Testing rhythmic and timing patterns of Italian

Antonio Romano
Laboratorio di Fonetica Sperimentale «Arturo Genre»
Dipartimento di Lingue e Letterature Straniere e Culture Moderne
Università di Torino

Abstract

This paper is aimed at testing on data of Italian the speech rhythm model proposed by O'Dell & Nieminen (1999) and Barbosa (2006) which predicts temporal patterns as the result of the coupling of two oscillators. Within this framework, the duration of the inter-stress interval is a function of the number of syllables and of the coupling strength of two clocks.

The model was tested on a set of specifically designed sentences read by 2 native Italian speakers. Data were then labeled to take durational measures. Finally, a regression analysis was carried out on Inter-Stress Intervals and the rhythmic properties of the sentences were also assessed in terms of deltas and PVIs for comparison purposes.

In compliance with what is usually claimed for Italian, the results of the regression analysis for both speakers indicate a roughly linear growth in both cases. Likewise, the deltas and the PVIs indicate low values of consonantal and vocalic variability for these samples.

Index Terms: speech rhythm, syllable, foot, compression, timing.

1. Introduction

As is well known, the traditional dichotomy opposing «stress-timed» and «syllable-timed» languages has been introduced by Pike (1945) and then extended by Abercrombie (1967) to «all the languages of the world». The label «stress-timed» has been used to refer to languages whose timing seems to be dominated by stress patterns, while the label «syllable-timed» has been used to refer to languages whose timing seems to be dominated by segmental or syllabic time patterns. The categorization of languages according to such principles is justified by their traditions in versification and by auditory impressions on the part of listeners. Abercrombie (1967) claimed that syllable-timed languages are characterized by isochrony at the syllable level, while stress-timed languages are said to show isochrony at the foot level. Several authors have contradicted these hypotheses by experimental tests and languages were not found to exhibit the expected metrical regularity in connected speech.

In other words (as it is summarized in Farnetani & Kori 1986) «while there is general agreement, in current theoretical views, that the perceived rhythm of language is based on the alternation of strong and weak elements in speech [...], there are instead opposite views about the relevant elements underlying the classical division of languages in syllable-timed and stress-timed» (p. 33). Since instrumental studies based on measurements of syllable and foot duration failed to find any sort of isochrony in speech, perceived isochrony seems to have no phonetic correlates.

Different positions have been expressed stating for instance that perceived isochrony has its bases in speech production and that tendencies towards isochrony are manifested, at the acoustic level, by systematic changes of segmental durations as a function of syllable composition and/or foot size and experimental work carried out by Farnetani & Kori (1990) has shown that in Italian, as Swedish or English, a tendency towards isochrony exists both at syllable level and at word level.

But studies on language rhythm also followed a distinct approach, aiming at modeling speech rhythm rather than measuring it. It seems to us that, except for studies such as Allen (1975) and few others, the two approaches have been and still are kept separated: in papers dealing with timing, rhythm measuring is generally not mentioned, just as timing (at segmental or syllabic level) is generally disregarded in studies on rhythm typology, as if speech timing were to be referred to co-articulatory properties between segments in intra-syllabic environments rather than to durational properties at the level of rhythmic patterns (feet or prosodic words, sentences etc.; see Romano & Mairano 2010 for further details).

In the light of this debate, in this paper, we raise relevant issues: a) whether there are different timing models or a unique model with local preferences and b) whether rhythm emerges from other structural properties or rather is a primitive linguistic variable (even unconsciously) controlled in production. We then discuss whether rhythm is a phonological variable or the phonetic consequence of other phonological events related to a syllabic or a more general prosodic patterning (similar concerns are in Barry *et alii* 2003 and Krull & Engstrand 2003).

1.1. Foot and syllable as rhythmic units

Under this light, a rhythm model should be grounded on the reformulation of language differences observed in terms of reduction of vowels (V) and consonants (C) in a gestural overlap hypothesis framework. As it has been proposed by Bertinetto & Bertini 2008 and 2010, this could lead to a revisited dichotomy not involving the stress-/syllable-timing axis but contrasting more controlling languages vs. more compensating languages.

This model has its roots in earlier studies of '80s and bases the assessment of speech rhythm on intra-syllabic durational stability vs. compression. These properties have been studied for Italian by several authors.

