RHYTHM METRICS FOR 21 LANGUAGES

Paolo Mairano[a] *& Antonio Romano*[b]

[a]GIPSA-Lab / DPC, Grenoble, France; [b]LFSAG, University of Torino, Italy
paolo.mairano@gipsa-lab.grenoble-inp.fr; antonio.romano@unito.it

ABSTRACT

The most frequently used rhythm metrics have been applied to V and C durations for 57 samples read by speakers of 21 languages to check for a rhythm categorization. Even though the variability between samples of the same language is remarkable, overall data reflect a scalar distribution of languages belonging to the traditional categories of stress-timing and syllable-timing. Nevertheless, it is impossible to determine which combination of metrics provides the best representation as there is no rigid framework with which results can be compared.

Keywords: rhythm typology, rhythm metrics, speech rhythm

1. THEORETICAL BACKGROUND

In the last decade, research in speech rhythm has focused on rhythm metrics (initially called acoustic correlates of speech rhythm), that is to say on variables derived from durational measurements of consonantal and vocalic intervals. [12] proposed a set of three acoustic correlates (ΔC, ΔV and %V) of the structural properties held to be responsible of stress-timing and syllable-timing by [4] and [7]. Their values seemed to be able to discriminate languages belonging to the three traditional rhythm categories (stress-timed, syllable-timed and mora-timing).

A similar approach had been independently developed by Low and co-workers in the late '90s and finally led to the publication of [9], who proposed the pairwise variability index (PVI) as a rhythm measure: it differs from the deltas suggested by [12] in that it takes in consideration the temporal succession of segments. The two authors calculated the consonantal raw PVI and the vocalic normalised PVI on samples of 18 languages and claimed that results provided a better categorization of languages reflecting the traditional rhythm classes.

Several studies tested these measures on some languages or language varieties and the general results can be resumed as follows:

- some language samples occupy intermediate positions (see [9]);
- rhythm metrics are heavily influenced by several factors, mainly speech tempo (see [8]) and speech style;
- their values reflect a high inter-speaker variability;
- their results can be influenced by choices in the segmentation of audio samples in vocalic and consonantal intervals.

For these reasons, several authors have proposed normalizations or other modifications of the original formulae. For example, [8] proposed the Varcos (a normalization of the deltas), while [5] proposed the CCI (Control and Compensation Index, a normalization of the PVI by the number of segments that compose it).

As it has been remarked by various authors, most studies concentrated on small language samples by a limited number of speakers and languages, mainly because manual segmentation is very time-consuming. Recently, some authors (e.g. [3]) applied these measures on larger corpora by adopting automatic segmentation procedures; however, the spectrum of languages included is usually fairly limited (e.g. 2 in [3]). So, we propose to give a contribution in this sense.

2. METHODOLOGY

2.1. Data

We included 57 native speakers of 21 languages, all reading translations of *The North Wind and the Sun*. Most items were recorded in the sound-proof booth of our laboratory, while items marked with an asterisk (see below) were taken from the Illustrations of the IPA ([10] or various issues of the *Journal of the IPA* – a thorough account of the sources is not given for obvious lack of space, but complete bibliographic references for the IPA illustrations can be found online at *http://www.sil.org/~olsonk/ipa.html*).

Samples consisted of: 2 Arabic speakers, 2 Mandarin Chinese speakers (1 from Chao Yang and 1 from Hong Kong), 1 Czech speaker, 1*

Danish speaker, 1* Dutch speaker, 5 speakers of English varieties (RP*, GA*, AusE, NZE* and IndE), 1* Estonian speaker, 2 Finnish speakers, 2 French speakers (1* standard speaker and a Canadian speaker), 2 German speakers, 1* Greek speaker, 10 Icelandic speakers, 6+1* Italian speakers, 1 Japanese speaker, 1* Polish speaker, 3 Portuguese speakers (1* of European Port. and 1+1* of Brazilian Port.), 6 Romanian speakers (from Brasov, Bucharest, Bucovina, Moldavia, Muntenia and Oltenia), 2 Russian speakers, 5 Spanish speakers (1* of Castilian Spanish, 1 from Granada - Spain, 1 from Bogotá, 1 from Caracas and 1 from Lima), 1* Swedish speaker, 1 Turkish speaker.

