Isochrony reconsidered.
Objectifying relations between Rhythm Measures and Speech Tempo

Michela Russo and William J. Barry

University of Paris 8/UMR 7023-C.N.R.S., France and University of Saarbrücken, Germany
mrusso@univ-paris8.fr; wbarry@COLI.Uni-SB.DE

Abstract

Recently, new ways of measuring rhythmic differences have been proposed. These are derived from syllable structure and prosodic differences between languages. The measurements are all *durational*: the *variability* of vocalic and consonantal intervals is calculated and plotted on two axes. Very small amounts of read material have been analysed. But the structural basis of the measurements implies dependency on the nature of the speech material: representativity problem. *Tempo* measurements are a function of phone- and syllable duration and must therefore be confounded with rhythm measures. This study presents a comparative rhythm analysis of spontaneous Italian and German. Some of these problems, in particular Speech Tempo, are discussed.

1. Introduction

Rhythm has traditionally been associated with the idea of isochronic intervals between sequences of speech units – either syllables or feet [15]. But failure to verify such intervals objectively [5, 11] has led to a reassessment of what lies behind the auditory impressions which prompted the original differentiation of languages as syllable- or stress-timed.

Dauer [7, 8] proposed a number of structural properties which can differ in their manifestation across languages, and which support the originally assumed dichotomy while also allowing mixed language types between the two extremes. Following this structural reassessment, the search for an objective basis has recently also been reorientated, taking the duration of the vocalic and consonantal components of the syllable as the basis of their calculations [9, 10, 17, 18, 19].

Simple measures of the vowel intervals and the inter-vocalic intervals have been suggested and tested which reflect phonetic consequences of most of the syllable-based structural properties proposed by Dauer [7, 8]. Thus, different syllabic complexity and a differing tendency for reduction of unstressed syllables results in a separation of languages within a two-dimensional space (vowel axis vs. consonant axis) in a way which appears to reflect the traditional rhythm types.

Furthermore, a satisfying scatter of measures has been found for different languages between what one might consider prototypical extremes. As might be expected with a multifactorial structure, not all languages fulfill all criteria, so that mixed rhythm types, as suggested on structural grounds in [16], are confirmed by measurement.

However, closer consideration of the principles involved in quantifying rhythm and categorizing languages according to rhythm type uncovers a number of as yet unrsolved problems. Representativity of the corpus used for quantification is an overriding issue. If it is large enough, has enough different speakers, enough different utterances spoken in enough different styles of speech, it must by definition claim to be representative of the language. But studies in rhythm have not so far been able to process such amounts and varieties, nor have the factors just enumerated been explicitly considered.

One aspect of speaking style, namely speech tempo has been discussed to some extent, and has led to diverging approaches to the calculation of vowel variation [10, 14], but there has been little consideration of the systematic relationship between tempo and rhythm measures (but cf. [1, 2, 3, 20]).

2. Aims and hypotheses

The overall aim of this study is to illuminate the relationship between currently used rhythm measures and established tempo measures. At the same time, the rhythmic similarity and difference (as reflected in these measures) between Italian and German will be examined. In particular, the variation in the measures as a function a) of speaker and b) tempo is of interest.

From the structural properties of the two languages, it has to be hypothesized that rhythm measures for Italian and German will be significantly different:

-Simple *vs.* Very varied (including very complex) syllable structures;

-No vowel quantity opposition vs. Short-long vowel distinction

-No lexical schwa vs. Lexical schwa and reduction of unstressed vowel durations

This prediction can also be derived from the traditional allocation of Italian to the syllable-timed group of languages and of German to the stress-timed group.

However, the existence of geminate consonants and stressed-vowel lengthening in Italian suggest that syllable variability (and consequently the consonantal and vocalic variability measures used as „rhythm measures“) should be greater than expected for unequivocal syllable-timing languages.

If rhythm measures are assumed to reflect perceptual characteristics of the language which are used by children during language acquisition [17], a further hypothesis ought to be that neither speaker differences nor tempo should radically affect inter-language differences.

