acoustic signal which potentially could be observed as regular and therefore could be associated with rhythmic characteristic of Malay was the inter-intensity minima durations. This will be discussed in the subsequent section.

6.6.2. Inter-intensity minima duration

Another candidate for regularity was the intensity contour. The intensity contour of the speech signal might be of interest since the maximum peaks of the intensity corresponded to vowels. As described in 3.3.2, there were no consonant clusters in the onset or the coda of the Malay syllable; therefore the recurring of these maximum peaks can be observed as somewhat constant. Nevertheless, the problem with these peaks was that they tend to have gentle curves or sometimes more rounded which made it more difficult to identify the stable section of these peaks.

To capture the variation of the intensity contour, the intensity minima was used instead. This was due to the fact that intensity minima were sharper and practically the peak minima were easier to identify than the peak maxima. Thus, it was relatively straightforward to analyse the durations of the intensity minima intervals.

Minimum intensity duration was associated with the type of the combination of consonant-vowel in a syllable or syllables in adjacent position. A low intensity corresponded to the manner of articulations like plosives, nasals and fricatives; while a high intensity usually corresponded to vowels or semi-vowels.

From the reading material, it could be observed that there were combinations of nasal-vowel-nasal, plosive-vowel-nasal, and plosive-vowel-plosive for CVC type. As well as nasal-vowel, plosive-vowel, fricative-vowel for CV type. These combinations in a sentence created a constant rise and fall of intensity on the intensity contours. Similar contour patterns were observed across speakers (Figure 6.20).

To capture this variation, the inter-intensity minima intervals were analysed. The boundary between one intensity minimum peak to the following intensity minimum peak was drawn (Figure 6.20). Then, the durations of these intervals were recorded in SPSS and Excell spreadsheet. Similar to the analysis of the syllable durations, the results of the full dataset and the data after the exclusion of the 5% of the maximum and minimum extreme outliers will be presented.

Figure 6.20 The intensity contours of sentence 1 produced by subjects 1 (penonton 1) and 2 (penonton 2)

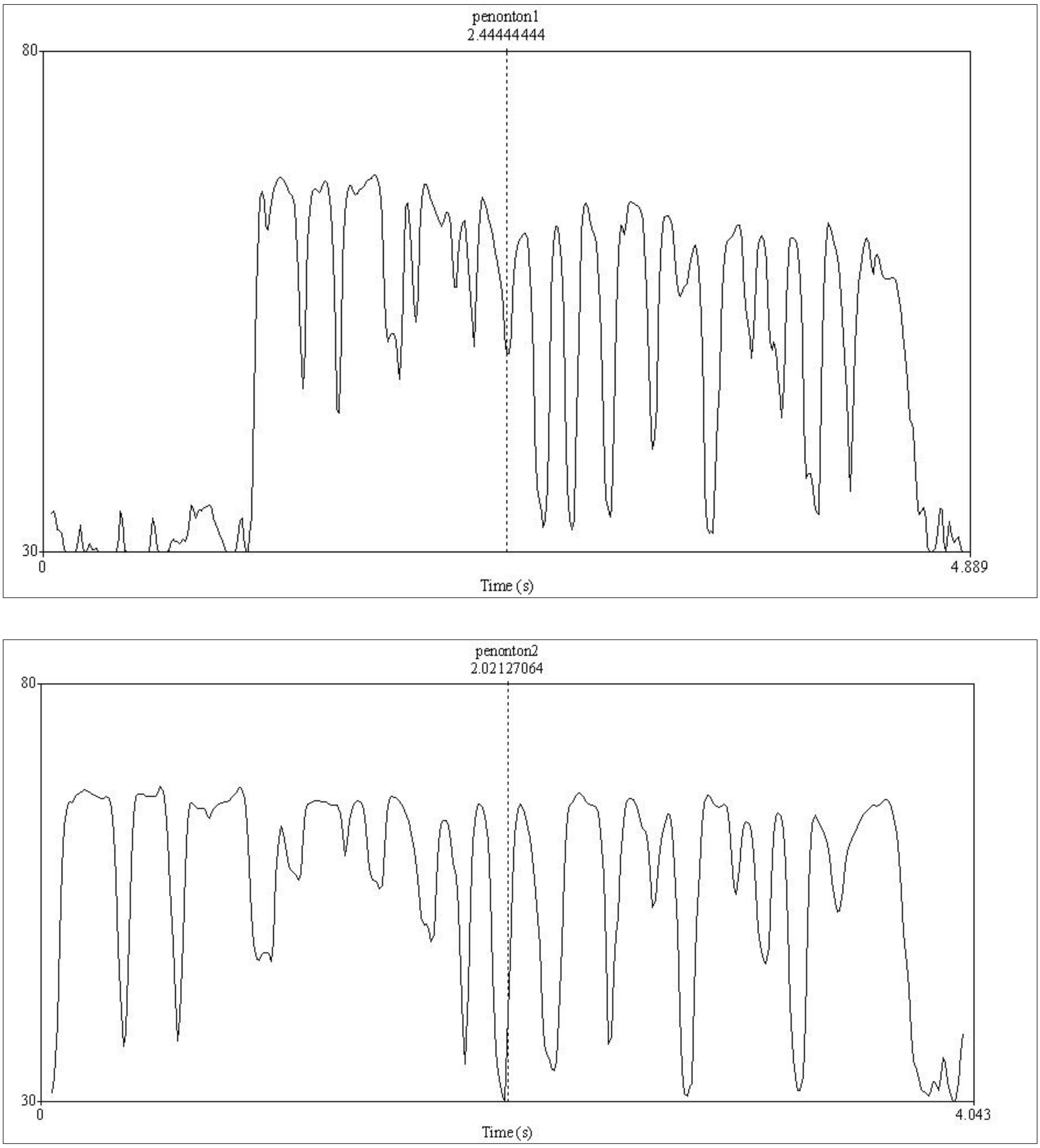
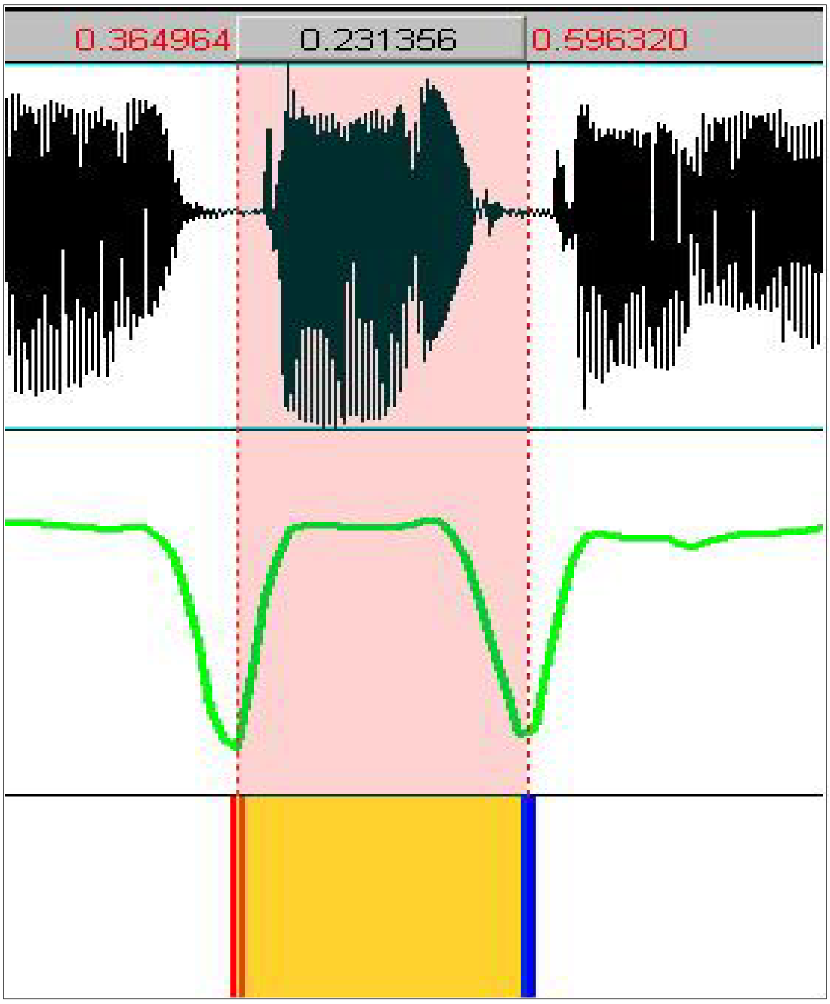


Figure 6.21 Example of boundaries created to determine the length between minimum intensities. The top panel shows the waveform, the middle panel shows the intensity trace and the bottom panel shows the segmentation lines



The sample size for this analysis across the sentences was 2944. The mean and standard deviation of the full dataset of the inter-intensity minima duration are given in Table 6.15.

Table 6.15 Mean and standard deviation of the duration of inter-intensity minima of the full dataset

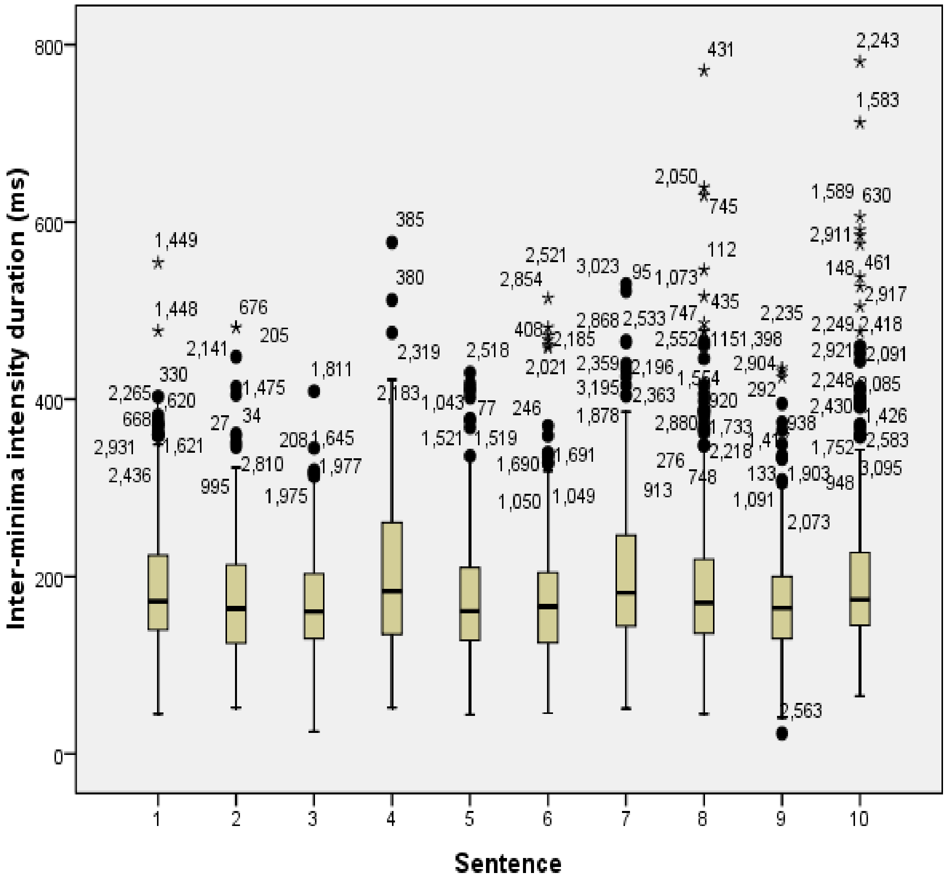
Mean (ms)	Standard Deviation
185.9	79.00

The boxplot illustrates the range of inter-intensity minima duration along with its outliers across the sentences (Figure 6.22). The variability within and across the sentences could be observed. Sentence 4 had the widest inter-intensity minima duration range but with the least outliers across the sentences. The possible explanation for this finding was sentence 4 '*Lolo mengukur benteng dibina Taman Tema Kulim itu*'- /lolo məŋuko benteŋ dɪbɪnə tɐmɐn temɐ kulɪm ɪtu/ has the combination of sounds which consisted of lateral-vowel and nasals-vowel hence the intensity produced by the subjects did not have obvious minimum intensity dip that which did not meet the requirement to be included in the calculation of the inter-intensity minima duration. As a result of this, making the range appeared to be wider in the boxplot.

On the other hand, sentence 9 '*Sepupu Tok Ki mengejek pengemis hodoh dan bodoh itu*'-/səpupu tok kɪ meŋedʒek pəŋəmɪs hodoh dɐn bodoh itu/- has the narrowest interquartile range in comparison to other sentences. There was also variability within this sentence, which was demonstrated with the outliers. A possible explanation for the narrow durational range for this sentence was because 56.3% of the sounds comprised plosive-vowel combinations. As described earlier, the intensity of plosive sounds were lower than that of the vowel, hence the inter-intensity minima contour for this particular sentence was distinct. Table 6.17 summarises the percentages of the sound combination across the sentences.

Meanwhile, sentence 8 was reported to have more outliers then the rest of the sentences which indicate, once again, the issue of variability across the subjects when producing the sentence '*Sam menyemak kos menanam dan mengedar pokok kopi di Puchong*' there are some plosive sequences in this sentences, therefore the intensities can vary from one subject to another.

Figure 6.22 Boxplot of the duration of inter-intensity minima of the full dataset across the sentences



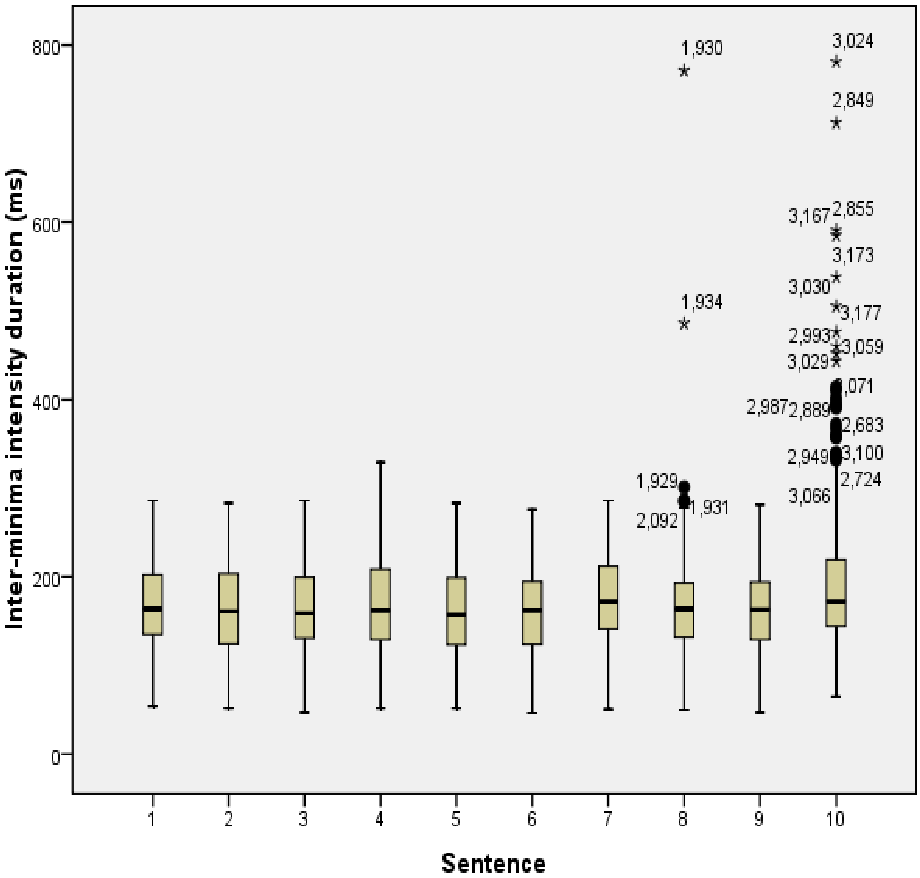
The sample size for this analysis was 2650 across the sentences. The mean and standard deviation of the inter-intensity minima duration after the 10% exclusion of the extreme values are given in Table 6.16.

