

Do Rhythm Measures Tell us Anything about Language Type?

Barry, W.J.[†], Andreeva, B.[†], Russo, M.^{‡†}, Dimitrova, S.[‡] and Kostadinova, T.[‡]

[†] Institute of Phonetics, University of the Saarland (IPUS), Germany

[‡] University of Paris 8/UMR 7023-CNRS, France

[‡] Sofia University, Bulgaria

E-mail: (wbarry/andreeva)@coli.uni-sb.de, michela.russo@univ-paris8.fr, snezhinad@hotmail.com

ABSTRACT

Recent instrumental approaches to measuring rhythm that have been applied with a view to capturing traditional rhythm typology differences are examined. The plausibility of their language classification results and the rationale behind the measures are discussed. A new, modified measure is suggested. The effects of language, different speaker groups, material and different speaking style – in particular tempo – on the various measures is examined on the basis of extensive corpora.

1. INTRODUCTION

In recent language-rhythm studies ([1, 2]), a number of different „rhythm measures“ have been shown to separate languages in a way which appears to confirm empirically the traditional rhythm types, namely syllable-, stress- and mora-timed, where previous instrumental approaches have failed (cf., for example the discussion in [2]). Critically based on the structurally determined vocalic and inter-vocalic (consonantal) intervals of utterances, these measures can – logically – be expected to increase in the reliability of their reflection of language-inherent properties as the size of the database from which they are derived increases. Conversely, they must vary with any smaller subset of material as a function of any factor affecting the vocalic and consonantal intervals, e.g., the syntactico-lexical structure, the speaker selection and the style of speech, an important aspect of which is speech tempo. The studies so far have not taken these factors into account, although very different values have been found for one and the same language. The problem of tempo fluctuation has received some attention [2], though only in support of a particular approach to rhythm calculation. Finally, in contrast to syllable- or stress-isochrony, the underlying rationale of the recent instrumental measures are not immediately interpretable in terms of the auditory impressions which prompted the traditional division of languages into syllable- and stress-timed [3, 4] Thus, plausible as results so far appear, it is unclear to what extent differences between languages reflect significant „rhythmic“ properties. This paper aims to illuminate some of the factors influencing the measures.

This exploration is based on over 5,000 (inter-pause stretches (ips) of segmentally labelled Bulgarian, German and Italian recordings. The Bulgarian data are a) *Bu1*: part of the BABEL database (project COP 1304) and b) *Bu2*:

recordings of highly controlled prompted speech made at IPUS, Saarbrücken. The German data is from the Kiel Corpus [5] of a) read: *Gr* and b) spontaneous speech: *Gsp*. The Italian consists of spontaneous speech recordings from the AVIP regional Italian database [6], from Bari: *Ba1* and *Ba2* (two labeling variants of the same data, see below), Naples: *Na* and Pisa: *Pi*. These 7 speaker groups provide a basis for a number of different comparisons: (i) *Gr* – *Gsp*: read-speech vs. spontaneous-speech, (ii) *Bu1*/*Bu2* – *Gr*: Bulgarian vs. German read speech, (iii) *Ba*/*Na*/*Pi* – *Gsp*: Italian vs. German spontaneous speech.

2. THE RHYTHM MEASURES

We consider first the rationale behind the two groups of measures under scrutiny, those used by Ramus [1, 7] and those used by Grabe and Low (G&L) [2].

Ramus uses three measures: %V, ΔV and ΔC ; G&L use two measures, a normalised nPVI-V and a raw rPVI-C measure. With the exception of %V, which is simply the proportion to which an utterance is vocalic and hence a reflection of overall syllable complexity, the measures all address the *variability* of the vocalic and consonantal interval durations within a stretch of speech (ips in this study). However, they differ in the *type* of variability they capture.

Ramus' Δ -values are the *standard deviation* of the vowel or consonantal intervals, i.e., a global variability measure which reflects nothing of the sequential patterning of durations which might logically be seen as underlying any auditory impression of rhythm. G&L's PVI measure (Pairwise Variability Index) *does* take the sequential variability into consideration by averaging the durational difference between consecutive vowel or consonantal intervals:

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

A normalised version of the PVI formula, used for PVI-V calculations and devised to correct for tempo fluctuations, relates the difference between consecutive intervals to the mean duration of the two intervals:

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

However, far from correcting for tempo changes (since it is a purely local normalisation), the normalised measure actually reduces stress- and accent-dependent as well as

phonological length differences, i.e., structural factors assumed to underlie stress- vs. syllable-timing.

