Rhythm metrics on syllables and feet do not work as expected

Rhythm metrics on syllables and feet do not work as expected

Paolo Mairano[1], *Antonio Romano*[2]

[1] GIPSA-Lab, Université Stendhal Grenoble 3, France
[2] LFSAG, Dip. Scienze del Linguaggio, Università di Torino, Italy
paolo.mairano@gipsa-lab.grenoble-inp.fr, antonio.romano@unito.it

Abstract

The aim of this paper is to explore the possibility of using rhythm metrics on the traditional units of speech rhythm (the syllable and the foot), instead of applying them to consonantal and vocalic intervals. Despite [14] had already proven that the standard deviation of syllables and feet did not provide a satisfactory representation of the traditional rhythm classes, some recent studies obtained encouraging results. In particular, [2] applied the PVI to English and Estonian syllables and feet, and a similar approach is intrinsic in the YARD index (cf. [15]) though only at syllable level.

We computed the deltas and the PVIs on syllables (measured as the distance between two successive vocalic onsets) and feet (measured as the distance between the onsets of two stressed vowels) for 30 samples of 14 languages.

The results do not confirm expectations and do not seem to support the use of these units for the study of speech rhythm in these terms.

Index Terms: speech rhythm, rhythm metrics, rhythm correlates, syllables, feet.

1. Introduction

This paper deals with rhythm metrics, that is to say with durational measures that have been devised in order to account for the impression that the speech rhythm is either based on stresses or on syllables. A rhythm categorization of languages was mainly put forward by [1], who claimed that stress-timed languages (such as English, Russian and Arabic) present isochrony at the foot level, while syllable-timed languages (such as French, Telugu and Yoruba) present isochrony at the syllable level. These theories were contradicted by experimental evidence published in many studies. [14] calculated the standard deviation of syllable and foot durations in the 6 languages quoted by [1] and found no evidence of a tendency to equalize stresses in stress-timed languages, nor to equalize syllables in syllable-timed languages.

Other authors (namely [6] and [3]) proposed (a) that the rhythm of languages should be characterized in terms of a continuum rather than into 2 (or 3 if we consider mora-timing) discrete classes and (b) that structural properties of the languages (mainly vowel reduction and syllable structure) should play a role in speech rhythm. These theories have provided the roots for the authors who, subsequently, proposed the so-called rhythm metrics.

Rhythm metrics such as the deltas (see [13]), the varcos (see [7]) and the PVIs (see [9]) were conceived as acoustic correlates of speech rhythm. They are meant to provide a representation of the degree of the durational variability of consonantal and vocalic intervals, which is considered to be higher for languages allowing for a complex syllable structure and for vowel reduction (i.e. for stress-timed languages). These measures have been proven to provide a scalar representation of languages on a rhythm continuum. Some authors (e.g. [8]) have also applied them to voiced and voiceless intervals, with results that are comparable to the ones obtained on vocalic and consonantal durations.

More recently, despite the failure of [14], other authors tried to apply rhythm metrics to the units which are at the base of the traditional syllable-timed vs. stress-timed dichotomy, i.e. the syllable and the foot. In particular, [2] applied the PVIs to foot and syllable durations for English (supposedly stress-timed) and Estonian (supposedly syllable-timed). They found that English presents high variability at the syllable-level and low variability at the stress level, while Estonian presents low variability at both levels. They concluded that Estonian tends to both stress-timing and syllable-timing.

Likewise, the YARD index proposed by [15] is a modification of the rPVI formula that is applied to z-transformed syllable durations. The authors tested this index on English, French, German and Italian samples obtaining promising results.

2. The experiment

An experiment was carried out in order to test the application of rhythm metrics on the traditional units for the study of speech rhythm (the syllable and the foot) on a wider sample of languages.