Table 6.16 Mean and standard deviation of the duration of inter-intensity minima of the dataset after the exclusion of 10% of the extreme values

Mean (ms)	Standard Deviation
171.56	58.51

The boxplot of this dataset (Figure 6.23) illustrates the analysis further. Even when the exteme outliers were excluded, varibility across sentences could still be observed. Sentence 10 is the most variable compared to other sentences. The high variability might be due to the way in which the sounds were combined in this sentence. There were combinations of nasal-vowels (28%) and plosive-vowels (56%) (Table 6.17) in sentence 10 which was consistent with the intensity contour pattern thus resulting the duration of the intensity minima to be variable. However the median values across the sentences were relatively consistent.

Figure 6.23 Boxplot of the duration of inter-intensity minima of the dataset after the exclusion of 10% of the extreme values across the sentences



A one-way ANOVA was carried out in order to study the effect of different sentences on the inter-intensity minima durations. The result showed a significant effect of sentences on this measure ($F_{(9,3266)} = 10.88$, $p < 0.05$), a post hoc comparison showed that sentence 10 was significantly higher than all of the sentences except for sentence 7, and sentences 5 and 6 were significantly lower than sentence 7. This result further supported the variation illustrated on the boxplot (Figure 6.23).

Another one-way ANOVA was also carried out to study the effect of subjects on the inter-intensity minima durations. The results revealed significant effect of the subjects on the duration of the inter-intensity minima ($F_{(19, 3256)} = 5.43$, $p <0.05$). Post hoc comparison showed subject 1 was significantly lower than subject 14, subject 3 was significantly higher than subject 4, meanwhile subject 4 was significantly lower than subjects 7, 12, 14 and 18, subject 10 was significantly higher than subjects 1, 4 and 9, while subject 11 was significantly higher than subjects 1 and 4, subject 15 was significantly higher than subjects 1, 4, 5, 6, 8, 9, 13, 16, 17 and 19 whereas on the other hand, subject 20 was significantly lower than subjects 10, 11, 14 and 15.

The variability of inter-minima duration across the subjects could be attributed to the individual differences in spacing the intensity when they produced the sentences. Table 6.17 illustrates the percentages of the combination of consonant-vowel in the reading materials.

Table 6.17 Percentages of the sound combinations across the sentences

	Sentences									
Sounds combination	1	2	3	4	5	6	7	8	9	10
Plosive-vowel	73.8	66.7	50.0	47.1	46.2	36.4	41.2	52.9	56.3	56.0
Nasal-Vowel	15.7	26.7	18.8	35.3	30.8	27.2	41.2	41.2	25.0	28.0
Fricative-Vowel	10.5	6.6	31.5		15.4	36.4	17.6	5.9	12.5	8.0
Lateral-vowel				17.6						
Affricate-vowel					8.0				6.2	8.0

Based on the significant variability found across the sentences and the subjects for the calculation of inter-intensity minima duration, it was appropriate to conclude that this feature was not regulated in Malay. At the same time, it was also sensitive to the functions of variables such as sentences and subjects. Hence, this feature was not able to capture the rhythmic characteristics of the language.

Subsequent section explores the final feature proposed in which could be associated to the rhythmic characteristics of Malay which was the single vowel duration.

6.6.3. Single vowel duration

The third alternative explored to identify the regulated feature in which can possibly best relate to Malay rhythm is the single vocalic duration. The rationale for investigating this acoustic measure is based on the initial observation in section 6.6.1 in which there was a consistency of the duration of one of the syllable types namely the V type syllable across the data (see 6.6.1.3).

Furthermore, another rationale for exploring this feature was derived from the work done earlier (Chapter 4) in which it was found that there was no sentence prominence in Malay. This could possibly reflect the regularity of single vowel duration. Therefore, it was appealing to this research to analyse the single vowel duration of the current data as this may be a potential prominent feature to describe the rhythmic characteristics of Malay

Unlike those rhythm metrics proposed by Ramus et. al (1999) and Grabe and Low (2002) This analysis focussed only on the measurements of single vowel duration and not sequences of vowels in an intervals either across syllables or words as in the ΔV and the *n*PVI. At the same time, this analysis excluded all the single vocalic intervals in sentence final position; this was done to avoid the durational effect due to final lengthening.

The methodological differences in defining the vocalic intervals for analysis between rhythm metrics and the single vowel duration are given in Table 6.18. Any vocalic intervals which were formed by a single vowel were included in both the analysis. The single vocalic interval also included vowel sequences with glottal stop insertion between them. Meanwhile, a vowel interval which included sequences of two or more vowels across word or syllable boundaries was excluded from the current analysis. Vowel sequences in which there was a glide inserted in between them were also excluded from this analysis. Other types of the vowel intervals not included in this analysis was vowels in vhv sequence as found in sentence 5. It was excluded due to the difficulty to segment the boundaries of the glottal fricative in intervocalic position. The following section presents the results of the single vocalic duration analysis of the data from the second experiment.

Table 6.18 A summary of the comparison between the criteria used to categorise the vocalic intervals of the rhythm metrics and of the single vocalic intervals

	The rhythm metrics (%V &ΔV) (Categorised as vocalic intervals)	The single vocalic duration (Categorised as vocalic intervals)
Single vowel in an interval (A vowel in between two consonants including the vowels separated by the [ʔ] insertion)	Included	Included
Vowels sequences across word/syllable boundaries (Vowels which were continuous, or there was a glide insertion)	Included	Excluded
In vhv sequence (In which it was difficult to segment the boundaries of the consonant in the intervocalic position)	Included	Excluded
Vowels in final position	Excluded	Excluded

The sample size for the single vocalic duration across the sentences was 2626. The overall mean of the full dataset of the single vocalic duration was 86.14 ms while standard deviation was 35.59. The means and standard deviations across sentences and subjects are given in Tables 6.19 and 6.20:

Table 6.19 Mean and standard deviation of the single vowel duration of the full dataset across the sentences

Sentence	Mean (ms)	Standard Deviation
1	84.24	34.20
2	83.31	24.38
3	77.84	31.20
4	88.00	31.33
5	85.81	32.19
6	84.28	32.72

7	95.52	49.46
8	87.36	34.42
9	88.61	36.55
10	87.81	39.28

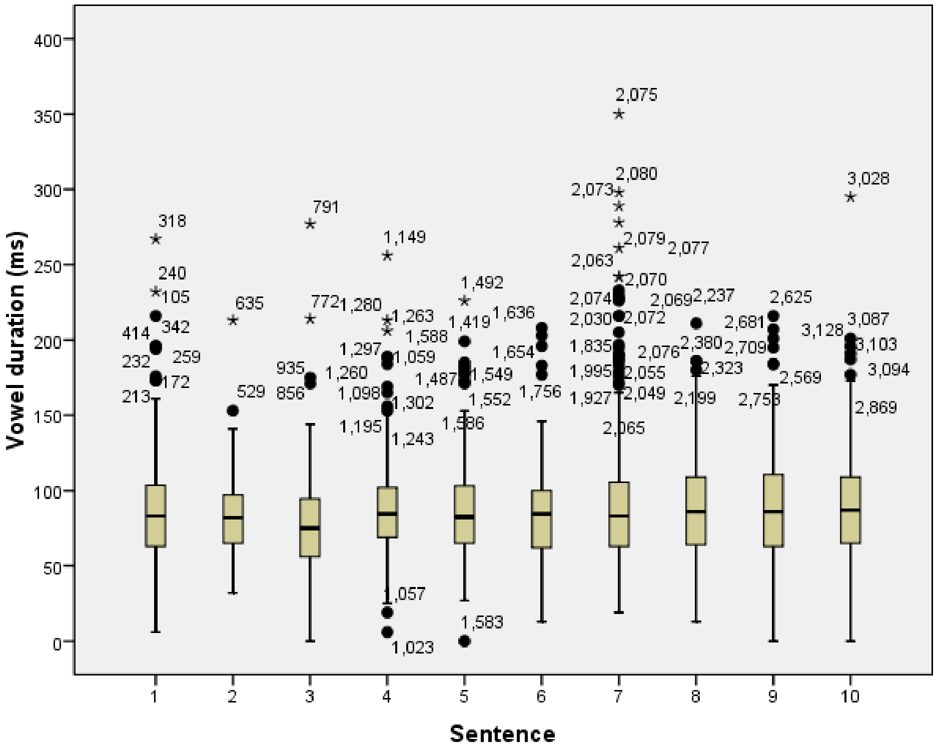
Table 6.20 Mean and standard deviation of the duration of the single vowel of the full dataset across the subjects

Subject	Mean (ms)	Standard Deviation
1	82.39	30.43
2	81.98	30.20
3	84.01	35.94
4	74.94	31.63
5	78.20	46.84
6	85.30	29.00
7	86.23	31.65
8	88.85	31.89
9	68.00	27.83
10	87.64	35.61
11	101.00	40.34
12	86.01	34.66
13	79.26	34.16
14	85.74	40.96
15	108.94	44.08
16	89.53	34.92
17	93.45	30.61
18	83.35	30.51
19	89.80	34.50
20	88.36	30.16

The boxplot (Figure 6.24) illustrates the relative consistency of means and interquartile range across the sentences despite the outliers occurring in all the sentences. The outliers reflected the durational variability of the single vowel duration within and across the sentences.

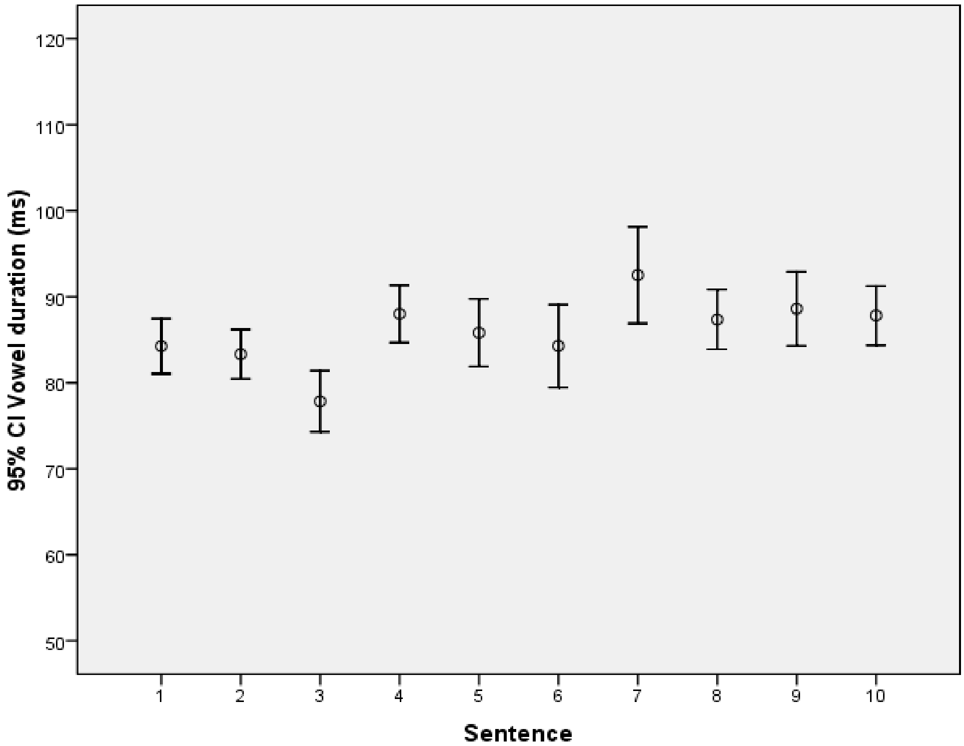
A one-way ANOVA carried out to study the effect of the sentences on the single vocalic duration revealed a significant effect of the sentences on the single vocalic duration ($F_{(19, 3240)} = 10.30$, $p<0.01$). The post hoc test revealed that sentence 7 was significantly higher than the other sentences. This outcome was not surprising since the speakers exhibited variability in the manner of which the sentences were produced. The differences across the speakers with regard to the proportions of consonantal and vocalic durations within the sentences could influence the durational values of the intervals as shown in the acoustic analysis in section 6.5.

Figure 6.24 Boxplot of the duration of single vowel of the full dataset across the sentences



Further investigation on the full dataset was carried out from which the error bars of the full dataset were generated (Figure 6.25) each of the bars occupies an approximately equally narrow range at 95% confidence interval. This outcome suggested that the duration of the single vowel within the mean values of these sentences were relatively consistent. Single vowel duration might be considered as a feature which is regulated and which forms the basis of the rhythmic property of Malay.

Figure 6.25 Error bars of the duration of single vowel of the full dataset across the sentences



In order to capture the central tendency of the single vocalic duration as displayed in the error bar (Figure 6.25), the decision to exclude the edges of the distribution was made in order to control the data from the influence of variables such as sentences and subjects. To extract the central tendency of the single vowel duration; 5% of both the maximum and minimum outliers were eliminated (10% of the overall data). The sample size for this dataset was 2364. The overall mean and standard deviation of vocalic intervals after the exclusion were 83.91 ms and 20.25 respectively. The mean values and standard deviation of each of the sentences and the subjects are given in Table 6.21 and Table 6.22.

Table 6.21 Mean and standard deviation of the duration of single vowel of the dataset after the exclusion of 10% of the extreme values across the sentences

Sentence	Mean Value (ms)	Standard Deviation
1	83.64	20.45
2	81.68	19.19
3	77.74	20.27
4	85.31	18.52
5	82.98	19.74
6	83.99	19.43
7	83.00	19.54
8	86.58	21.04
9	84.39	21.43
10	87.34	20.71

Table 6.22 Mean and standard deviation of the duration of single vowel of the dataset after the exclusion of 10% of the extreme values across the sentences across the subjects

Subject	Mean Value (ms)	Standard Deviation
1	85.09	20.43
2	83.76	18.66
3	83.19	21.18
4	78.43	19.52
5	79.06	18.55
6	82.39	20.28
7	83.00	20.36
8	84.56	20.31
9	73.27	17.47
10	82.56	19.68
11	90.26	18.96
12	85.20	21.28
13	80.69	20.00
14	81.88	20.99
15	91.97	20.66
16	86.30	19.14
17	87.01	19.41
18	85.62	19.39
19	88.24	22.79
20	85.17	18.75

Despite the influence of the variables such as sentences and subjects, the boxplot (Figure 6.26) and the effect size showed that single vocalic duration was relatively consistent.

