MEASURES OF SPEECH RHYTHM IN EAST-ASIAN TONAL LANGUAGES

Antonio Romano[a], *Paolo Mairano*[b] *& Lidia Calabrò*[c]

[a]LFSAG - University of Torino, Italy; [b]GIPSA-Lab - Université Grenoble3, France;
[c]Department of Western Languages - Chulalongkorn University, Bangkok, Thailand
antonio.romano@unito.it; paolomairano@gmail.it; lidia.g@chula.ac.th

ABSTRACT

The most frequently used variables for language rhythm categorisation have been measured for some tonal languages. Cantonese, Mandarin, Thai and Vietnamese are languages which ground their prosodic properties on phonological tones and are supposed to reserve secondary role to stress whose functions are controversially bounded to prominence patterns of constituency. Therefore, a distinctive characterisation in terms of speech rhythm is not usually acknowledged for them.

In spite of these phonological expectancies, our results reflect a clear-cut distribution of these languages along timing categories corresponding to languages traditionally considered as stress-timed or syllable-timed (e.g. English vs. Spanish). The differences shown by the speech samples we analysed seem to be due to phenomena related to specific syllable compression patterns and vowel reduction / lengthening strategies affected by tone and more general prosodic rules.

Keywords: rhythm class hypothesis, tonal languages, stress, vowel length

1. INTRODUCTION

Recent publications (e.g. [6, 13]) have given a new impulse to the experimental research on linguistic rhythm by proposing new typological indexes, the so-called rhythmic metrics, in order to account for the *Rhythm Class Hypothesis* (*RCH*, [4]). Several studies have followed, mainly aiming at classifying different languages along the axes of diagrams based on duration measures (for a critical view see [5]). The proposed metrics can be calculated also for tonal languages, whose syllable properties and prominence patterns are not described in traditional phonological accounts (see [11] for a discussion) and are not easy to detect. This raises relevant questions on how to relate the *RCH* to principles of unrhythmicity, ambi-rhythmicity or hetero-rhythmicity, or – vice versa – on how to account for emergent rhythmic properties to more general structures governed by other prosodic variables.

We shall concentrate on results obtained for samples of four languages – Cantonese, Mandarin, Thai and Vietnamese – whose rhythmic properties, together with the ones of other tonal languages, are controversially analysed by the various authors who investigated them [9, 11, 12] or who analysed Asian accents of English (e.g. [10]).

2. METHODOLOGY

2.1. Data and pre-processing

Data consisted in 8 productions by 2 native speakers per language (mainly university students) reading translations of *The North Wind and the Sun* (texts of Cantonese, Mandarin and Thai were taken from [7, 8, 15], the Vietnamese translation was original). Speech samples have been recorded in good quality fieldwork conditions (details in Figure 2), except for one of the two Thai speakers (taken from [15]). They are fairly homogeneous in terms of tempo and are similar to the ones already used in [9].

Two operators separately analysed each sample and segmented it in CV intervals with *Praat* following specific phonetic criteria [14]. Some general prosodic properties and fluency measures are illustrated in Table 1.

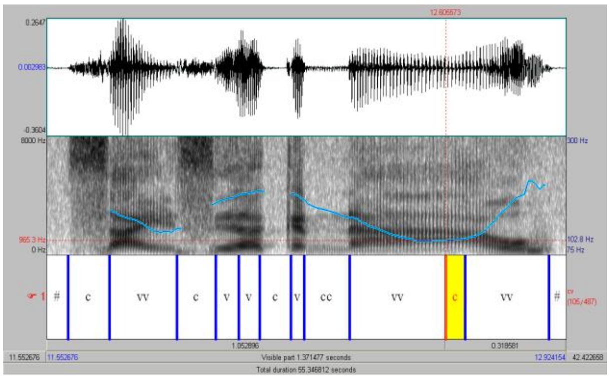
Table 1: General prosodic properties of data (each pair of speakers is separated by a backslash).

	Cantonese	Mandarin	Thai	Vietnamese
No. of #	24 / 24	24 / 22	37 / 26	13 / 16
No. V int.	150 / 145	157 / 153	191 / 187	177 / 175
%V	49,1 / 48,5	51,2 / 51,2	48,4 / 45,7	40,1 / 40,2
Syll. rate	4,15 / 4,24	5,14 / 5,41	4,07 / 4,02	4,53 / 4,41

Differences among samples of the same language are generally marginal, except for the two Thai speakers (one of them made more pauses yet maintaining a similar tempo). Speech rate for Mandarin speakers was slightly faster than for the others, whereas Vietnamese speakers produced fewer pauses.