It should be pointed out that, while capturing sequential variation, the PVI fails to maximise this possible advantage over the Ramus measures because vowel and consonant variation are calculated *separately*. The combined effect of vocalic and consonantal structure on an auditory rhythmic pattern is therefore not taken into consideration. A logical extension of the existing PVI measures is therefore applied in this study. It takes the consonant and vowel intervals together, thus capturing the varying complexity of consonantal + vowel groupings in sequence within an interpause stretch. We use the label *PVI-CV* for this measure.

3. DIFFERENCES BETWEEN LANGUAGES

The scalar model of rhythm implied by the measures just discussed is theoretically grounded in the structural discussion by Rebecca Dauer [8], in which rhythm is seen as the total effect created by the interaction of a number of phonetic and phonological segmental and prosodic properties. Considering those criteria, which are potentially different in the three languages under scrutiny, should allow a first prediction of their relative position on the „syllable-timed“ – „stress-timed“ continuum:

The *duration* criterion for stress-timing hinges on whether stressed syllables, in particular stressed vowels, are considerably longer than unstressed ones. This is not convincingly the case for Bulgarian (cf. [9]). It is usually considered to apply to German, where the vowel-length opposition is neutralised in unstressed syllables; it is certainly the case in Italian, where allophonic vowel lengthening occurs in stressed (open) syllables.

The *syllable complexity* criterion points to both Bulgarian and German being stress-timed, and Italian being more syllable-timed. Bulgarian and German can both have as many as three consonants in the onset and the coda, whereas standard analyses of Italian allow a maximum CC onset and single consonant coda. Dimitrova [9] points out that very complex syllables are rare and that Bulgarian speech is dominated by much simpler structures. However, despite a preponderance of simpler syllable structures in our data, the overall C/V ratio for Bulgarian (1.39) and German (1.54) is higher than for Italian (1.18), supporting the separation of Italian from the other two.

The *vowel-realisation* criterion traditionally offers support for a separation of Bulgarian and German as stress-timed from Italian as syllable-timed. However, the picture seems to be less clear in reality. Bulgarian certainly has a reduced unstressed vowel system, though the unstressed vowels appear to undergo a raising rather than a centralising process [9, 10]. German has a minimally reduced system (unstressed long /a:/ and short /a/ are neutralised) and a loss of the length opposition. However, there is no phonological „schwa-isation“ of unstressed vowels, although there is a considerable loss of timbre phonetically. Italian is considered to have no vowel reduction. However, a recent study [11] showed that there is neutralisation of

the mid-close vs. mid-open opposition and considerable phonetic centralisation in unstressed position. In many areas of Italy, word-final unstressed vowels tend towards schwa [12].

In summary then, it is not easy to predict where the three languages should be placed relative to one another on the rhythm continuum. Only the syllable-complexity criterium offers clear differentiation along traditional lines. Direct or indirect comparisons between Italian and German by Ramus [1, 7] and Grabe and Low [2] show Italian in an intermediate position between Spanish and English or German, but values derived from a more extensive database [12, 13] show that, depending on speech material, speaker and tempo, both German and Italian can vary greatly in their position within the “rhythm space”.

4. DO THE RHYTHM MEASURES SEPARATE THE LANGUAGES?

Results have been reported elsewhere for read German, and for spontaneous Italian and German [12, 13]. The read-speech results confirmed a speaker and material dependency of the measures. For spontaneous speech, it was essentially shown that both the Ramus and the G&L measures are able to separate German and Italian to some degree, though the G&L normalised PVI-V measure was insensitive to differences highlighted by ΔV . When differentiated for tempo, the location of the three tempo classes within the “rhythm space” varied strongly, ranging from a high variability “stress-timed” position for both languages to a position for “fast” German very close to that found for Spanish in other studies. But differences between German and Italian were largely maintained across the tempo classes except for the slow rate. Two measures that were reliable across all *three* tempo classes were Ramus’ ΔV and the complexity-linked %V.

