# Comparing native and non-native speech rhythm using acoustic rhythmic measures: Cantonese, Beijing Mandarin and English

*Peggy P.K. Mok[1] and Volker Dellwo[2]*

[1]Department of Linguistics and Modern Languages, The Chinese University of Hong Kong
[2]Department of Phonetics and Linguistics, University College London
peggymok@cuhk.edu.hk; v.dellwo@ucl.ac.uk

## Abstract

This study investigates the speech rhythm of Cantonese, Beijing Mandarin, Cantonese-accented English and Mandarin-accented English using acoustic rhythmic measures. They were compared with four languages in the BonnTempo corpus: German and English (stress-timed) and French and Italian (syllable-timed). Six Cantonese and six Beijing Mandarin native speakers were recorded reading the North Wind and the Sun story with a normal speech rate, telling the story semi-spontaneously and reading the English version of the story. Both raw and normalised rhythmic measures were calculated using vocalic, consonantal and syllabic durations (ΔC, ΔV, ΔS, %V, VarcoC, VarcoV, VarcoS, rPVI_C, rPVI_S, nPVI_V, nPVI_S). Results confirm the syllable-timing impression of Cantonese and Mandarin. Data of the two foreign English accents poses a challenge to the rhythmic measures because the two accents are syllable-timed impressionistically but were classified as stress-timed by some of the rhythmic measures (ΔC, rPVI_C, nPVI_V, ΔS, VarcoS, rPVI_S and nPVI_S). VarcoC and %V give the best classification of speech rhythm in this study.

## 1. Introduction

Speech researchers have traditionally classified languages into different rhythmic groups: syllable-timed, stress-timed and mora-timed. The original suggestion was that there are quasi-isochronous durational units in the speech signal for such classification: syllable for syllable-timing, inter-stress intervals for stress-timing, and mora for mora-timing. However, no acoustic evidence for such isochronous units to support the rhythmic class hypothesis could be found (see [4, 7] for a review). Dauer [4] proposed that instead of having different isochronous units, stress-timed languages and syllable-timed languages differ in several important aspects: syllable structure, vowel reduction and stress. Stress-timed languages have more variation in syllable length and structure, more reduced unstressed syllables, more variation in the phonetic realisation of stress and more stress-related rules than syllable-timed languages. These features combine with one another to give the impression of stress-timing versus syllable-timing. Languages can be more or less stress-timed or syllable-timed, with a continuum between the two.

On the acoustic level, several measurement procedures have been proposed which could reflect the auditory impression of different rhythmic classes: %V (percentage of vocalic durations in speech), ΔC, ΔV (standard deviations of consonantal and vocalic durations respectively) by Ramus et al. [10] and Pairwise Variability Index (PVI) of vocalic and consonantal durations by Grabe & Low [7]. These measures take only the duration of vowels and consonants as the basis for rhythmic classifications. Their results show that %V and ΔC, the normalised vocalic PVI and the raw consonantal PVI can categorise different languages into distinct rhythmic clusters, but it is unclear how languages falling between these clusters should be classified rhythmically.

This study investigates the speech rhythm of Cantonese and Beijing Mandarin using the above acoustic measures. Cantonese has a very simple syllable structure with no lexical stress and no phonological vowel reduction. Every syllable carries a lexical tone. In emotionally neutral sentences, each syllable receives roughly equal emphasis [1]. Impressionistically, Cantonese is a typical syllable-timed language, but so far no study has examined its rhythm using acoustic measures.

The speech rhythm of Mandarin is less clear. Mandarin is similar to Cantonese in that it also has lexical tones and a very simple syllable structure. However, unstressed syllables (the so-called 'neutral tone') occur frequently in Mandarin. Duration of such toneless syllables is dramatically reduced and their vowel qualities are also reduced to schwa-like [3]. Grabe & Low [7] found that Mandarin has the lowest vocalic PVI values among all the languages in their study suggesting that it is a typical syllable-timed language. However, they looked at Singaporean Mandarin in which unstressed syllables occur much less frequently than in Beijing Mandarin because Singaporean Mandarin is heavily influenced by other southern Chinese languages. It is possible that there may be subtle differences between the rhythms of these two Mandarin accents. Benton et al. [2] compared Beijing Mandarin and American English using rhythmic measures with over 50 speakers in each language. They found that the rhythmic values for Mandarin and English are significantly different, but there was considerable diversity between individual speakers of both languages. They also did not state explicitly whether Mandarin is a syllable-timed language. In addition, given that rhythmic differences between languages are continuous rather than categorical [4], there can be a significant variation among languages belonging to the same rhythm class [5, 6]. Therefore, comparison with more languages is necessary in order to investigate the speech rhythm of Beijing Mandarin.