Special care was used in labelling hyatuses as well as vowel-like realisation of consonants. In particular, we carefully evaluated the emergence of consonantal approximant phases in the realisation of double vowels and extra-long Thai diphthongs in pre-pausal conditions (unless resulting in nuclei with lengths of more than 500 ms, see an example in Figure 1).

Figure 1: Labelling as C of the temporal reduction in the amplitude and energy contour occurring within extra-long Thai final diphthong –าว /aːʊ/ → [aːʕaʊ].



The same criterion was applied, however, to the other languages used as a reference, accounting for specific lengthening conditions. Thai vowel and diphthong lengthening has been verified for four different speakers who all seem to double long vowels according to prosodic grouping (within tone and intonation units, see [1] against e.g. Mandarin which has no phonologically long vowels but has similar diphthongs whose length is much more controlled in this position.

Nevertheless, differences persist between individual assessments and this is the reason why we aimed to spot variance analysis in the charts summarising our measurements.

2.2. Measurements and calculations

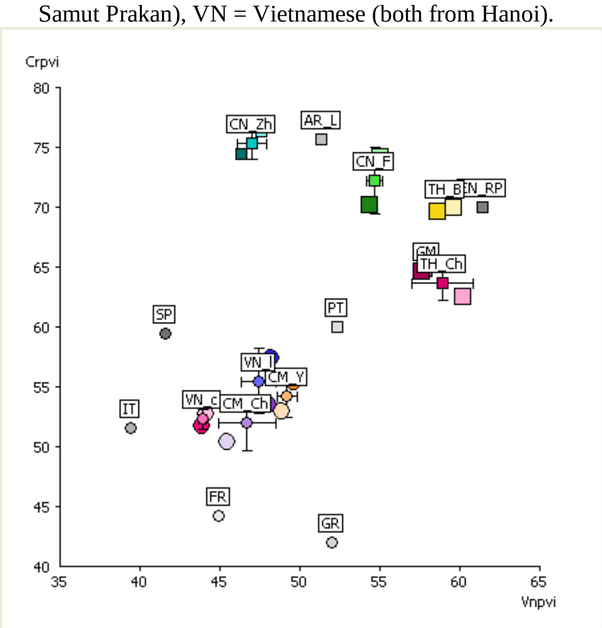
Even though intervals in our data are homogeneous in terms of speech rate, the samples have been processed following the *Delta* and *PVI* principles (see [6, 13]). Values were computed with *Correlatore*, a programme specifically developed for computing rhythm metrics which also outputs a graphic representation on bi-dimensional charts and allows variance analysis (for details see www.lfsag.unito.it/correlatore/index_en.html).

3. RESULTS

3.1. The *PVIs*

Figure 2 shows *nPVI* and *rPVI* values obtained for each speech sample analysed (big squares and circles represent the results obtained by each segmentator, whereas small squares and circles with standard deviation bars represent the mean). Results are compared to those found by [14] for a selection of syllable- and stress- timed languages.

Figure 2: *PVI* chart for our data (*IPA* data marked with *; see [14] for details). AR = Arabic (Lebanese), CM = Mandarin (Chaoyang, Liaoning, & Yuncheng, Shanxi), CN = Cantonese (Foshan & Zhuhai, Guangdong), GM = German, EN = English*, FR = French*, GR = Greek*, IT = Italian, PT = Portuguese*, SP = Spanish*, TH = Thai (Bangkok* & Samut Prakan), VN = Vietnamese (both from Hanoi).



On the one hand, this chart clearly shows that Mandarin Chinese and Vietnamese samples cluster with Spanish, Italian, French & Greek which are supposedly syllable-timed languages. All samples of these languages are positioned within the alleged syllable-timed area, with values in agreement with the ones already available from other studies. In particular *nPVI* for Mandarin spans from 45.4 to 49.6, overlapping with the ones reported by [9] (which are much higher than the ones in [6]). Mandarin resulted in *rPVI* values spanning from 50.4 to 55.5, in agreement with [6] and [9]. Furthermore, measures of the *Deltas* are consistent to the ones in [12], confirming that Mandarin tends towards syllable-timing.