2.2. The segmentation

Each sample was segmented in CV intervals with *Praat* by the two authors. Since previous studies have shown the weight on final values given by segmentation choices, we shall state our criteria:

- on-glides were considered as consonantal;
- off-glides were considered as vocalic;
- syllabic consonants were considered as vocalic;
- we relied on the second formant to establish the end of vocalic intervals;
- initial voiceless plosives were attributed a hold phase of 70 ms;
- devoiced vowels were considered as consonantal segments (e.g. in Japanese);
- glottal stops were labelled only if they had a phonological role in the language;
- epenthetic vowels were labelled if their duration was longer than or equal to 20 ms.

As has been said above, the CCI is a modification of the rPVI formula by which the duration of each vocalic and consonantal interval is divided by the number of phonological segments that compose it and aims at measuring the level of segmental compensation allowed by different languages. This meant that special conventions had to be used for the calculation of the CCI, namely:

- geminate consonants were considered as a double interval (e.g. in Italian);
- phonologically long vowels (e.g. in Finnish) and diphthongs were considered as a double interval;
- vowels in hiatus were considered as two independent vocalic intervals;
- deleted segments were counted in the total number of segments of an interval (in order to account for compensation phenomena).

2.3. The calculation of the metrics

Rhythm metrics were computed with a script specifically developed for this purpose (see [11]). The calculation was carried out in 2 different ways, i.e. globally (including all the values at once) and locally (averaging results obtained for each sentence).

3. RESULTS

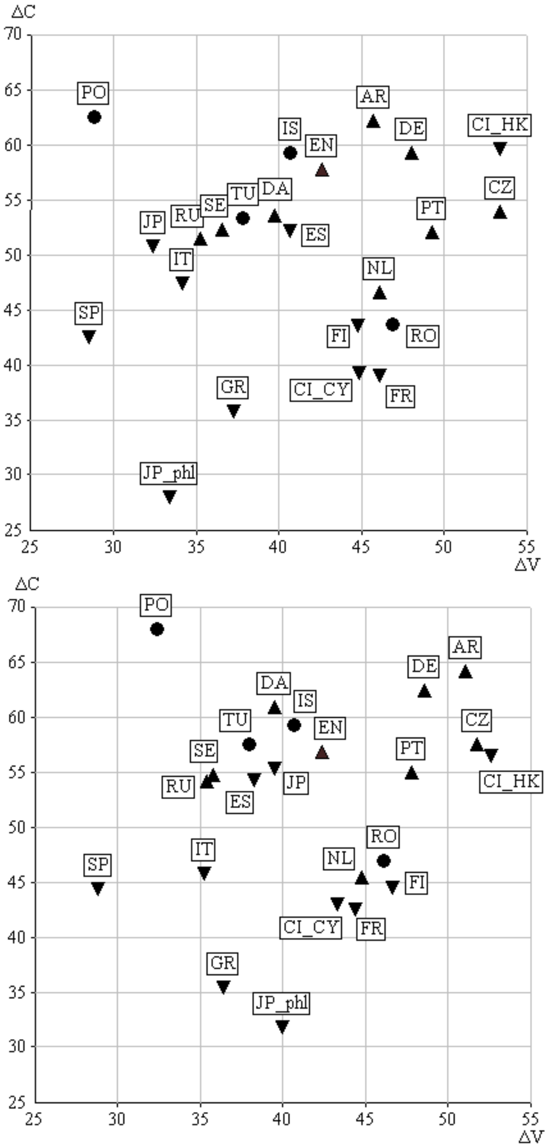
3.1. The deltas

Figure 1 shows the final results of ΔC and ΔV (the mean of the values obtained by each author). It can be seen that in both charts Italian, Greek and Spanish (supposedly syllable-timed languages) are situated in the south-western corner showing low delta values, whereas English, German, Portuguese, Arabic and Czech (supposedly stress-timed languages) tend to cluster in the north-eastern corner with high delta values. Polish is isolated in the north-western corner, with high values of consonantal variability (reflecting the complex consonantal clusters allowed by this language) and low vocalic variability (reflecting the lack of phonological vowel reduction). These results are in compliance with those presented by [12].