2.1. The data

Data consisted of translations of *The North Wind and the Sun* for 14 languages read by 30 speakers in total. Samples are heterogeneous as for reading style, given that they come from different sources (see below). We are aware of the limits of such an approach, but we privileged the need to have comparable samples for a relatively high number of languages. We shall specify more in detail the data included. Unmarked items were recorded at our laboratory in a sound-proof booth, while items marked with an asterisk were taken from the *Illustrations of the IPA* (either from [10] or from various articles published in the Journal of the IPA – a complete reference is not provided for obvious reasons, however the official list is provided at the following internet address: http://www.sil.org/~olsonk/ipa.html):

- 1 Arabic speaker (from Lebanon)
- 1 Czech speaker (standard)
- 1* Dutch speaker (standard)
- 5 speakers of English varieties (RP*, GA*, AusE, NZE* and IndE)
- 1* Estonian speaker (standard)
- 2 Finnish speakers (standard)
- 2 French speakers (1* standard speaker and a Canadian speaker)
- 2 German speakers (standard)
- 1* Greek speaker (standard)
- 3+1* Italian speakers (standard)

- 1 Japanese speaker (standard)
- 1* Polish speaker (standard)
- 3 Portuguese speakers (1* of European Portuguese and 1+1* of Brazilian Portuguese)
- 5 Spanish speakers (1* of Castilian Spanish, 1 from Granada - Spain, 1 from Bogotá, 1 from Caracas and 1 from Lima).

2.2. The methodology

All samples have been labeled by the first author with *Praat*. Since we wanted to compare the results of the metrics calculated on consonantal and vocalic intervals with the results of the metrics calculated on syllables and feet, we had to devise a labeling method that allowed to retrieve the durations of consonantal and vocalic intervals, as well as of syllables and feet. The definition of syllables and feet being far from straightforward, we decided for sake of simplicity that syllable durations would be deduced by the CV segmentation (see below), while stresses would be marked on all prominent vowels with an asterisk, according to impressionistic criteria.

The criteria followed for the CV segmentation are not reported here as they can be consulted in our preceding publications, e.g. [11]. As for the segmentation of syllables and feet, we measured "syllable durations" as the distance between two successive vocalic onsets (from now on *inter-onset distance*). Similarly, feet were measured as the distance between the vocalic onsets of stressed vowels (from now on *inter-stress distance*).

Finally, the values of the metrics on consonantal and vocalic intervals were computed with a program specifically developed for this purpose at our laboratory (see [12]) and available online at the following address:

http://www.lfsag.unito.it/correlatore/index_en.html

Instead, the values of the metrics on inter-onset and inter-stress distances were computed with a *Praat* script specifically developed by the first author. All the results were put on similar charts for comparison.

2.3. The hypothesis

According to the traditional theories presented in 1, syllables should be of roughly equal durations in syllable-timed languages, whereas they are supposed to be more variable in stress-timed languages. Instead, stresses should occur at regular intervals in stress-timed languages (thus resulting in roughly regular feet) and at irregular intervals in syllable-timed languages (thus resulting in irregular feet).

Even though such a restrictive view has been proven to be faulty, we could still hypothesize durational variability at *syllable level* to be smaller for syllable-timed than for stress-timed languages: consequently, metrics applied to inter-onset durations are expected to yield to lower values for syllable-timed languages than for stress-timed languages. Conversely, we could hypothesize durational variability at *stress level* to be smaller for stress-timed languages than for syllable-timed languages: therefore, metrics applied to inter-stress durations are expected to yield to lower values for stress-timed languages.

2.4. The interest of such a representation

A representation based on inter-onset and inter-stress durations would provide an interesting alternative to vocalic and consonantal intervals. In effect, considering only vocalic and consonantal durations means sticking to the first (segmental) aspect of speech rhythm: a representation based on these parameters can only make a distinction between syllable-timed and non syllable-timed (but not necessarily stress-timed) languages, because nothing is measured at stress level.

The main advantage of this approach is precisely that it integrates both levels of rhythm, namely the syllable level and the stress/accent level.