Even though our samples of Vietnamese sound quite different from those of Mandarin Chinese from the point of view of rhythm, they cluster together: Vietnamese is characterised by a more complex phonotactics which does not seem to be reflected by *rPVI* values. Moreover, it shows frequent coda glottalisation which determines a significant number of checked syllables with a fairly stable vowel duration: this accounts for *nPVI* values reflecting a comparable degree of variability to our Mandarin samples.

Surprisingly, the realisation of prominence hierarchies accounting for the internal organisation of poly-morphemic words is not associated with significant phenomena of vowel reduction or lengthening.

On the other hand, we found Cantonese and Thai in the upper right corner of the chart, clustering with languages like English, German, Arabic and, to a lesser extent, Portuguese. The first two reference languages are well-known examples of stress-timing, whereas the other two are variously described in the literature. Arabic, for instance, has no evidence of stress at all and yet it has been reported to cluster with stress-timed languages (at least some varieties, see [4]).

The same applies to Cantonese (*pace* [12]). Not only do we found a low *%V* for this language, but all our samples present high values of *DeltaC* and *rPVI* (even though with a limited number of consonant cluster types, which means that syllables can be easily stretched and compressed depending on tone type). Finally, vowel reduction seems not to determine higher values of *nPVI*, probably because it has no counterpart in vowel lengthening of more prominent syllables (vowel length is phonologically controlled).

Results on Thai reflect our auditory impressions as a stress-timed language. In this case, a well defined function of stress (as well as tonal features) has been acknowledged by various authors (e.g. [15]). More specifically, acoustic correlates of speech rhythm in Thai are studied by [6] with the same method: both *nPVI* and *rPVI* values found for this language overlap with those obtained by us, thus confirming the positioning of Thai among stress-timed languages.

3.2. Other figures

The number of single intervocalic consonants (C) analysed is almost the same for all the languages in the sample (with mean lengths slightly longer for Vietnamese). Clusters of two C are less frequent in Mandarin and Cantonese (less than 29% of C intervals; in Vietnamese and Thai the number of C clusters represents more than 44%). As it can be observed in Table 2, Cantonese and Thai preserve the mean longest CC int. and standard deviation in compliance with higher *rPVI* (see Figure 1). Only Vietnamese seems to compress them to one and a half the length of a mean single C int. (note that 5-7% of C int. in Vietnamese and Thai have more than two C). Vowels (V) duration values are also very interesting. The percentages of V clusters (mainly depending on the frequency of diphthongs) are in fact almost the same for the four language samples, but their mean lengths vary significantly for single V int. of Cantonese and VV int. for Thai. Vietnamese has the shortest (and quite stable) double nuclei, whereas Thai has the longest ones which cause a high degree of variability (cp. *nPVI* values in Figure 2).

Table 2: Mean values (μ) and standard deviation (σ) in *ms* for vowel and consonant clusters for the four languages (CM = Mandarin, CN = Cantonese, VN = Vietnamese, TH = Thai; *n* = number of items, % on the whole number of intervals).

		CM	**CN**	**VN**	**TH**
C	*n*	*114 (72%)*	*102 (67%)*	*89 (49%)*	*104 (53%)*
	μ	**73**	**85**	**99**	**87**
	σ	32	34	30	44
CC	*n*	*42 (26%)*	*44 (29%)*	*84 (46%)*	*85 (44%)*
	μ	**136**	**189**	**151**	**168**
	σ	43	61	33	41
V	*n*	*115 (74%)*	*100 (67%)*	*120 (68%)*	*122 (65%)*
	μ	**80**	**100**	**77**	**81**
	σ	33	47	34	28
VV	*n*	*39 (25%)*	*48 (33%)*	*56 (32%)*	*65 (34%)*
	μ	**147**	**152**	**118**	**174**
	σ	42	51	42	47

Summing up, increasing the number of C, Cantonese shows a relevant growing of mean C int. duration (resulting in higher *rPVI* values), whereas higher *nPVI* values in Thai may depend on a more remarkable variation in the duration of VV intervals.

The reasons why Thai and Cantonese seem to behave as stress-timed languages are therefore perhaps not related to stress (and rhythm) properties but to distinct timing constraints at a segmental (or syllabic) level.

