SPEECH RHYTHM AND TIMING: STRUCTURAL PROPERTIES AND ACOUSTIC CORRELATES

Antonio Romano
Laboratorio di Fonetica Sperimentale 'Arturo Genre'
Dipartimento di Scienze del Linguaggio – Università degli Studi di Torino
antonio.romano@unito.it

1. ABSTRACT

In my intervention to the Round Table I summarised results from a selection of recent contributions to the research on rhythm and speech timing coming from two Italian laboratories: the *Laboratorio di Linguistica* of the *Scuola Normale Superiore di Pisa* and the *Laboratorio di Fonetica Sperimentale 'Arturo Genre'* of the University of Turin.

In my short presentation I emphasised reference to papers by Pier Marco Bertinetto and Chiara Bertini (Bertinetto & Bertini, 2008, forthcoming; Bertini & Bertinetto, 2009) and Paolo Mairano and Antonio Romano (Mairano & Romano, 2007, 2010), with an introduction explaining the reasons of my own interests in it.

2. PERSONAL INTERESTS AND MOTIVATIONS

Since the early '90s, even though mainly working on intonation structures for my *PhD* (Romano, 1999, 2001), I reserved a relevant interest in speech timing (not neglecting references to other connected fields) and I have been studying the basic literature on this topic (early intuitions of Pike and Abercrombie, more specific contributions by Allen, Bertinetto, Dauer, Fowler, Lindblom, Miller, Roach and many others) oriented to the description of diverging tendencies shown by languages in terms of speech timing and rhythm in production and perception.

Since my research focused on the prosodic properties of Italian dialects, in an early description of the suprasegmental properties of Southern Italian varieties I made an attempt to include considerations on rhythm, following suggestions from Bertinetto (1977), Dauer (1983), Bertinetto (1989) and Schmid (1996).[1] For that research I had the opportunity to study a considerable amount of literature that had appeared in the early '90s on timing and rhythm of Italian dialects (which is now sometimes neglected by scholars working on Southern Italian varieties).[2]

---

[1] Sallentinian and Apulian dialects are well told apart on the basis of rhythmic properties (a description is now in Romano 2003). In Molinu & Romano (1999) we also took into account measures and experimental work on syllables and other relevant structural properties of some of these dialects.

[2] For instance, many papers published by John Trumper's research team (see, e.g., Mendicino & Romito, 1991, and Romito & Trumper, 1993) refer to evaluation methods proposed by Lindblom & Rapp (1973). Results as well as other research indications suggested by these instrumental works pointed out significant differences in this linguistic domain. As proved by Schmid (2004, 2008), Italian varieties (even within the same broad area, such as

However, all my studies on this topic were carried out before the proliferation of new analysis techniques after the proposal by Ramus, Nespor & Mehler (1999). Research on rhythm exploiting this new approach started in our laboratory in more recent years, when I had the chance to meet Paolo Mairano, who helped me to rapidly get an insight on how rhythm research had evolved in the last decade, and when we found similar interests with the staff of the Laboratory of Linguistics in Pisa.

3. THE RELATIONSHIP BETWEEN TIMING AND RHYTHMIC PATTERNS

The very topic of my presentation started with a rapid hint to the relationship that ties timing and stress- or syllable- patterns. After a short discussion on terminological issues, the terms of a scientific divide were reviewed.

3.1 Terminological issues and cause-effect doubts

We know that the label 'stress-timed' is reserved to languages whose timing is dominated by stress patterns and that the corresponding label 'syllable-timed' refers to languages whose timing is regulated by segmental time patterns depending on syllabic constraints.

A natural typology of languages where foot vs. syllable seems to dominate the metric organisation of speech is confirmed by the observation of different versification traditions in the world's languages and by their variable predisposition to fit in musical-rhythmic frames (cp. Pamies, 1999, referring to Sachs, 1953), but is mainly based on linguists' intuitions which are often influenced by impressions.[3] It has been claimed that languages of the former type exhibit isochrony at the foot level, while languages of the latter type are said to exhibit isochrony at the syllable level (Abercrombie, 1967; cp. Crystal, 1994).

In fact, these alleged isochronies have not been verified in speech and, as it has been proved by several authors (cp. Roach, 1982), languages do not show this kind of metrical regularity when we observe them in connected speech. This way, the isoaccentual/ isosyllabic divide we refer to, in Italian for instance, seems to have little to do with real phonetic rhythmic cues – 'isochrony' is used by linguists and dialectologists to account for vocalic lengthening patterns and to explain distribution rules observed in some dialects.[4]

Furthermore, once that metrical regularities are not confirmed in connected speech, even the stress-/syllable-time classes should be reanalysed in broader terms (as proposed by Stephan Schmid during this workshop), namely distinguishing *stress-based* (STB)

---

mid-southern ones, represented in his data by Neapolitan and Bitontino) show rather different rhythmic patterns and may not be accounted for without due distinctions.

[3] The original distinction emerges from auditory cues. It is well stated everywhere that we owe these definitions to Lloyd James (1940: 25), distinguishing "machine-gun" vs. "morse-code" languages, and to Pike (1945: 35) who refers to syllable-timed vs. stress-timed languages.

[4] As for other less used labels, such as 'isochronicity', perhaps we have a more useful candidate in order to illustrate this structural phonetic regularity on acoustic grounds. A critical view of these language properties, based on instrumental research carried out on spontaneous speech as well as on pre-planned speech, is now in Giordano (2008), also reviewing various sources not confirming such regularities in Italian varieties.

vs. *syllable-based* (SYB) languages independently of stress or syllable phonological constraints which could be compensated at a phonetic level (see the examples of Arabic evaluated by Ghazali *et al.*, 2002; see §5.1).[5]

Several questions persist, however, which cannot be answered merely by a terminological discussion. We still do not know whether there are different timing models or a unique model with local preferences. We still do not know whether rhythm does emerge from other structural properties or rather is a primitive linguistic variable (even unconsciously) controlled in production.[6] In other terms, as discussed by Krull & Engstrand (2003), we do not even know if rhythm is a phonological variable or the phonetic consequence of other phonological events.

3.2 Other issues related to scientific divides

Along this line, I pointed out another issue whose evidence comes again from personal notes. I carried out my *PhD* research in Grenoble (France) in the years when research on rhythm was gaining ground (thanks to Ramus' *PhD*) but I attended a series of conferences on timing where rhythm measuring was not mentioned, probably because it was considered to pertain to different research scopes.[7] What's more, this happened in a laboratory next to the one where Plinio Barbosa was preparing his own *PhD*:[8] the Ramus' approach was not considered a key contribution by the research team with whom Barbosa was working with and we may say that, somehow, it is still considered at least not directly related to linguistic rhythm modelling.

The two main approaches to speech rhythm (modelling and measuring) are still kept separated, as is also proved by connections to research on speech perception and production. That could be shown by another anecdote. In 1995 I was appointed research assistant in the Research Centre where Kate Demuth was working to her *bootstrapping* model whose main account was to be published in the following years (Demuth, 1996), without a reference to rhythm evaluation methods used in Europe at that time (in laboratories of the same town). Yet, in the following years the interest for rhythm in speech perception and language acquisition awoke newly and many results were published by several

---

[5] See the discussion in Vayra *et al.* (1984) and references below (§2.3). Perception issues are raised by Allen (1975).