3. The Results

3.1. Metrics calculated on CV durations

The values of the metrics computed on vocalic and consonantal durations can be seen in figure 1 (deltas above, PVIs below).

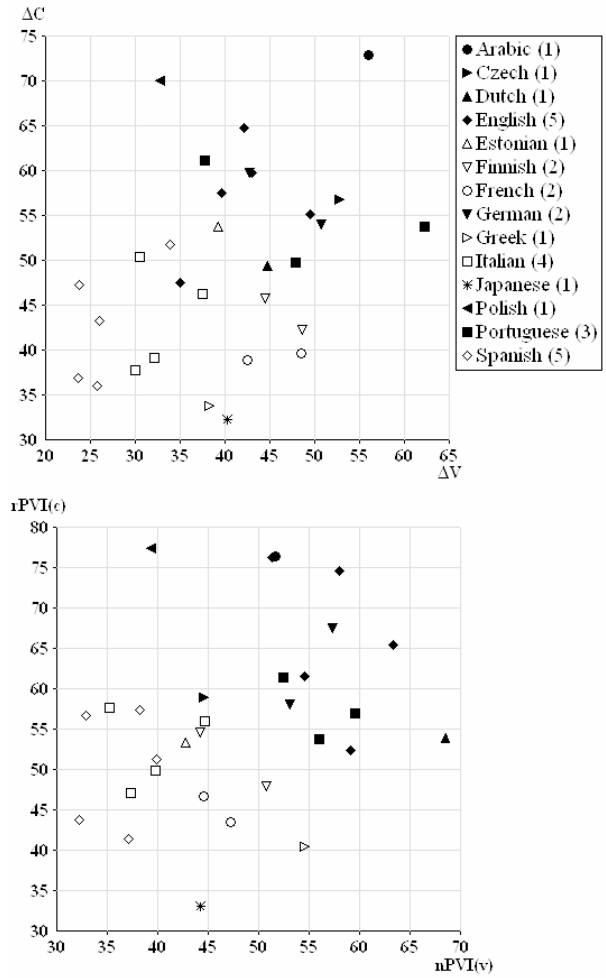

Figure 1: *Values of the deltas (above) and of the PVIs (below) calculated on consonantal (y-axis) and vocalic (x-axis) durations. Supposedly stress-timed languages are represented by black shapes, supposedly syllable-timed languages are represented by white shapes. Values are in ms.*

In compliance with expectations set by rhythm metrics' authors (see [13] for the deltas and [9] for the PVIs) and confirmed by many studies in the literature, stress-timed languages (black shapes in the chart) roughly occupy the north-eastern corner of the chart (high deltas and PVIs), whereas syllable-timed languages (white shapes in the chart) roughly occupy the south-western corner of the chart (low deltas and PVIs). This indicates a higher variability for both

consonantal and vocalic intervals for supposedly stress-timed languages than for syllable-timed languages. Japanese, the only supposedly mora-timed language in the data analyzed, exhibits very low values of variability and is collocated at the extremity of the syllable-timed area (cf. [9] for similar results).

3.2. Metrics calculated on syllables and feet

The values of the metrics computed on inter-onset and inter-stress distances can be seen in figure 2 (deltas above, PVIs below). It has to be remarked that the difference between the deltas and the PVIs is mainly that the latter take into account the temporal succession of segments by considering only successive values, while the former consider all possible pairs of values (a discussion of details of the formulae is beyond the scope of this paper, see for instance [12]).

As has been stated above, according to the traditional theories about stress-timing and syllable-timing, one would expect stress-timed languages (black shapes in the chart) to cluster in the south-western corner (low inter-stress variability and high inter-onset variability), and syllable-timed languages (white shapes in the chart) in the south-eastern corner (with opposite value of inter-stress and inter-syllable variability).

However, the results are far from reflecting expectations. First of all, the north-western and south-eastern corners are empty in both charts: instead, languages are placed in the north-eastern and south-western corners.