The syllable as a rhythmic unit has been investigated by Farnetani & Kori (1986) whose results indicate that the syllable is a base temporal unit in Italian, both for compression effects and for position effects. The shortening of the nucleus and of consonants in the syllable onset in correspondence of an increase in syllable complexity suggests that the controlled temporal unit extends all over the syllable. They also showed that compression effects are primarily anticipatory: the syllable undergoes more systematic and greater durational variations affecting the following, rather than the preceding context. Other data give further support to the hypothesis that the unit which tends to be constant in duration is the entire

V-to-V temporal interval (from the onset of a vowel to the onset of the following one).

The results of other experiments on the foot (or word) as a rhythmic unit indicate that rightward extension has systematic shortening effects on the first stressed syllable (and also affects the preceding unstressed vowels).

These results are partially in contrast with previous findings by Bertinetto (1983), Vayra *et alii* (1984) and Marotta (1985) who studied patterns of temporal compression in Italian by means of similar protocols but on different speech samples.

At the foot level, Farnetani & Kori (1983, 1990) based the study of the relative prominence among syllables on duration measures. The results indicated that both rhythmic principles and syntactic structures have systematic effects on syllable duration, as durational ratios of stressed to unstressed vowels increase from a minimum of 1.10 to a maximum of 2.28 when the Inter-Stress Interval (ISI) increases up to 3 syllables. The subsequent extensive study by Marotta (1985) on mono- and polysyllabic words confirmed these tendencies.

1.1. The double oscillator model

As mentioned above, after the failure of the search for isochrony, different methods were devised in order to account for the specific rhythm patterns of languages. In particular, several authors tried to analyse this kind of regularities in terms of more sophisticated speech models.

An approach that seems to have yielded interesting perspectives consisted in defining the Inter-stress interval (ISI=I) as a function of the number of syllables (n) and trying to assess the values of different parameters such as a and b in the following formula:

$$I(n) = a + b \cdot n$$

where a is a constant and b is a parameter describing the growing ratio of I versus n. With this formula, the two extreme ways of establishing the priority in rhythmic regulation of different languages are:

- an absolute stress-timing, when b is naught and, therefore, the inter-stress interval is a constant (b=0 → I=a; see figure 1.i);

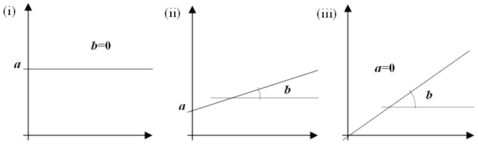

Fig. 1. The growth of inter-stress intervals for (i) absolute stress-timed languages (on the left), (iii) for absolute syllable-timed languages (on the right) and (ii) for a mixed-timed language (in the mid) (from Romano & Mairano 2010).

- an absolute syllable-timing, when a is naught and the inter-stress interval is directly proportional to the number of syllables (a=0 → I=$b \cdot n$; see figure 1.iii);
- yet, languages usually tend to exhibit an intermediate form (see figure 1.ii).

This approach is fully explained by Barbosa (2006), but has been used earlier (with some variations) by other authors, such as Bertinetto (1983), and later by, for instance, Engstrand & Krull (2003). In particular, Bertinetto (1983) also applied this method to an intra-syllabic domain: in some of his charts he defined syllable duration as a function of the number of

segments composing a syllable. Results for his data of Italian confirmed that the inter-stress interval increases as a function of the number of syllables composing it and, likewise, syllable duration increases as function of the number of segments composing it. As already reported, the author concluded then that Italian exhibits a tendency towards syllable-timing.

This approach is also the basis of the model that has been re-proposed (see the relevant literature on previous studies, e.g. in Barbosa 2006) and which predicts temporal patterns as the result of the coupling of two oscillators (see O'Dell & Nieminen 1999). The duration of the inter-stress interval is described as the function of the number of syllables and of two clocks whose contributions are regulated by a coupling strength (called *r*-parameter). So, *a*, *b* and *I* of the preceding equation are re-defined as in the following formula:

$$I(n) = \frac{r}{r\omega_1 + \omega_2} + \frac{1}{r\omega_1 + \omega_2} n$$

where ω_1 is the oscillation velocity of the accentual oscillator, ω_2 is the velocity of the syllabic oscillator and *r* is the coupling strength. When the value of the coupling strength (*r*) is 1, then *a* of the original equation is equal to *b* and both oscillators have the same influence; but when *r* is greater than 1 ($r > 1$) the overarching accentual-oscillator is dominant whereas when *r* is lesser than 1 ($r < 1$) it is the subordinated syllabic-oscillator which is dominant.