A one-way ANOVA was conducted to analyse the differences of the vocalic duration between these sentences. The results showed significant results ($F_{(9,2616)}$= 5.05, $p< 0.05$) and the post hoc comparison indicated sentence 3 was significantly lower than sentences 1, 4, 8, 9 and 10 while sentence 2 was significantly lower than sentence 10. At the same time, the distribution of the boxplot of the single vocalic duration across the sentences in Figure 6.27 demonstrated that the variability reported in the ANOVA was not visually apparent.

Figure 6.26 Boxplot of the duration of single duration of the dataset after the exclusion of 10% of the extreme values across the sentences

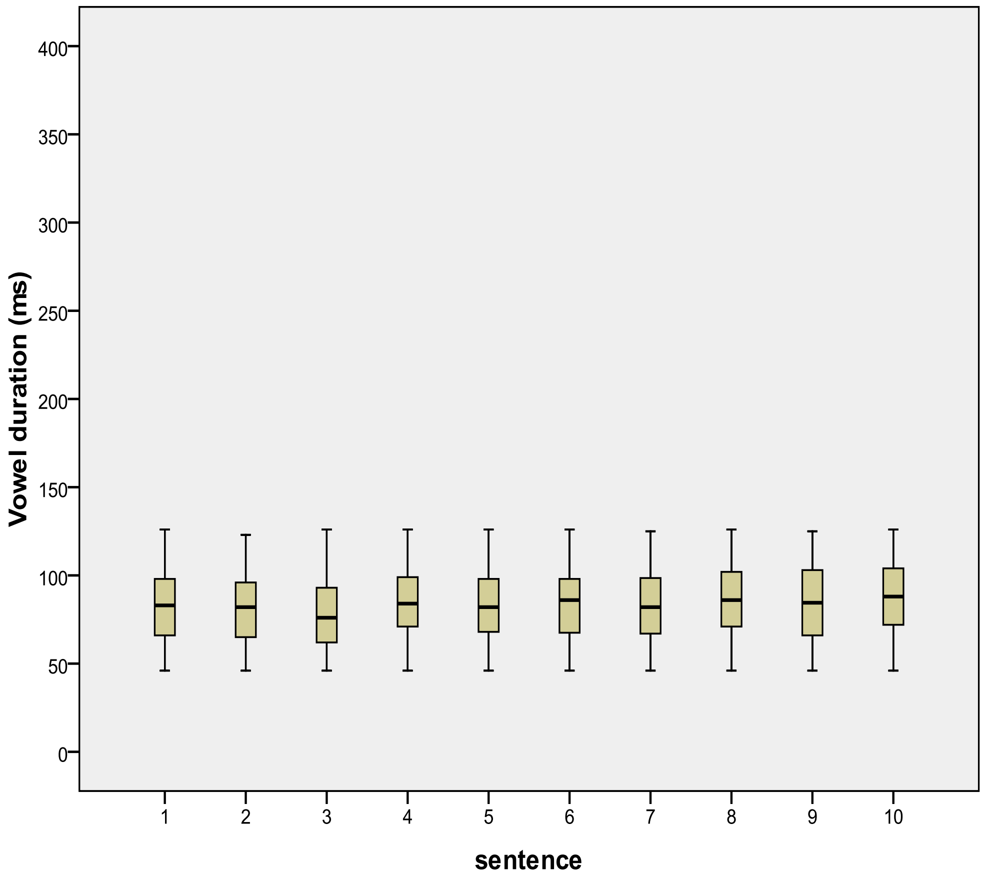
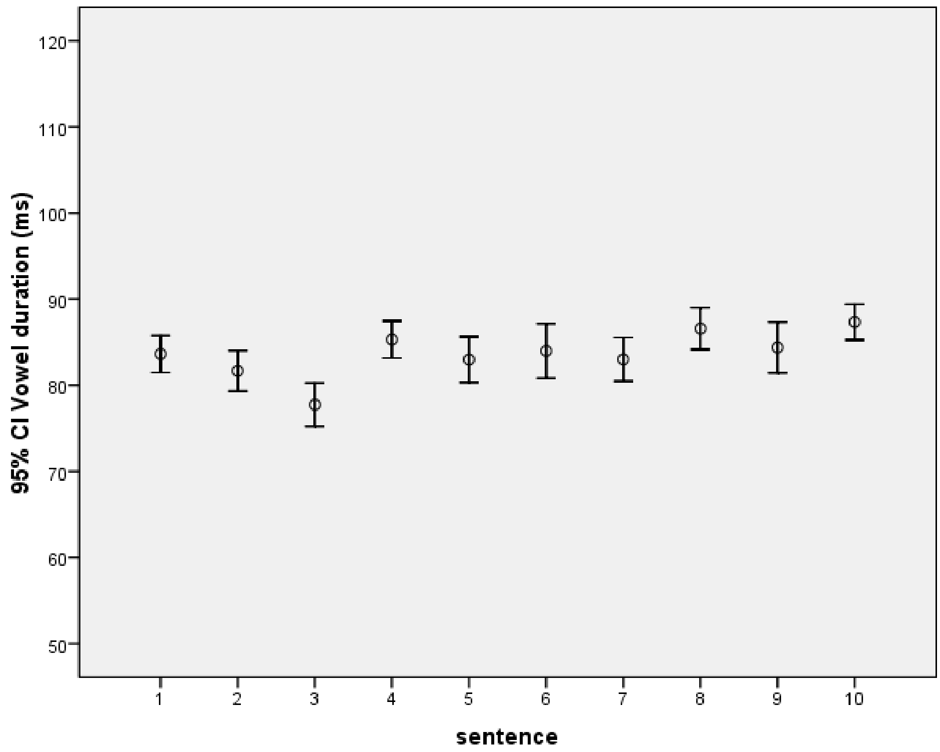


Figure 6.27 compares the means across the sentences in which a highly uniform range of the error bars at 95% confidence interval were observed. The narrow range reflects the relative consistency of the data within

each of the sentences. After the exclusion of the extreme values, an even more regular pattern of consistency of the single vowel duration was observed. This further substantiate the findings of the single vocalic duration of the full dataset in which there was relative consistency across the sentences despite the influence of sentences and subjects.

Figure 6.27 Error bars of the duration of single vowel of the dataset after the exclusion of 10% of the extreme values across the sentences



With regard to the effect size calculated using partial eta squared, 17% of the variability of the single vocalic duration was explained or predicted by the sentence affliation of the vowels concerned. A small effect size further highlighted that the differences was consistent with the finding in which the data were centrally pooled towards a narrow range of the single vocalic duration albeit with some speaker and sentence differences found in this experiment.

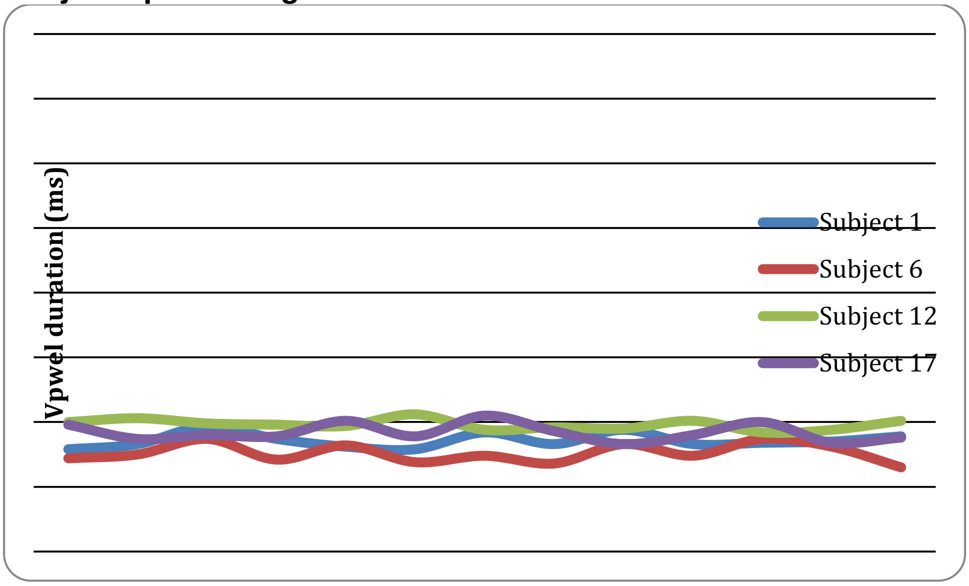
A comparison of the datasets after the 10% exclusion extreme values was made between the single vowel duration and the duration of CVC syllable type. It revealed that the latter still had variability within the sentence as well as across the sentences (Figure 6.15) even after the exclusion; this showed that the CVC syllable type was not regulated. On the other hand, greater consistency of the former suggested that this feature was regulated across the dataset.

Another comparison of the datasets after 10% of the exclusion of the extreme values was made between the single vowel duration and the duration of CV syllable type. There was relative similarity as well as differences between these two measures. Similarly, these two features seemed to have relatively consistent range across the sentences (Figures 6.17 and 6.27), however on a closer inspection, the duration of CV syllable type were varied within each of the sentences and across the sentences, some outliers were also identified in this dataset. Whereas, on a detailed analysis of the boxplot of the single vowel durations, there was no variability found within the sentences as well as across the sentences.

On these two accounts of comparison between the results of the single vowel duration and syllable duration, it was evident that the regularity of the acoustic measures of the speech was best captured by the set of specific single vowel duration. The analysis in this section showed that the findings in a syllable analysis of just the V syllable type (section 6.7.3) which was quite consistent across sentences and subjects could be applied to all syllable types if only the single vowel duration of these syllables were looked at.

To exemplify the relative consistency of the single vowel duration further, the line graph (Figure 6.28) illustrates all the single vowel durations of four subjects (subjects 1, 6, 12 and 17) of sentence 2- /pətɐnɪ bebɐs məmbɪnə benteŋ dekɐt dusunɲə/. It can be observed that within each of the subjects, the durations are relatively regulated.

Figure 6.28 The line graph of single vowel durations of four subjects producing Sentence 2



This research also acknowledged the fact that, to some extent, the excluded vowel duration (Table 6.18) might also contributed to capturing the rhythmic features of Malay.

To further substantiate the regularity of the single vowel duration, a comparison of the results of the single vocalic duration from this research with those of other languages from both end of the rhythmic continuum was also made. This was to test the extent to which the variability of the single vocalic duration of the languages in the syllable-time end of the continuum is lower than those languages in the stress-timed end of the continuum. Table 6.27 summarises the comparison.

Jacewicz et. al (2007) compared the vowel duration of three American English dialects, namely Inland North, Midlands and South. They analysed four short vowels and a diphthong in their study, however for the present purposes I focus only on the four vowels /ɪ/, /ɛ/, /æ/ and /e/. A total of 54 subjects were involved in this study, they were 18 native speakers of each of the dialects consisting of nine male and nine female subjects.

The results showed that main sentence stress was systematically varied to create different emphasis conditions for each of the target words. There was some variability across the dialects; the given results are presented in Table 6.23 in which the standard deviations of all the single vocalic durations were higher than that of the single vocalic durations of Malay in this present study. Higher standard deviations compiled in Jacewicz et. al (2007) were due to the fact that there were stressed and unstressed vowels in the varieties of American English. Fry (1955), Dauer (1983), Couper –Kuhlen (1986) Arvaniti (1994), Berkovits (1994) and de Jong (2004) prooved that the durational values of each of the vowels were influenced by the stress they received. At the same time, American English was traditionally positioned in the stress-timed end of the continuum, the greater variability across its vowels were inevitable.

Another example of a study on vowel duration of another variety of English was carried out by McClure (1977), he researched Scottish English following Aitken’s Law (1962) which stated that all vowels and diphthongs were long in stressed open syllables, between voiced fricatives and /r/ and before morpheme boundaries and they were short elsewhere In order to test this law on Scottish English, McClure recorded himself reading target words with /i/, /ɪ/, /e/, /ɛ/, /a/, /ʌ/, /ɔ/, /o/ and /u/. The target words were repeated twice in isolation and once embedded in sentences. Table 6.23

shows the results of the single vowel durations produced in sentence context only since it is the most consistent point of comparison to that of the current study of Malay. The results were derived from calculating the mean and the standard deviation of all the vowel durations recorded because McClure (1977) only provided all the durational information of all the target vowels without giving the overall mean or standard deviation. The mean duration of 208.23 ms and standard deviation of 79.76 were higher than those of Malay respectively, this reflected greater variability in the single vowel duration in Scottish English similar to those of the varieties of American English found in Jacewicz et. al (2007). The variety of Scottish English also was rhythmically considered to be in the stress-timed end of the continuum. This was further evidence that there were differences of the single vocalic duration of languages grouped in different end of the continuum when the comparison was made between the results of the single vocalic duration of these studies and that of the Malay in this current data.

Spanish which was regarded to be in the same syllable timed end of the continuum of Malay was also looked at. Cervera, Miralles, and Gonzalez-Alvarez (2001) investigated and compared Spanish vowel produced by Laryngectomised subjects with that of a group of controlled subjects who have normal speech. The control group comprised 10 male subjects. They produced twenty-four Spanish two-syllable words (CVCV). For the purpose of comparison with the current Malay data, only the results of the single vocalic duration produced by the control group were used.

The list of the words reflected frequent syllable structure and stressed patterns in Spanish. Table 6. shows the results of the analysis of the vowel duration for the control group, it could be seen that four out of the five vowels had relatively narrow standard deviations. These results were consistent with the findings from this study. Furthermore the overall mean standard deviation for Spanish taken from Cervera et. al (2001) showed that the variability of vowel duration for Spanish was small (27.20). Hence, the single vocalic durations of Malay and Spanish had a narrow range compared to the other languages in the opposite end of the continuum.

It was important to include that whilst Malay lacks clear correlates of stress (see Chapter 4); Spanish has a clear lexical stress. It is also acknowledged that the study by Cervera et. al (2001) and the present study are not strictly comparable. This was because the former analysed the duration of single vowels in isolated context while the latter analysed the duration of single vowels in connected speech. However, as it is observed, both the datasets showed greater consistency across single vowel durations.

Following the analysis across the languages, single vocalic duration could be regarded as a regulated feature for the languages in the syllable-timed end of the continuum given that there was less variability of the single vocalic duration compared to languages on the stressed-time end of the continuum.

Table 6.23 Summary of the comparison of the single vocalic durations between languages

<table>
<tr><th rowspan="2">Language (traditional classification)</th><th rowspan="2">Summary of Results</th><th colspan="2">Overall Results</th></tr>
<tr><th>Mean Duration (ms)</th><th>Standard Deviation</th></tr>
<tr><td>Malay (syllable-timed language)[1]</td><td></td><td>83.91</td><td>20.25</td></tr>
<tr><td>American English (stress-timed)[2]</td><td>Four vowel durations in two consonant contexts ms (sd)
<table>
<tr><th></th><th>b_dz context</th><th>b_ts context</th></tr>
<tr><td>/ɪ/</td><td>137 (46)</td><td>112 (41)</td></tr>
<tr><td>/ɛ/</td><td>154 (48)</td><td>128 (37)</td></tr>
<tr><td>/æ/</td><td>192 (52)</td><td>154 (40)</td></tr>
<tr><td>/e/</td><td>212 (51)</td><td>173 (42)</td></tr>
</table></td><td>157.3</td><td>44.63</td></tr>
<tr><td>Scottish English (stress-timed)[3]</td><td>Nine vowels were studied in 7 different contexts</td><td>208.3</td><td>79.76</td></tr>
<tr><td>Spanish (syllable-timed language)[4]</td><td>Results from the control group
<table>
<tr><th></th><th>Ms</th><th>SD</th></tr>
<tr><td>/i/</td><td>81</td><td>26</td></tr>
<tr><td>/e/</td><td>78</td><td>28</td></tr>
<tr><td>/a/</td><td>89</td><td>23</td></tr>
<tr><td>/o/</td><td>83</td><td>19</td></tr>
<tr><td>/u/</td><td>75</td><td>40</td></tr>
</table></td><td>81.2</td><td>27.20</td></tr>
</table>

[1]Results from the second experiment of the current research
[2]Jacewicz et. al (2007)
[3]McCure (1977)
[4]Cervera et. al (2001)

To confirm further the decision to consider single vocalic duration as the regular feature this signified the rhythmic patterns of Malay. The dataset of the single vocalic duration of the first experiment as seen in Chapter 4 was also analysed.