At first sight, one is tempted to say that a distinction between supposedly stress-timed and supposedly syllable-timed languages is however maintained, since English, Dutch, Arabic and Czech cluster in the north-eastern side, whereas the other languages occupy the south-western side. One could therefore hypothesize that syllable-timed languages are simply characterized by a higher degree of regularity at all levels, whereas stress-timed languages are characterized by a higher degree of variability at all levels (syllabic and accentual). This could also be put in relation to what was proposed by [3], who states that intra-syllabic and inter-syllabic compensation are facets of the same property and that it may perhaps be possible to distinguish between duration controlling and duration compensating languages. His claim is also based on the observation that it would be implausible for a language to exhibit flexibility at the syllable level and control at the foot level (or vice versa): "it seems much more sensible to imagine that both levels obey the same tendency" (p. 123). Yet, the German, Polish and Portuguese samples (supposedly stress-timed) all cluster together with syllable-timed languages, exhibiting low levels of variability at both levels. It is indeed difficult to account for the low values of Δsyll and pvi(syll), since these languages exhibit fairly high values of ΔC/ΔV and rPVI(c)/nPVI(v) (see figure 1). One possible explanation could suggest that short and reduced vowels occur in these languages when the following consonantal interval is complex and that, vice versa, long and fully articulated vowels occur when the following consonantal interval is simple. However, these are mere speculations and need to be tested on data.

Another noteworthy remark concerns the fact that samples of the same language cluster together or within a limited distance. This in contrast to the scenario offered by the ΔC/ΔV and rPVI(c)/nPVI(v) charts, where samples of the same languages can be scattered far from each other (e.g. the 3 Portuguese samples in figure 1), as is also proven by our previous results showing high inter-speaker variability (see [11, 12]).

Interestingly, in the above chart (displaying the results of the deltas) reported in figure 2, samples are roughly scattered along a line. This seems to suggest that the variability of inter-onset intervals is directly proportional to the variability of inter-stress intervals. This behavior is less pronounced (but still visible) in the chart below, showing PVI values: this may suggest that these results are heavily influenced by speech rate, so we decided to compute the varco and the nPVI which are, respectively, the normalized versions of the delta and the rPVI.

