9

Rhythm in Singapore and British English: a comparative study of indexes

Fiona Ong Po Keng, David Deterding and Low Ee Ling

Introduction

Indexes have been devised in an attempt to measure the rhythmic patterns of speech based on acoustic studies. Recent indexes developed by Ramus, Nespor and Mehler (1999), Low, Grabe and Nolan (2000) and Deterding (2001) have been employed in research studies including medical research, language acquisition and development, and sociolinguistics.

This paper, based on Ong (2004), investigates the reliability of the indexes by examining inter-rater variability and also the correlation between the results obtained through measurement and perceptual impressions. Problems in the measurement process that can substantially affect the output are also discussed.

There is a substantial prosodic difference between spontaneous and scripted speech. Here, Singapore English (SgE) and British English (BrE) conversational data from the NIECSSE is used, as it represents a reasonably natural occurrence of discourse, and also incorporates a range of rhythmic patterns.

Basis of acoustic studies in rhythm investigations

Speech rhythm is difficult to define, as it encompasses both perceptual and physical domains, and this partly explains the different approaches and limited success in early attempts to measure speech rhythm (eg Roach 1982).

However, more recent indexes developed by Ramus et al (1999), Low et al (2000) and Deterding (2001) have produced empirical evidence to support the plausibility of rhythmic categorisation of different languages.

Deterding D, Brown A & Low E L (eds) English in Singapore:
Phonetic Research on a Corpus © *McGraw-Hill Education (Asia) 2005*

Each of these indexes focuses on a different aspect in the phonological realms identified by Dauer (1983, 1987) as contributing to rhythmic typology.

Ramus et al deal with variations in vocalic and consonantal duration in an attempt to classify a range of languages on a scale of stress- and syllable-timing. Low and her co-authors, on the other hand, measure variability in vowel duration, on the grounds that the incidence of full and reduced vowels generally predicts rhythm well (Dauer 1983; Cruttenden 2001:251). In contrast, Deterding measures variability in syllable duration, as syllable structure and syllable weight play crucial roles in rhythm typology.

Since the index of Ramus and his associates is designed to classify different languages, not variation within the same language, it will not be discussed further here. This paper will concentrate on a comparison of the indexes of Low et al and Deterding.

Data

This study is based on the conversational speech of five SgE female (F1 to F5), five SgE male (M1 to M5) and five BrE speakers from the interviews in the NIECSSE. Three utterances with at least seven syllables were selected from each speaker, giving a total of 45 utterances. Each utterance consists of a stretch of continuous speech that has no hesitations or pauses, either silent or filled, and no extra lengthening of syllables.

Perceptual investigation

For the perceptual investigation, the 45 utterances were randomly ordered and presented to 32 listeners, 27 of whom were students while the other 5 were one SgE and four BrE lecturers at NIE. All the subjects had at least a basic understanding in phonetics.

To aid listeners in using the scale of rhythm, the sentence ‘This is an utterance to show the different kinds of rhythm’ was recorded twice to represent the extremes of a syllable-based and a stress-based utterance. These two examples were played to the subjects prior to the main perception test.

Each of the 45 utterances was then played three times, and the listeners estimated the rhythm of each utterance by circling a number on a scale of 1 to 9 on the answer sheet, 1 indicating most syllable-based and 9 indicating most stress-based.

In addition, 22 of the listeners were asked to judge the number of syllables they felt there were in a range of words played from the data, including *Morocco* and *Portugal*. These 22 listeners were in the same class and were able to commit the time to complete this extra test. The purpose of the test was to corroborate the raters' measurements and to evaluate the robustness of each index against listeners' perceptions.

Acoustic measurements and problems

Three raters, one BrE (rater 1) and two SgE (raters 2 and 3), were involved in this study to obtain two sets of values for each index. Rater 1 and Rater 3 did the measurements for the Variability Index (VI) of Deterding's index, while Rater 2 and Rater 3 were involved for the Pairwise Variability Index (PVI) of Low and associates' index.

Problems in syllabification

Both indexes depend crucially on breaking the utterances into syllables. A major contention arose concerning the perception of the number of syllables, because phonological processes such as compression and possible omission of final suffixes resulted in subjective decisions in many cases. For example in the first utterance in Table 9.1, there is a discrepancy concerning the number of syllables in *basically*, and in the second utterance, the discrepancy concerns whether there is an *-ed* suffix on *participated*.