A surprising result was a generally *higher* vowel and consonant variability for spontaneous Italian than for German in the Ramus measures, and a higher consonant (but not vowel) variability in the G&L PVI measures. In rhythmic terms this would make spontaneous Italian more stress-timed than German. The possibility that this was an artefact stemming from a particular labelling strategy is examined below on the basis of the Bari data, which explicitly labels lengthened vowels and sonorants which are used by some (Italian) speakers instead of hesitation pauses. The Bari data is therefore presented in two versions: Ba1 calculates the values with the lengthened segments included as speech sounds; Ba2 interprets the lengthened segments as “filled pauses” and calculates the inter-pause stretches accordingly.

Figure 1 shows a clear separation on the ΔV axis of the Italian speaker groups from both Bulgarian sub-groups and from read and spontaneous German. Spontaneous German lies between read German and Bulgarian and is not significantly different from either. On the ΔC axis, the Bari speakers are significantly different from all the other groups at the top end of the variability scale, and the Bulgarian groups at the bottom, but the Pisa and Naples

groups do not differ from either read or spontaneous German.

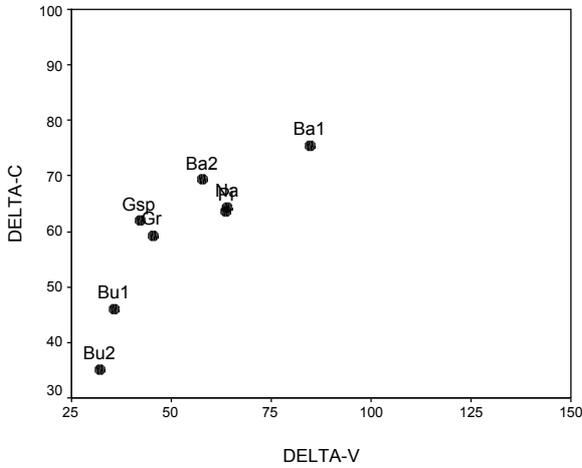


Fig 1: Rhythm plot of Ramus ΔV and ΔC values

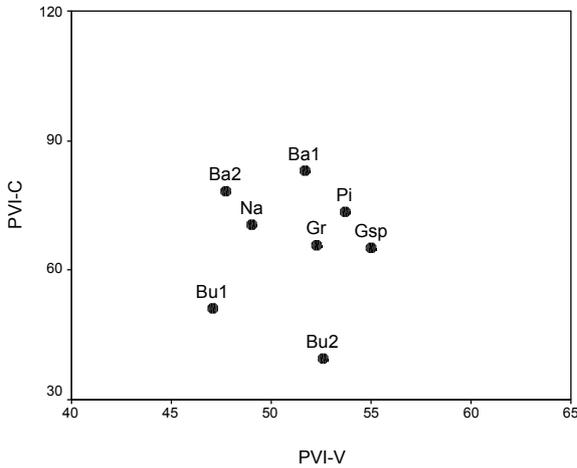


Fig. 2: Rhythm plot of G&L PVI values.

The G&L normalised PVI-V values do not separate the speaker groups in any plausible language-linked manner cf. also table 1 below), but the PVI-C differentiates the groups in almost exactly the same way as Ramus' ΔC.

Measure	Significant Language-group Differences
%V	(Ba = Pi) > (Pi = Na) > Bu2 = Bu1 = Gr > Gsp
ΔV	Ba = Na = Pi > (Gr = Gsp) > (Gsp = Bu1) > (Bu1 = Bu2)
ΔC	Na > Pi = Gsp = Ba = Gr > Bu1 > Bu2
PVI-V	(Gsp = Ba = Bu1 = Gr) > (Bu1 = Gr = Na = Pi) > (Na = Pi = Bu2)
PVI-C	Na > Pi = Ba = Gr = Gsp > Bu1 > Bu2
PVI-CV	Na = Ba = Pi > Gr = Gsp > Bu1 > Bu2

Table 1: Speaker-group differences for the “rhythm” measures on the basis of Scheffé post-hoc comparisons..

In neither figure do the positions of Ba1 and Ba2 indicate that the high variability for Napoli and Pisa speakers found in our previous studies results from some speakers' segment-lengthening speech habits. The “cleaned” values indicated by Ba2 are still higher than either read or spontaneous German for both Ramus values and for G&L PVI-C (but not for the normalised PVI-V).