2.1. Speech material and analysis methods

The Italian spontaneous speech database AVIP/API (*Archivio Varietà Italiano Parlato*, ftp: //ftp.cirass.unina.it) and the spontaneous-speech part of the German Kiel Corpus (IPDS, cf. [12, 13]) were used for the analysis. The AVIP/API recordings are of Map-Task dialogues in which one speaker tries to elicit from the other a route round a map. Recordings were made of speakers from the Bari, Naples and Pisa areas. The German dialogues are negotiations for a proposed meeting between the two persons based on the entries and gaps in their respective calendars. The recordings were made in Kiel and the speakers are from the broad Northern German area. In total, there are 13 Italian speakers (6 from Naples, and 7 from Pisa). A matching number of German speakers were selected.

Both corpora had been segmented and labelled, providing the segmental identities and durations which form the basis of the rhythm and tempo measures. Pauses, hesitations and other interruptions had also been annotated, so it was possible to identify prosodically uninterruptd “inter-pause stretches” (ips). The ips is the utterance unit used for calculating individual rhythm measures, which are then grouped and averaged over speaker and language (and additionally over regional group for Italian). Since rhythm requires a sequence of syllables to be perceived, a lower limit of 4 syllables for an ips was set. All shorter ips were excluded from the analysis. This lower limit is ultimately arbitrary, but the choice was based on the observation that ips below that limit often manifested very extreme tempo and rhythm measures which were not found for ips of four syllables and more.

Tempo measures were calculated on the basis of both syllables per second and phones per second. These two measures always correlate highly across any corpus, and it is often rather prematurely assumed that they capture essentially the same phenomenon. In the two corpora under scrutiny they correlated to the extent of R = 0.894 for the Italian data and 0.815 for the German data, i.e., very highly significantly. These degrees of correlation mean that 79.9% and 66.3% , respectively, of the tempo variation within each measure is predictable from the other measure. Conversely, of course, it means that in the Italian data 20.1%, and for German 33.6% of the variation is *independent* of the other measure. So there is clearly scope for further consideration in this area of speech variability.

To calculate the relationship between tempo and rhythm measures, the tempo range of each language group was divided into three equal parts and the individual ips allocated to the appropriate tempo class.

Rhythm measures were calculated according to [17, 19] and [10].

1. The Ramus measures [cf. 17, 19] are (i) the proportion of vowels in the ips (%V), (ii) the standard deviation of the Vowel duration in the ips (ΔV) and (iii) the standard deviation of the intervocalic consonantal interval (ΔC).

2. The Grabe and Low measures [cf. 10] correspond in essence to the Ramus variability measures, but are calculated in pairwise steps through the ips rather than globally across the ips. They are therefore called „Pairwise Variability Indices“ (PVIs). The difference (i) between consecutive vowels and (ii) between consecutive intervocalic intervals) are averaged over the ips, giving a vocalic and consonantal variability measure. In the case of the vowel intervals, the difference is related to the sum of the two vowels. This “normalisation” is claimed to be necessary (and possible) for the vowel intervals in order to counteract shifts in tempo because vowels vary more than consonants with tempo, and there is never more than one vowel in a vowel interval. The two PVI formulae are as follows:

(i) Non-normalized consonantal PVI:

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

(ii) Normalized vowel PVI:

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

A number of other measures were calculated, which were considered to offer illumination of the rhythm-tempo relationship. Among these were:

- The ratio of number of consonants/number of vowels (as a rough measure of syllable complexity in an ips or in a corpus) and
- The ratio of vowel-duration / consonant-duration (as a measure of the temporal structuring of the syllable).

2.2. Results

For the comparisons of Italian and German performed on the data, possible Italian regional differences were allowed for by testing for effects of and differences betwen speaker origin across four speaker groups: Napoli, Pisa and German.

2.3. Language differences

Before considering tempo effects, we examine the primary hypothesis that Italian and German will differ sgnificantly in their rhythm measures. Not all rhythm measures showed a difference.

In one-way ANOVAS to test the influence of language background with the Ramus measures (deltaV, deltaC and %V) the two vowel measures show a clear language effect (F = 155.8, df. 2; p < 0.001 for ΔV; F = 458.8, df. 2; p < 0.001 for %V), but the ΔC measure do not differ significantly (F = 1.17, df. 2; p = 0.312). A Sheffé post-hoc test shows that both vowel measures separate Italian from German (p < 0.001), and in the ΔV (but not in %V) the Pisa and Naples speaker groups are also different (p < 0.05).