Studies of the '80s-'90s carried out for Swedish and English (quoted by Barbosa 2006) have evaluated *r* on different corpora with changing tempos and have assessed values around 2 against typical values obtained for Italian or Greek ($r \approx 0.9$). Barbosa (2006) tested the same mathematical model for different speech rates for Brazilian Portuguese finding values about 1.5. However, *r* did not systematically decrease for increasing speech rates, so that a shift towards syllable-timing for rapid tempos was not confirmed (see Dellwo & Wagner 2003 for different results obtained with a different approach).

1. The experiment

We tested this model on a corpus of Italian sentences like the ones analysed by Marotta (1985). Similar sentences were measured and tested for Italian by other authors (Bertinetto 1983, Vayra *et alii* 1984, Farnetani & Kori 1986, 1990) bringing evidence of limited compression phenomena for Italian and discussing the controversial but discriminant role of stiffness parameters related to syllable and segment durations (similar outcomes are summarised for Spanish and French by other authors, see Romano & Mairano 2010).

A regression analysis was carried out on ISIs and the rhythmic properties of the sentences were assessed in terms of deltas (cp. Ramus *et alii* 1999) and PVIs (cp. Grabe & Low 2002).

1.1 The data

The author recorded a corpus of 27 sentences designed for testing different properties at two different speech rates. Two speakers read in a sound-proof booth the following three series of sentences, which were characterized by a growing

number of inter-stress syllables (9×2 utterances ×2 rates ×2 speakers). Speaker 1 has southern origins whereas Speaker 2 is a speaker of northern Italian.

Perciò pésa tùtto di nuovo.
Perciò pésalo tùtto di nuovo.
Perciò pésamelo tùtto di nuovo. (1st series)

Perciò spósta tùtto di nuovo.
Perciò spóstalo tùtto di nuovo.
Perciò spóstamelo tùtto di nuovo. (2nd series)

Perciò tàta tùtto di nuovo.
Perciò tàtata tùtto di nuovo.
Perciò tàtatata tùtto di nuovo. (3rd series)

As compared to Marotta (1985), we included sentences with different segmental structures (*sposta* instead of *pesa*) and even a nonsense word with a growing number of syllables as it happens in reiterant speech (*tàta*, *tàtata*, *tàtatata*). Sentences were pronounced by two male speakers with a mean syllable rate of 7.00 σ/s with local minima down to 5.73 and maxima up to 8.71 (both for speaker 2, who showed a more variable tempo by contrasting slow vs. fast speech rate). The mean syllable rates for the two speakers were 7.75 and 6.25 σ/s.

1.1 The methodology

The sentences were segmented and phonetically labelled with Praat (Boersma 2001).

Durations were extracted from separate Textgrids, taking measurements with two methods: A) from the onset of a stressed syllable to the onset of the following stressed syllable, thus obtaining σ-to-σ measures and

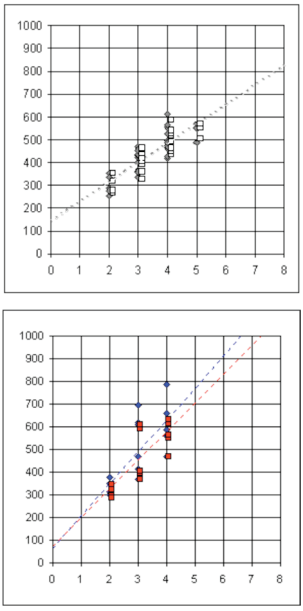

Fig. 2. The growth of Inter-stress Intervals in the data of both speakers (Speaker 1 and 2; ◊ for measures at foot level (A) vs. for V-to-V measures (B)) . Duration measures in ms on the y-axis and the number of syllables (n) on the x-axis.

B) from the stressed vowel (without the syllable onset) to the following stressed vowel, thus obtaining V-to-V measures.

1. The results

The results are summarised for both speakers in figure 2 together with the two regression lines giving estimates of a roughly linear growth in both cases (A and B). For data on the upper plot (speaker 1) *a* has quite high values (slightly lower for B: 148 vs. 145) whereas *b* is almost the same for the two measures (rounded to 85). Data on the lower plot (speaker 2) have *a* around 69 (A) and 78 (B) and a higher *b*: 141 (A) and 127 (B).

For speaker 1, this yields *r* values of 1.14 and 1.44, respectively, as if the accentual

oscillator were dominant at phrase level (this is not very surprising according to Bertinetto 1983 and other authors above). The sensitivity of the measures to the segmental content of syllables is evident when analysing separately the three series: the low value of the coupling strength (0.14) was obtained for the third series and accounts for very variable results (in contrast with e.g. $r = 2.48$ of the first series): these variations are greater than the ones induced by changes in speech rate.