However, other languages do not confirm expectations. In particular, French shows high values of ΔV, whereas Japanese shows high values of ΔC. While the former is difficult to account for, the latter can be explained by considering that Japanese devoiced vowels were labeled as consonantal segments: this has the obvious consequence that the devoiced V, the preceding C interval and the following C interval are all joined in one long C interval. This is of course bound to have an effect on ΔC. In effect, we re-segmented the Japanese sample by labeling devoiced vowels as V and found very different results (see JP_phl in the charts). It has to be noted that some devoiced vowels were also observed in Romanian and for the Canadian French speaker.

Moreover, it may be noticed that, ΔC values seem to better separate supposedly stress-timed languages from supposedly syllable-timed languages than ΔV.

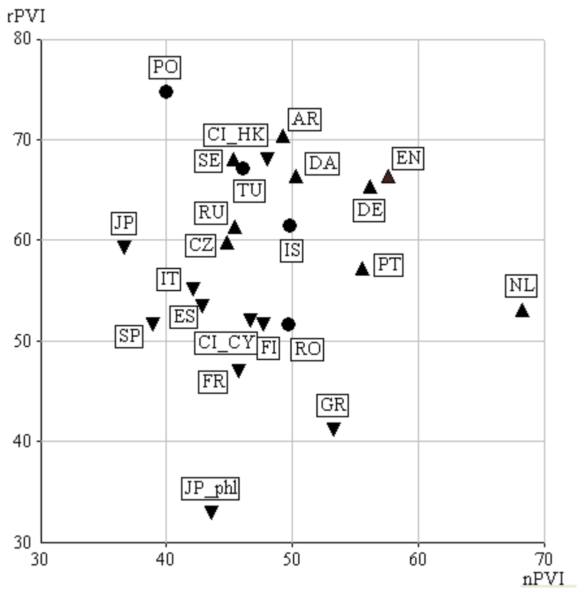
Figure 1: Results for the deltas calculated globally (above) and locally (below). Languages traditionally classified as stress-timed are marked as triangles pointing upwards, while languages traditionally classified as syllable-timed or as mora-timed are marked as triangles pointing downwards. AR=Arabic, CI=Mandarin, CZ=Cech, DA=Danish, DE=German, EN=English, ES=Estonian, FR=French, GR=Greek, IS=Icelandic, IT=Italian, JP=Japanese, NL=Dutch, PO=Polish, PT=Portuguese, RO=Romanian, RU=Russian, SE=Swedish, SP=Spanish, TU=Turkish.



3.2. The PVIs

The scenario provided by the PVIs (see figure 2) is similar but slightly superior in that it corrects some inconsistencies. In particular, French, Finnish, Estonian, Romanian and Chao Yang Mandarin cluster with other supposedly syllable-timed languages in the south-eastern corner (very probably as a result of the normalization of V intervals introduced by the nPVI). Polish and Japanese_phl once more occupy a very isolated position, but this time Dutch is also far away from the other samples, showing extremely high levels of V variability (in spite of the normalization).

Figure 2: Results for the PVIs calculated globally (in compliance with what has been proposed by [9], we applied the rPVI formula to C intervals and the nPVI formula to V intervals).