6.6.3.1 Single vocalic duration of the first experiment.

The decision to carry out an identical analysis on a different dataset, namely the data collected for the word stress experiment reported in Chapter 4 was made to further substantiate the claim that the single vowel duration of Malay was relatively consistent. This consistency captured the rhythmic characteristic of Malay.

To recap, the dataset of the first experiment included two sections which were the production of the target words in isolation and the production of the target words in connected speech. This material was produced by three subjects. For the purpose of this current investigation, the analysis was carried out on the connected speech section of the experiment which consisted of twelve sentences. The criteria for the selection of the single vocalic duration were exactly the same used for the second experiment in 0.

The advantage of using the dataset from the first experiment was there was greater influence of sentences due to them not being controlled with regard to the sound combinations. The sample size also was smaller than that of the second experiment because only three subjects were included in the first experiment.

Similar to those of the previous analysis, this data was also analysed at the full dataset level. The sample size for this analysis was 315. The overall mean was 77.81 ms (SD 25.70). The means of each of the sentences and subjects are given in Table 6.24 and Table 6.25 respectively. Initial observations of these tables revealed to be promising in which the single vowel duration of the full dataset was relatively consistent across the sentences and the subjects.

Table 6.24 Mean and standard deviation of the duration of single vowel of the full dataset of the first experiment across the sentences

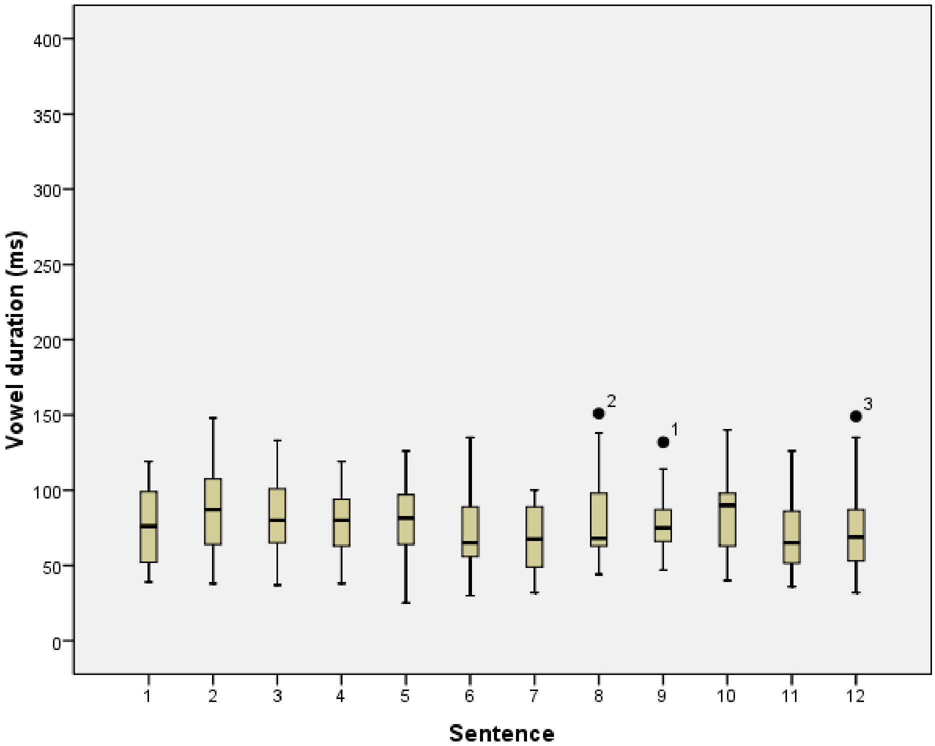
Sentence	Mean (ms)	Standard deviation
1	75.45	25.06
2	86.89	27.25
3	82.14	27.90
4	78.33	21.67
5	79.13	23.75
6	74.19	27.89
7	67.45	22.58
8	79.41	27.80
9	79.05	20.37
10	83.57	26.85
11	70.15	24.39
12	74.86	32.04

Table 6.25 Mean and standard deviation of the duration of single vowel of the full dataset of the first experiment across the subjects

Subject	Mean (ms)	Standard deviation
1	75.55	24.33
2	77.55	24.44
3	80.32	28.16

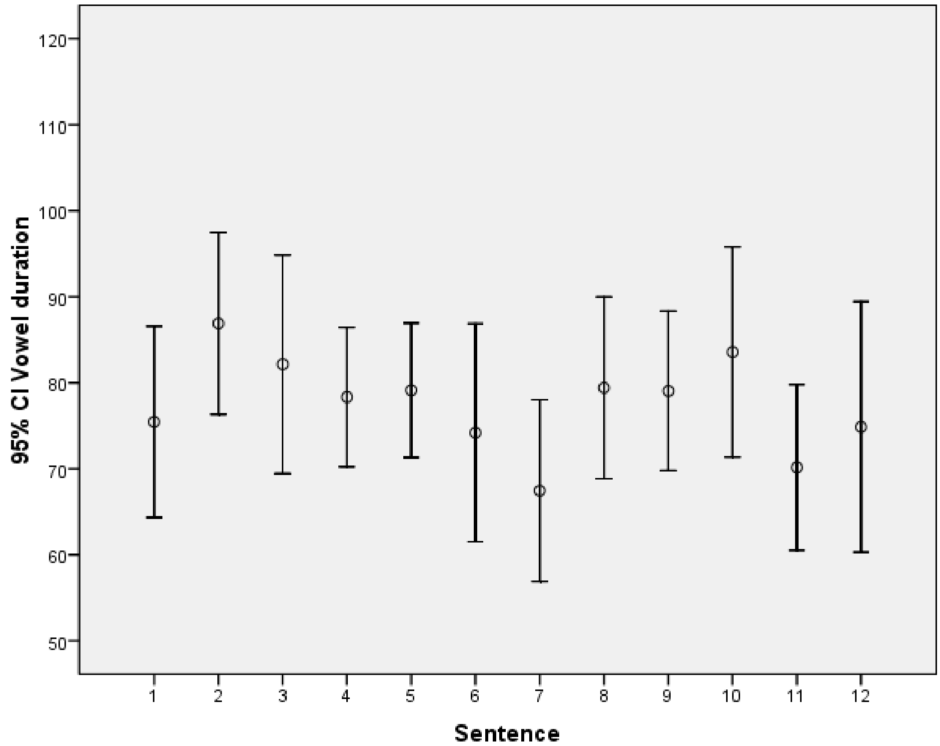
From the boxplot in Figure 6.29, it can be seen that the range for the single vowel duration was reasonably regular. While there are outliers on the graph which suggest variability; generally the means and the interquartile range were consistent across the sentences. A one-way ANOVA was conducted to compare the effect of the sentences on the single vocalic duration in this experiment. There was no significant effect of the sentences on the single vocalic duration ($F_{(11,287)} = 1.09$, $p > 0.05$). This finding suggested consistency of the single vocalic duration across the sentences.

Figure 6.29 Boxplot of the duration of the single vowel duration of the full dataset after of the first experiment across the sentences



To compare the means across the sentences, error bars at 95% confidence level were generated (Figure 6.30). It was observed that these sentences have relatively similar range of confidence intervals; which could reflect consistency across the twelve sentences.

Figure 6.30 Error bars of the single vowel duration of the full dataset of the first experiment across the sentences



To extract more of the central tendency pattern of this feature, a dataset which excluded 5% of both the maximum and minimum extreme values were also analysed (the sample size for this current data was 284 across the sentences). The mean of this newly generated dataset was 77.05 ms (SD 20.70). Table 6.26 and Table 6.27 show the means and standard deviations of the sentences and subjects.

Table 6.26 Mean and standard deviation of the single vowel duration of the dataset after the exclusion of 10% of the extreme values of the first experiment across the sentences

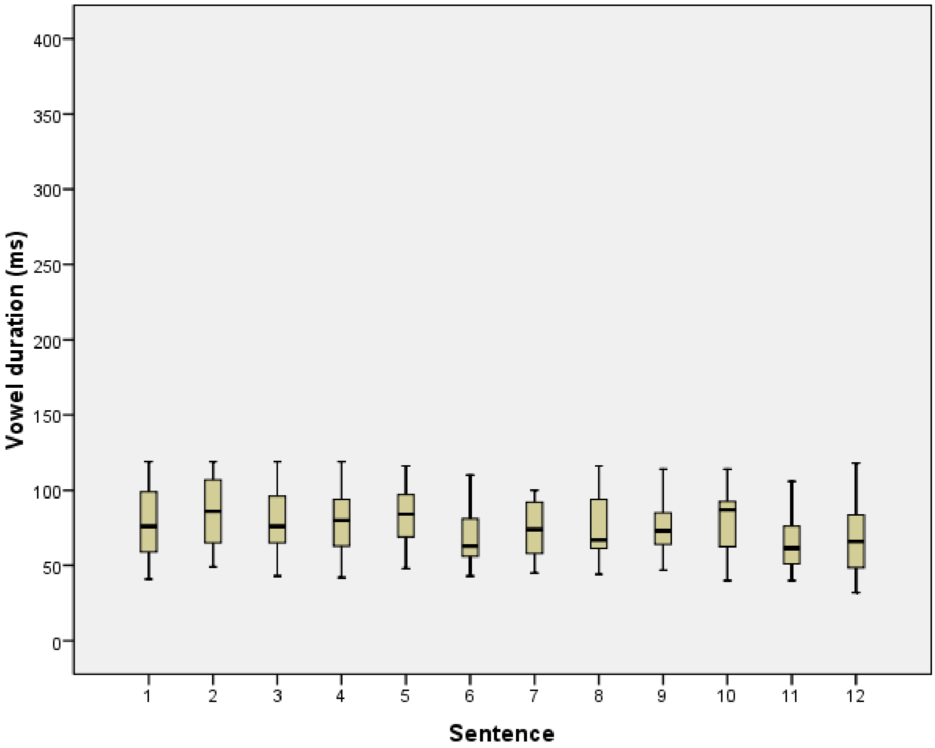
Sentence	Mean (ms)	Standard deviation
1	77.19	24.29
2	84.60	22.14
3	79.56	19.51
4	79.92	20.65
5	82.21	17.95
6	70.22	19.71
7	73.35	18.92
8	74.59	21.89
9	76.42	17.25
10	78.37	22.26
11	65.85	19.30
12	67.79	24.10

Table 6.27 Mean and standard deviation of the single vowel duration of the dataset after the exclusion of 10% of the extreme values of the first experiment across the subjects

Subject	Mean (ms)	Standard deviation
1	75.67	20.02
2	75.57	21.62
3	78.07	21.60

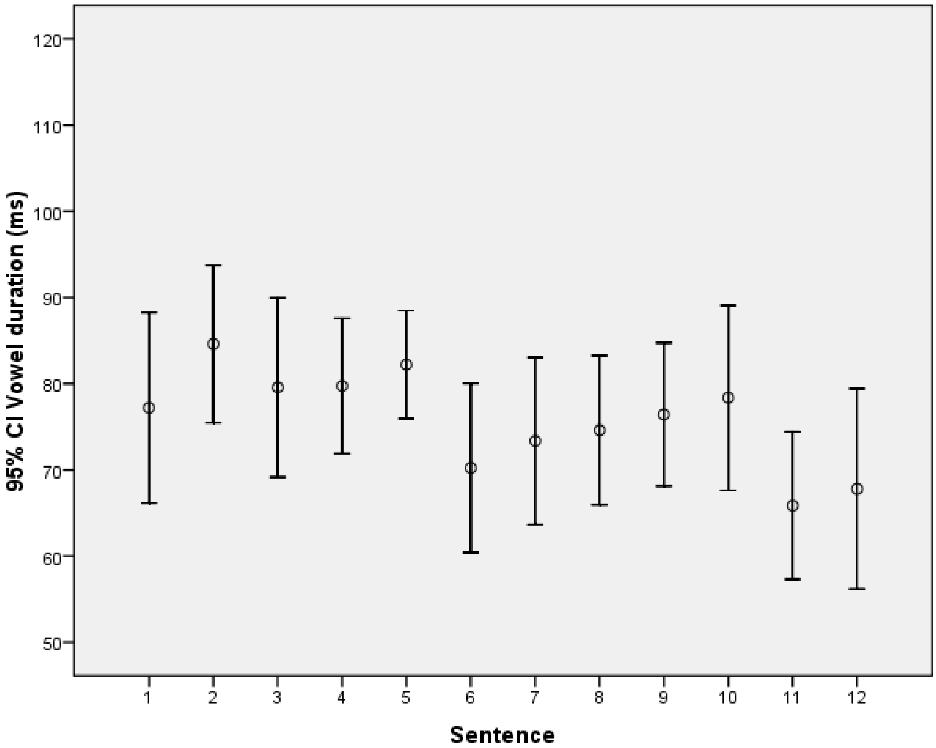
Figure 6.31 illustrates the boxplot of the dataset of the first experiment when the 5% of the minimum and maximum extreme values were excluded. It shows more consistency across the sentences. A one-way ANOVA also was conducted to compare the effect of different sentences on the vocalic duration. No significant effect of the sentences was found on the durational values ($F_{(11,254)}$= 1.73, p> 0.05). No variability across the data suggested that the single vocalic duration was consistent, similar to those of the data from the second experiment.

Figure 6.31 Boxplot of the single vowel duration of the dataset after the exclusion of 10% of the extreme values of the first experiment across the sentences



After the extreme outliers were excluded, error bars at 95% confidence interval across the twelve sentences were generated (Figure 6.32). The ranges of the means were relatively consistent across the sentences despite the variability of the confidence intervals. The variability of the durational values could be attributed to the ways in which the three subjects produced the sentences.

Figure 6.32 Error bars of the single vowel duration of the dataset after the exclusion of 10% of the extreme values of the first experiment across the sentences



In sum, the findings of the analysis of the first experiment dataset suggested that single vowel duration is a possible acoustic measure due to its high degree of regularity across speaker and sentences. The consistency of both the data suggests that the single vocalic duration seems like a good candidate for a feature that is regulated in connected speech in Malay and which may therefore be a key property of Malay rhythm.