The values which characterise data of speaker 2 are less variable and span from 0.48 to 0.70 accounting for a stable dominance of the syllabic oscillator.

Both speakers account for Italian as a syllable-based language and, in partial contrast with previous studies (such as Farnetani & Kori 1990), the particular sample of sentences we analysed does not show any isochrony at word level.

The fact that these sentences belong to a syllable-based language is well reflected in rhythmic charts as shown in figure 3 (note that utterances at different speech rate are gathered in each corresponding series). Values of vocalic and consonantal intervals place both samples in the south-western corner of the charts both for delta and PVI metrics. These samples of Italian cluster fairly well with values obtained for the same speakers on the basis of the analysis of a larger set representative samples (see Romano & Mairano 2010).

On delta charts, Speaker 1 generally shows more variation of C intervals (in one case rising towards values calculated for Spanish metrics) whereas Speaker 2 tends to let V intervals vary more (occasionally moving towards values obtained for French). This is less evident in PVI charts where differences among the three series which could be related to speech rate are supposed to be neutralised or at least reduced. Anyway, the global positioning in the PVI chart also confirms that these samples cluster within the syllable-based area.

Even though no significant syllable compression was generally observed,

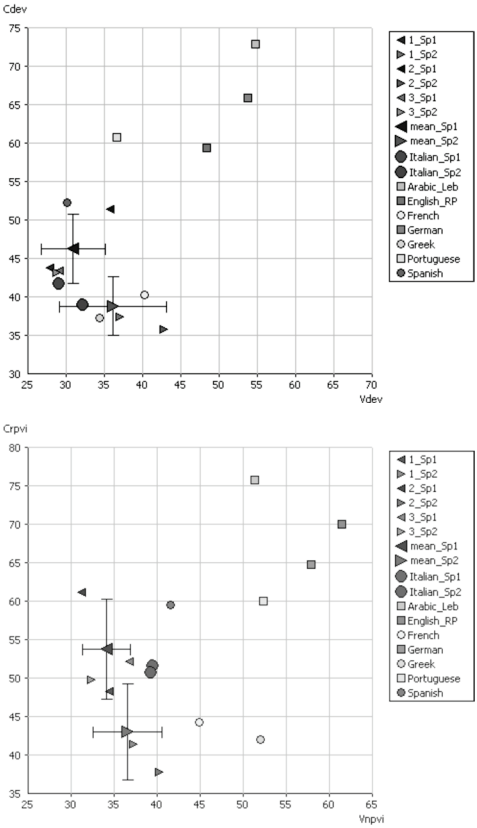

Figure 3. Charts of Deltas and PVIs calculated on the sentences of Speaker 1 and 2 and compared to values obtained on larger samples of the same 2 speakers (Italian_Sp1-2) and to data of other languages in Romano & Mairano (2010). Plots are made with Correlatore a program available online (http://www.lfsag.unito.it/correlatore/index_en.html).

an increase in the stressed/unstressed duration ratio appears as a function of *n* depending on speech rate. For unchecked syllables (series 1 and 3), at the faster rate Speaker 1 showed a mean ratio of 1.28 → 1.66, whereas the mean ratio for Speaker 2 was 1.71 → 2.12. The mean values for the slower rate were instead higher (1.29 → 1.91 for Speaker 1 and 1.76 → 2.59 for Speaker 2). This is generally associated to a fairly stable duration of the stressed vowel and slight reduction of the unstressed vowels and suggests a kind of inter-syllabic compensation.

Our results, however, do not provide evidence of syllable isochrony either. Even though the corpus was not meant to bring evidence in this field, we measured a certain segmental stiffness rather than a syllabic one. When the stressed syllable was checked, the nucleus was shortened but the whole duration significantly increased (t=2.33, df=37, 0.01<p<0.05).

Moreover, we found that the specific segmental content strongly affects intra-syllabic compensation, reflecting constraints due to the intrinsic timing of Cs and Vs. Unstressed /ta/ syllables obtained indeed higher durations than syllables like /me/ or /lo/ (t=2.90, df=19, 0.005<p<0.01) and that accounts for the differences between the results obtained by Vayra *et alii* (1984), Marotta (1985) and Farnetani & Kori (1986) who used corpora based on words characterised by fairly different choices in terms of segmental content.