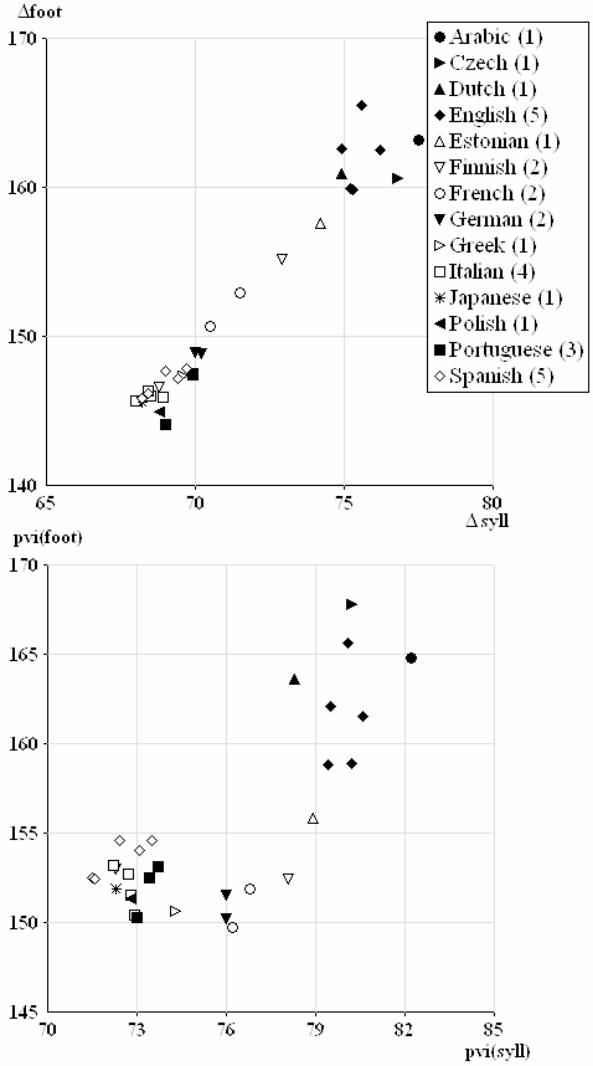

Figure 2: *Values of the deltas (above) and of the rPVI (below) calculated on inter-stress (y-axis) and inter-onset (x-axis) durations. Supposedly stress-timed languages are represented by black shapes, supposedly syllable-timed languages are represented by white shapes. Values are in ms.*

3.3. Normalized metrics

The values of normalized metrics computed on inter-onset and inter-stress distances can be seen in figure 3 (varcos above, nPVIs below). In both cases, the normalization is obtained by dividing durations (in ms) by the mean duration of syllables or stresses.

As it can be seen, the results are comparable to those presented in figure 2, still presenting a distinction between, on one side, English, Dutch, Arabic and Czech, and, on the other side, all other samples. Furthermore, in the chart above, data tend again to join along a line.

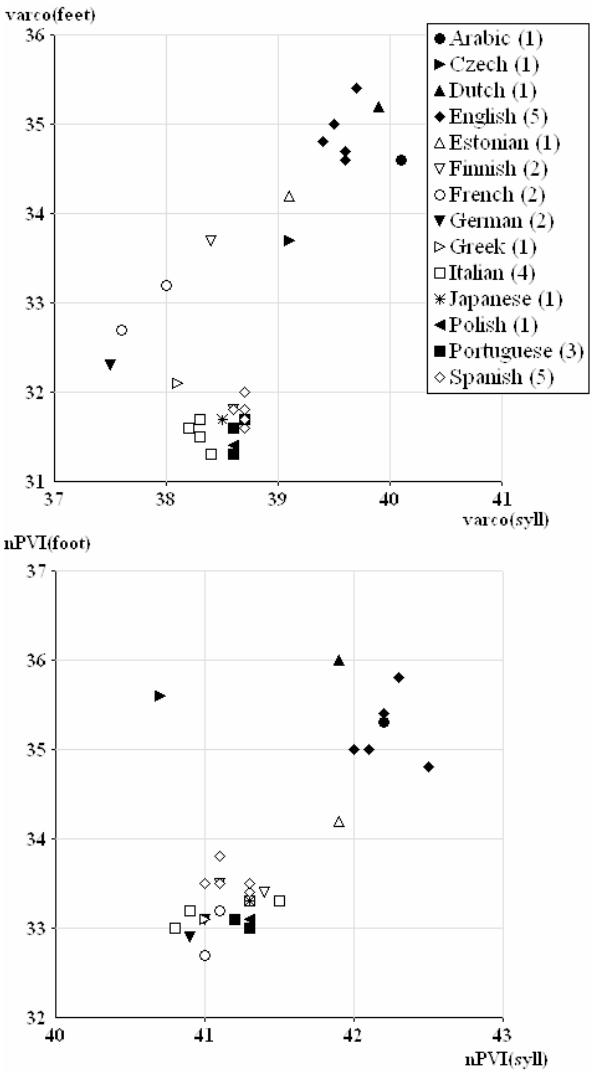

Figure 3: *Values of the varcos (above) and of the nPVI (below) calculated on inter-stress (y-axis) and inter-onset (x-axis) durations. Supposedly stress-timed languages are represented by black shapes, supposedly syllable-timed languages are represented by white shapes.*

4. Conclusions

The results of the deltas and the PVIs on inter-onset (syllable) and inter-stress (foot) durations have not confirmed predictions made on the basis of traditional theories.

In particular, stress durational variability in stress-timed languages is not lower than in syllable-timed languages: on the contrary, it seems to be higher in most cases. Instead, the durational variability of syllables in stress-timed languages might indeed be higher than in syllable-timed languages (yet, German, Polish and Portuguese samples contradict even this claim).