[6] Generally speaking, rhythm is explicitly mentioned as an organising principle of linguistic timing by Lenneberg (1967). Rhythmical regularities present in the phonology of languages have been extensively studied in terms of prominence patterns by various scholars (see e.g. Liberman & Prince, 1977; Hayes, 1984; Nespor, 1993).

[7] 'Timing' may be referred to co-articulatory properties between segments in intrasyllabic environments as well as to internal durational properties of syllabic or rhythmic patterns, or even to temporal relations between units in a macro-rhythmic environment (feet or prosodic words, sentences...). Research made within intra-/inter-syllabic environments is summarised in works by Rudolph Sock (who was an invited speaker of one of these conferences; see, e.g., Sock *et al.*, 1996).

[8] Furthermore, my *PhD* supervisor, Michel Contini, was in the dissertation panel of Barbosa's *PhD*.

scholars, including Demuth, showing rhythm as a base for *bootstrapping*[9]. The relevance of rhythm in language acquisition has been proved at least since Mehler *et al.* (1981) and Miller (1984), while several studies of the '90s (see e.g. Nazzi, Bertoncini & Mehler, 1998) added experimental evidence. And again, in the same years, MacNeilage's theory 'frame, then content' was brought to completeness, leading to an articulatory acknowledgement of the role of rhythm in speech production and language acquisition (mainly in relation to mandibular oscillations; cp. Rhardisse & Abry, 1995) without explicit mutual connections between these studies (see MacNeilage & Davis, 1990, MacNeilage, 1998).
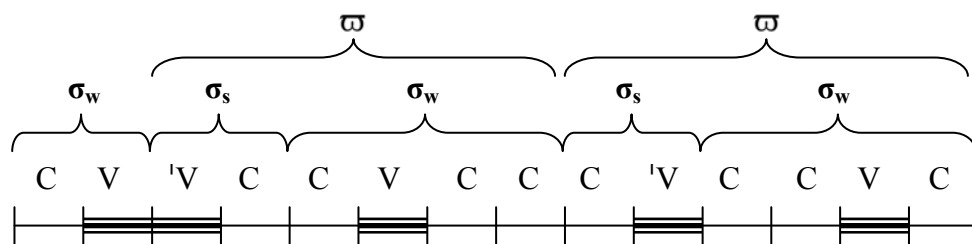
Figure 1: Foot analysis of a fictitious weak-strong syllable
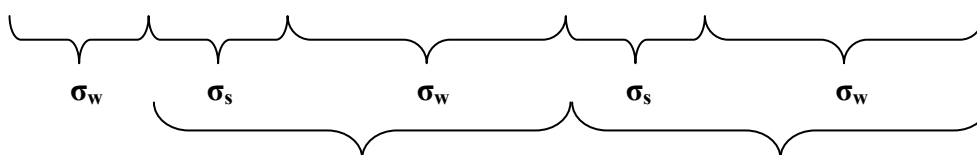sequence and different syllable types

Figure 2: Units which would be taken in consideration for the fictitious utterance in
Figure 1 by a phonological evaluation – durations of syllables and feet (interstress patterns)
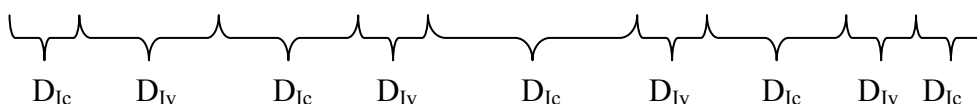
Figure 3: Units which are taken in consideration for the fictitious utterance in Figure 1 by
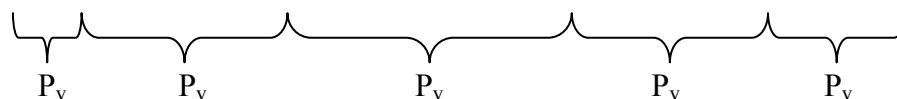recent rhythmic measures – durations of vocalic and consonantal intervals.

Figure 4: Units which are taken in consideration for the fictitious utterance in Figure 1 by
traditional rhythmic measures – durations of intervals between vocalic pulses

[9] Within the framework of generative approaches, one may find interesting phonetic data in Frota, Vigário & Martins (2002) and Frota, Vigário & Freitas (2003). For a general reference see Fikkert (2007).

This shows as, in those years, significantly different approaches to the study of time variables in speech focused on rhythm modelling and timing evaluation (even though they did not always found a joint interest in it for their respective research fields).

3.3 Different approaches

As it has been said, the history of speech rhythm research (summarised in the bibliography collected by P. Roach) could be divided in a *before* and an *after*. In the *before*, to which authors sometimes go back for new proposals aiming at testing more robust models, we found experimental studies on the resistance of syllable- or interstress-patterns to the compression.

In these studies, measures were carried out at the syllable level, by evaluating foot or syllable durations and their resistance to tempo variation or with increasing lengths of utterances (Figure 2).[10]

Despite this dominant approach, at that time, there were already studies pointing out the relevance of vocalic pulses in rhythm perception (see Allen, 1975; Barbosa, 2006; also cp. Miller, 1984). They proposed measures as in the third method here recalled (based on V-to-V intervals, related to the so-called *P-centre* method, Figure 4).

In Figure 3, we show instead measures suggesting Consonantal and Vocalic Intervals as relevant units, following Ramus and colleagues proposals which determined the starting point of the *after*.

4. SPEECH RHYTHM MEASUREMENTS BASED ON DURATION

Within the framework defined by Ramus and colleagues, Mairano & Romano gave a few early contributions in rhythmic typology since 2006 – and that, without developing new models or theories on speech rhythm. Nevertheless, we consider them to be fairly original because they raise relevant questions.

We started by measuring stretches of V and C as other colleagues were doing in different European laboratories and, in the same period, we shared with Pier Marco Bertinetto our impressions about the different rhythmic metrics and doubts about which kind of speech make-up to use in order to test them. We observed a lack of consistency in certain basic assumptions made by rhythm specialists and the need to be more explicit on working hypotheses often implicit in recent papers, namely the specific choices made in the segmentation of speech files and during the statistical processing. Among sources of variability in the estimation of rhythm metrics, we investigated the agreement between different operators in the classification of V-C items.

---

[10] We may refer to, e.g., Marotta (1985), who studied sentences like: *Perciò pésa((me)lo) tutto di nuovo... Perciò pesate((me)lo) tutto di nuovo...* Similar sentences were measured and tested by other authors among which I would like to mention A. Pamies (see references in Pamies, 1999) who brought evidence on the reduced compression properties of Spanish and French and discussed the discriminating role of stiffness parameters related to syllable and segment durations. Similar outcomes are discussed for Italian by Bertinetto, (1977, 1983) and Farnetani & Kori (1986, 1990).