Table 9.1: Discrepancies in the number of syllables due to compression and deletion of a suffix. Word boundaries are assumed to be syllable boundaries. Word-internal syllable boundaries are shown by '.'

File	Rater	Utterance	Syllables	Process
M4-c:21	R2	that's the pre.mise of the sto.ry ba.si.ca.lly and	13	Compression
	R1/R3	that's the pre.mise of the sto.ry ba.si.cally and	12	
BF-1a:39	R2	not that I par.ti.ci.pate in that sort of thing	12	Existence of *-ed* suffix
	R1/R3	not that I par.ti.ci.pa.ted in that sort of thing	13	

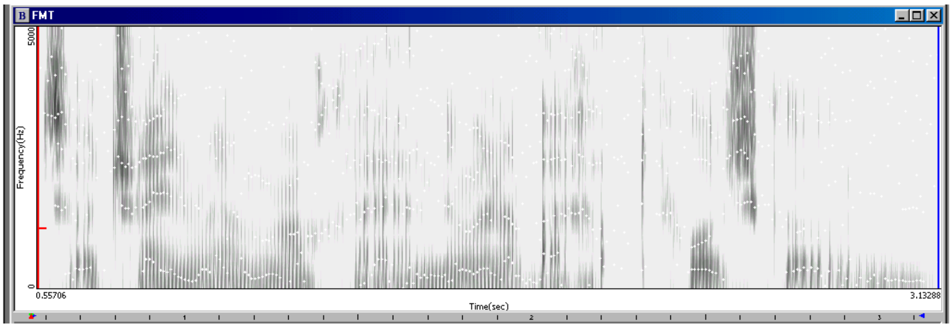
Another discrepancy in the perceived number of syllables was due to the plausibility of syllabic consonants. For example, in *Morocco* pronounced as [mrɒkoʊ] (BM2-a:16) and also *Portugal* pronounced as [pɔːtʃgəl] (BM-2a:31), the possibility of [m] and [ʃ] representing syllables resulted in variation between the raters. This indeterminacy was supported by the syllabification test for these two words, the responses for which are shown in Table 9.2.

Table 9.2: Responses for number of syllables in *Morocco* and *Portugal*

File	Word	Transcription	Syllables	Responses
BM2-a:16	*Morocco*	/mrɒ.koʊ/	2	10
		/mə.rɒ.koʊ/	3	12
BM2-a:31	*Portugal*	/pɔːtʃ.gəl/	2	9
		/pɔː.tʃʊ.gəl/	3	13

The spectrogram in Figure 9.1 further illustrates that there is no vowel, especially in the [ʃ] of *Portugal*, to help in determining the number of syllables.

Figure 9.1: Spectrogram of *took a train around Morocco and then went up back up to Portugal* (BM2-a:16)



tʊkə tr eɪn əraʊnmrɒ k oʊ ən ðenwenʌpbækʌpt p ɔː tʃ g ə l

Syllable boundaries

One possibility to minimise subjectivity is to adhere strictly to pre-determined principles wherever possible. For Deterding’s index, the

Maximal Onset Principle (MOP) was enforced in all cases. This proved useful especially in cases of potential ambisyllabicity and in instances of simplification due to connected speech process.

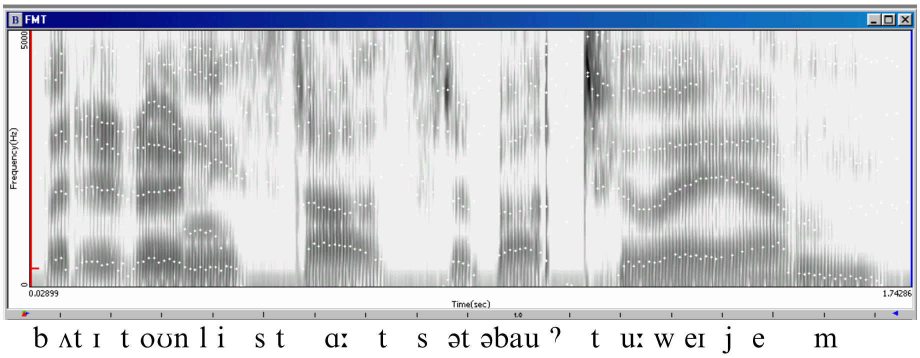
However, such strict adherence to the MOP was at times counter-intuitive. Furthermore, even if one were to follow the British pronunciation suggested in a standard pronouncing dictionary such as Roach, Hartman and Setter (eds 2003), there are variants which give rise to alternative ways of syllabifying words such as *Australia* and *escape* as shown in Table 9.3.