The Grabe and Low PVI measures (PVI-V and PVI-C) both show a highly significant speaker-group effect (F = 18.11, df 2; p < 0.001 for PVI-V; F = 9.46, df 2; p < 0.001 for PVI-C). However, the group differentiation, as indicated by the Scheffé post-hoc tests, is seen to deviate from that found in the Ramus values. The Naples speaker group differs significantly in its vowel variability measure (PVI-V) from both the Pisa and the German group (p < 0,001), whereas the German and Pisa groups do not differ. Consonantal variability (PVI-C), on the other hand, does not distinguish the Pisa and the Naples groups (p = 0.364) but separates the German group from both the Pisa group (p < 0.001) and the Naples group (p < 0.05). Fig. 2 shows the vowel- and consonant-PVI values for the speaker groups (values from the previous studies are given for Spanish, English and Polish for comparison with the Naples, Pisa and German):

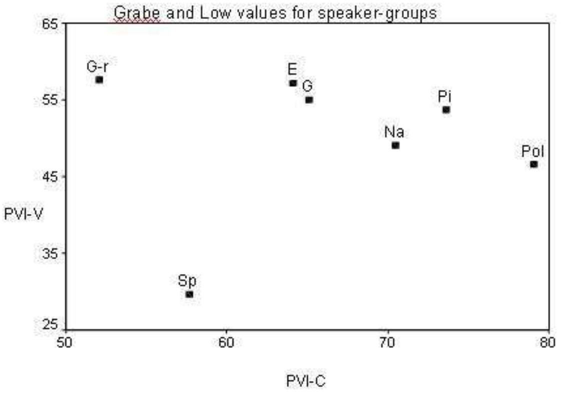

Fig. 1. *Grabe and Low PVI values (PVI-V and PVI-C) for the Naples (N), Pisa (P), read German (G-r), German, English and Spanish speaker groups.*

The evidence from figure 2 suggests that the rhythm values are not reliable indicators of a language's, or even of a regional variety's rhythmic status in typological terms, but more a reflection of the language material that occurs in the utterances produced, and of the style in which that utterance is produced (cf. [20] – two speakers reading the same texts varied significantly, but in different dimensions for different texts).

2.4. Speaker differences

One-way ANOVAs for speaker as independent factor, ungrouped for language background give a highly significant effect for all measures (see table 1):

Table 1*: Results of one-way ANOVAs for effect of speaker on rhythm measures*

Measure	**F-value**	**Degrees of freedom**	**Significance level**
%V	68.2	16	< 0.001
Delta-V	26.5	16	< 0.001
Delta-C	8.1	16	< 0.001
PVI-V	5.7	16	< 0.001

Post-hoc tests give very different groupings of speakers. For the Ramus measures, only %V separates the speakers cleanly along language lines and makes no distinctions between Pisa and Naples speakers. ΔV also shows a systematic ordering of the speaker values according to language background, and separates the 4 German speakers from 4 of the 7 Italian speakers for at least one of their recording sessions, one of the German speakers from all 7 Italian speakers for one recording and from two speakers for both recordings. The average speaker values for ΔC showed no consistent ordering corresponding to language background. Fig. 3 shows the average speaker values for the two Ramus measures ordered according to magnitude:

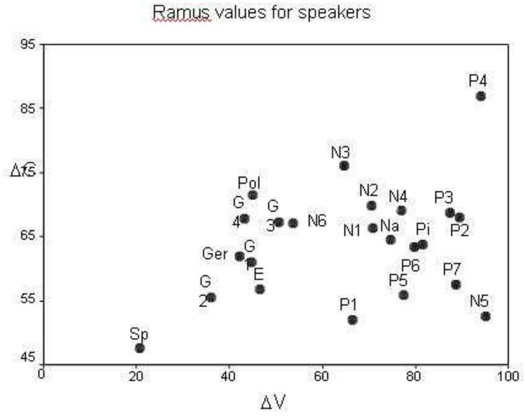

Fig. 2 *Ramus measures for German (G), Naples (N) and Pisa (P) speakers.*

Neither of the Grabe and Low PVI measures approach the speaker-differentiation power of %V or ΔV nor their apparently systematic ordering of the speakers according to language background. Of the two measures PVI-C differentiates better than PVI-V, appearing to reflect an order which bears slightly more relationship to language background than PVI-V or ΔC [cf. 1, 2, 20].