In particular, in Mairano & Romano (2007a, 2007b, 2008a):

1) we wondered whether the classification of segmental units should be done phonologically or phonetically and we raised questions about the classification of velarised and vocalised laterals, rhotic elements in coda, syllabic sonorants and so on;

2) we proposed to test the sensitivity of the metrics to segmentation choices made by different operators (discussing at the same time problems concerning speaker variation and speech rate conditions);

3) we observed different methods for determining the metrics and we tested the differences in the results obtained joining together the metrics from different interpausal units or, instead, keeping separate statistics for them.

We showed how these elements influence the final values of the metrics and introduce relevant changes in the rhythmic topology of languages. Figure 5 shows an example relating to the first issue (see Mairano & Romano, 2007a). In this example (from German) the syllabic sonorant (the second nasal segment) has been analysed as vocalic even though it has less energy than the other two surrounding nasals which are told apart as consonants on a phonological basis. Thus, we classified the syllabic sonorant in *seinen Mantel* as a V. But how many automatic tools used for segmentation would have detected it and classified it like that?
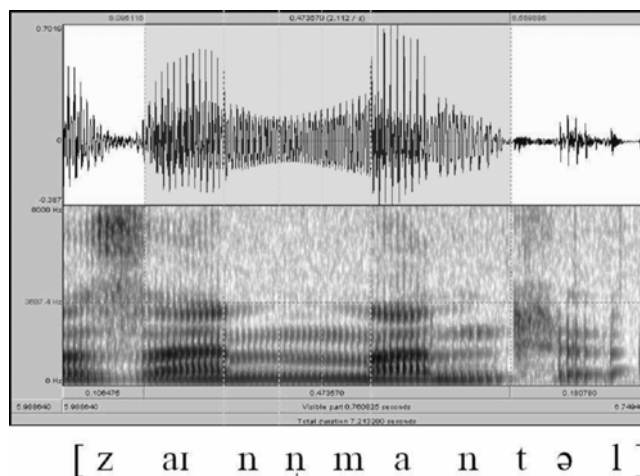

[z aɪ n n̩ m a n t ə l]

Figure 5: Which segmental cues justify the classification of a segment as V or C?

Another relevant question we raised is then: should rhythm be evaluated in terms of expectations or of real facts?

Still working on the traditional typology which classifies languages along the stress-/syllable-timed axis, we tried to assess how much the results might depend on operators' segmentation and classification choices during measurement tasks (see Figure 6; cp. Maira-no & Romano, 2007a & b). We confirmed a specific, sample-dependent, language conditio-ning and we discussed this topic at the same time. However, we enlarged the scope of

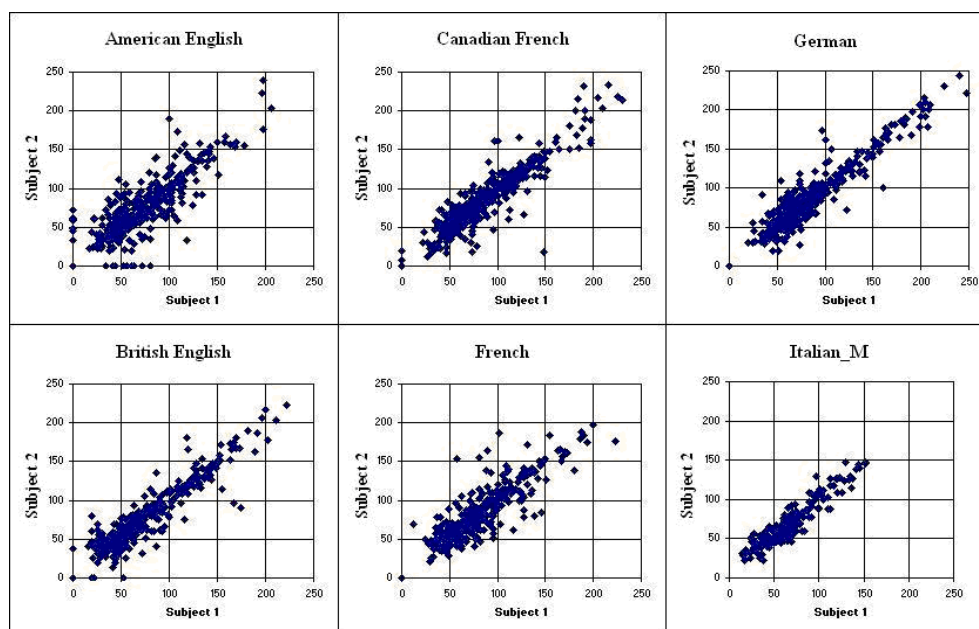similar arguments in a poster presented at the last *ICPhS* in Saarbrücken (see Mairano & Romano, 2007b).

Figure 6: How much do operators agree in segment classification and measurement?

In these plots (from different language samples) two operators separately judged the duration of intervals. Values are variously scattered (sometimes significantly) from the regression line. A few segments have been measured as consonants by one operator (especially /r/ realisations in American English) whereas the other one considered them as belonging to rhotacised and diphthongised vocalic nuclei (duration as C=0).

An aspect on which authors are not explicit when presenting their results is how the metrics are computed, whether on the entire production (which we call A-mode) or by averaging partial results on each interpausal interval (B-mode). The question is particular relevant when dealing with speech samples coming from spontaneous dialogues which are naturally segmented at least by turn-taking and because the two speech production lines are intertwined.

We tested the effects of the two ways of computing metrics in our speech make-up (continuous monologue productions) and we showed significant changes (see Figure 7 for a comparison between the deltas computed for the same six samples: the rhythmic topology of languages changes when switching from the A-mode (on the left) to the B-mode (on the right).[11]

---

[11] Plots are made with the software *Correlatore* (by P. Mairano, available at http://www.lfsag.unito.it/correlatore/index_en.html).
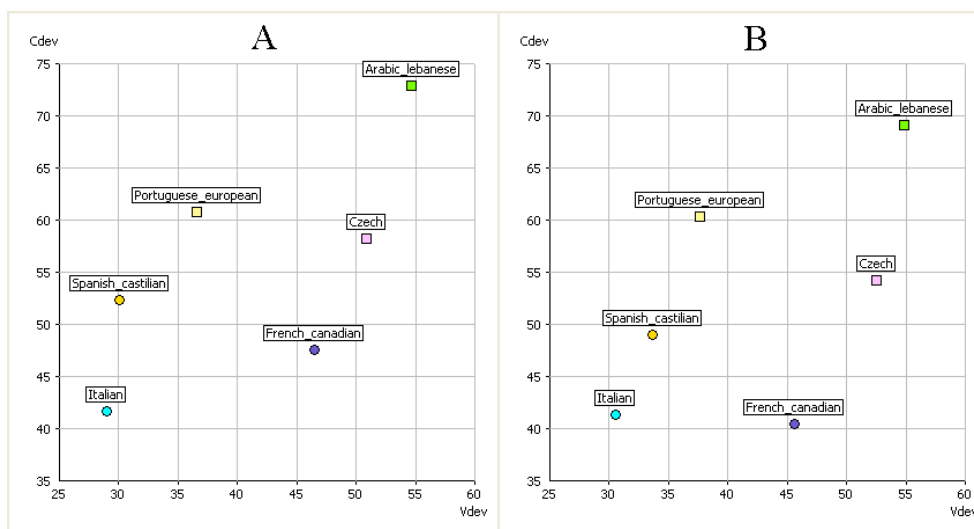
Figure 7: Effects of joint (A) vs. separate (B) statistic computations
for the metrics proposed by Ramus *et al.* (1999)

Differences have been tested also on samples uttered by different speakers for the same language. Sensible differences are detectable for languages such as Italian, for speakers from different regions, even when dealing with standard-like samples and have been confirmed for ten speakers of Icelandic (which has virtually no geographic variation at all) showing a fairly strong dependence on speaking styles.