Table 9.3: Variations in syllabification of *Australia* and *escape* as shown in Roach et al (eds 2003:42, 185)

Word	Variations in syllabification
Australia	ɒs.treɪ.liə, ɔː.streɪ.liə
escape	ɪ.skeɪp, es.keɪp, ə.skeɪp

Another concern in determining syllable boundaries arises because conventional word boundaries are generally not observed in connected speech (Calvert 1980:164). They are often blurred in processes like elision, linking, assimilation and coalescence (Brown 1990:57–88). For example, the presence of the linking [w] and [j] was found in the segment *two AM* pronounced as [tuːweɪjem] as shown in the spectrogram in Figure 9.2.

Figure 9.2: Spectrogram of linking [w] and [j] in *two AM* (BF1-a:35)



This caused two problems. Firstly, it is uncertain if these linking elements should be regarded as real consonants as they are not present in the underlying string of phonemes. Secondly, if they are acknowledged as consonantal segments, according to the MOP, they should be placed at the onset of the second syllable. However, due to [w] and [j] being semi-vowels, it was difficult to identify the exact location of the end of the first syllable and onset of the second syllable.

Reduction and deletion of vowels

Adherence to the MOP solves many problems resulting from perceptual inconsistencies. However, issues remain for Low et al's index if there is no vowel to be measured, especially in the case of syllabic consonants.

The original guideline in Low et al (2000) is for a voiceless plosive to be measured from the onset of the plosive, and this was devised for polysyllabic words in read data. For the current study, it was suggested that this guideline be applied to all function words containing a voiceless plosive preceding a /ə/ since there is a tendency for the /ə/ to be deleted in fast speech.

This adaptation of the guideline was eventually extended to other instances where the /ə/ was deleted, such as in *just* (BF2-a:26) where /dʒ/ was measured to represent the vocalic segment and the /st/ in the coda was not included. Similarly, on the assumption that the /m/ in *Morocco* and the /tʃ/ in *Portugal* are syllabic, one could measure the duration of this syllabic consonant. However, judgement about whether a vowel has been deleted or not remains an issue, and this highlights the need for additional guidelines to be created to reduce the subjectivity inherent in the measurement process.

Results

Inter-rater variability

Table 9.4 shows the degree of agreement (Pearson's r value) between two raters for each index. There is a greater agreement between the raters for Deterding's index than for Low et al's index, which implies that the former is easier to use, partly because clear principles and guidelines were already in place for use in conversational speech. In comparison, the results suggest that Low et al's index may require additional guidelines to cater for phonological processes that occur in conversational speech and thereby avoid some of the subjectivity in the measurements.

Table 9.4: Correlation coefficients of measurements between two raters

Raters	Index	Pearson's r value
R1 and R3	Deterding (2001)	0.74
R2 and R3	Low et al (2000)	0.60

Comparison of mean values: Deterding's index

The mean values for the two raters for Deterding's index are shown in Table 9.5 and they are significantly different (t=2.83, df=43, paired-sample, two-tailed, p<0.01). This suggests that there is a systematic shift in judgement between the two raters when dealing with the measurement of syllables.

Table 9.5: Comparison of mean (VI) for R1 and R3 for Deterding's index

Rater	Mean VI
Rater 1	0.54
Rater 3	0.59

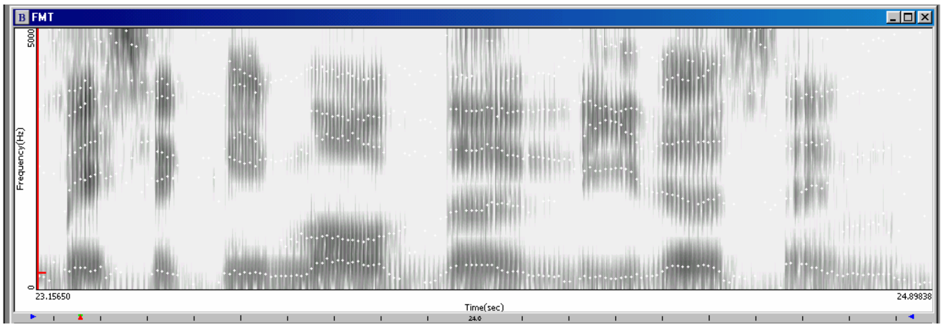
The mean values of the two raters are shown separately for the BrE and SgE speech in Table 9.6, and this reveals that there is no significant difference in the mean for the BrE speech (t=0.45, df=13, paired-sample, two-tailed, ns) but there is a significant difference for the SgE speech (t=3.20, df=28, paired-sample, two-tailed, p<0.01).