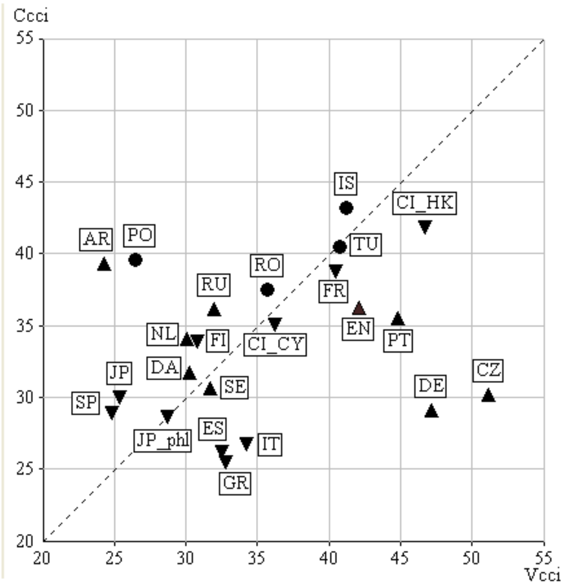
3.3. The CCI

The CCI is a modification of the rPVI formula that divides each interval by the number of segments that compose it in an attempt to measure the amount of compression allowed by a language at the segmental level. The predictions for this index are different from the deltas and the PVIs. Controlling (syllable-timed) languages are expected to align along the bisector (variability of V segments and C segments should be comparable); instead, compensating (stress-timed) languages are expected to fall below the bisector reflecting a higher variability for V segments than for C segments – see [5] for details. The results of our data for this index are shown in figure 3.

It can be seen that most languages tend to align along the bisector. Notable exceptions include, on the one hand, German and Czech, which, as expected, fall below the bisector, and, on the other hand, Arabic and Polish, which, surprisingly, occupy a position that was thought to be implausible (see [5]). To a lesser extent, English, Portuguese, Italian and Greek also fall below the bisector: while this was expected for English and

Portuguese, the same cannot be said of Italian and Greek (however, the results for Italian may be explained by considering the lengthening of stressed vowels).

Figure 3: Results for the CCIs calculated globally.



4. CONCLUSION

It has been shown that all rhythm metrics provide an acceptable representation of speech rhythm that is mainly consistent with expectations based on perceptive impressions. Moreover, other studies have managed to obtain a rhythm categorization even by applying the deltas and the PVIs to voiced and devoiced intervals and to syllable and stress durations (see [1]). This somehow suggests that the different formulae are challenging more for their rationales and theoretical perspectives than for the practice. Each of them has advantages and drawbacks: for example, metrics that normalize for speech rate run the risk of neglecting relevant phenomena; yet, they usually provide a better representation precisely because they neutralize differences in speech rate.

We would like to stress the fact that rhythm metrics exclusively reflect the first level of speech rhythm, *i.e.* the segmental one: they do not measure anything at the second (accentual) level. Therefore, researchers should be careful at classifying languages on the basis of what they see on these charts. It can be inferred that low deltas or PVIs characterize languages tending towards syllable-timing or segmental control. However, languages that show high values of deltas or PVIs cannot be said to tend towards stress-timing because nothing has been measured at stress level. Rather, it can be said that these languages do not tend to syllable-timing, but there is no proof at all that they tend towards stress-timing. The CCI is clearer than the other metrics as for this, as its explicit aim is merely that of describing intra-syllabic behavior; therefore, languages aligning along the bisector are only said to show segmental control, whereas languages clustering below the bisector are only considered to compensate at the segmental level.

In line with other studies (see Bertinetto & Bertini, 2010), we suggest that the two levels of speech rhythm (syllable and stress/accent) allow for a bi-dimensional categorization of languages based on control/compensation at each level.

5. REFERENCES

[1] 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.

[2] Barry, W.J., Andreeva, B., Russo, M., Dimitrova, S., Kostadinova, T. 2003. Do rhythm measures tell us anything about language type? *Proc. of the 15th ICPhS* Barcelona, 2693-2696.

[3] Benton, M., Dockendorf, L., Jin, W., Liu, Y., Edmonson, J.A. 2007. The continuum of speech rhythm: computational testing of speech rhythm of large corpora from natural Chinese and English speech. *Proc. of the XVIth ICPhS* Saarbrücken, 1269-1272.

[4] Bertinetto, P.M. 1989. Reflections on the dichotomy 'stress' vs. 'syllable-timing'. *Revue de Phonétique Appliquée*, Mons, 99-130.

[5] Bertinetto, P.M., Bertini, C. 2008. On modeling the rhythm of natural languages. *Proc. of Speech Prosody 2008*, Campinas, Brazil, 427-430.

[6] 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.

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

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

[9] 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* Berlin: Mouton de Gruyter, 7, 515-546.

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

[11] Mairano, P., Romano, A. 2010. Un confronto tra diverse metriche ritmiche usando Correlatore. *Proc. of the 5th AISV Congress*, Zurich, Switzerland, 79-100.

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