The discussion about this point was hinted at in our previous papers, but final results were discussed at the EASR 2008 workshop in UCL (Mairano & Romano, unpublished, poster presentation), together with the issue about which spoken corpora should be used.

Further discussions took place in Granada during the presentation of Russo & Barry (2008a) about the kind of speech make-up on which to test models. The questions are:

1) Are we able to perceive linguistic rhythm by listening to the speech of our interlocutor?

2) If yes – as it must be, since we began discussing about this topic from Pike's and other linguists' intuitions –, in what kind of linguistic productions are we able to detect enough cues to classify the linguistic rhythm of our interlocutor and of his/her language?

We got fairly good results by computing the simple metrics proposed by Ramus *et al.* (1999) on the short narratives offered by the *IPA Illustrations* in which there are enough rhythmic cues in order to allow a trained listener to guess a rhythmic classification. 'Fairly good results' means that the positioning of languages in a continuous space is in accordance with listeners' intuitions.

5. SPEECH RHYTHM MODELLING

Speech rhythm modelling received considerable attention in the last years. In this section of my intervention, I made reference to some recent contributions, accounting for linear regression studies (summarised in Barbosa, 2006) and new multi-layer models (discussed in Bertinetto & Bertini, forthcoming). The duration of the stress group (Interstress Interval) is defined as a function of the number of syllables (*n*). The well known formula is the following one:

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

(1)

where *a* is a constant and *b* is a parameter describing the growing ratio of *I* versus *n*.
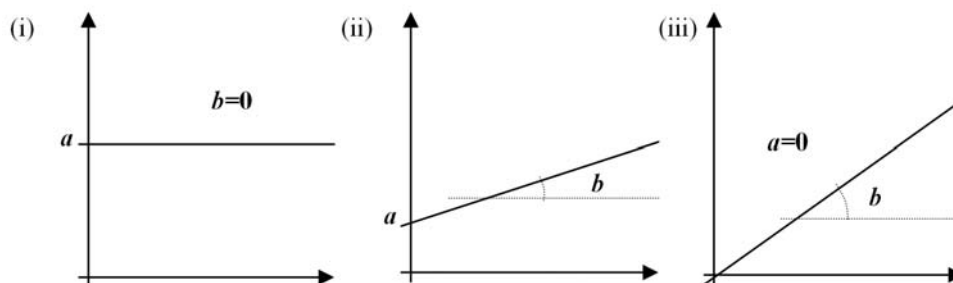

Figure 8: The growth of Interstress Intervals for (i) absolute stress-timed languages (on the left), (iii) for absolute syllable-timed languages (on the right) and (ii) for a mixed-timed language (in the mid)

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 Interstress Interval is a constant (*b=0* → *I=a*; see Figure 8 (i));

- an absolute syllable-timing, when *a* is naught and the Interstress Interval is directly proportional to the numer of syllables (*a=0* → *I=bn*; see Figure 8 (iii));

- but languages usually tend to show an intermediate way (see Figure 8 (ii)).

5.1 The double oscillator model

For long-term qualitative descriptions of language timing, a model has been re-proposed – see the relevant literature on previous studies, e.g. in Barbosa (2006) – who predicts temporal patterns as the result of the coupling of two oscillators (see O'Dell & Nieminen, 1999).

The duration of the Interstress 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 that *a*, *b* and *I* of the preceding equation are re-defined as in the following formula (where ω_1 is the oscillation velocity of the accentual oscillator, ω_2 is the oscillation velocity of the syllabic oscillator and *r* is the coupling strength):

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

(2)

When the value of the coupling strength (*r*) is 1, then the *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 (Fant, Eriksson and others cited 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* ≈ 0.9).

Barbosa (2006) tested the same mathematical model for different speech rates for Brazilian Portuguese finding values about 1.5. But for increasing speech rates *r* did not systematically decrease, thus not confirming the prediction of more syllable-timed behaviours of the same language for rapid tempos (see Dellwo & Wagner, 2003, for different results).

5.2 The Control-Compensation model

The Control-Compensation model (*CC*) proposed by Pier Marco Bertinetto and Chiara Bertini (2008, 2009 and forthcoming) finds its origins in earlier studies of '80s (see e.g. Bertinetto & Vékás, 1991) and reintroduces the double oscillator model in view of providing a unified, fully explicit and more predictive theory.

The model is grounded on the reformulation of language differences observed in terms of reduction of vowels (V) and consonants (C) in a gestural overlap hypothesis framework. That leads to a revisited dichotomy not involving the stress-/syllable-timing axis but contrasting more controlling languages (*CTRL*) vs. more compensating languages (*CMPS*).

The rhythmic metrics proposed aim at accounting for intrasyllabic durational stability vs. compression in a dynamic model inspired by the *PVI* model (see Grabe and Low, 2002). Therefore the *CCI* model introduces the number of segments (*n*) in the metrics.

The formulae defining the two indexes *CCI(V)* and *CCI(C)* are:

$$CCI(V) = \frac{100}{n_{I_V} - 1} \cdot \sum_{k=1}^{n_{I_V}-1} \left| \frac{d_k}{n_k} - \frac{d_{k+1}}{n_{k+1}} \right| \quad \text{and} \quad CCI(C) = \frac{100}{n_{I_C} - 1} \cdot \sum_{k=1}^{n_{I_C}-1} \left| \frac{d_k}{n_k} - \frac{d_{k+1}}{n_{k+1}} \right|$$

(4)   (5)

with n_{IV} and n_{IC} the numbers of Vocalic and Consonantal Intervals in the speech sample and n_k the number of segments in the *k* interval.

The CCIs are applied to two levels of organisation: a phonotactic level, called 'Level-I', which is based on the coupling of the vocalic and consonantal oscillators (as it is also suggested by Goldstein *et al.*, 2007); a phrasal level, called 'Level-II', which is based on the coupling of the accentual and syllabic oscillators (see O'Dell & Nieminen, 1999).

The formula defining the Interstress Interval was applied to the Level-I oscillator relating the duration of inter-V-onset intervals – from one V-onset to the next (as suggested by Keller & Port, 2007; Barbosa, 2006) – to the number of intervening consonants. The Intervocalic Interval is regulated by r_I.