In addition, although American and British English are typical stress-timed languages, other English accents can belong to a different rhythm class, e.g. Singaporean English and Taiwan English are syllable-timed [8, 9]. Measuring syllable durations, Setter [11] found that English spoken by Hong Kong Cantonese speakers exhibits much less variation than by British English speakers, which contributed to the syllable-timing impression of Cantonese English. However, she did not investigate its rhythm using rhythmic measures developed by [7, 10]. The rhythm of English spoken by Beijing Mandarin speakers is also little explored, so it is

worth examining the rhythm of these two English accents using acoustic rhythmic measures.

In addition to investigating the speech rhythm of Cantonese, Beijing Mandarin, Cantonese-accented English and Mandarin-accented English using the above-mentioned acoustic rhythmic measures, we also compared these languages with four languages in the BonnTempo Corpus [5]: German and English (stress-timed), French and Italian (syllable-timed).

## 2. Method

### 2.1. Speakers

Six native Hong Kong Cantonese speakers and six native Beijing Mandarin speakers (three male, three female) were used. They were either undergraduate or postgraduate students at the Chinese University of Hong Kong and were paid to participate in the experiment. None of them reported any speech or hearing problem.

We compared these speakers with previously published data (BonnTempo corpus, see [5]). The number of languages and speakers are as follows: German (15), British English (7), French (6) and Italian (3). German and English represent examples of stress-timed languages, French and Italian examples of syllable-timed languages.

### 2.2. Materials and procedures

The North Wind and the Sun story was used as the experimental material for Cantonese and Mandarin speakers. The recording took place in a sound-treated room at The Chinese University of Hong Kong. Recordings were made directly to disk with a sampling rate of 22050 Hz. The speakers practised reading the story as many as times as they liked before the actual recording. They were recorded reading the story with three self-selected speech rates: normal, fast and slow. Then, they were recorded telling the story themselves without reading the script for semi-spontaneous speech. Finally, they read the English version of the story with a normal speech rate for their foreign-accented English. Only data for normal speech rate (reading in both Chinese and English) and semi-spontaneous speech (telling the story) is reported in this paper.

The speech material in the BonnTempo Corpus consists of read speech based on a short passage from a novel in German, which was translated into the other languages by native speakers of the target language (English, French and Italian). Five speech rates were used: very slow, slow, normal, fast, very fast. Again, only data for normal speech rate is used for comparison in this study.

### 2.3. Labelling

All Cantonese and Mandarin sound files were labelled manually into syllabic, consonantal and vocalic intervals using Praat and were cross-checked by the first author, a native Cantonese speaker who also speaks Mandarin. Syllable intervals were labelled as phonological syllables by reference to acoustic cues and careful listening, unless no acoustic cues of the syllable can be found as in the case of elision. Segmentation criteria followed those in [7] except that a 50 ms closure duration was added to all post-pausal initial stops for consistency. The story was divided into several sentences. Any silent pause within a sentence was excluded from further analysis. Pre-pausal or utterance-final syllables were not excluded because they may be language-specific and may contribute to the perceived rhythmic pattern. The sound files in the BonnTempo Corpus were labelled in a similar way [5].

### 2.4. Calculation of rhythmic measures

Durations (ms) of syllabic, consonantal and vocalic intervals were extracted using a Praat script. The following rhythmic measures were calculated for each sentence by each speaker, which were then averaged for each speaker.

- ΔC: the standard deviation of consonantal durations
- ΔV: the standard deviation of vocalic durations
- ΔS: the standard deviation of syllabic durations
- %V: the proportion of vocalic durations within a sentence

Since ΔC and ΔV have repeatedly been demonstrated to interact with the average segment duration, we applied a normalisation procedure by calculating the coefficient of variation [6].

- VarcoC: (ΔC / mean consonantal duration) × 100
- VarcoV: (ΔV / mean vocalic duration) × 100
- VarcoS: (ΔS / mean syllabic duration) × 100

In addition, two sets of PVI values, raw and normalised, were calculated using the formulas in [7]. Raw PVI was calculated for consonantal (rPVI_C) and syllabic (rPVI_S) durations, while normalised PVI was calculated for vocalic (nPVI_V) and syllabic (nPVI_S) durations.