Table 9.6: Mean (VI) of BrE and SgE speakers for Deterding's index

Rater	BrE speakers	SgE speakers
Rater 1	0.60	0.51
Rater 3	0.62	0.58

This seems surprising as one might expect that the clearer syllables of SgE would be easier to measure. A possible explanation is that there is a systematic bias in the perception of SgE speech.

Figure 9.3: Spectrogram of *basically more of the leaders and* (M5-b:26)



b eɪ s ɪ k l i m ɔːɒ v ð ə l iː d ə z æ n

This issue is illustrated by the judgements of the boundary between *more* and *of* in the utterance shown in Figure 9.3. The two words are merged together, so estimation of the boundary between them is inevitably subjective. Rater 1 assumed durational equality for the two syllables, maybe because of a belief that the syllable-based rhythm of SgE speech results in a tendency to avoid weak forms in words such as *of*. In contrast, Rater 3 was influenced by the general convention of long and short vowels /ɔː/ and /ɒ/, and judged that the long vowel in *more* dominated a shorter vowel in *of*.

Comparison of mean values: Low et al's index

While the correlation between raters may be lower for Low et al's index, the mean values for the two raters as shown in Table 9.7 are not significantly different (t=1.49, df=43, paired-sample, two-tailed, ns).

Table 9.7: Comparison of mean (PVI) for R2 and R3
for Low et al's index

Rater	Mean PVI
Rater 2	48.4
Rater 3	51.4

The fact that the overall means for R2 and R3 are closer than R1 and R3 might be because R2 and R3 are both Singaporeans, while R1 is British. This suggests that a difference in ethnicity may contribute to a bias from preconceived beliefs, resulting in a systematic shift in the results. Note that the discrepancy illustrated in the boundary between the words *more of* in Figure 9.3 involves the duration of vowels, so it would affect the results of both indexes equally.

Investigation of the perceptual impression of rhythm

The scatter plots for the output of the two indexes against the results of the perception test are shown in Figure 9.4. For both indexes, there is quite wide variation, but the alignment between the index and the perception test is slightly better for Low et al's index, partly because of the large number of items with a high score for the index but a low score for the perception test for Deterding's index.

Figure 9.4: Scatter plots for Deterding's and Low et al's index against the perception test

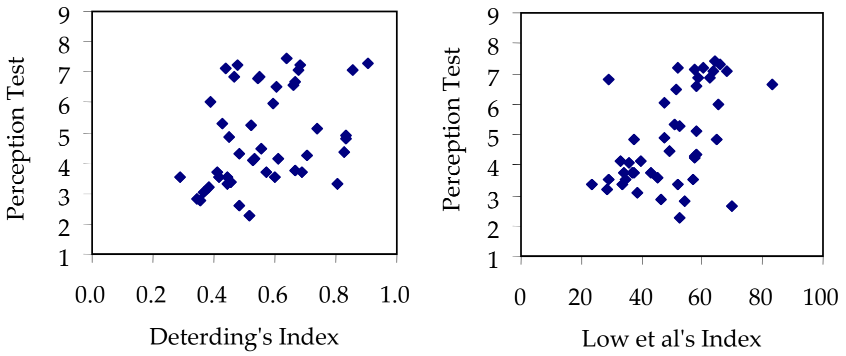
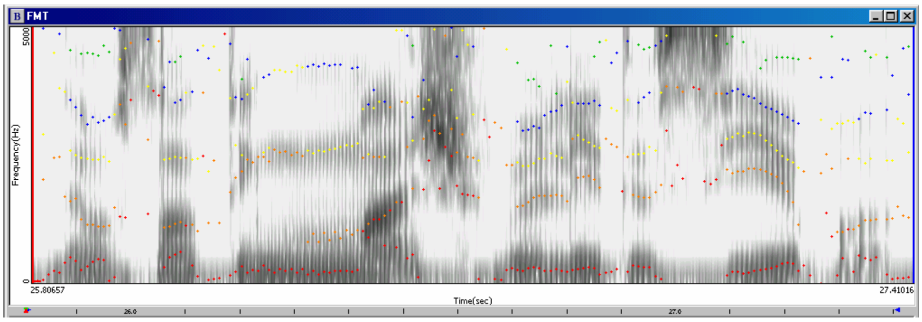


Table 9.8 shows the correlation (Pearson's r value) between each of the indexes and the perception data, and it confirms that Low et al's index has a stronger relationship with the perception of rhythm than Deterding's index. This suggests that the measurement of vowels instead of syllables better embodies rhythmic patterns in speech.