6.7. Summary

The second experiment of this thesis involved 20 native speakers of the standard Malay variety producing a list of ten sentences. The overall calculations of both types of rhythm metric (Ramus et. al, 1999 and Grabe & Low, 2002) for this dataset indicated that Malay was grouped in the same zone as other languages identified as syllable-timed. To recap, these were the results of the rhythm metrics of this data: ΔC=34.6, %V=51.1, ΔV=36.6, *r*PVI=29.5 and *n*PVI=44.8 (see sections 6.2 and 6.3). Ramus et. al (1999) pointed out that ΔV was not a suitable indicator of rhythm, therefore it will not be discussed in this section, although it was discussed in a detailed in section 6.2.3..

An analysis was made of these rhythm metrics and it was found that there was a significant variability of all of these metrics as a factor of sentences and subjects (see sections 6.2 and 6.3). There was no consistency of the values of the overall metrics across these independent variables. The significant variability found in this data was consistent with the findings of Dellwo (2004) and Arvaniti (2009) in which they found that the same languages have different scores of the rhythm metrics depending on the subjects, materials and speech rate. Hence, the overall results of the metrics were not able to explain the variability that was found across these independent variables.

The interactions between these metrics and speech rate were also investigated in section 6.4. It was revealed that ΔC had a significant negative correlation with speech rate. A slower speech rate gave a higher value of ΔC than that of the fast speech rate. While for the interaction of the rest of the metrics and the speech rate indicated a non-significant weak correlation. These findings were consistent with those of Dellwo (2004). This relationship indicated that to some extent the values of the metrics were influenced by the speech rate.

These findings supported the evaluation of these metrics by Dellwo (2003), Delwo and Wagner (2004), Asu and Nolan (2005), Wiget et. al (2010), White and Matys (2007), Arvaniti (2009), Fletcher (2010) in which they also found the variability of the rhythm metrics across speakers, materials and speech rate. In their reviews, they suggested that the parameters proposed by Ramus et. al (199) and Grabe and Low (2002) concentrated more on the timing of the languages but do not exactly capture the prominent features that may contribute to the rhythmic structure of any given languages.

To understand the reasons for this variability, further acoustic analysis was carried out in section 6.5. The objective was to compare the production of the material by the subject who has extreme scores of the rhythm metrics to the subject who has similar scores of the rhythm metrics to those of the overall values. Subjects 14 and 7 were identified to best fit this description respectively. The acoustic analysis was done on these subjects' productions of sentence 5. The overall duration of sentence 5 produced by subject 14 was longer than that of subject 7. A similar result was also found when the analysis was further made on the consonantal interval duration and vocalic interval duration. This demonstrated that the durational values were influenced by the speaking styles of the subjects. The variability of the input subsequently gave a great impact on the calculation of the rhythm metrics.

Further acoustic analysis was also carried out on sentence 4 in section 6.5.2. This sentence was recognised as having the most outliers on all of the rhythm metrics discussed previously. The presence of outliers in the rhythm metric scores indicated that there were different durational values of the intervals produced by the subjects which lead to relatively high variability of the rhythm metric scores within the same sentence. It was revealed that one of the consonantal intervals of sentence 4 which had the combination of nasal and voiced plosive had more variability in terms of the durational values than other consonantal intervals within the same sentence. The outcome of this investigation suggested that the segmental features of the intervals could manipulate the variability of the durational values which in return affected the calculations of the rhythm metrics.

The investigation carried out revealed that there were problems with these rhythm metrics proposed by Ramus et, al (1999) and Grabe and Low (2002). It was found that the metrics were too sensitive to the factors of subjects, sentences, speech rate and the segmental make-up of the material. This drawback motivated this research to go a step further to find any evidence of prominent feature in the acoustic signal which were regulated across the data in order to describe the Malay rhythm. This feature should be able to withstand the influence of the independent variables such as the subjects, sentences and speech rate.

Three features were analysed in this part of the experiment. They were syllable duration, inter-intensity minima duration and single vocalic duration. A similar analysis was carried out on these features to that of the rhythm metrics. Of all the three features, single vocalic duration appeared to be less variable as factors of subjects and materials. Further to this, a similar analysis of the single vocalic duration was also carried out on the dataset of the first experiment. The results of the first experiment also indicated that this feature was also relatively consistent throughout the dataset. This result seemed to suggest that single vocalic duration is a feature that is regulated and which may lead Malay having the rhythmic characteristics as it is.

The next chapter will attempt to assess the implication of the current study for the general theory of speech rhythm. At the same time, it will also consider the results of this study in order to draw some conclusion regarding the rhythmic property of the standard Malay variety in particular.

Chapter 7

Discussion and Concluding Comments

7.1. Introduction

This thesis has set out to provide a better understanding of the rhythmic characteristics of Malay by looking at works over the last decade which has tried to refine our understanding of differences between languages which have traditionally thought of as being stress-timed vs. syllable-timed. The main approach taken by Ramus, et. al (1999), Grabe and Low (2002) and Dellwo (2003) is to look at acoustic metrics for capturing the rhythmic properties of languages. Subsequently, the aim of the present research was to look at Malay in that context; and to see whether the study of Malay could contribute back to the more general field of research on speech rhythm.

7.2. Summary of the studies

Two experiments were conducted in order to achieve these objectives. The first experiment was conducted to test instrumentally the claim that there is no variation in phonetic parameters, which is consistent with what have been found in other languages in relation to the realisation of lexical stress (Zuraidah et. al, 2008). This study looked at three word types, namely disyllabic monomorphemic words, polysyllabic monomorphemic words and polysyllabic morphologically complex words. Stress is an important attribute related to the rhythmic properties of a language as this feature contributes to the perceptual characteristic of rhythm formed by the listeners. The importance of stress to the study of rhythm was highlighted in Dauer (1982), Ramus et al. (1999) and Grabe & Low (2002). Fry (1955) pointed out that duration, loudness and pitch were the three acoustic parameters known to be related to stress in other languages; however F_0 was not looked at in the present study because previous works (Zuraidah et. al., 2008 and Teoh, 1994) suggested that it was not relevant to Malay speech. This experiment, following key studies in the past by Fry (1955), House (1955) and Cutler (2005), looked at aspects of the vowels within the target words.

The results of analysing the vowel durations were twofold. First, in words in isolation, the final vowel of the target words was longer than the other vowel positions in the disyllabic and polysyllabic monomorphemic words whereas the penultimate vowel duration was longer for the polysyllabic morphologically complex words. Second, in a sentence context, the durational effects of these target words disappeared altogether and there were no significant effect of the vowel positions on the durations.

The lengthening of the final vowels in the disyllabic and polysyllabic monomorphemic word types in isolation could be attributed to final lengthening (Fletcher, 2010). The results of the first experiment revealed that in isolation, the final vowel of these types of words were generally 70% longer than other vowels in different position. These findings were consistent with final lengthening found in other languages such as French. Meanwhile the lengthening of the penultimate vowel of the polysyllabic morphologically complex words could reflect that the timing structure of Malay words in isolation was influenced by the morphological structure of the words as this vowel was also identified as the final vowel of the word stem (see Chapter 4).

Meanwhile, the analysis of vowel intensity showed no statistically significant differences across any of the vowel positions of the target words. The findings for the vowel intensity applied across all of the three word types in both the isolated and embedded contexts.

A correlation test revealed that in both isolated and sentence contexts, these two acoustic features were independent of each other. In single words, duration on its own differentiated final syllables from penultimate syllables (for disyllabic and polysyllabic monomorphemic words) or penultimate syllables from final syllables (for polysyllabic morphologically complex words). Meanwhile, no significant difference was found across syllables in respect to peak of intensity.

On a bigger picture, looking at the data of continuous speech, the results of the first experiment suggest that there is no systematic alignment of both duration and intensity to lexical stress.

A second experiment was carried out to investigate the acoustic correlates of Malay rhythm. It involved the analysis of ten sentences produced by twenty native speakers of Malay. The methodology of the data collection was reported in Chapter 5.

This second experiment centred on two specific objectives: The first aim was to compare the results of the rhythm metrics obtained from this Malay data with those of other languages found in Ramus et. al (1999) and Grabe and Low (2002). The analysis reported in Chapter 6 revealed that Malay was positioned in the same zone as other languages which were traditionally classified as syllable-timed. The results confirmed previous auditory studies of

Malay rhythm (Farid, 1980; Baskaran, 1982; Suhaila, 1994; and Teoh 1994) which suggest Malay is probably best described as a syllable-timed language.

The results of the samples of analysed positions Malay towards the syllable-timed end of the continuum which Ramus et al. (1999) and Grabe and Low (2002) provided, nevertheless, when a detailed analysis looking at variability across sentences, subjects and as a function of speech rate was carried out (see sections 6.21 and 6.31), it was found that those mean figures did not really capture the range of variability produced by speakers of Malay. Therefore, this raises questions around the validity of these measures if they are thoroughly sensitive to subject and sentence variability; and it also raises the question whether is there any other feature which might be less prone to the confounding factors which has been identified such as the speaker and sentence variability

Hence, the second aim of the second experiment was to investigate whether there were other features within the speech signal, beyond the rhythm metrics, which appeared to be more related to the characteristics of Malay rhythm. These other features should be robust to variation arising from factors such as inter-speaker and inter-sentence variability. Three possible candidates were analysed to achieve this objective; they were syllable duration, inter-intensity minima duration and the duration of single vowel. Following this analysis, it was found that single vowel duration as presented in section 6.7.3 was the most consistently regulated feature in the speech signal which could be associated to Malay rhythm. A similar analysis on the dataset from the sentence section of the first experiment also exhibited that single vocalic duration was relatively consistent across the subjects and sentences. The next section will evaluate the findings of both the experiments in relation to the current literature of speech rhythm.

7.3. Characterising Malay Speech Rhythm

Pike (1947) proposed a dichotomy of speech rhythm based on stress and syllable timing and since then a big concern has been to test the validity of that dichotomy; to find in the case of so called stress-timed languages, isochrony of feet and in the case of syllable-timed languages, isochrony of syllables (example: Roach, 1982 and Dauer, 1983). This was further demonstrated in the syllable duration analysis in section 6.7.1 in which it was found that even within the same syllable structure, isochrony was not detected. Notwithstanding the fact that researchers have struggled to find isochrony, nevertheless this kind of distinction between the syllable-timed vs. stress-timed has continued to be influential as a framework of capturing some aspects of different languages in terms of their rhythm. Thus, one of the questions concerning this thesis was where Malay is located in respect to this dichotomy. Critics like Roach (1982), Dauer (1983) and Grabe and Low (2002) have also argued that not only does the categorical approach proposed by Pike (1947) provide an inaccurate measure of rhythm but it was also too simple and not able to classify the rhythmic properties of other languages which were not adequately described using Pike's traditional labels.

Following this criticism, Ramus et. al (1999) and Grabe and Low, (2002) moved away from the position of having to force languages to be in one category or another, to adopt a new position whereby languages could be more or less different from each other along a continuum defined by the acoustic measures. However, at one end of the continuum they found languages which are more or less stress-timed, and at the other end are languages that are more or less syllable-timed. Furthermore, Ramus et al. (1999) and Grabe and Low, (2002) argued that the results of the rhythm metrics obtained by calculating the input from the durational values of the consonant and vocalic intervals were able to characterise: rhythm instrumentally and cross-language differences in rhythmic characteristics of speech in a more refined way in which is more practical from the instrumental analysis point of view.

In the present study, the results for Malay obtained by the same metrics positioned it at the same point of the continuum as other languages that were traditionally described as syllable-timed. However, some limitations of these rhythm metrics were identified. A comparison of the metric scores of this data with those of the Malay reported by Grabe and Low, (2002) was made, the two analyses provide very different results; the same language is positioned in a different part of the acoustically defined continuum. Malay in Grabe and Low, (2002) was in the stress-timed end of the continuum whereas Malay from the present data was in the syllable-timed end of the continuum. This result was not surprising since variability across sentences and speakers were found within the current dataset as presented in the scatter plots (Figure 6.1 and Figure 6.5). The scatter plots illustrated that the rhythm metrics values of some subjects were consistent with those of the languages in the stress-timed end of the continuum while the values of the rhythm metrics of other subjects were consistent with those of the languages from the syllable-timed end of the continuum. Hence, it was rather ineffective to position the languages in a continuum rhythmically depending solely on their overall scores for the rhythm metrics.

The likes of Asu and Nolan, (2005); Dellwo, (2004); Arvaniti, (2009); Wiget et. al (2010) criticised these metrics for not adequately capturing aspects of the rhythmic property of the languages studied. Furthermore the rhythm metrics might provide a snapshot of these languages in general but they fail to provide a sense of the wider complexity in respect to the realisation of rhythmic property and how they vary from speakers and contexts. The same language could be positioned in different places in the continuum. These discrepancies are down to the speakers, the materials and the speech rate. This was clearly demonstrated in the analysis of the second experiment in which there was significant variability across the subjects the materials and the speech rate.

To some extent, the continua proposed by Ramus et. al (1999) and Grabe and Low, (2002) are useful; if analysed carefully, they reveal a lot of variability across languages which reflect real differences between languages concerned. A further advantage of these metrics is it does not require investigators to put languages to one traditional category or another. On the other hand, the continua were based on production and they do not capture the perceptual dimensions along which languages differ in respect of their rhythm.

Further drawback of these rhythm metrics was they originated from the measurement based on segmental analysis. Arvaniti (2009) argued that there were two problems related to the use of segmental duration to account for rhythm. Firstly, segmental durations were not as simple and straightforward as they seemed. This was due to the fact that some phonological features of a particular language could influence the durational values of the intervals. Fletcher (2010:525) listed factors such as '... vowel height where close vowels tend to be shorter than open vowels, vowel lengthening before voiced plosives, phonological contrasts between short and long vowels in which could influence durational values of the segment' which could contribute to the segmental durations. In addition to those factors, the syllable structure of a language also could be a factor that influenced the durational values of the intervals, for example the complex syllable structure also had a tendency to have shorter segment compared to simple syllable structures (Arvaniti, 2009).

Secondly, Arvaniti (2009) pointed out that it was impractical to assume that the segmental durations behaved in a standard manner across languages. Features like final lengthening could be applied to Malay (e.g. Deterding (2003), current study), however this feature was not applicable to languages like Greek (Arvaniti 2000).

Although the present study was carefully put together and controlled in terms of the selection of the sounds and the consistency of the segmentation processes throughout the data; these controls still did not hinder the fact that variability of the rhythm metrics results could not be avoided. Acoustic analysis of the subjects indicated that the individual differences in a way which sentences were read influenced the durational measurements.

It was also found that there was significant variability across the sentences in the data. Some of the sentences exhibited more syllable-timed scores whereas some of them exhibited more stress-timed scores. This finding was consistent with the claim brought forward by Arvaniti (2009) in which she observed that where a particular utterance is aligned within the continuum varies on the syllable structure and the segmental make up of the sentences with the consequence that within the same language, a sentence could be positioned closer towards the syllable-timed end of the continuum or the stress-timed end of the continuum.

Despite the differences between the approaches used in the rhythm metrics proposed by both Ramus et. al (1999) and Grabe and Low (2002) and the dichotomy proposed by Pike (1947), there were underlying similarities between these different approaches in which they equated rhythm as a feature of timing. The two issues related to these approaches are firstly, timing was looked at from the production point of view and they did not investigate the perceptual dimension, which is quite important because of what is perceived as being regular is not similar with the acoustic dimension. This has been shown by the works on P-centres by Morton, et al. (1976), as discussed in section 2.3 of this thesis. Secondly, languages do not necessarily demonstrate a fixed set of rhythmic properties as being demonstrated by Arvaniti (2009) in section 2.5. Within the same language, some utterances may be more syllable-timed than the other utterances.