In the light of these considerations, as pointed out by [2], we claim that assessments on the *RCH* should not omit to account for the number of segments included in the intervals and their specific timing conditions. Measurements failing to consider the internal (segmental) organisation of C and V int. miss to account for compression effects at different levels which are responsible for potentially distinct categorisations.

4. CONCLUSION

In this paper we applied some recently developed rhythm measures to four tonal languages.

The methods currently used to classify languages by their rhythm are based on durational measures and aim at testing the *RCH* (which was conceived on the basis of auditory impressions of samples of Western languages).

Labels such as “syllable-timed” or “stress-timed” are traditionally used to refer to (a) languages whose timing seems to be regulated by segmental time patterns depending on syllabic constraints and to (b) languages whose timing seems to be dominated by stress patterns.

Since stress has been attributed little weight by the literature on tonal languages, isochrony at syllable level has often been alleged for them. Still, stress has been noticed by some authors to have a role in the hierarchical organisation of constituents [11, 15].

We included 4 distinct tonal languages by applying the *Deltas* and *PVIs* to 8 speech samples fairly similar in terms of speech rate. Results have been compared with data of Western languages published in previous studies [14].

Mandarin and Vietnamese cluster among syllable-timed languages, whereas Cantonese and Thai are positioned among stress-timed languages. Stress and inter-syllabic compression patterns are supposed to be at the origin of this completely different trend, but also a quantitative analysis of the internal (segmental) structure of C and V int. (like the one carried out on our samples; see Table 2) could help explaining these phenomena.

Cantonese and Thai positioning could be imputed to compression or expansion rules governing inter-stress distance in these two languages (something which still has to be demonstrated). Yet, it remains to be explained why the different auditory impressions of samples of these two languages are not accounted for by metrics.

Our results reflect the general properties observable in these languages, but obviously depend on the particular selection and arrangement of data analysed. We propose, therefore, to test rhythm metrics on a larger corpus of tonal languages and, as already suggested in [14], to enlarge the reference framework to a multi-layer model [2].

5. REFERENCES

[1] Abramson, A.S. 2001. The stability of distinctive vowel length in Thai. In Tingsabadhm, M.R.K., Abramson, A.S. (eds.), *Essays in Tai Linguistics*, Bangkok: Chulalongkorn University Press, 13-26.

[2] Bertinetto, P.M., Bertini, C. 2010. Towards a unified predictive model of natural language rhythm. In Russo, M. (ed.), *Prosodic Universals. Comparative Studies in Rhythmic Modeling and Rhythm Typology*. Rome: Aracne, 43-78.

[3] Dellwo, V., Wagner, P. 2003. Relations between language rhythm and speech rate. *Proc. 15th ICPhS* Barcelona, 471-474.

[4] Ghazali, S., Hamdi, R., Barkat, M. 2002. Speech rhythm variation in Arabic dialects. *Proc. of Speech Prosody* Aix-en-Provence, France, 331-334.

[5] Gibbon, D., Gut, U. 2001. Measuring speech rhythm. *Proc. of Eurospeech 2001* Aalborg, Denmark, 95-98.

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

[7] IPA 1999. *Handbook of the IPA*. Cambridge: Cambridge Univ. Press.

[8] Lee, W.S., Zee, E. 2003. Chinese, Standard (Beijing). *Journal of the IPA* 33(1), 109-112.

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

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

[11] Lu, B., Duanmu, S. 2002. Rhythm and syntax in Chinese: A case study. *Journal of Chinese Language Teachers Association* 37(2), 123-136.

[12] Mok, P.P.K., Dellwo, V. 2008. Comparing native and non-native speech rhythm using acoustic rhythmic measures: Cantonese, Beijing Mandarin and English. *Proc. of Speech Prosody* Campinas, Brazil, 63-66.

[13] Ramus, F., Nespor, M., Mehler, J. 1999. Correlates of linguistic rhythm in the speech signal. *Cognition* 73(3), 265-292.

[14] Romano A., Mairano P. 2010. Speech rhythm measuring and modelling: pointing out multi-layer and multi-parameter assessments. In Russo, M. (ed.), *Prosodic Universals: Comparative Studies in Rhythmic Modeling and Rhythm Typology*. Roma: Aracne, 79-116.

[15] Tingsabadh, M.R.K., Abramson, A.S. 1999. Thai. *Handbook of the IPA*. 147-150.