We suggested that these results might reflect the hypotheses put forward by [3] that languages obey to the same tendency at both levels, i.e. they show either flexibility or control at both syllable and stress level. However, this hypothesis needs to be further explored because, as has been said, the values shown by German, Polish and Portuguese do not seem to support this claim.

Finally, despite the failure of this first attempt, we agree with [4] and believe that a model integrating both levels of speech rhythm (the segmental/syllabic one and the stress/accentual one) is highly desirable. Rhythm metrics applied to consonantal and vocalic (or to voiced and voiceless) intervals are not enough to give a full account of speech rhythm as they only consider the first (segmental/syllabic) level.

5. Acknowledgements

The authors would like to thank all the speakers who accepted to be recorded for this experiment.

6. References

[1] Abercrombie, D. (1967) *Elements of General Phonetics*, Edinburgh: University Press.

[2] Asu, E. L. & Nolan, F. (2006) Estonian and English rhythm: a two-dimensional quantification based on syllables and feet. *Proc. of Speech Prosody 2006*, Dresden (Germany).

[3] Bertinetto, P. M. (1989) Reflections on the Dichotomy 'Stress' vs. 'Syllable-timing'. *Revue de Phonétique Appliquée*, Mons, 99-130.

[4] Bertinetto, P.M. & Bertini, C. (2010) Towards a unified predictive model of Natural Language Rhythm. In: M. Russo (ed.) *Prosodic Universals. Comparative Studies in Rhythmic Modeling and Rhythm Typology*. Rome: Aracne.

[5] Boersma, P. & Weenink, D. (2010) *Doing phonetics by computer*. [computer program]

[6] Dauer, R. M. (1983) Stress-timing and Syllable-timing Reanalysed. *Journal of Phonetics*, 11, 51-62.

[7] Dellwo, V. & Wagner, P. (2003) Relations between language rhythm and speech rate. *Proc. of the XVth ICPhS*, Barcelona (Spain), 471-474.

[8] Galves, A., Garcia, J.E., Duarte, D., Galves, C. (2002) Sonority as a Basis for Rhythmic Class Discrimination. *Proceedings of Speech Prosody 2002*, Aix-en-Provence (France), 11-13 April 2002.

[9] Grabe, E., & Low, E. L. (2002) Durational Variability in Speech and the Rhythm Class Hypothesis. In: Gussenhover, C., Warner, N. (eds.), *Papers in Laboratory Phonology 7*, Berlin: Mouton de Gruyter, 515-546.

[10] IPA (1999) *Handbook of the International Phonetic Association*. Cambridge: University Press.

[11] Mairano, P. & Romano, A. (2007) Inter-Subject Agreement in Rhythm Evaluation for Four Languages (English, French, German, Italian). *Proc. of the XVIth ICPhS*, Saarbrücken (Germany), 6-10 August 2007, 1149-1152.

[12] Mairano, P. & Romano, A. (2010) Un confronto tra diverse metriche ritmiche usando Correlatore. In: S. Schmid, M. Schwarzenbach & D. Studer (eds.) *La dimensione temporale del parlato – Proc. of the 5th AISV National Congress*, Zurich (Switzerland), 4-6 February 2009, Torriana (RN): EDK Editore, 79-100.

[13] Ramus, F., Nespor, M. & Mehler, J. (1999) Correlates of Linguistic Rhythm in the Speech Signal. *Cognition*, 73/3, 265-292.

[14] Roach, P. (1982) On the Distinction between 'Stress-timed' and 'Syllable-timed' Languages. In D. Crystal, *Linguistic controversies*, London: Edward Arnold, 73-79.

[15] Wagner, P. & Dellwo, V. (2004) Introducing YARD (Yet Another Rhythm Determinator) and Re-Introducing Isochrony to Rhythm Research. *Proc. of Speech Prosody 2004*, Nara (Japan), 23-26 March 2004.