Three potential features investigated were syllable duration, inter-intensity minima duration and single vocalic duration. Although these three candidates were related to timing as opposed to other aspects of rhythm; these three are not an exhaustive set of all the different features which could have been looked at but they are plausibly connected to the productive aspects of rhythm in Malay. Subsequently the next section describes the relevance of these candidates to capture aspects of the rhythmic property of Malay.

The first candidate was syllable duration (see 6.7.1). Since Malay was identified as being at the syllable-timed end of the continuum, the rationale of selecting this feature was to test the claim of isochrony of the syllable intervals brought forward by Pike (1947). Traditionally languages described as syllable-timed should have equal syllable duration. The results showed that even within the same syllable type, significant variability was apparent. This finding was consistent with those in Roach (1982) in which he found the languages studied did not have isochronous syllable durations.

The second candidate was inter-intensity minima duration. The rationale for opting for this feature was because in Malay, there was a preference for CVCV syllable structure. The repetition of this type of syllable produced a distinctive pattern of intensity contour as observed in section 6.7.2. The intensity contours which differentiated between the peak and the minima intensity in the production of CV syllables were more prominent in the production of plosive/fricative and vowel sequences; however the contours corresponding to the nasal vowel combination showed a flat intensity pattern. Therefore, the regularity of this feature was appealing if the consonants in the CVCV syllable structure in the material corresponded to plosives or fricatives.

The third candidate was the single vowel duration. This acoustic measure was selected following the results of the first experiment, in which it was found in sentence contexts the only difference in vowel duration encountered was in phrase-final contexts; whereas, there was no alternation between stressed and unstressed vowels within a phrase that has been found in other languages (see section 4.3). Therefore, it seemed that Malay did not have stressed and unstressed vowels in the connected speech which allowed me to postulate that there might be less variability of vocalic duration in the data. Unlike the calculation of the %V and the nPVI, this calculation only considered the durations of single vowel intervals rather than the sequences of vocalic intervals.

The results of this analysis were promising. Single vocalic duration was relatively consistent across the speakers and sentences. Less variability was found across the subjects and the sentences as demonstrated in section 6.7.3. Despite the durational differences across speakers, it was found that within the subjects, the single vocalic duration was relatively consistent. Furthermore, to support this candidate another analysis was conducted on the single vocalic duration of the sentence data from the first experiment (section 6.7.3.1). This feature was found to be relatively consistent across this dataset too. Based on the evidence from these two experiments, single vocalic duration was a feature which seemed to be robust to the factors of subject and sentence.

It can be argued that by excluding certain types of vocalic intervals the approach taken to this measurement pre-disposed the outcome and arguably affected its validity (section 6.6.3). However, the selection of certain single vowel interval (as summarised in Table 6.18) is reasonable since the exclusion of the final vowel is straightforward to justify final lengthening. Meanwhile, the inclusion of vowel sequences with the insertion of glottal stop is valid because it actually relates to the morphophonological structure of Malay.

Hence, this feature was promising but by no means conclusive. The next step in future is to consider the possibility of other non-durational signals of the Malay speech as a possible feature to describe Malay rhythm.

The consistency of the single vocalic duration as the most suitable acoustic measure is coherent with the evidence from the phonological features of Malay. The lack of stress in connected speech was one of the contributing factors of the consistency of the single vocalic duration. As demonstrated in the first experiment (chapter 4) along with Zuraidah et. al (2008), vowel duration was found not to be correlated with stress in this language. The results of the first experiment were consistent with the second experiment as there were no vowel alternation between stressed and unstressed vowels within the sentence context. With the exception of the vowel duration in pre-pausal positions, which were excluded from the analysis, the single vocalic intervals elsewhere have relatively consistent duration.

In support of this consistency further, Tajul (2000) stated that Malay does not permit syllable reduction in connected speech. The observation by Tajul (2000) was found to be true on the vocalic segment of the syllable as demonstrated in the current study. The vowel durations were less variable across the data (see section 6.7.1.3). Thus, it would be interesting in future work to look at whether is there any differences in formant realisation to substantiate the claim made about the Malay vowel in natural speech.

Subsequently, the phonemic system of Malay has characteristics which are conducive to maintenance of reasonably consistent single vowel duration, as found in the results of this study. As discussed in section 3.3, vowel length was not phonemically contrasted in Malay. This feature influenced the durational values of the vowel since the speakers did not have to distinguish between short and long vowels in terms of the length in order to get their messages across. Hence, the speakers had the tendency to produce a regular single vocalic interval durations in their utterances which was evident in the acoustic signal. As a result, less variability of these single vowel durations throughout the data was found.

Glottal stop insertion, as shown in the data, contributed to the relatively consistent single vocalic duration. To substantiate further, the result from this data was similar to the findings of the acoustic study in Esposito and Scarborough (2004) in which a glottal stop was inserted in between the sequences of short vowels in Pima, an Uto- Aztecan language to distinguish them from long vowels. The results revealed that both the vowel durations in pre and post glottalic position were relatively consistent. Glottal stop insertion has the consequence of removing situation in which there might be elongated vowels. The consequence of this is to enhance the rhythmic properties of Malay which in the present study seems to suggest might be closely associated with the durations of individual vowels.

It is observed that Malay phonological word further governed connected speech since glottal stop insertion was also extended to between word boundaries particularly words with vowel sequences across the stems. When vowels come into context phonologically, Malay inserts a glottal stop and as a result instead of VV sequence, VCV sequence is produced with shorter vowels instead of two vowels that come together. This reinforces this particular rhythmic property. The results suggest that if rhythm is about having some features of the speech signal, which have some cyclic characteristics, then single vocalic duration appears to be a good candidate in Malay and warrant further investigation.

Referring to the example of the morphologically complex word earlier, the separation of the two vowels sequence with the insertion of the glottal stop altered the sequences of the syllable structures from a CV.V sequence to a CV.CV sequence. This supported the notion that Malay preferred the CVCV syllable structure. This was evident with the fact that one of its phonotactic rules did not allow Malay to have consonantal clusters in both the onset and the coda (Teoh, 1994). On another level the preference for CVCV syllable structure influenced the intensity contour as demonstrated in section 6.7.2. However the intensity-minima duration of the contour exhibited large variability across the data, in which it was found that the intensity contour observed, was only applicable when there was certain combinations of the phonemes like nasals or plosives.

One reason why single vocalic duration is an interesting parameter is because it seems to fit with the characteristics in relation to phonological realisation of words in Malay. The results of single vowel durations were better than the results of rhythm metrics, such as %V and *n*PVI in capturing the regularity of the acoustic measures. This was because the former was able to withstand the functions of sentences, subjects and speech rate. Furthermore, the single vowel duration was a feature in which could be considered as a language-specific feature due to its consistency to the phonological words and the phonetic realisation of Malay.

This is the first study, to our knowledge, to examine other alternative acoustic measures as indicators of rhythm in attempt to capture the grouping prominent of the speech. Furthermore, the study of rhythm also has never included the function of single vocalic duration especially in those languages that were considered to be on the syllable-timed end of the continuum.

7.4. Generalisations from the results

The current study of Malay reinforces the claimed made by Dellwo and Wagner (2005), Arvaniti (2009) and Fletcher (2010) on the limitations of the rhythm metrics proposed by Ramus et al. (1999) and Grabe and Low (2002). To summarise the drawbacks of these rhythm metrics: firstly, while they are capable of differentiating languages along a graded scale which can readily mapped to the acoustic signal, what they don't do is draw out the factors which are at the root of the cross-language differences which are observed. Secondly, results obtained using these acoustic metrics suggest that they are sensitive to variability arising from the speakers from whom they are measured and from the materials which those speakers are producing, all of which in turn has made it very difficult to stably define the rhythmic characteristics of a particular language (section 7.3).

Following these limitations, in this thesis it was decided to go beyond simply reporting how Malay was positioned in respect of the pre-existing acoustic metrics in order to investigate other phonetic features of Malay that might be plausibly associated with rhythmic properties of that language. This is in line with Arvaniti (2009) who claims that rhythm is about grouping and prominence and it is language-specific. The approach taken by the current research (i.e. looking for which parameters seem to provide the basis for the particular rhythmic characteristics of a language rather than only applying pre-conceived metrics) is one which could be applied to the studies of rhythm for other languages as it provides a useful means of investigating the language-specific properties which underpin thr rhythmic structure of a particular language. From the methodological point of view, it was found that for at least one language there are other measurements which seem more robust to speaker-material variability factors which were the cause for concern to

Arvaniti (2009) and which would be worth further exploration on other languages.

Although a major caveat for this study is it only looks at read speech, nevertheless, some of the features of read speech are shared by other speech. It would be interesting to see to what extent the results could be found to apply in other more natural styles.

The subjects from this study were native speakers of the standard variety of Malay. Thus, this raises the question of to what extent do these findings can be generalised to other varieties There is very little to go on with respect to other varieties of Malay in Malaysia; but they have some similarities to the standard variety investigated in this study; for example the glottal stop insertion (Teoh, 1994). Moreover, there have been no studies of the phonetic realisation of prominence in other varieties of Malay but auditory analysis suggests that other varieties are no different from that of Standard Malay (Teoh, 1994 and Tajul, 2000). However, due to the similar characteristics, there is the possibility that the findings in this studies can be generalised to them.

This study looked at the native speakers of Malay; however Malaysia being a multilingual country, Malay spoken by second language speakers such as the Chinese and the Indians may have different rhythmic characteristics probably due to the influence of their first language.

7.5. Limitations of the study

Having discussed the results of this research, it is important to note a number of limitations which applied to the study:

In experiment 1, the stress experiment focused only on two phonetic correlates: duration and intensity while F_0 was not included in the analysis of the recorded speech. From the number of previous studies of some languages, as pointed out in section 2.2, F_0 has been shown to be relevant in characterising the phonetic correlates of stressed syllables. However, after consulting the limited literature on Malay stress, (most of them based on perceptual observation: Farid (1980), Baskaran (1982), and Suhaila (1994)) only duration and intensity may have some roles in differenciating stressed and unstressed syllables (section 3.4), thus, the measures of F_0 were not looked at in the first experiment. However, it is hoped that future work will factor in F_0 alongside duration and intensity to understand Malay stress further. Subsequently, more work is needed on how duration, intensity, F_0 and other features are used to help Malay speakers convey contrastive stress to accomplish the pragmatic function associated with given vs. new information.

Experiment 1 was designed as a pilot study, therefore only drew on the performance of three subjects. Future work would obviously benefit by looking at larger number of individuals. Interestingly, although there were only three subjects, there were consistent patterns across speakers.

Another limitation lies in the fact that this present study only analysed read speech. Natural continuous speech was not included in this study. But in mitigation of this the read speech nevertheless embodied many features that also occur in the natural speech such as the insertion of glottal stops.

Essentially, this study focussed only on the production aspect of speech rhythm because without solid understanding of the rhythmic properties of Malay it would be difficult to design valid perceptual experiments. Thus, looking at the production is only a part of the story on speech rhythm because there is no way of telling if any of these durational consistencies are perceivable by the listeners of Malay

7.6. Recommendations for further research

Following the generalisation discussed in section 7.4, further studies are required to establish whether the acoustic measure used in this study is a good indicator of Malay rhythm. More experimental investigations focussing particularly on single vocalic duration should be carried out in order to test and replicate the findings of the current study. Future analysis also should include other varieties of Malay as well as other speaking styles.

At the same time, in light of the consistency of the single vocalic duration, this research has thrown up other question in need of further investigation. The issue of whether this particular acoustic measure is perceptually relevance to Malay rhythm should be addressed following Arvaniti (2009:58) in which the psychological meaning of rhythm as a perception of a repetitive pattern of speech stimuli has to be satisfied.

Finally, other non-durational parameters such as intensity and F_0 must also be considered thoroughly in the future in order to better capture the aspects of rhythmic properties of Malay.

7.7. Concluding comments

In conclusion, this study has carried out an acoustic analysis of the rhythmic properties of Malay. The results of the rhythm metrics of this dataset further highlights the drawbacks of the measures that investigators have come to use in recent years. This present study found a parameter that seems to work better for Malay with greater consistency. Similar to recent studies in speech rhythm, this study took a direct approach of measuring and analysing durational segment of the acoustic signal. This study analysed the duration of a specific set of single vowel duration. Unlike the existing rhythm metrics, single vowel duration was found to be relatively consistent across the sentences and subjects.

APPENDIX

First Experiment

Section A – Word list

Simple monomorphemic words	Longer monomorphemic words	Morphologically complex words
Batu /bɐtu/ *A stone*	Kepada /kəpɐdə/ *To*	Dibatukan /dɪbɐtukɐn/ *Turned into a stone*
Muka /mukə/ *A face*	Paduka /pɐdukə/ *Highness*	Kemukakan /kəmukəkɐn/ *To present*
Padu /pɐdu/ *Solid*	Pedati /pədɐtɪ/ *Cartwheel*	Memakukan /məmɐkukɐn/ *To nail*
Paku /pɐku/ *A nail*	Pidato /pɪdɐto/ *A speech*	Membidakan /məmbɪdɐkɐn/ *To bid*

