

# On the syllable-timing of Cantonese and Beijing Mandarin

Peggy Pik Ki Mok

Department of Linguistics and Modern Languages, The Chinese University of Hong Kong  
peggymok@cuhk.edu.hk

## Abstract

This study investigates the speech rhythm of Cantonese and Beijing Mandarin using some recently developed acoustic rhythmic measures. The two languages 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 and telling the story semi-spontaneously. Both raw and normalised rhythmic measures were calculated using vocalic, consonantal and syllabic durations ( $\Delta C$ ,  $\Delta V$ ,  $\Delta S$ , %V, VarcoC, VarcoV, VarcoS, rPVI\_C, rPVI\_S, nPVI\_V, nPVI\_S). Results confirm the syllable-timing impression of Cantonese and Beijing Mandarin, and suggest that Cantonese may have the most typical syllable-timed rhythm among the languages in this study, probably due to its lack of lexical stress. This study also shows that, in addition to consonantal and vocalic durations, syllable durations can potentially be useful in distinguishing speech rhythm.

## 1. Introduction

Speech researchers have traditionally classified languages into different rhythmic groups: syllable-timed, stress-timed and mora-timed [1, 17]. English and German are typical stress-timed languages; French and Italian are typical syllable-timed languages and Japanese is a typical mora-timed language. This rhythm class hypothesis was based on the notion of isochrony, i.e. there are units of equal or near-equal duration in the speech signal for such classification: syllables for syllable-timed languages, inter-stress intervals (feet) for stress-timed languages and mora for mora-timed languages. However, many experimental studies could not find concrete evidence for such isochronous units in the speech signal to support the rhythmic class hypothesis (see [8, 13, 17] for a review). For example, the syllable durations of syllable-timed languages are equally variable as stress-timed languages [16], while durations of inter-stress intervals in stress-timed languages are not more variable than in syllable-timed languages [8]. Beckman [3] and Laver [14] concluded the early attempts to find acoustic correlates of speech rhythm by suggesting that speech rhythm is merely perceptual, since no reliable evidence could be found for isochrony.

Nevertheless, despite the lack of isochronous units, Dauer [8] and Roach [16] pointed out that stress-timed languages and syllable-timed languages differ in several important phonological 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, rather than any isochronous unit, combine with one another to give the impression of stress-timing versus syllable-timing. In addition, contrary to the early assumption of categorical

distinction of speech rhythm, they suggested that languages can be more or less stress-timed or syllable-timed, with a continuum between the two.

The above insights are captured by several recently developed acoustic measures of speech rhythm which could reflect the auditory impression of different rhythmic classes: %V (percentage of vocalic durations in speech),  $\Delta C$ ,  $\Delta V$  (standard deviations of consonantal and vocalic durations respectively) by Ramus *et al.* [15] and Pairwise Variability Index (PVI) of vocalic and consonantal durations by Grabe & Low [13]. These measures depart from the search of isochronous phonological units; instead, they consider the variability in speech. They take only the duration of vowels and consonants as the basis for rhythmic classifications. Due to the various phonological differences mentioned above, stress-timed languages would have higher variability of consonant and vowel durations than syllable-timed languages. Their results show that %V and  $\Delta C$  (Figure 1), the normalised vocalic PVI and the raw consonantal PVI (Figure 2) can categorise different languages into distinct rhythmic clusters, while languages having less typical or unknown rhythm may fall between these clusters. Subsequent studies also confirm that these acoustic measures can be used to distinguish languages with different speech rhythm, e.g. [18].

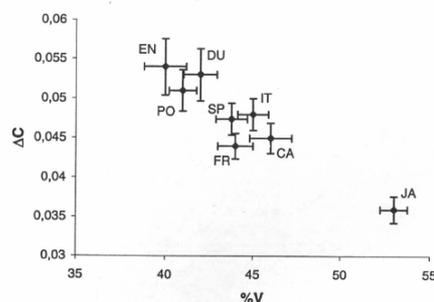


Figure 1: Results from Ramus *et al.* [13].