Figures 3 and 4 show tempo-differentiated distributions of the language groups in the Δ and PVI “rhythm spaces”. In both cases the *Slow Ba1* measure lies outside the range of the x- and y-axes; it is excluded from the figure to improve resolution and legibility.

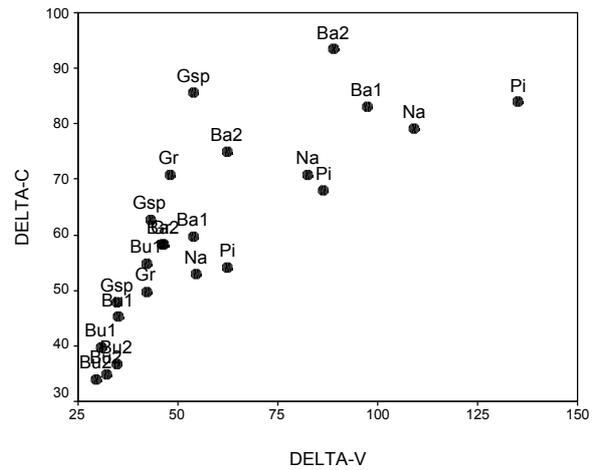


Fig. 3: Ramus tempo-differentiated values

With the exception of Bu2, the Ramus measures show a clear value shift to lower variability with increasing tempo, i.e. towards traditional “syllable-timed” positions. It can also be observed that there is no difference between read and spontaneous speech with regard to the direction of the shift, but the shift is less for read (Gr) than for spontaneous German (Gsp), and is even more reduced for Bu1 and Bu2. The degree of shift appears to correlate with the degree of tempo variation.

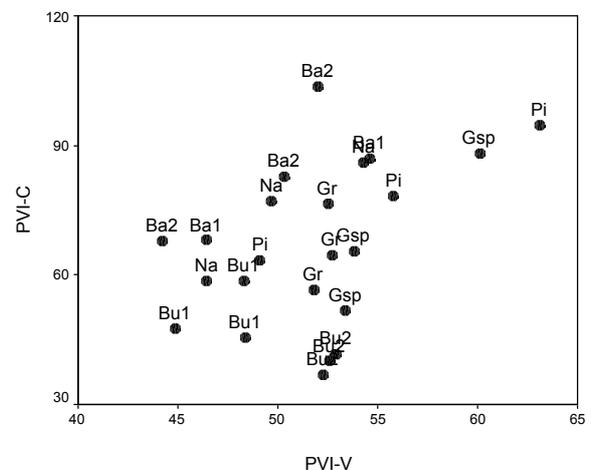


Fig 4. G&L tempo-differentiated values here

Fig 4 shows the same sort of reduction of variability in the PVI-C results across all languages, again with a difference

in degree for read in comparison to spontaneous speech. The PVI-V measure also shows a systematic shift with tempo, but as in Fig. 2, there is no language-oriented difference between speaker-groups. Both sets of measures show the apparently strong tendency of Italian towards positions in the rhythm space associated with “stress-timing”. The comparative results for the Bari speakers (Ba1 vs. Ba2) show this to be a real effect.

Figure 5 shows a “rhythm plot” (Slow Ba1 is again excluded) using the new PVI-CV and the Ramus %V measure – the two measures which differentiate the languages most plausibly and effectively (cf. table 1).

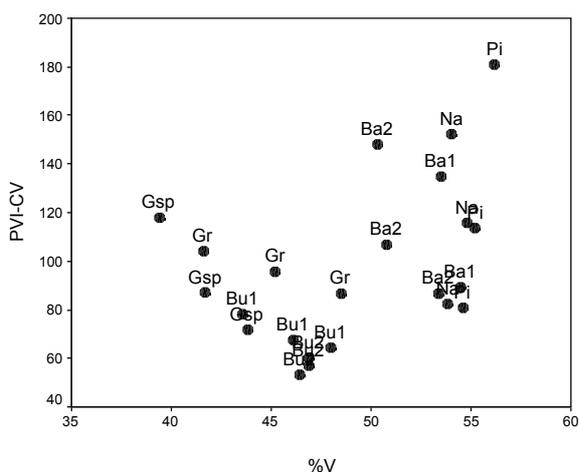


Fig. 5: Tempo differentiated PVI-CV plotted against %V

The tempo-differentiated plot shows that Bu1, Gr and Gsp shift considerably towards Italian on the %V axis with increasing tempo, whereas the values for the Italian speaker groups remain relatively constant. In contrast the shift of PVI-CV values is stronger for the Italian groups than for German and Bulgarian though the Ba, Na and Pi values for the Fast tempo class remains higher (more “stress-timed”) than for Fast Gsp and Fast Bul.