Section B- The sentences

1. Dia suka memakai **batu** permata

/dɪ**j**ə sukə məmɐkɐɪ **bɐtu** pərmɐtə/

*She/he likes to wear **stone** jewelleries*

2. Si Tenggang telah **dibatukan** oleh ibunya

/sɪteŋgɐŋ təlɐ **dɪbɐtukɐn** ole ɪbuɲə/

*Si Tenggang's mother has had him **turned into a stone***

3. Dia **kemukakan** agenda baru

   /dɪjə **kəmukəkɐn** ɐʤenda bɐru/

   *He/She **presents** a new agenda*

4. Surat itu diberi **kepada** dia

   /surɐt ɪtu dɪbərɪ **kəpɐdə** dɪ**j**ə/

   *The letter is given **to** him*

5. Saya telah **memakukan** anyaman itu di dinding

   /sɐjə təlɐ **məmɐkukɐn** ɐɲɐmɐn ɪtu dɪdɪndɪŋ/

*I have **nailed** the weaved mat to the wall.*

6. Dia telah **membidakan** wangnya

   /dɪjɐ təlɐ **məmbɪdɐkɐn** wɐŋɲə/

   *She/he has **bid** all of her/his money*

7. Berseri-seri **muka** dia

   /bərsərɪ sərɪ **mukə** dɪjə/

   *Her/his **face** is glowing*

8. Isi **padu** pepejal itu diukur

   /ɪsɪ **pɐdu** pəpədʒɐl ɪtu dɪuko/

   *The **volume** of the material is measured*

9. Seri **Paduka** Baginda

   /sərɪ **pɐdukə** bɐgɪndə/

   *His/her royal **highness***

10. Jangan biar **paku** berselerak

    /dʒɐŋɐn bɪjɐ **pɐku** bərsəlerɐk/

    *Don't let the **nails** scatter around*

11. Kuda itu mempunyai **pedati** kuning

    /kudə ɪtu məmpuɲɐɪ **pədɐtɪ** kunɪŋ/

    *The horse has a yellow **cartwheel***

12. Hujah-hujah **pidato** pelajar itu amat bernas

    /hudʒɐ hudʒɐ **pɪdɐto** pəlɐdʒɐ ɪtu ɐmɐt bərnɐs/

    *The student's **speech** is very commendable*

Section C- Reading List

1. Batu

/bɐtu/

A stone

2. Dia suka memakai **batu** permata

/dɪ**j**ə sukə məmɐkɐɪ **bɐtu** pərmɐtə/

*She/he likes to wear **stone** jewelleries*

3. Dibatukan

/dɪbɐtukɐn/

Turned into a stone

4. Si Tenggang telah **dibatukan** oleh ibunya

/sɪteŋgɐŋ təlɐ **dɪbɐtukɐn** ole ɪbuɲə/

*Si Tenggang's mother has had him **turned into a stone***

5. Kemukakan

/kəmukəkɐn/

To present

6. Dia **kemukakan** agenda baru

/dɪjə **kəmukəkɐn** ɐdʒenda bɐru/

*He/She **presents** a new agenda*

7. Kepada

/kəpɐdə/

To

8. Surat itu diberi **kepada** dia

/surɐt ɪtu dɪbərɪ **kəpɐdə** dɪ**j**ə/

*The letter is given **to** him*

9. Memakukan

/məmɐkukɐn/

To nail

10. Saya telah **memakukan** anyaman itu di dinding

/sɐjə təlɐ **məmɐkukɐn** ɐɲɐmɐn ɪtu dɪdɪndɪŋ/

*I have **nailed** the weaved mat to the wall.*

11. Membidakan

/məmbɪdɐkɐn/

To bid

12. Dia telah **membidakan** wangnya

/dɪjɐ təlɐ **məmbɪdɐkɐn** wɐŋɲə/

*She/he has **bid** all of her/his money*

13. Muka

/mukə/

A face

14. Berseri-seri **muka** dia

/bərsərɪ sərɪ **mukə** dɪjə/

*Her/his **face** is glowing*

15. Padu

/pɐdu/

Solid

16. Isi **padu** pepejal itu diukur

/ɪsɪ **pɐdu** pəpəʤɐl ɪtu dɪuko/

*The **volume** of the material is measured*

17. Paduka

/pɐdukə/

Highness

18. Seri **Paduka** Baginda

/sərɪ **pɐdukə** bɐgɪndə/

*His/her royal **highness***

19. Paku

/pɐku/

A nail

20. Jangan biar **paku** berselerak

/ʤɐŋɐn bɪjɐ **pɐku** bərsəlerɐk/

*Don't let the **nails** scatter around*

21. Pedati

    /pədɐtɪ/

    Cartwheel

22. Kuda itu mempunyai **pedati** kuning

/kudə ɪtu məmpuɲɐɪ **pədɐtɪ** kunɪŋ/

*The horse has a yellow **cartwheel***

23. Pidato

    /pɪdɐto/

    The speech

24. Hujah-hujah **pidato** pelajar itu amat bernas

/huʤɐ huʤɐ **pɪdɐto** pəlɐʤɐ ɪtu ɐmɐt bərnɐs/

*The student's **speech** is very commendable*

Second Experiment

1. Penonton tennis sedang duduk di petak penyokong itu disepak Enot

/pənonton tenɪs sədɐŋ duduk dɪ petɐk pəɲokoŋ ɪtu dɪsepɐk enot/

The tennis spectator who is sitting on the grandstand is kicked by Enot.

2. Petani bebas membina benteng dekat dusunnya

/pətɐnɪ bebɐs məmbɪnə benteŋ dekɐt dusunɲə/

The farmer is free to build a wall near his field.

3. Emosi sangat penting untuk kesihatan individu

/emosɪ sɐŋɐt pəntɪŋ untuk kəsɪhɐtɐn ɪndɪvɪdu/

Emotion is important in an individual well being.

4. Lolo mengukur benteng dibina Taman Tema Kulim itu.

/lolo məŋuko benteŋ dɪbɪnə tɐmɐn temɐ kulɪm ɪtu/

Lolo is measuring the wall erected by Kulim Theme Park.

5. Tujuh ekor ketupat sotong enak dimasak Munah.

/tuʤuh eko kətupɐt sotoŋ enɐk dɪmɐsɐk munɐh/

Munah is cooking seven calamari with glutinous rice.

6. Eni menonton TV sehingga subuh.

/enɪ mənonton tɪvɪ səhɪŋgə subuh/

Eni watches the television till dawn.

7. Meme menuduh Hendon menyebar fitnah tanpa tahu punca

/meme mənuduh hendon meɲebɐ fɪtnɐh tampa tɐhu punʧə/

Meme accuses Hendon of spreading the news even without knowing the truth.

8. Sam menyemak kos menanam dan mengedar pokok kopi di Puchong.

/sem məɲemɐk kos mənɐnɐm dɐn məɲeda pokok kopɪ dɪ putʃoŋ/

Sam is checking the cost of growing and distributing coffee in Puchomg.

9. Sepupu Tok Ki mengejek pengemis hodoh dan bodoh itu.

/səpupu tok kɪ meɲedʒek pəŋəmɪs hodoh dɐn bodoh itu/

Tok Ki's cousin is bullying the ugly and stupid pauper.

10. Keno mengenakan jeket berbunga kuning dan kasut putih menghidupkan enjin motor itu.

/keno meɲenənəkɐn dʒeket bebuŋə kunɪŋ dɐn kasut puteh meŋhɪdupkɐn endʒen moto/

Keno, who is wearing the yellow floral jacket with white shoes, starts the motorcycle.

First Experiment

1. Dia suka memakai batu permata

/dɪə sukə məməkɐɪ bɐtu pərmɐtə/

Vowel sequence	**PVI calculation**	**Single vowel duration calculation**
/ɪə/	Included as a vocalic interval	Excluded – Two single vowels in a sequence. Across syllable boundary: /ɪ.ə/
/ɐɪ/	Included as a vocalic interval	Included- a diphthong

2. Si Tenggang telah dibatukan oleh ibunya

/sɪteŋgɐŋ təlɐh dɪbɐtukɐn ole ɪbuɲə/

Vowel sequence	**PVI calculation**	**Single vowel duration calculation**
/e ɪ/	Included as a vocalic interval	Excluded – Two single vowels in a sequence. Across word boundary: /e ɪ/

3. Dia kemukakan maklumat baru

/dɪə kəmukəkɐn mɐklumɐt bɐru/

Vowel sequence	**PVI calculation**	**Single vowel duration calculation**
/ɪə/	Included as a vocalic interval	Excluded – Two single vowels in a sequence. Across syllable boundary: /ɪ. ə/

4. Surat itu diberi kepada dia

/surɐt ɪtu dɪbərɪ kəpɐdə dɪə/

Vowel sequence	**PVI calculation**	**Single vowel duration calculation**
/ɪə/	Included as a vocalic interval	Excluded – Two single vowels in a sequence. Across syllable boundary: /ɪ.ə/

5. Saya telah memakukan anyaman itu di dinding

/sɐjə təlɐh məmɐkukɐn ɐɲɐmɐn ɪtu dɪdɪndɪŋ/

Vowel sequence	**PVI calculation**	**Single vowel duration calculation**
-no vowel sequence -		

6. Dia telah membidakan wangnya

/dɪɐ təlɐh məmbɪdɐkɐn wɐŋɲə/

Vowel sequence	PVI calculation	Single vowel duration calculation
/ɪə/	Included as a vocalic interval	Excluded – Two single vowels in a sequence. Across syllable boundary: /ɪ.ə/
/wɐ/	Included as a vocalic interval	Excluded as it is a sequence of glide single and vowel-it is difficult to segment a glide followed by a vowel.

7. Berseri-seri muka dia

/bərsərɪ sərɪ mukə dɪə/

Vowel sequence	PVI calculation	Single vowel duration calculation
/ɪə/	Included as a vocalic interval	Excluded – Two single vowels in a sequence. Across syllable boundary: /ɪ.ə/

8. Isi padu pepejal itu diukur

/ɪsɪ pɐdu pəpədʒɐl ɪtu dɪuko/

Vowel sequence	PVI calculation	Single vowel duration calculation
/ɪu/	Included as a vocalic interval	Excluded – Two single vowels in a sequence. Across syllable boundary: /ɪ.u/

9. Seri Paduka Baginda

/sərɪ pɐdukə bɐgɪndə/

Vowel sequence	PVI calculation	Single vowel duration calculation
-no vowel sequence-		

10. Jangan biarkan paku berselerak

/dʒɐŋɐn bɪjɐ pɐku bərsəlerɐk

Vowel sequence	PVI calculation	Single vowel duration calculation
/ɪjɐ/	Included as a vocalic interval	Excluded – a sequence of vowel and glide sequence

11. Kuda itu mempunyai pedati kuning

/kudə ɪtu məmpuɲɐɪ pədɐtɪ kunɪŋ/

Vowel sequence	PVI calculation	Single vowel duration calculation
/ə ɪ/	Included as a vocalic interval	Excluded – Two single vowels in a sequence. Across word boundary: /ə ɪ/
/ɐɪ/	Included as a vocalic interval	Excluded- Two single vowels across a syllable boundary, /ɪ/ is a suffix and morphologically change the respective word.

12. Hujah-hujah pidato pelajar itu amat bernas

/hudʒɐh hudʒɐh pɪdɐto pəlɐdʒɐ ɪtu ɐmɐt bərnɐs

Vowel sequence	PVI calculation	Single vowel duration calculation
/u a/	Included as a vocalic interval	Excluded – Two single vowels in a sequence. Across syllable boundary: /ɪ.u/

Second experiment

1. Penonton tennis sedang duduk di petak penyokong itu disepak Enot

/pənonton tenɪs sədɐŋ duduk dɪ petɐk pəɲokoŋ ɪtu dɪsepɐk enot/

2. Petani bebas membina benteng dekat dusunnya

/pətɐnɪ bebɐs məmbɪnə bentəŋ dekɐt dusunɲə/

3. Emosi sangat penting untuk kesihatan individu

/emosɪ sɐŋɐt pəntɪŋ untuk kəsɪhɐtɐn ɪndɪvɪdu/

4. Lolo mengukur benteng dibina Taman Tema Kulim itu.

/lolo məŋuko bentəŋ dɪbɪnə tɐmɐn temɐ kulɪm ɪtu/

Vowel sequence	PVI calculation	Single vowel duration calculation
-Sentences 1-5: no vowel sequence-		

5. Tujuh ekor ketupat sotong enak dimasak Munah.

/tuʤu eko kətupɐt sotoŋ enɐk dɪmɐsɐk munɐh/

Vowel sequence	PVI calculation	Single vowel duration calculation
/u e/	Included as a vocalic interval	Excluded – Two single vowels in a sequence. Across syllable boundary: /u e/

6. Eni menonton TV sehingga subuh.

/enɪ mənonton tɪvɪ səɪŋgə subuh/

Vowel sequence	PVI calculation	Single vowel duration calculation
/əɪ/	Included as a vocalic interval	Excluded – Two single vowels in a sequence. Across syllable boundary: /əɪ/

7. Meme menuduh Hendon menyebar fitnah tanpa tahu punca

/meme mənuduh hendon meɲebɐ fɪtnɐh tampa tɐhu punʧə/

8. Sam menyemak kos menanam dan mengedar pokok kopi di Puchong.

/sem məɲemɐk kos mənɐnɐm dɐn məŋeda pokok kopɪ dɪ puʧoŋ/

Vowel sequence	PVI calculation	Single vowel duration calculation
-Sentences 7 & 8: no vowel sequence-		

9. Sepupu Tok Ki mengejek pengemis hodoh dan bodoh itu.

/səpupu tok kɪ meŋedʒek pəŋəmɪs hodoh dɐn bodo itu/

Vowel sequence	PVI calculation	Single vowel duration calculation
/o i/	Included as a vocalic interval	Excluded – Two single vowels in a sequence. Across syllable boundary: /o i/

10. Keno mengenakan jeket berbunga kuning dan kasut putih menghidupkan enjin motor itu.

/keno meŋenənəkɐn dʒeket bebuŋə kunɪŋ dɐn kasut puteh meŋhɪdupkɐn endʒen moto itu /

Vowel sequence	PVI calculation	Single vowel duration calculation
/o i/	Included as a vocalic interval	Excluded – Two single vowels in a sequence. Across syllable boundary: /o i/

References

Abercrombie, D. (1967) *Elements of general phonetics.* Chicago: Aldine.

Abercrombie, D. (1973) 'A Phonetician's view of verse structure ', in Jones, W. E. and Laver, J.(eds) *Phonetics in Linguistics.* London: Longman.

Arvaniti, A. (1994) 'Acoustic features of Greek rhythm structure', *Journal of PHONETICS,* 22, pp. 239-298.

Arvaniti, A. (2009) 'Rhythm, timing and the timing of rhythm', *Phonetica,* 66, (1-2), pp. 46-63.

Asu, E. L. and Nolan, F. (2005) 'Estonian rhythm and the pairwise variability index*', Fonetik* Goteborg, pp. 29-32.

Attridge, D. (1982) *The rhythm of English poetry.* Essex: Longman.

Barry, W. J., Andreeva, B., Russo, M., Dimitrova, S. and Kostadinova, T. (2003) 'Do rhythm measures tell us anything about language type*', 15th ICPhS.* Barcelona, pp. 2693-2696.

Benton, M., Dockendorf, L., Jin, W., Liu, Y. and Edmondson, J. (2007) 'The continuum of speech rhythm: Computational testing of speech rhythm of large corpora from natural Chinese and English speech*', 16th ICPhS.* Saarbrucken, pp. 1269-1272.

Berkovits, R. (1994) 'Durational effect in final lengthening, gapping and contrastive stress', *Language and Speech,* 37, (3), pp. 237-250.

Blevins, J. (1995) 'The syllable in phonological theory', in Goldsmith, J. A.(ed), *The Handbook of Phonological Thoery.* Oxford: Blackwell Book.

Boersma, P. and Weenink, D. (2006) *Praat: doing phonetics by computer* (Version 5.0.25)

Bradlow, A. R. (1995) 'A comparative study of English and Spanish vowels', *The Journal of the Acoustical Society of America,* 97, (3), pp. 1916-1924.

Bradlow, A. R. and Bent, T. (2002) 'The clear speech effect for non-native listeners', *The Journal of the Acoustical Society of America,* 112, (1), pp. 272-284.

Braunschweiler, N. (1997) 'Intergrated cues of voicing and vowel length in German: A production study', *Language and Speech,* 40, (4), pp. 353-376.

Brown, A. (1988) 'The staccato effect in the pronunciation of English in Malaysia and Singapore', in Foley, J.(ed), *New Englishes: The case of Singapore.* Singapore: Singapore University Press.

Campbell, W. N. and Isard S. D. (1991) 'Segment durations in a syllable frame', *Journal of Phonetics,* 19, pp. 37-47.

Cervera, T., Miralles, J. L. and Gonzalez-Alvarez. (2001) 'Acoustical analysis of Spanish vowels produced by Laryngectomized subjects', *Journal of Speech, Language and Hearing Research*, 44, pp. 988-996.

Cho, T. (2004) 'Prosodically conditioned strengthening and vowel to vowel coarticulation in English', *Journal of PHONETICS,* 32, pp. 141-176.