In summary, given the uncontrolled nature of the speech material analysed, we cannot claim that the rhythm measures reflect any language-inherent properties that might distinguish Italian and German. The only measure which offers a material-independent separation of the Italian and the German speakers is the proportion of vowels as a simple percentage of the ips duration (%V). This is to be expected, given the observation that Italian syllable structure is much simpler, on average, than German syllable structure; i.e., there are fewer consonants per syllable, and therefore proportionally more vowels. This can be quantified with two measures: 1. The average ratio of consonants (Cn/Vn) to vowels in ips is 1.181 for Italian (1.179 for Naples, 1.183 for Pisa) and 1.52 for German. Translating this into a durational ratio: 2. The vowel-duration to consonant-duration ratio (Vdur/Cdur) is 1.217 for Italian (1.201 for Naples, 1.232 for Pisa) and 0.738 for German.

.

3. Tempo variation and rhythm measures

In terms of phones/sec the four German speakers spoke significantly faster (range of speaker averages13.48 – 15.24 ph/sec) than all the Italian speakers (range range of speaker averages 8.21 – 10.22 ph/sec), and in terms syllables/sec they spoke significantly faster (range range of speaker averages 5.34 – 6.08 syl/sec) than all but two of the Italian speakers (range of speaker averages 3.17 – 4.77 syl/sec). Therefore, to assess the links between articulation rate and rhythm measures (and the links with syllable complexity), the division of the ips into three tempo groups was performed separatly for Italian and German.

Calculating rhythm values within the three tempo groups for each language, we observe that they vary considerably. The extent of the rhythm value shifts is given in table 2a (phone rate) and table 2b (syllable rate):

Table 2a: *Variability in rhythm values as a function of phone rate*

Lang/Tempo	**%V**	**ΔV**	**ΔC**	**PVI-V**	**PVI-C**
N/1 (slow)	55.7	135.3	82.8	59.5	97.2
N/2 (medium)	54.1	82.2	70.4	49.4	75.4
N/3 (fast)	54.2	53.8	54.2	46.3	59.6
P/1 (slow)	55.9	123.9	89.3	57.5	101.5
P/2 (medium)	55.0	87.1	65.2	56.1	74.5
P/3 (fast)	54.8	60.4	52.3	50.1	61.9
D/1 (slow	39.5	55.2	88.6	57.5	92.7
D/2 (medium)	41.7	43.6	65.0	52.5	68.7
D/3 (fast)	43.7	32.8	50.9	48.1	56.4

Table 2b: *Variability in rhythm values as a function of syllable rate*

Lang/Tempo	%V	ΔV	ΔC	PVI-V	PVI-C
N/1 (slow)	64.0	158.9	60.4	60.0	69.4
N/2 (medium)	54.9	83.7	70.1	50.1	75.7
N/3 (fast)	52.0	51.2	59.8	45.6	65.8
P/1 (slow)	58.8	132.6	84.3	59.9	97.3
P/2 (medium)	55.5	86.0	65.0	54.9	75.2
P/3 (fast)	53.5	61.9	56.1	50.6	64.8
D/1 (slow	43.8	59.2	87.3	58.9	62.2
D/2 (medium)	42.4	43.8	64.8	51.7	54.9
D/3 (fast)	40.7	31.5	52.0	48.2	51.3

4. Conclusions

Considering tempo as expressed by syllable rate, we find theat Ramus' %V parameter is by far the most tempo-resistant *and* the most language-distinguishing measure. This observation offers some support for Ramus assumptions regarding the role of the vocalic segments of utterances in the language acquisition process [20] and is in line with the even more simplified approach suggested by [19], where sonorant vs. non-sonorant parts of the utterances rather than vocalic and consonantal parts are calculated.

The reduction in %V for the Naples and Pisa speakers with increasing phone rate (in contrast to syllable rate) agrees with the long-accepted observation that vowels are more sensitive to rate change than consonants. The lack of change for syllable rate is an indication that the tempo measure in terms of syllables/sec also reflects a covariation of syllable complexity with syllable rate. This is borne out by the highly significant (negative) correlation of Cn/Vn ratio with syllable tempo (Pearson R = –0.305, N 841, $p < 0.001$).