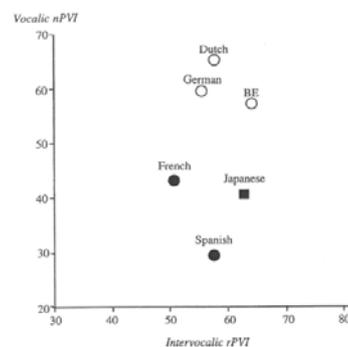


Figure 2: Results from Grabe & Low [11].

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 [2]. Impressionistically, Cantonese is a typical syllable-timed language, but so far no experimental study has examined its rhythm using acoustic measures. Moreover, syllable-timed languages like Spanish and Italian have regular lexical stress while Cantonese does not. It is also unclear how Cantonese speech rhythm compares with other syllable-timed languages.

The speech rhythm of Beijing Mandarin is less clear. Mandarin is similar to Cantonese in that it also has lexical tones and a very simple syllable structure. Impressionistically, it sounds syllable-timed. 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 [5, 7]. The frequent occurrence of unstressed syllables is characteristic of stress-timing. Cao [6], by measuring syllables and feet durations, concluded that there is no evidence to regard Beijing Mandarin as syllable-timed because no isochronous units could be found. However, since the notion of isochrony was shown to be inadequate for speech rhythm classification, her conclusion should be considered tentative.

Other studies have used acoustic rhythm measures to investigate Mandarin speech rhythm. Grabe & Low [13] 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 rhythm of these two Mandarin accents. Benton *et al.* [4] 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. Also, they did not state explicitly whether Mandarin is a syllable-timed language, although the text implied that it is. In addition, given the continuum between stress- and syllable-timing, there can be statistically significant variation among languages belonging to the same rhythm class [9, 11]. Therefore, comparison with more languages using rhythmic measures is necessary in order to investigate the speech rhythm of Beijing Mandarin.

Besides investigating the speech rhythm of Cantonese and Beijing Mandarin using the above-mentioned acoustic rhythmic measures, this study also compares the two languages with four languages in the BonnTempo Corpus [11]: German and English (stress-timed), French and Italian (syllable-timed) for a clearer picture of Cantonese and Mandarin speech rhythm.

## 2. Method

### 2.1. Speakers

Six native Hong Kong Cantonese speakers and six native Beijing Mandarin speakers (three male, three female) were recorded. 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.

These speakers were compared with previously published data (the BonnTempo corpus, see [11]). 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. Only data for reading in a normal speech rate and telling the story semi-spontaneously is reported in this paper. Analysis of the data with different speech rates is underway.

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 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 [13] 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.

### 2.4. Calculation of rhythmic measures

Durations (ms) of syllabic, consonantal and vocalic intervals were extracted using a Praat script. Altogether eleven rhythmic measures were calculated for each sentence by each speaker, which were then averaged for each speaker. Details of the measures are as follows:

- $\Delta C$ : the standard deviation of consonantal durations
- $\Delta V$ : the standard deviation of vocalic durations
- $\Delta S$ : the standard deviation of syllabic durations
- %V: the proportion of vocalic durations within a sentence

Since  $\Delta C$  and  $\Delta V$  have repeatedly been demonstrated to interact with the average segment duration, a normalisation procedure was applied by calculating the coefficient of variation [9].

- VarcoC:  $(\Delta C / \text{mean consonantal duration}) \times 100$
- VarcoV:  $(\Delta V / \text{mean vocalic duration}) \times 100$
- VarcoS:  $(\Delta S / \text{mean syllabic duration}) \times 100$

In addition, two sets of PVI values, raw (1) and normalised (2), were calculated using the following two formulas from [13]. These indexes express the level of variability in successive intervals. Raw PVI, taking the absolute difference in duration between each pair of successive units, was calculated for consonantal (rPVI\_C) and syllabic (rPVI\_S) durations. Normalised PVI uses the mean duration of each pair of successive units to normalise for speech rate variations. Normalised PVI was calculated for vocalic (nPVI\_V) and syllabic (nPVI\_S) durations.

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

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

(where  $m$  = number of items;  $d$  = duration of the  $k$ th interval)

### 3. Results

#### 3.1. %V, $\Delta C$ , VarcoC, nPVI\_V and rPVI\_C

Following [15], [9] and [13], Figure 1 to 3 show  $\Delta C$  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.