When r_I is greater than 1 the vocalic oscillator prevails on the consonantal oscillator. This allows to predict that:

- the consonantal oscillator should emerge as dominant along with tempo increases, for the consonants comprised between two vocalic gestures cannot be compressed beyond a certain threshold, whereas vowels allow for more compression;

- in *CTRL* languages, however, due to the relative incompressibility of unstressed Vs, the vocalic oscillator should partly compensate this effect.

These predictions have all been confirmed, at present, in the simulation of Bertinetto & Bertini (forthcoming), even though with a number of adapted considerations for each sample analysed.

Nevertheless, the groupings in rhythmic classes have been reanalysed in terms of an interplay of Level-I and Level-II leading to four ideal groups and results are encouraging (see Table 1 and Bertinetto & Bertini, forthcoming, for a detailed discussion).

TYPE	LEVEL-I	LEVEL-II	EXAMPLE
1	CTRL	CTRL	*Italian*: relatively simple phonotactics, fairly rigid word stress pattern
2	CMPS	CMPS	*English*: fairly complex phonotactics, fairly mobile word stress pattern, density of secondary stresses yielding further prominence sites
3	CMPS	CTRL	*Polish*: complex phonotactics, fairly rigid word stress pattern
4	CTRL	CMPS	*Chinese*: simple phonotactics, uncertain word stress pattern

Table 1: The four ideal groups emerging from the interplay of Level-I and Level-II
(adapted from Bertinetto & Bertini, forthcoming)

Furthermore, the model fulfils a number of epistemological requirements and works as a good starting point for future improvements.

6. COMPARISON BETWEEN METRICS

Discussions on topics like the ones in the §4 are leading to our growing interest towards rhythm models. However, the research which has been carried out in Turin was not aimed until now to test models. We started to apply the delta calculations (see Ramus *et al.*, 1999) on duration measurements made on a multi-language sample, in order to verify how metrics based on duration of consonantal and vocalic intervals were good correlates of known or expected rhythmic types.

Those are the grounds on which we hope to have given a significant contribution until now, including in our measurements a relevant sample of languages and discussing how metrics capture language clustering around STB and SYB rhythmic poles.

6.1 The discriminating power of the deltas

We calculated %V, ΔV, ΔC for different translations of the *The North Wind and the Sun* read by several speakers. This choice has been made in order to focus on controlled

samples instead of spontaneous productions. We argue that wild auto-segmentation proce-dures should be avoided where possible since the gain in time is counterbalanced by a loss of precision. Moreover, we think there is no urge to get huge amounts of data, as simple listening tests suggest that humans need only a few seconds to distinguish between languages belonging to different rhythm classes (even when they do not know how to express that; see Ghazali et al., 2002).

We propose here a selection of 29 speakers (4 for German and Italian, 3 for English, 2 for (varieties) of each of the following languages: Chinese, French, Finnish, Icelandic, Portuguese and Romanian, and only one for Arabic, Czech, Japanese, Russian, Spanish and Turkish; see §5.3 and Appendix for details).

Language	%V	ΔV	ΔC
French_european	49,4	41,04	39,86
Romanian_muntenian	40,4	32,70	41,33
Italian_1	45,5	28,96	41,71
Finnish_2	46,7	48,99	43,10
Chinese_Mandarin	51,2	43,67	43,88
Icelandic_10	45,4	37,87	46,40
Icelandic_4	46,5	38,27	46,40
Italian_2	48,2	42,00	46,81
Finnish_1	48,6	52,90	46,93
Romanian_moldavian	43,4	40,29	47,44
French_canadian	51,2	46,51	47,57
Italian_3	46,3	39,62	48,65
Portuguese_brazilian (SP)	49,2	46,96	50,39
Italian_4	42,2	30,62	52,14
Spanish_castilian	42,0	30,07	52,31
Japanese_fast	46,0	35,87	53,62
English_GA	42,2	43,75	55,36
Japanese_slow	48,0	39,38	55,93
Chinese_Cantonese	46,7	51,36	57,47
Turkish	44,9	37,96	57,50
Russian	38,3	36,23	57,92
Czech	44,9	50,81	58,21
English_RP	40,7	51,82	58,85
English_Aus	40,6	42,71	59,66
Portuguese_european	43,9	36,60	60,77
German_IPA	46,9	47,00	60,79
Zurich_German_	50,0	52,68	65,57
German_1	46,1	53,66	65,93
German_2	42,8	47,30	69,85
Arabic_lebanese	46,8	54,69	72,92

Table 2: Delta values for the 30 language samples analysed (ranked on ΔC basis)

Results are fairly encouraging, allowing a reasonable matching between calculated values and impressionistic expectations. As we show in Table 2, samples rank quite well in two classes (in the upper half SYB languages and in the lower half STB languages; cp. plots Figures 9 and 10).[12] That happens if samples share the same general recordings quality and capture a similar degree of fluency and spontaneity for the represented language.
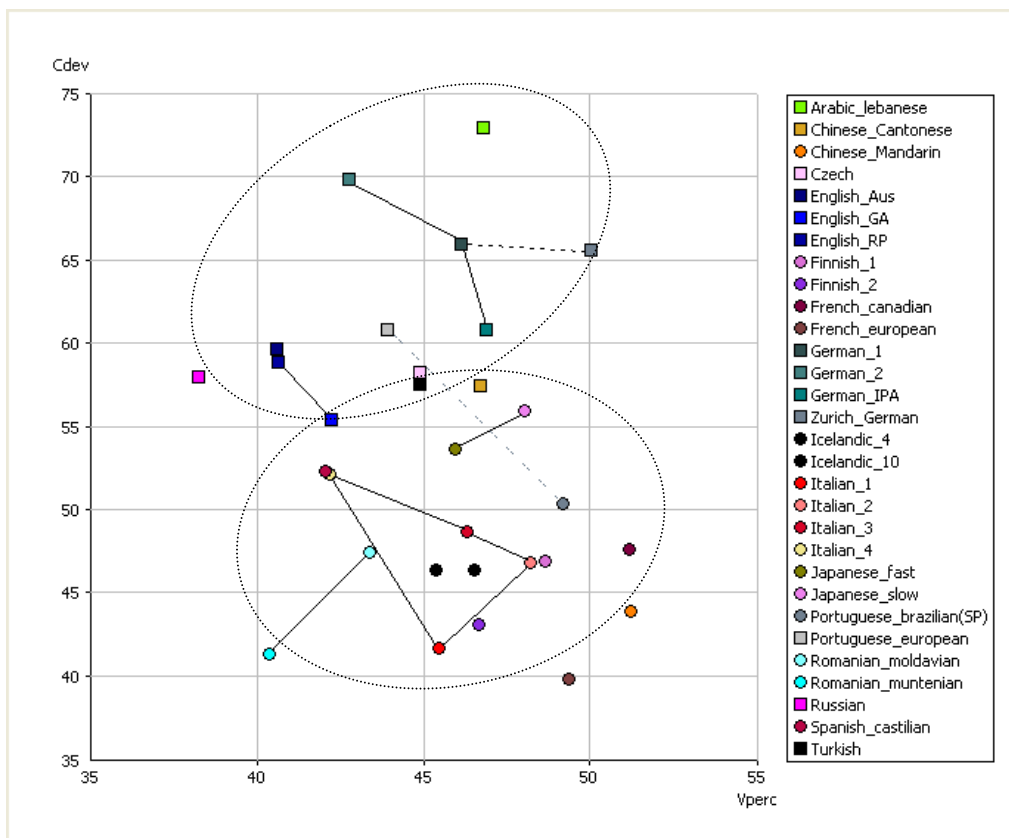

Figure 9: ΔC vs. %V plot for the language samples in Table 2

The metrics calculated for different samples of German (improperly including Zurich German), English, Italian, Portuguese, Romanian and Japanese have been connected in order to allow an easier reading for languages represented by more than one sample.