The contrast between syllable-rate and phone-rate in the direction of the small change of %V for the German speakers reflects an even stronger effect of syllable-structure simplification at higher articulation rates and a considerably higher negative correlation of Cn/Vn ratio with syllable tempo (Pearson R = –0.407, N 635, $p < 0.001$).

The shifts in the values of Ramus delta values and the Grabe and Low PVI values also support the interpretation that rhythm values are to a considerable part a function of articulation rate: As the rate increases (either syllable rate or phone rate), the degree of variability of the vowel and consonant intervals within an ips decreases.

5. References

[1] Barry W. J.; Russo, M., 2003. Measuring rhythm. Is it separable from speech rate?. *AAI Workshop, Prosodic Interfaces*, A. Mettouchi; G. Ferré (ed.). Nantes: Université de Nantes, UFR Lettres et Langage, AAI (Acoustique, Aquisition, Interprétation), 15-20.

[2] Barry, W. J.; Russo, M. 2004. Isocronia oggettiva o soggettiva? Relazioni tra tempo articolatorio e quantificazione ritmica. In *Il parlato Italiano*, F. A. Leoni; F. Cutugno; M. Pettorino; R. Savy (ed.). Napoli, 13-15 febbraio 2003, Napoli: D'Auria, cdrom A02.

[3] Barry, W. J; Andreeva, B.; Russo, M.; Dimitrova, S.; Kostadinova, T., 2003. Do Rhythm Measures Tell us Anything about Language Type?. *15th International Congress of Phonetic Sciences*, M.-J. Solé; D. Recasens; J. Romero (ed.). Barcelona: Causal Productions Pty Ltd, 2693-2696.

[4] Bertinetto, P.M., 1981. *Strutture prosodiche dell'italiano*. Firenze: Accademia della Crusca.

[5] Bolinger, D., 1965. *Forms of English: Accent, morpheme, order*. Cambridge, Massachusetts: Harvard University Press.

[6] Dasher, R.; Bolinger, D., 1982. On pre-accentual lengthening. *Journal of the International Phonetic Association* 12, 58-69.

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

[8] Dauer, M. R., 1987. Phonetic and phonological components of language rhythm. *11th International Congress of Phonetic Sciences*. Tallinn, Estonia: U.S.S.R. / Academy of Science of the Estonian S.S.R., vol. 5, 447-450.

[9] Gibbon, D.; Gut, U., 2001. Measuring speech rhythm. *Eurospeech 2001-Scandinavia*, P. Dalsgaard; B. Lindberg; H. Benner (ed.). Aalborg: Kommunik Grafiske Losninger A/S, vol. 1, pp. 95-98.

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

[11] Hoequist, C. Jr., 1983. Durational correlates of linguistic rhythm categories. *Phonetica* 40, 19-31.

[12] IPDS, 1994. *The Kiel Corpus of Read Speech*. Kiel: Institut für Phonetik und digitale Sprachverarbeitung, vol. 1, CD-ROM #1.

[13] IPDS, 1994-1997. *The Kiel Corpus of Spontaneus Speech*. Kiel: Institut für Phonetik und digitale Sprachverarbeitung, vol. 1-3, CD-ROM #2-4.

[14] Low, E.L.; Grabe, E., 1995. Prosodic patterns in Singapore English. *XIIIth International Congress of Phonetic Sciences*, K. Elenius; P. Branderud (ed.). Stockholm: KTH and Stockholm University, vol. 3, pp. 636-639.

[15] Pike, K. L. 1945. *The intonation of American English*, Ann Arbor: University of Michigan Press.

[16] Nespor, M., 1990. On the rhythm parameter in phonology. In *Logical Issues in Language Acquisition*, I.M. Roca (ed.), Dordrecht: Foris, 157-175.

[17] Ramus, F., 1999. *Rythme des langues et acquisition du langage*. Doctoral dissertation. Paris: EHESS.

[18] Ramus, F., 2002. Acoustic correlates of linguistic rhythm: Perspectives. *Speech Prosody*, B. Bel; I. Marlien (ed.). Aix-en-Provence: Laboratoire Parole et Langage, 115-120.

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

[20] Russo, M.; Barry, William J., 2004. In che misura l'italiano è 'iso-sillabico'? Una comparazione quantitativa tra l'italiano e il tedesco. In *Generi, Architetture e forme testuali*, P. D'Achille (ed.). Firenze: Cesati, 387-401.