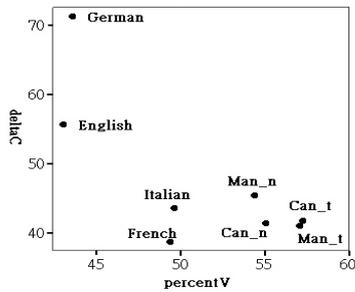


Figure 1:  $\Delta C$  and %V of all the languages.

It can be seen in Figures 1 and 2 that  $\Delta C$  and %V seem to give a clearer separation of stress-timed languages and syllable-timed languages than VarcoC and %V, although both sets of measures show distinct clusters of languages. In both cases, Cantonese and Mandarin pattern with syllable-timed Italian and French more than stress-timed German and English. The %V values of Cantonese and Mandarin are higher than Italian and French, suggesting that the two languages may be even more syllable-timed 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. The %V values of semi-spontaneous speech of Cantonese and Mandarin (\_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 [ $t(5) = -3.754, p = 0.013$ ]. In addition, Mandarin has a higher VarcoC value than Cantonese for both versions, but independent-samples t-tests show that this difference is insignificant (\_n: [ $t(10) = -0.913, p = 0.383$ ]; \_t: [ $t(10) = -1.175, p = 0.267$ ]).

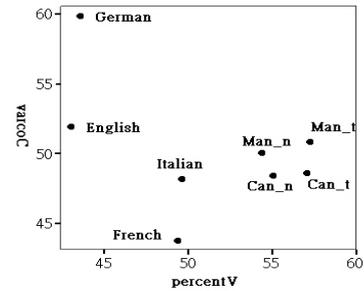


Figure 2: VarcoC and %V of all the languages.

The nPVI\_V and rPVI\_C parameters differentiate syllable- and stress-timed less clearly than  $\Delta C$  and %V. In Figure 3, the stress-timed German is indistinguishable from syllable-timed languages in the nPVI\_V parameter. Cantonese and Mandarin again pattern with syllable-timed Italian and French. 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 also do not have more extreme values than Italian and French.

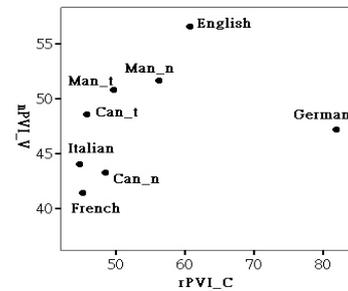


Figure 3: nPVI\_V and rPVI\_C of all the languages.

#### 3.2. Indexes of syllable durations

In addition to calculating various indexes for consonantal and vocalic durations, this study also calculates such indexes for syllable durations. Tables 1 and 2 show the values of  $\Delta S$ , VarcoS, rPVI\_S and nPVI\_S of all the languages in this study in descending order. It is interesting to note that 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  $\Delta S$ . 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 Beijing Mandarin. Cantonese is ranked the lowest by three out of the four measures, indicating that Cantonese may sound even more syllable-timed than Italian and French, echoing the results of %V.

Table 1:  $\Delta S$  and VarcoS of all the languages.

| Language | $\Delta S$ | Language | VarcoS |
|----------|------------|----------|--------|
| English  | 88.74      | English  | 51.87  |
| 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 |
|----------|--------|----------|--------|
| English  | 115.50 | English  | 69.67  |
| 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

The main focus of this study is to investigate the speech rhythm of Cantonese and Beijing Mandarin. All rhythmic measures confirm the syllable-timing impression of the two languages. The results also suggest that Cantonese may have an even stronger syllable-timed rhythm 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 [13] 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 the perception of syllable-timing. Results from the present study and [13] suggest that the presence of lexical stress not only contributes to the distinction between stress-timing and syllable-timing, but can also affect the degree of syllable-timing.

Although both Cantonese and Beijing Mandarin have a syllable-timed rhythm, the rhythm of natural Mandarin speech sounds more variable than that of Cantonese impressionistically. The results of syllable durations in both read speech and semi-spontaneous speech confirm this impression of the two languages. Further analysis using naturally occurring speech materials in both languages with acoustic rhythmic measures is underway for a more thorough investigation of their speech rhythm. In addition, since Ramus *et al.* [15] showed that listeners can distinguish speech rhythm

by listening to highly reduced synthesized speech (flat ‘sasasa’ speech), it will be interesting to investigate whether listeners can distinguish Cantonese and Mandarin speech rhythm perceptually using highly reduced synthesized speech based on naturally occurring speech data.