This difference in tempo-dependent behaviour can be interpreted in terms of a reduction in syllable complexity in German and Bulgarian with a concomitant increase in the vocalic proportion. This is plausible in the light of the greater underlying syllable complexity in these languages. Italian, on the other hand, with an underlyingly simpler syllable structure appears to reduce (its very considerable) syllabic durational variability without much structural change, and consequently without much change in the vocalic proportion.

5. CONCLUSIONS

The application of the Ramus and G&L measures to a larger collection of speech data indicates that they are not inherently suited to support typological statements about language rhythm. However, their logical dependence on realised syllabic structure make them sensitive to style, particularly tempo-induced processes. Despite fundamental differences in their underlying rationale, both measures of consonantal variability (ΔC and PVI-C) appear to cap-

ture the same phenomena. ΔV and the normalised PVI-V measure, on the other hand, are not comparable. The most plausible language-linked differentiation is achieved with a new PVI-CV measure and Ramus’ %V, which together capture different aspects of the consonantal-vocalic relationship, the former globally, the latter sequentially. Tempo-differentiated analyses clearly show a tendency for languages to converge with increasing tempo towards what has been defined as a “syllable-timed” position.

ACKNOWLEDGMENTS

Our grateful thanks to Caren Brinckmann for her invaluable and patient programming support.

REFERENCES

- [1] F. Ramus, *Rythme des langues et acquisition du langage*, Thèse de doctorat, EHESS, 1999.
- [2] E. Grabe and E.L. Low, “Durational Variability and the Rhythm Class Hypothesis”, *Papers in Laboratory Phonology VII*, to appear.
- [3] J. A. Lloyd, *Speech Signals in Telephony*, London, Sir Isaac. Pitman & sons, 1940.
- [4] K.L. Pike, *The intonation of American English*, Ann Arbor, University of Michigan Press, 1945.
- [5] IPDS, *The Kiel Corpus of Read Speech, CD-ROM #1 and The Kiel Corpus of Spontaneous Speech, CD-ROM #2-4*, Kiel, Institut für Phonetik und digitale Sprachverarbeitung, 1994-1997.
- [6] AVIP = Archivio Varietà dell’ Italiano Parlato, <ftp://ftp.cirass.unina.it>.
- [7] F. Ramus, M. Nespore and J. Mehler, “Correlates of Linguistic Rhythm in the Speech Signal”, *Cognition*, 73, 265-292, 1999.
- [8] R.M. Dauer, “Stress-timing and syllable-timing reanalyzed”, *Journal of Phonetics*, vol. 11, pp. 51-62, 1983.
- [9] S. Dimitrova, “Bulgarian Speech Rhythm: Stress-Timed or Syllable-Timed?”, *Journal of the International Phonetic Association*, vol. 27.1, 27-33, 1998
- [10] T. Pettersson and S. Wood, “Vowel Reduction in Bulgarian”, *Folia Linguistica*, vol 22, 239-262, 1988.
- [11] S. Calamai, “Vocali fiorentine e vocali pisane a confronto”, *Proceedings of the congress “Il Parlato Italiano”*, Napoli, 13-15 febbraio, 2003, to appear.
- [12] W.J. Barry and M. Russo, “In che misura l’italiano è iso-sillabico? Una comparazione quantitativa tra l’italiano e il tedesco”, *Proceedings of the VII Convegno della Società Internazionale di Linguistica e Filologia Italiana “Generi, architetture e forme testuali”*, Roma 1-5 ottobre, 2002, to appear a.
- [13] W.J. Barry and M. Russo, “Measuring rhythm. Is it separable from speech rate?”, *Proceedings of the International AAI Workshop “Prosodic Interfaces”*, Nantes 27-29 mars, 2003, to appear c.