In some particular cases, one may ask whether the fact the e.g. Arabic ranks at the bottom of this list means that it is a stress-based language. In this respect we believe that the use of such labels per se does not constrain a judgement on the phonological level; only a phonetic indirect evaluation is concerned (further comments in §5.2).

---

[12] Plots are made with *Correlatore* (by P. Mairano, available at http://www.lfsag.unito.it/correlatore/index_en.html).
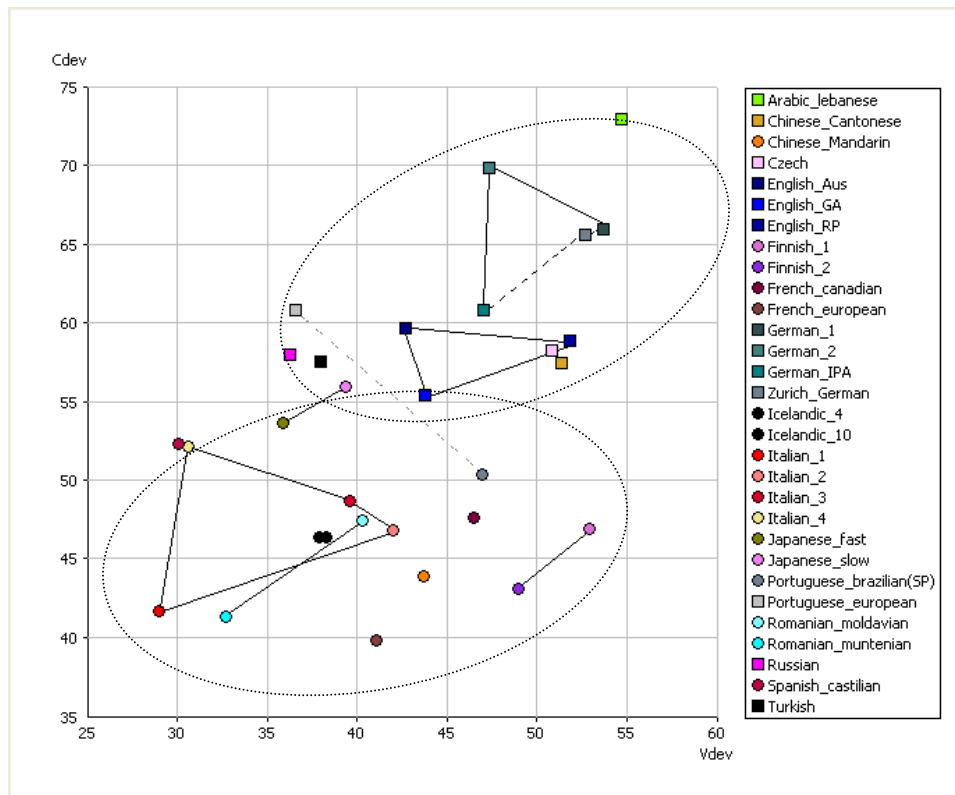
Figure 10: ΔC vs. ΔV plot for the language samples in Table 2[13]

We did not study the effects caused by speech rate (which have been extensively treated in Dellwo & Wagner, 2003, and others) as our samples are fairly homogeneous in this respect (5-6.5 syllables/s) as well as about prosodic segmentation (12-23 silent pauses). The only exception is Japanese which is represented by two samples realised at different speech rates (slower, 3.8 syllables/s, and faster, 4.6 syllables/s). We kept the two samples because – like for other samples we observed before (e.g. Italian in Mairano & Romano, 2007a) –, according to predictions in Dellwo & Wagner (2003), faster speech rates move the results of language metrics towards the SYB pole. In the case of our Japanese samples, which lie in the mid of the transition region along the SYB-STB continuum (against evidence proposed in other studies, e.g. Ramus *et al.*, 1999), this parameter determines a critical placement (cp. Barbosa, 2006). See Mairano & Romano (2010) and Figure 11 for a discussion of the sensitivity of metrics to a different treatment of devoiced close vowels in this language.

---

[13] The metrics calculated for different samples of German, English, Italian, Portuguese, Romanian and Japanese have been connected as in Figure 9.

6.2 Comparison between different metrics - Formulae

Our contributions are restricted to formulae adaptation and metrics comparison for a growing sample of languages (now including Romanian varieties, Czech, Russian, Turkish, Japanese and Chinese, see §5.1) with the definition of a robust evaluation procedure (see above). Like White *et al.* (2007), we believe that – depending on speech rate and style – rhythm metrics may yield rhythmic discrimination of languages along a continuum between the two poles and, perhaps better, in a multidimensional space.

As it is well known (cp. §5.1), Ramus *et al.* (1999) proposed three rhythm metrics (ΔC, ΔV and %V) which are intended to distinguish between two (or four) poles in the space of possible rhythmic organisations (in three charts: ΔC vs. %V, ΔV vs. %V and ΔC vs. ΔV).

Other rhythmic metrics have been proposed in order to decrease the sensitivity of these representation to speech rates and speech styles. The PVIs (*nPVI(V)* and *rPVI(C)* by Grabe & Low, 2002) and the Varcos (*VarcoV* and *VarcoC* by Dellwo & Wagner, 2003) attempt to normalise the effects of speech rate on rhythmic parameters, while the CCIs (controlling and compensating indexes, *CCI(V)* and *CCI(C)* by Bertinetto & Bertini, 2008), a modification of the PVIs, are an attempt to measure the degree of compensation that a language allows for and are inspired by previous work by Fowler (1977) and based on Bertinetto & Vékás (1991; see above §4.2).

We summarise here the formulae for those metrics that have been applied to consonantal and vocalic intervals:[14]

- Deltas (Ramus *et al.*, 1999)

$$\Delta V = \sqrt{\frac{\sum_{i=1}^{n_V}(D_{V_i} - \overline{D}_V)^2}{n_V - 1}} \quad \text{and} \quad \Delta C = \sqrt{\frac{\sum_{i=1}^{n_C}(D_{C_i} - \overline{D}_C)^2}{n_C - 1}}$$

(6) (7)

- Varcos (Dellwo & Wagner, 2003; Dellwo, 2006)

$$\text{VarcoV} = \frac{\Delta V \cdot 100}{\overline{D}_V} \quad \text{and} \quad \text{VarcoC} = \frac{\Delta C \cdot 100}{\overline{D}_C}$$

(8) (9)

---

[14] Dellwo *et al.* (2007) use voiced-unvoiced parameters which seem promising except for the fact that they are prone to show sensitivity to cases of voicing undecidibility and, on a phonological level, they need to be tested on those languages where voice contrasts are not pertinent. Gibbon & Gut (2001) refer to the RIM (Rhythmic Irregularity Measure) defined by Scott *et al.* (1986) and apply their *Rhythm Ratio* (RR, an early modification of the PVI) to syllables and vowels.