Couper-Kuhlen, E. (1986) *An introduction to English prosody*. London: Edward Arnold.

Crystal, D. (2003) *Dictionary of linguistics and phonetics.* Cambridge: Cambridge University Press.

Crystal, T. H. and House, A. S. (1988) 'Segmental duration in connected-speech signals: Syllabic stress', *Journal of the Acoustical Society of America,* 83, pp. 1574-1585.

Cutler, A., Mehler, J. and Norris, D. (1986) 'The syllable's differing role in the segmentation of French and English', *Journal of Memory and Language,* 25, pp. 385-400.

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

Dauer, R. (1987) 'Phonetic and phonological components of language rhythm', *11th International Congress of Phonetic Sciences.* Talinn, pp. 447-450.

de Jong, K. J. (2004) 'Stress, lexical focus and segmental focus in English: Patterns of variations in vowel durations ', *Journal of Phonetics,* 32, pp. 493-516.

de Manrique B. A, M. and A, S. (1983) 'Segmental duration and rhythm in Spanish', *Journal of PHONETICS,* 11, pp. 117-128.

Dellwo, V. (2006) ' Rhythm and Speech Rate: A Variation Coefficient for ΔC', in Karnowski, P. and Szigeti, I.(eds) *Language and language-processing.* Frankfurt am Main: Peter Lang., pp. 231-241.

Dellwo, V. (2008) 'The role of speech rate in perceiving speech rhtyhm', *Proceedings of Speech Prosody.* Campinas, pp. 375-378.

Dellwo, V. and Koreman, J. (2008) 'How speaker idiosyncratic is measureable speech rhythm?', *IAFPA.* Lausanne, pp.

Dellwo, V. and Wagner, P. (2003) 'Relations between language rhythm and speech rate', *15th Preceedings of ICPhS.* Barcelona, pp. 471-471.

den Os, E. (1988) *Rhythm and tempo of Dutch and Italian: A contrastive study.* PhD thesis. Rijksuniversiteit.

Desain, P. and Windsor, L. (eds.) (2000) *Rhythm Perception and Production.* The Netherlands: Swets & Zeitlinger.

Deterding, D. (2001) 'Letter to Editor: The measurement of rhythm: A comparison of Singapore and British English', *Journal of Phonetics,* 29, pp. 217-230.

Dilley, L. and Shattuck-Hufnagel, S. (1996) 'Glottalisation of word-initial vowels as a function of prosodic structure', *Journal of PHONETICS,* 24.

Dimitrova, S. (1998) 'Bulgarian speech: stress-timed or syllable-timed?', *Journal of the International Phonetic Association,* 27, pp. 27-33.

Dogil, G. (1999) 'The phonetic manifestation of word stress in Lithuanian, Polish and German and Spanish', in van der Hulst, H.(ed), *Word prosodic systems of the languages of Europe.* Berlin: Mouton de Gruyter, pp. 273-310.

Duez, D. (2009) 'Consonant and vowel duration in Parkinsonian French speech', *Folia Phoniatr et Logopaedica,* 61, (4), pp. 239-246.

Erickson, M. L. (2000) 'Simultaneous effects on vowel duration in American English: A covariance structure modelling approach', *Journal of Acoustical Society America,* 108, (6), pp. 2980-2995.

Esposito, C., M and Scarborough, R. (2004) 'An acoustic and electroglottographic study of VʔV in two indigenous American languages.', *Poster presented at the Meeting of the Acoustical Society of America.* San Diego, CA

Fant, G., Kruckenberg, A. and Nord, L. (1991) 'Durational correlates of stress in Swedish, French and English', *Journal of Phonetics* 19, pp. 351-365.

Field, A. (2009) *Discovering statistics using SPSS Third Edition.*Third ed London: Sage Publication.

Fletcher, J. (2010) 'The Prosody of Speech: Timing and Rhythm', in Hardcastle, W. J., Laver, J. and Gibbon, F. E.(eds) *Handbook of Phonetic Sciences (2nd edition).* . Second edLondon: Wiley-Blackwell.

Fry, D. B. (1955) 'Duration and intensity as physical correlates of linguistic stress', *Journal of the Acoustical Society of America,* 27, (4), pp. 765-768.

Goldsmith, J. A. (1990) *Autosegmental and metrical phonology: An introduction.* Oxford: Blackwell Publishing.

Grabe, E. and Low, E., L. (2002) 'Durational variability in speech and the rhythm class hypothesis. ', in Gussenhoven, C. and Warner, N.(eds) *Laboratory Phonology 7.* Berlin: Mouton De Gruyter.

Greenberg, S., Carvey, H., Hitchcock, L. and Chang, S. (2003) 'Temporal properties of spontaneous speech: A syllabic centric perspective', *Journal of Phonetics,* 31, (465-485).

Guaitella, I. (1999) 'Rhythm in speech: what rhythm organisations reveal about cognitive process in spontaneous speech versus reading aloud', *Journal of Pragmatics*, 31, pp. 509-523.

Halle, M. and Vergnaud, J. (1987) *an essay on stress.* London: MIT Press.

Harrington, J. (2010) 'Acoustic phonetics', in Hardcastle, W. J., Laver, J. and Gibbon, F. E.(eds) *Handbook of Phonetic Sciences (2nd Edition).* London: Wiley-Blackwell.

Hayes, B. (1991) 'Weight of CVC can be determined by context', in Cole, J. and Kisseberth, C.(eds) *Perspectives in Phonology.* Stanford: CSLI.

Heldner, M. and Strangert, E. (2001) 'Temporal effects of focus in Swedish', *Journal of Phonetics*
29, pp. 329-361.

Hoequist, C. J. (1983) 'Syllable duration in stress-, syllable-, and mora-timed languages', *Phonetica,* 40, pp. 203-237.

House, A. S. (1961) 'On vowel duration of English', *The Journal of the Acoustical Society of America,* 33, (9), pp. 1174-1178.

Jacewics, E., Fox, R. A. and Salmons, J. (2007) 'Vowel duration in three American English dialects', *American Speech,* 82, (4), pp. 367-385.

Jones, O., P (2006) 'Durational variability and stress-timing in Hawaiian'*, 11th Australian International Conference on Speech Science & Technology*

Kahn, D. (1976) *Syllable based generalization in English Phonology.* PhD thesis. MIT.

Keane, E. (2006) 'Rhythmic characteristics of formal and colloquial Tamil', *Language and Speech,* 49, pp. 299-332.

Kelly, M. H. (2004) 'Word onset patterns and lexical stress in English', *Journal of Memory and Language,* 50, pp. 231-244.

Kohler, K. J. (2003) 'Domains of temporal control in speech and language. From utterance to segment', *15th ICPhS*. Barcelona

Ladd, D. R. (1996) *Intonational phonology*. Cambridge: Cambridge University Press.

Ladefoged, P. (1967) *Three areas of experimental psychology*. London: Oxford University Press.

Ladefoged, P. (1982) *A course in phonetics.* New York: Harcourt Brace Jovanovich.

Ladefoged, P. (2003) *Phonetic data analysis: An introduction to fieldworks and instrumental techniques.* Oxford: Blackwell Publishing.

Ladefoged, P. (2005) *A course in phonetics.* Texas: Harracout College Publisher.

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

Lin, H. and Wang, Q. (2007) 'Mandarin rhythm: An acoustic study', *Journal of Chinese Language and Computing*, 17, (3), pp. 127-140.

Lindblom, B., Aguele, A., Sussman, H. and Eiir Cortes, E. (2007) 'The effect of emphatic stress on vowel coarticulation', *The Journal of the Acoustical Society of America*, 121, pp. 3802-3813.

Markus, D., Bond, D. and Stockmal, V. (2005) 'Measures of native and non-native rhtyhm in quantity sensitive language', *Language and Speech*, 48, pp. 55-63.

Massaro, D. W. and Idson, W. L. (1978) 'Temporal course of perceived vowel duration', *Journal of Speech and Hearing*, 21, pp. 37-55.

Mataim, B. (2001) *The Phonotactics of Brunei Malay: an optimality theoretic account.* PhD thesis. University of Essex.

McClure, J. D. (1977) 'Vowel duration in a Scottish accent', *Journal of International Phonetic Association*, 7, pp. 10-16.

Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., and Amiel-Tison, C. (1988). ‘A precursor of language acquisition in young infants,’Cognition **29**, 143–178.

Miller, J. L. and Grossjean, F. (1997) 'Dialect effects in vowel perception: The role of temporal information in French', *Language and Speech*, 40, (3), pp. 277-288.

Mochizuki-Sudo, M., Imagawa, H. and Kiritani, S. (1991) 'Naturalness judgments for stressed vowel duration in English and Japanese', *Annual Bulletin of Research Institute of Logopedics and Phoniatrics*, 25, pp. 99-102.

Morrison, G. S. and Escudero, P. (2007) 'A cross-dialect comparison of Peninsular-and Peruvian-Spanish vowels*', 16th ICPhS.* Saarbrucken, 6-10 August. pp. 1505-1508.

Morrison, G. S. and Nearey, T. M. (2006) 'A cross-language vowel normalisation procedure', *Canadian Acoustics,* 34, (3), pp. 94-95.

Morton, J., Marcus, S. and Frankish, C., (1976) 'Perceptual Centres (P-centres)', *Psychological Review*, 83(5), pp. 405-408

Nakatani, L. H., O'Connor, K. D. and Aston, C. H. (1981) 'Prosodic aspects of American English speech rhythm', *Phonetica,* 38, pp. 84-106.

Nazzi, T., Bertoncini, J., and Mehler, J. (1998) 'Language discrimination by newborns: towards an understanding of the role of rhythm' Journal of Experimental Psychology and human perception performance. 24, 756–766.

Nik Safiah, K. (1981) *Beberapa persoalan linguistik Bahasa Melayu*. Kuala Lumpur: Dewan Bahasa Pustaka.

Nik Safiah, K. (1993) *Tatabahasa dewan edisi baru*. Kuala Lumpur: Dewan Bahasa Pustaka.

Nooteboom, S. G. (1972) *Production and perception of vowel duration; a study of durational properties of vowels in Dutch*. PhD thesis.

Ouellet, M. and Tardif, B. (1996) 'From segmental duration properties to rhythmic structure: a study of interactions between high and low level constraints*', 4th International conference on spoken language processing*. Philadelphia,

Park, K., Lee, O. and Kim, J. (2003) 'English rhythm in foreign language education for Korean learners*', 8th Proceedings of Pan Pacific Association of Applied Linguistics.* Japan, pp. 368-372.

Pike, K. (1946) *The intonation of American English. 2nd Edition.*Second ed Ann Arbor: University of Michigan Press.

Ramus, F. (2002) 'Acoustic correlates of linguistic rhythm: Perspectives*', Speech Prosody.* Aix-en-Provence, 11th-13th April. pp. 115-120.

Ramus, F., Dupoux, E. and Mehler, J. (2003) 'The psychological reality of rhythm classes: perceptual studies*', 15th ICPhS.* Barcelona

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

Roach, P. (1982) 'On the distinction between 'stress-timed' and 'syllable-timed' Languages', in Crystal, D.(ed), *Linguistic Controvercies.* London: Edward Arnold.

Ross, T., Ferjan, N. and Arvaniti, A. (2008) *1st Southern California workshop on Phonetics/Phonology.* 1st November.

Ross, T., Ferjan, N. and Arvaniti, A. (2008) 'Speech rhythm and its quantification in L1', *1st Southern California workshop on Phonetics/Phonology.* 1st November. pp.

Rouas, J., Farinas, J., Pellegrino, F. and Andre-Obrecht. (2005) 'Rhythmic unit extraction and modelling for automatic language identification', *Speech Communication,* 47, pp. 436-456.

Sarmah, P., V, G. D. and Wiltshire, C. (2009) 'Thai English rhythm and vowels', *English World-Wide,* 30, (2), pp. 196-217.

Schütz, A. J. (1985) *The Fijian Language.* Honolulu: University of Hawaii Press.

Scott, S. K. (1998) 'The point of P-centres', *Psychological Research,* 61, (1), pp. 4-11.

Sievers, E. (1881) *Grundzüge der Phonetik.* Leipzig: Breitkopf and Hartel.

Smiljanic, R., Viau, J. and Bradlow, A. R. 'Rhythmic characteristics of English clear and conversational speech', *10th Conference on Laboratory Phonology.* Paris, 29th July- 1st August.

Stetson, R. H. (1928) 'Motor Phonetics ', *Archives néerlandaises de phonétique expérimentale, Tome 3.,* III.

Strange, E., Weber, A., Levy, E. S., Shafiro, V., Hisagi, M. and Nishi, K. (2007) 'Acoustic variability within and across German, French and American English vowels: Phonetic context effects', *The Journal of the Acoustical Society of America,* 122, (2), pp. 1111-1129.

Suomi, K. (2007) 'On the tonal and temporal domains of accent in Finnish', *Journal of Phonetics* 35, pp. 45-55.

Tajul, A. K. (2000) *Phonological word in Standard Malay.* PhD thesis. Newcastle University.

Teoh, B. S. (1994) *The sound system of Malay revisited.* Kuala Lumpur: Dewan Bahasa Pustaka.

Thomas, E. R. and Carter, P. M. (2006) 'Prosodic rhythm and African American English', *English World-Wide,* 27, (3), pp. 331-355.

Turk , A., Nakai, S. and Sugahara, M. (2006) 'Acoustic segment durations in prosodic
research: a practical guide', in Sudhoff, S., Lenertova, D., Meyer, R., Pappert, S., Augurzky, P., Mleinek, I., ., R. N. and Schliesser, J.(eds) *Methods in Empirical Prosody Research.* Berlin: de Gruyter, pp. 1-28.

Turk, A. E. and Sawusch, J. (1997) 'The domain of accentual lengthening in American English', *Journal of Phonetics,* 25, pp. 25-41.

Volin, J. and Studenovsky, D. (2007) 'Normalization of Czech Vowels', *16th ICPhS.* Saarbrucken, 6-10 August.

Waltl, P. and Hoole, P. (2008) 'An EMG study of the German vowel system', *8th International Seminar on Speech Production.*

Wenk, B. J. and Wioland, F. (1982) 'Is French really syllable-timed?', *Journal of Phonetics,* 10, pp. 193-216.

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

White, L. and Mattys, S. L. (2007) 'Rhythmic typology and variations in first and second languages', in Prieto, P., Macaro, J. and Sole, M.-J.(eds) *Segmental and prosodic issues in Romance phonology.* Barcelona: John Benjamins, pp. 237-257.

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

Yunus, M. (1980) *The Malay sound system.* Kuala Lumpur: Fajar Bakti.

Zaharani, A. (1998) *Phonology and morphology interface in Malay: An optimality theoratical account.* PhD thesis. University of Essex.

Zaharani, A. (2004) 'Alignment and the syllable coda condition in Malay: An optimality account', *Gema,* 4, (2), pp. 1-32.

Zuraidah, M. D. and Knowles, G. (2006) 'Prosody and turn-taking in Malay broadcast interviews', *Journal of Pragmatics,* 38, (4), pp. 490-512.

Zuraidah, M. D., Knowles, G. and Yong, J. (2008) 'How words can be misleading: A study of syllable timing and 'stress' in Malay', *The Linguistics Journal,* 3, (2).