Conclusions

In this paper we attempted to merge rhythm modelling and measuring showing that a better account may be given by a joint approach. By extending the same kind of experiments carried out in the '80s (which sometimes lead to controversial results) to the new perspectives brought about by research in the last decade, we confirmed the initial claim of Bertinetto (1983), showing a linear lengthening of the foot with a growing number of syllables. A limited inter-syllabic compensation has even been confirmed between stressed and unstressed syllables but is counterbalanced by intra-syllabic control depending on a certain degree of segmental stiffness (this should however be verified in further studies with a more dedicated/specific corpus).

On the one hand, the regression analysis of our data within a rhythm model based on the coupling of two oscillators (see O'Dell & Nieminen 1999 and Barbosa 2006) revealed a variable coupling strength (r) among utterances by two speakers at two different speech rates. Even though the r values span form 0.48 to 1.14 for σ-to-σ measures (A method) and from 0.70 to 1.44 V-to-V measures (B method), overall values (characterising the whole corpus) are respectively assessed on 0.91 and 1.02. Thus, although the accentual oscillator seems to dominate for a particular set of sentences uttered by speaker 1, a general dominance of the syllabic oscillator emerges in the whole corpus, where differences introduced by various segmental choices are much more significant than those related to variations in speech rate.

On the other hand, rhythm metrics, based on a quantitative analysis of C and V intervals, brought to similar results,

placing the samples in the south-western corner of the charts which is usually occupied by syllable-based languages.

As a general conclusion, we may consider rhythm to be the phonetic consequence of several phonological events related to syllabic and prosodic patterning. We cannot exclude that, in some cases, the rhythmic alternation of different units takes over and determines the particular prosodic properties of an individual or of a specific variety. This is obviously not demonstrated here, but it could offer an explanation for the differences observed.

Acknowledgements

This paper is the result of a research work planned together with Paolo Mairano. I am indebted to him for the suggestions during the early steps of the experimental design and the help in the final stage of the redaction. All errors remain mine.

References

Abercrombie D. (1967). *Elements of General Phonetics*. Edinburgh, Edinburgh University Press.

Allen G.D. (1975). Speech rhythm: its relation to performance universals and articulatory timing. *Journal of Phonetics*, 3, 75-86.

Boersma P. (2001). Praat, a system for doing phonetics by computer. *Glot International*, 5 (9-10), 341-345.

Barbosa P.A. (2006). *Incursões em torno do ritmo da fala*. Campinas, Pontes.

Barry W.J., Andreeva B., Russo M., Dimitrova S. & Kostadinova T. (2003). Do rhythm measures tell us anything about language type? *Proc. of the XVth ICPhS*, Barcelona (Spain), 2693-2696.

Bertinetto P.M. (1983). Ancora sull'italiano come lingua ad isocronia sillabica. In: AA.VV., *Scritti linguistici in onore di G.B. Pellegrini*, II, Pisa, Pacini, 1073-1082.

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

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

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.

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

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

Engstrand O. & Krull D. (2003). Rhythmic intentions or rhythmic consequences? Cross-language observations of casual speech. *Proc. of the XVth ICPhS*, Barcelona (Spain), 2789-2792.

Farnetani E. & Kori Sh. (1983). Interaction of syntactic structure and rhythmical constraints on the realization of word prosody. *Quaderni del Centro di Studio per le Ricerche di Fonetica del CNR*, 2, 288-318.

Farnetani E. & Kori Sh. (1986). Effects of Syllable and Word Structure on Segmental Durations in Spoken Italian. *Speech Communication*, 5, 17-34.

Farnetani E. & Kori Sh. (1990). Rhythmic Structure in Italian Noun Phrases: A Study on Vowel Durations. *Phonetica*, 47, 50-65.

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

Krull D. & Engstrand O. (2003). Speech rhythm – intention or consequence? Crosslanguage observations on the hyper/hypo dimension. *Phonum, 9*, 133-136.

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 AISV2009*, Zurich, 2009), Torriana (Italy), EDK, 79-100.

Marotta G. (1985). *Modelli e misure ritmiche: la durata vocalica in italiano*. Bologna, Zanichelli.

O'Dell M. & Nieminen T. (1999). Coupled oscillator model of speech rhythm. *Proc. of the XIVth ICPhS*, San Francisco (USA), 1075-1078.

Pike K.L. (1945). *The Intonation of American English*. Ann Arbor, University of Michigan Press.

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

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

Vayra M., Avesani C. & Fowler C. (1984). Patterns of temporal compression in spoken italian. *Proc. of the Xth ICPhS* (Utrecht, The Netherlands, 1983), 2, 541-546.