- PVIs (Grabe & Low, 2002)

$$nPVI(V) = 100 \cdot \frac{\sum_{k=1}^{n_{I_V}-1} \frac{|d_k - d_{k+1}|}{(d_k + d_{k+1})/2}}{n_{I_V} - 1} \quad \text{and} \quad rPVI(C) = \frac{\sum_{k=1}^{n_{I_C}-1} |d_k - d_{k+1}|}{n_{I_C} - 1}$$

(10)   (11)

- CCIs (Bertinetto & Bertini, 2008)

$$CCI(V) = \frac{100}{n_{I_V} - 1} \cdot \sum_{k=1}^{n_{I_V}-1} \left| \frac{d_k}{n_k} - \frac{d_{k+1}}{n_{k+1}} \right| \quad \text{and} \quad CCI(C) = \frac{100}{n_{I_C} - 1} \cdot \sum_{k=1}^{n_{I_C}-1} \left| \frac{d_k}{n_k} - \frac{d_{k+1}}{n_{k+1}} \right|$$

(12)   (13)

6.3 Comparison between different metrics - Plots

Following an early evaluation based on deltas, we aimed at testing the variation of results according to a number of factors (as discussed above): we also calculated VarcoV, VarcoC, nPVI(V), rPVI(C), CCI(V) and CCI(C) for the language samples presented in §5.1.

All the metrics yielded to a certain amount of overlapping between rhythm classes and each of them showed sensitivity to slightly different phenomena related to speech timing (cp. Barry & Russo, 2004; Dellwo, 2008). Results are anyway encouraging, confirming our impressions on single samples (even when contrasting with other authors' predictions on rhythm from general features or specific measurements).[15]

In Figure 11 we have an example of a comparison between charts based on different metrics. In particular we may observe how languages are mapped on the CCI map in relation to the bisecting line, along which syllable-timed language are usually placed.

---

[15] Arbitrarily stating that all Germanic languages are stress-timed (against our impression on Icelandic, see Figure 9), Fikkert *et al.* (2004) surprisingly expect European Portuguese to be a syllable-timed language. The same happens for Czech (our Czech sample sounds more STB, against evidence summarised in Dankovicová & Dellwo, 2007; also see Volín, 2005). Furthermore, our sample of Cantonese lies in the STB area in agreement with the results published by Grabe and Low (2002) (*pace* Mok and Dellwo, 2008; also cp. Jian, 2004). As for the alleged inclusion of Arabic varieties in the STB class see Ghazali, Hamdi & Barkat (2002), even though ΔC gets aberrant values in some tables.
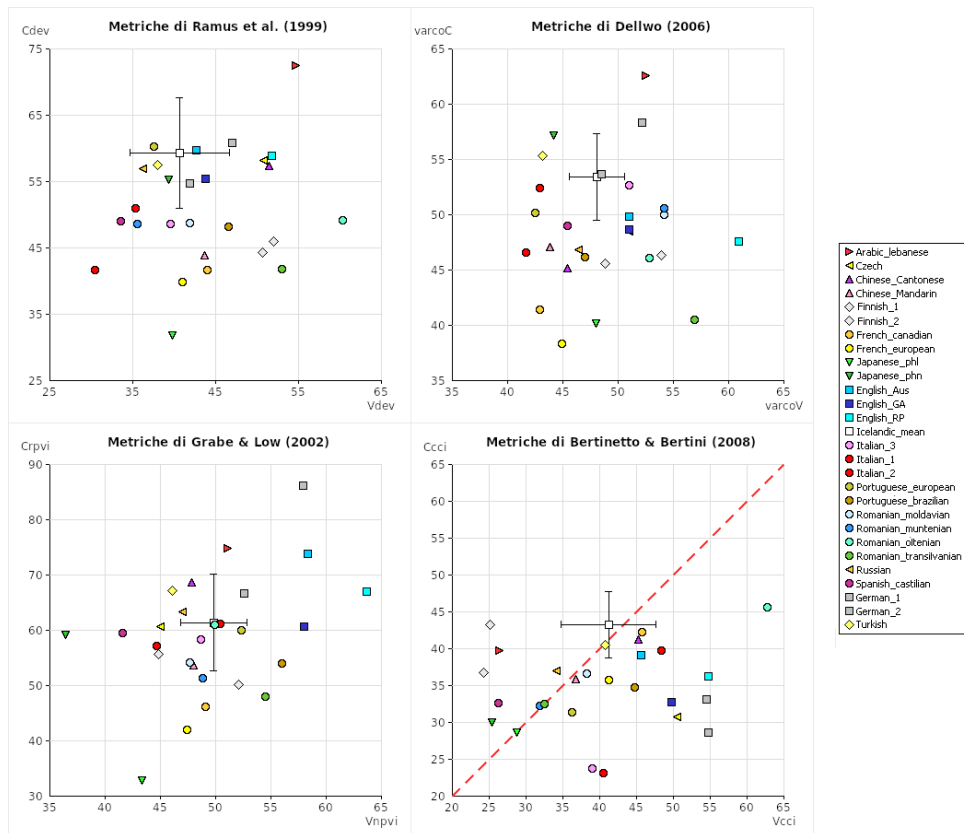
Figure 11: Plots for the same language samples with different metrics
(adapted from Mairano & Romano, 2010)

Dividing the Consonantal and Vocalic Intervals by the number of segments has a correction effect on languages with consonant gemination (pushing a couple of Italian samples towards more STB regions) and on languages with long vowels (occasionally with gemination too; thus dramatically changing the placement of Arabic and Finnish).

As for Italian varieties, we started to analyse data from various dialects. General rhythmic properties of Romance languages are discussed in Mairano & Romano (2009) whereas results on a selection of Piedmontese varieties are presented in Romano, Mairano & Pollifrone (2010).

7. NOT ONLY DURATION

Generally speaking, speech rhythm represents a complex prosodic phenomenon resulting from the co-operative effect of several elements determining strong-weak alternations and it is influenced by various factors (from timing to intonation, within a smaller or a larger scale). Regularity or irregularity in a sequence of pulses may be reflected in sequences of prominences raised by different parameters, among which we mainly expect

to find local variations in the energy, local peaks or movements in the pitch curve and alternating duration patterns (Allen, 1975).

The increasing interest in temporal correlates of rhythm, which seems reasonable above all in comparison with music rhythm, has recently limited the attention of some researchers to these facts and has brought them to neglect other parameters.