## 3. Results

### 3.1. %V, ΔC, VarcoC, nPVI_V and rPVI_C

Figure 1 to 3 show ΔC plotted against %V, VarcoC against %V and nPVI_V against rPVI_C respectively of the languages used in this study. In all the figures and tables below, Can = Cantonese, Man = Mandarin, _n = reading with a normal speech rate, _t = telling the story semi-spontaneously, CanEng = Cantonese-accented English, ManEng = Mandarin-accented English.

It can be seen in Figure 1 and 2 that the %V values of Cantonese and Mandarin are higher than Italian and French (comparing Italian and French with the normal version of Cantonese and Mandarin [$F(3,17) = 5.758$, $p = 0.007$]). Post hoc comparisons with Bonferroni adjustment shows that French is significantly different from both Cantonese ($p = 0.020$) and Mandarin ($p = 0.047$), while Cantonese and Mandarin are not significantly different.

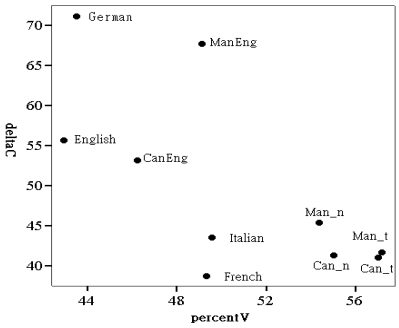


Figure 1*: ΔC and %V of all the languages.*

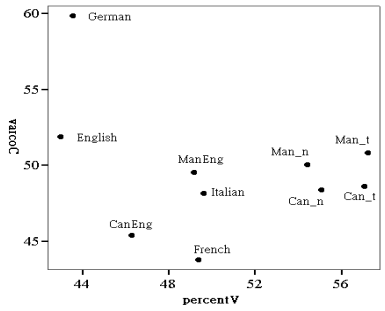


Figure 2: *VarcoC and %V of all the languages.*

The %V values of semi-spontaneous speech of the two languages (_t) are higher than read speech with a normal speech rate (_n). Paired-samples t-tests show that this stylistic difference is significant for both Cantonese [t(5) = -3.591, p = 0.016] and Mandarin [t(5) = -3.754, p = 0.013].

Figure 1 shows that the ΔC parameter groups both Cantonese English and Mandarin English with stress-timed languages English and German. However, VarcoC in Figure 2 groups the two English accents with syllable-timed languages.

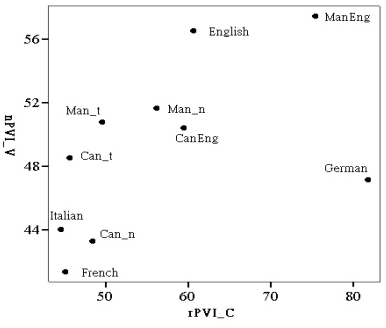


Figure 3: *nPVI_V and rPVI_C of all the languages.*

The nPVI_V and rPVI_C parameters differentiate syllable- and stress-timing less clearly than VarcoC and %V. In Figure 3, the syllable-timed Mandarin English is numerically closer to stress-timed languages, while Cantonese English is numerically closer to syllable-timed languages. In addition, the stylistic difference between read speech and semi-spontaneously speech in Cantonese and Mandarin observed in %V disappear in both nPVI_V and rPVI_C. The two languages are also not distinct from Italian and French.

### 3.2. Indexes of syllable durations

The average syllable durations (ms) of all the languages in this study in descending order are as follows: ManEng (250), CanEng (217), German (197), Mandarin_n (190), Cantonese_n (184), Cantonese_t (178), English (178), Mandarin_t (172), French (155), Italian (141). Pair-samples t-tests indicate that Cantonese and Mandarin speakers spoke significantly slower in their accented English than in their native language (both _n and _t versions), which results in more lengthened syllables (Can_n [t(5) = -2.68, p = 0.044], Can_t [t(5) = -4.578, p = 0.006], Man_n [t(5) = -8.028, p < 0.0005], Man_t [t(5) = -9.267, p < 0.0005]). The difference in syllable durations between read speech and semi-spontaneous speech is only significant in Mandarin [t(5) = 3.798, p = 0.013].