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, Benton *et al.* [4] 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 acoustic rhythmic measures were developed to capture the variability of consonantal and vocalic intervals in speech. These two kinds of intervals were used partly because of the failure of early attempts to find isochronous phonological units in the speech signal, and also partly because infants are able to distinguish languages based on their speech rhythm, see [15] for discussion. If infants who have no prior knowledge of a language’s phonological structure can distinguish languages with different speech rhythm, they must be attending to some basic acoustic properties in the speech signal. Dellwo *et al.* [10] also showed that stress- and syllable-timed languages can be distinguished on the basis of voiced and voiceless intervals alone. Nevertheless, the present study shows that the variability of phonological syllable durations can also potentially distinguish stress- versus syllable-time languages, echoing [12]. Of course, further studies with more languages are needed in order to verify this claim, as well as to investigate which rhythmic measures work best with syllable durations.

#### 5. Conclusions

This study confirms the syllable-timing impression of Cantonese and Beijing Mandarin with acoustic rhythmic measures. Results show that Cantonese may have an even stronger syllable-timed rhythm than Mandarin, French and Italian, probably due to its lack of lexical stress. In addition to consonantal and vocalic durations, this study also demonstrates that syllable durations can potentially be used to distinguish languages with different speech rhythm.

#### ACKNOWLEDGEMENTS

The author would like to thank Volker Dellwo for helpful discussion and the generous sharing of his data. Thanks also go to Thomas Tsoi for segmentation of the sound files.

## 6. References

- [1] Abercrombie, D. 1967. *Elements of General Phonetics*. Edinburgh: Edinburgh University Press
- [2] Bauer, R. S.; Benedict, P.K. 1997. *Modern Cantonese Phonology*. New York: Mouton de Gruyter.
- [3] Beckman, M. E. 1992. Evidence for speech rhythms across languages. In *Speech Perception, Production and Linguistic Structure*, Y. Tohura, E. Vatikiotis-Bateson & Y. Sagisaka (eds.). Tokyo: IOS Press, 457-463.
- [4] 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.
- [5] Cao, J. F. 1986. An analysis of Mandarin neutral tone syllables [in Chinese]. *Applied Acoustics* [in Chinese], 4, 3-8.
- [6] Cao, J. F. 2000. Rhythm of spoken Chinese — linguistic and paralinguistic evidences. *The 6th ICSLP*. Beijing, 2000, 357-360.
- [7] Chao, Y. R. 1968. *A Grammar of Spoken Chinese*. Berkeley: University of California Press.
- [8] Dauer, R. M. 1983. Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics* 11, 51-62.
- [9] Dellwo, V. 2006. Rhythm and Speech Rate: A Variation Coefficient for  $\Delta C$ . In *Language and Language-Processing*, Karnowski, P.; Sziget, I. (eds.). Frankfurt am Main: Peter Lang, 231-241.
- [10] Dellwo, V.; Fourcin, A.; Abberton, E. 2007. Rhythmical classification of languages based on voice parameters. *The 16th International Congress of Phonetic Sciences (ICPhS)*, Saarbrücken, Germany, 1129-1132.
- [11] 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.
- [12] Deterding, D. 2001. The measurement of rhythm: a comparison of Singapore and British English. *Journal of Phonetics*, 29, 217-230.
- [13] 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.
- [14] Laver, J. 1994. *Principles of Phonetics*. Cambridge: Cambridge University Press
- [15] Ramus, F.; Nespors, M.; Mehler, J. 1999. Correlates of linguistic rhythm. *Cognition* 73, 265-292.
- [16] Roach, P. 1983. On the distinction between stress-timed languages and syllable-timed languages. In *Linguistic Controversies: Essays in Honour of F.R. Palmer*, D. Crystal (ed.). London: Arnold.
- [17] Warner, N.; Arai, T. 2001. Japanese mora-timing: a review. *Phonetica*, 58, 1-25
- [18] White, L.; Mattys, S. L. 2007. Calibrating rhythm: first language and second language studies. *Journal of Phonetics* 35, 501-522.