Table 9.8: Correlation of values between indexes and the perception test

Index	Pearson's r value
Deterding	0.37
Low et al	0.51

Figure 9.5: Spectrogram of *There was nothing much really to see about* (M2-a:26)



ðeəwɒ z n ʌ θ ɪ ŋ m ʌ tʃ r ɪə l ɪ t ʊ s iː ə b aʊ ˀ

One possibility could be that, since vowels represent the sonorous peaks in the syllable, they are perceptually much more distinct (Ladefoged 2001:227), and listeners may be basing their estimates of the rhythm of the different utterances on the perception of contrastive strong and weak forms of vowels. However, the measurements using Deterding's index are often substantially affected by the duration of consonants such as fricatives, nasals and aspirated plosives, and it is conceivable that listeners by and large ignore these when estimating rhythm.

For example, in the utterance in Figure 9.5, the fricatives /z/, /s/ and affricate /tʃ/ are all quite long, and so are the nasals /ŋ/ and /m/, and as a result the variation in syllable duration was found to be quite high. However, this utterance was perceived to have syllable-based rhythm, possibly because the function word *to* has a full vowel /tʊ/. It seems, therefore, that the nature of the vowels, especially whether function words have a reduced vowel or not, may be far more important in the perception of rhythm than the acoustic duration of syllables.

Conclusion

One of the main issues affecting the results is the subjectivity in deciding on the number of syllables in an utterance. While fixed adherence to the Maximal Onset Principle in Deterding's index was successful in reducing the number of subjective decisions and thus resulted in better inter-rater consistency, it also resulted in some counter-intuitive syllable boundaries that could have contributed to the lower correlation between the measurements and the listeners' perception of rhythm.

In contrast, Low et al's index reflected rhythm better, but showed a reduced degree of inter-rater reliability. This is because the original guidelines were devised for read speech where the data was controlled. Although additional guidelines were implemented in an attempt to account for phonological processes, particularly the deletion of vowels, they were not easy to apply in conversational speech. Therefore, Low et al's measure needs to be tightened with more guidelines based on the considerations for the phonological processes that are inherent in speech.

Another significant finding that emerged is the possibility of a systematic bias in the measurements. It is possible that this arose because of differences in ethnicity and some bias from preconceived beliefs.

Finally, we might note that, though the correlation of Low et al's index with the perceptual judgements was better than that for Deterding's index, the agreement is still not very high, and there are some utterances that are perceived as syllable-based but measured as stress-based. Further work is clearly required to examine the effects of phonological processes on the application of rhythm indexes, and also to tighten the application of the indexes.

References

Brown G (1990) *Listening to Spoken English* (2nd edition) Longman, London.

Calvert D R (1980) *Descriptive Phonetics* (2nd edition) Thieme, New York.

Cruttenden A (2001) *Gimson's Pronunciation of English* (6th edition) Arnold, London.

Dauer R M (1983) 'Stress-timing and syllable-timing reanalyzed' *Journal of Phonetics* 11:51–69.

Dauer R M (1987) 'Phonetic and phonological components of language rhythm' *Proceedings of the XIth International Congress of Phonetic Sciences* Tallin, Estonia, pp 447–50.

Deterding D (2001) 'The measurement of rhythm: a comparison of Singapore and British English' *Journal of Phonetics* 29:217–30.

Ladefoged P (2001) *A Course in Phonetics* (4th edition) Harcourt College Publishers, Fort Worth.

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

Ong P K F (2004) 'Rhythm: a comparative study of indexes' Honours academic exercise, National Institute of Education, Singapore.

Ramus F, Nespor M & Mehler J (1999) 'Correlates of linguistic rhythm in the speech signal' *Cognition* 73(3):265–92.

Roach P (1982) 'On the distinction between "stress-timed" and "syllable-timed" languages' in D Crystal (ed) *Linguistic Controversies: Essays in Linguistic Theory and Practice in Honour of F R Palmer* Edward Arnold, London, pp 73–79.

Roach P, Hartman J & Setter J (eds 2003) [Daniel Jones] *English Pronouncing Dictionary* (16th edition) Cambridge University Press.