In addition to calculating various indexes for consonantal and vocalic durations, we also calculated such indexes for syllable durations. Table 1 and 2 show the values of ΔS, VarcoS, rPVI_S and nPVI_S of all the languages in this study in descending order. At first glance, none of the four measures seem to classify the languages satisfactorily according to the auditory impression of their rhythm. Both Cantonese English and Mandarin English have higher ΔS and rPVI_S values than English and German. For VarcoS and nPVI_S, Cantonese English and Mandarin English fall between English and German. However, it is interesting to note that if the data of Mandarin English and Cantonese English was excluded, except VarcoS, the other three measures all rank stress-timed English and German at the top, followed by other syllable-timed languages. The rPVI_S parameter seems to give the best separation between stress-timed and syllable-timed languages, followed by ΔS. Although nPVI_S gives the same order, there is only a small difference between German and Italian suggesting that there may not be a clear-cut separation. Finally, all four measures rank Mandarin higher than Cantonese meaning that there is more variation of syllable durations in Mandarin than Cantonese, in line with expectation because of the frequent occurrence of unstressed syllables in Mandarin. Cantonese is ranked the lowest by three out of the four measures.

Table 1: *ΔS and VarcoS of all the languages.*

| Language | ΔS | | Language | VarcoS |
|---|---|---|---|---|
| ManEng | 124.52 | | English | 51.87 |
| CanEng | 106.80 | | ManEng | 49.29 |
| English | 88.74 | | CanEng | 47.43 |
| German | 80.78 | | Italian | 46.73 |
| Man_n | 75.80 | | German | 43.53 |
| Man_t | 68.33 | | Man_t | 39.27 |
| Italian | 67.61 | | Man_n | 38.17 |
| Can_t | 62.90 | | French | 36.15 |
| Can_n | 57.48 | | Can_t | 34.70 |
| French | 55.30 | | Can_n | 30.71 |

Table 2: *rPVI_S and nPVI_S of all the languages.*

| Language | rPVI_S | | Language | nPVI_S |
|---|---|---|---|---|
| ManEng | 151.13 | | English | 69.67 |
| CanEng | 124.70 | | ManEng | 60.67 |
| English | 115.50 | | CanEng | 57.65 |
| German | 99.62 | | German | 56.42 |
| Man_n | 86.08 | | Italian | 54.78 |
| Italian | 82.68 | | French | 49.47 |
| Man_t | 79.37 | | Man_t | 45.95 |
| French | 75.89 | | Man_n | 45.02 |
| Can_t | 65.97 | | Can_t | 36.77 |
| Can_n | 63.62 | | Can_n | 34.32 |

## 4. Discussion

All rhythmic measures in this study confirm the syllable-timing impression of Cantonese and Beijing Mandarin. The results show that Cantonese has more extreme rhythmic values than Mandarin, French and Italian, which presumably is contributed by the absence of lexical stress in Cantonese. A similar situation is also found in Singaporean Mandarin which has far fewer unstressed syllables than Beijing Mandarin. The data in [7] shows that Singaporean Mandarin has the lowest nPVI_V value and the highest %V value among all the

languages in their study, suggesting that Singaporean Mandarin is the most typical syllable-timed language. It will be of interest to compare more syllable-timed languages with and without lexical stress to assess the effect of lexical stress in syllable-timing.

The significant difference in %V values of the two styles in Cantonese and Mandarin (read speech vs semi-spontaneous speech) implies that speakers may slightly change their rhythmic patterns according to speaking styles. This seems quite possible because read speech and spontaneous speech can differ in many aspects, including prosody. The stylistic difference in %V can also be partly explained by segmentation issues. Initial /j/ and /w/ were considered consonantal if there were acoustic cues for segmentation. However, in semi-spontaneous speech, many of these initial glides could not be separated from the following vowels so they could only be considered vocalic. This contributed to a higher percentage of vocalic portions in semi-spontaneous speech. On the other hand, results indicate that Mandarin speakers spoke faster in semi-spontaneous speech than read speech. Benton et al. [2] also showed that in Mandarin, genre (news broadcast vs interview) indeed gave significantly different values for various rhythmic measures, which parallels the stylistic difference found in this study. So far, most studies on speech rhythm use only one speaking style, either read speech or spontaneous speech. More studies comparing speaking styles are needed in order to further explore the relationship between speech styles and rhythm.

The data of Cantonese English and Mandarin English poses a challenge to the acoustic measures. The two English accents sound syllable-timed. VarcoC and %V show that they are closer to syllable-timed than to stress-timed languages, but ΔC, nPVI_V and rPVI_C values all suggest that they are closer to stress-timed languages. The parameters of ΔC, ΔS, nPVI_V, rPVI_C and rPVI_S can only categorise languages according to the auditory impression of speech rhythm classes if the data of the two English accents was excluded. This situation highlights the issue of using acoustic measures to determine speech rhythm of non-native speakers.

Results of the averaged syllable durations suggest that Cantonese and Mandarin speakers employed a slower speaking rate when they read in English, which is a common phenomenon of second language speakers. As a result, many of their syllables would be lengthened compared to the speech of native English speakers. Careful listening to the accented-English speech samples reveals that such lengthening is not simply due to final-lengthening because these lengthened syllables can occur in various positions within an utterance. As expected, difficult words were lengthened more than easy words, but simple words like 'North Wind' and 'Sun' could be lengthened too. Individual speakers differ in the degree of such selective lengthening, but they all reduced their speaking rate in English compared to their native language. These speakers, having a syllable-timed native language, did not reduce unstressed syllables like what native English speakers would do. Such selective lengthening contributes to a higher degree of pairwise variability and a larger standard deviation of various intervals, but in an opposite way compared to native English speakers (having many lengthened syllables vs having many reduced syllables). Impressionistically, the two English accents still sounds quite syllable-timed. A slower speaking rate and selective lengthening result in the discrepancy between listeners' impression and the conclusion based on some acoustic measures.

Normalisation procedures for speaking rate alone may not solve this problem because the higher variability in duration is contributed by both speaking rate and selective lengthening. Among the five normalised measures used in this study, only VarcoC shows evidence that the two English accents are grouped with syllable-timed languages. White & Mattys [12] also found some discrepancy between subjective impression of second language rhythm and the results based on acoustic rhythmic measures. Therefore, more studies on second language rhythm are needed in order to address this issue.

## 5. Conclusions

This study confirms the syllable-timing impression of Cantonese and Beijing Mandarin with acoustic rhythmic measures. Results show that Cantonese has more extreme rhythmic values than Mandarin, French and Italian because of the lack of lexical stress. A slower speaking rate and selective lengthening in Cantonese English and Mandarin English contribute to the discrepancy between subjective impression of their rhythm and the results based on rhythmic measures. The VarcoC and %V parameters give the best classification of speech rhythm in this study.

## 6. References


[1] Bauer, R.S.; Benedict, P.K. 1997. *Modern Cantonese Phonology*. New York: Mouton de Gruyter.

[2] Benton, M.; Dockendorf, L.; Jin, W.; Liu, Y.; Edmondson, J. 2007. The continuum of speech rhythm: computational testing of speech rhythm of large corpora from natural Chinese and English speech. *The 16th ICPhS*. Saarbrücken, 1269-1272.

[3] Chao, Y. R. 1968. *A Grammar of Spoken Chinese*. Berkeley: University of California Press.

[4] Dauer, R. M. 1983. Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics* 11, 51-62.

[5] Dellwo, V.; Aschenberner, B.; Dancovicova, J.; Wagner, P. 2004 The BonnTempo-Copus and Tools: A database for the combined study of speech rhythm and rate. *INTERSPEECH-2004 (ICSLP)*. Jeju Island, Korea, 777-780.

[6] Dellwo, V. 2006. Rhythm and Speech Rate: A Variation Coefficient for ΔC. In *Language and Language-Processing*, Karnowski, P.; Szigeti, I. (eds.). Frankfurt am Main: Peter Lang, 231-241.

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

[8] Jian, H. 2004. On the syllable timing in Taiwan English. *Speech Prosody 2004*, Nara, Japan, 247-250.

[9] Low, E. L.; Grabe, E.; Nolan, F. 2000. Quantitative characterisations of speech rhythm: syllable-timing in Singapore English. *Language and Speech* 43, 377-401.

[10] Ramus, F.; Nespor, M.; Mehler, J. 1999. Correlates of linguistic rhythm. *Cognition* 73, 265-292.

[11] Setter, J. (2006). Speech rhythm in world Englishes: the case of Hong Kong. *TESOL Quarterly* 40, 763-782.

[12] White, L.; Mattys, S. L. 2007. Calibrating rhythm: first language and second language studies. *Journal of Phonetics* 35, 501-522.