As for the questions raised on this topic by the chairman of the Zurich Round Table, Prof. W. Barry, a selection of relevant points have been discussed, among which I chose to answer to the following ones:

- Should rhythm measures be limited to duration? If so, why?

- If not, which other parameters should be included?

In Turin, since the beginnings, we have been wondering whether to include other acoustic correlates in our assessments of speech rhythm (see Mairano & Romano, this volume).

A convincing experiment we have attempted is based on a selection of manipulated sound samples that can be used for a simple informal listening test.

Synthetic stimuli are obtained from two original speech samples using the *AMPER-dat* scripts (for Matlab) which I designed during my *PhD* and which are now adopted within the *AMPER* project testing procedures (see references, *AMPER*).

The two natural speech samples we manipulated here are extracted from the two narratives published for English (RP) and French (Parisian) by the *IPA* (see Roach, 2004, and Fougeron & Smith, 1999). The original utterances are *And at last, the Northwind gave up the attempt* for English (see Figure 12) and *Finalement, elle renonça à le lui faire ôter* for French. The manipulation technique is applied to an already stylised version of the main prosodic content of the sound sample (mainly based on sequences of values for the three parameters, energy, f_0 and duration, for all its vowels).
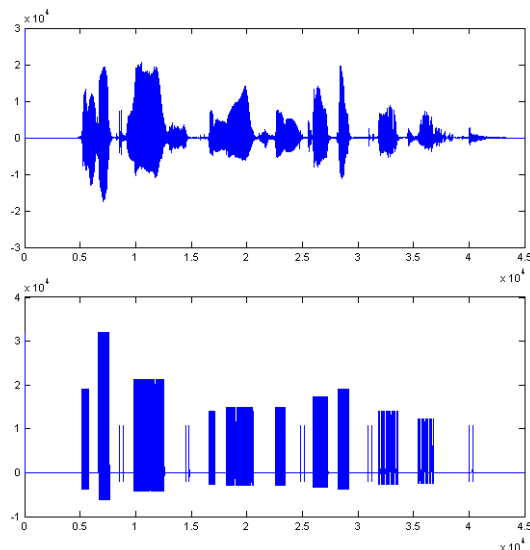

Figure 12: Amplitude plots of the original speech sound sample (up)
and the corresponding *.ton* sound file (down)

Figure 12 shows the contrast between the original speech sound sample and the corresponding *.ton* sound file, the latter being a sound sample alternating series of pulses (generated with duration, energy and pitch of each vowel in the original sample) and silences (occasionally broken by isolated pulses representing inner bursts in clusters of obstruents; see Table 3 for an example).

IPA_englishRP_narrative7_F.txt			size: 45725		11-feb-09
	duration [ms]	energy [dB]	fo1	fo2	fo3 [Hz]
1	35	78	263	260	267
2	57	82	301	288	256
3	0	0	50	50	50
4	165	79	185	178	189
5	0	0	50	50	50
6	29	74	204	192	187
7	148	75	174	173	172
8	53	75	195	179	170
9	0	0	50	50	50
10	79	77	171	180	185
11	59	78	208	204	189
12	0	0	50	50	50
13	106	74	144	152	143
14	84	71	154	137	121
15	0	0	50	50	50

values at:
5135 5414 5694 6646 7099 7552 8534 8534 8534 9834 11156 12478 14471 14471
14471 16606 16840 17075 18132 19318 20504 22529 22952 23375 24807 24807
24807 25937 26571 27206 28203 28672 29140 30886 30886 30886 31845 32695
33545 35395 36068 36740 39975 39975 39975

Table 3: An example of text file meant to generate a *.ton* file[16]

Figure 13 shows a comparison between the original prosodic content of a sound sample, in this case the same as in Figure 12 and its stylised version. The amplitude display (up) is associated with the natural course of f_0 (mid) and its close-copy stylisation obtained with the *AMPER-fox* scripts for Matlab (down). Dots on the lower line represent bursts in clusters of obstruents; the plot is based on values of Table 3.

At its origins, this stylisation procedure shares the basic assumptions of the best known close-copy stylisation defined by 't Hart, Collier & Cohen (1990). It provides a synthetic approximation of the natural course of the three prosodic parameters (see Figure 13 for f_0 and duration), meeting two criteria: the prosody of the final sample should be perceptually indistinguishable from the original, and it should be based on the smallest possible number of values stored in a text file (see an example in Table 3).

---

[16] Segments 3, 5, 9, 12, 15 are isolated pulses, conventionally indicated with 50 Hz values in pitch columns, reproducing internal bursts within C clusters or final (pre-pausal) bursts.
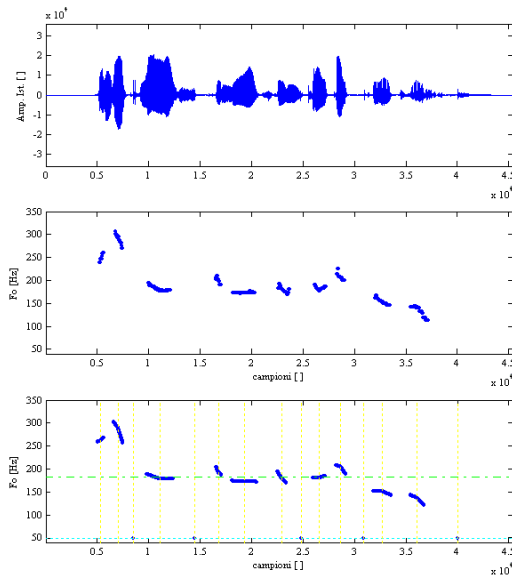
Figure 13: A comparison between the original prosodic content of a sound sample
and its stylised version

Our listening tests are based on *.ton* sound files which are series of pulses synthesised from a text file of such kind.

After the two natural sound files for the two samples above (English, supposed STB, and French, supposed SYB),[17] we propose in Figure 14 a selection of manipulated sound files obtained with fixed values for one of the three parameters for all the vowels in the original speech sample.

In Figure 14a. one may listen to the original speech samples and to the synthesised versions with close-copy stylisation of their prosodic content (see the corresponding diagrams for the three parameters D, E, P).

In Figure 14b. one may listen to the *.ton* files with manipulated duration: duration of all the vocalic nuclei in both the original speech samples are fixed to 100 ms (or fixed to the mean value of each sample, which is more natural but keeps the original distinct tempo for both of them). Surprisingly, one notices that, even when the information on the duration of vowels (shown by the higher bars in the diagrams) is lost, the distinct rhythm of the two samples is still present.

Sound files (and diagrams) in Figure 14c. demonstrate the very little contribution of energy in rhythm perception: the lower bars in the diagrams show the normalised E values: their rhythmic characterisation is very similar to that of the original samples.

---

[17] Calculations of deltas on these short samples give about %V=37, ΔV=46 and ΔC=80 for the English utterance (which sounds strongly STB and is therefore well told apart by these metrics) and %V=47, ΔV=18 and ΔC=35 for the French utterance (also prototypically SYB). In this case, we do not care about the languages concerned but simply about the rhythmic properties present in the samples.
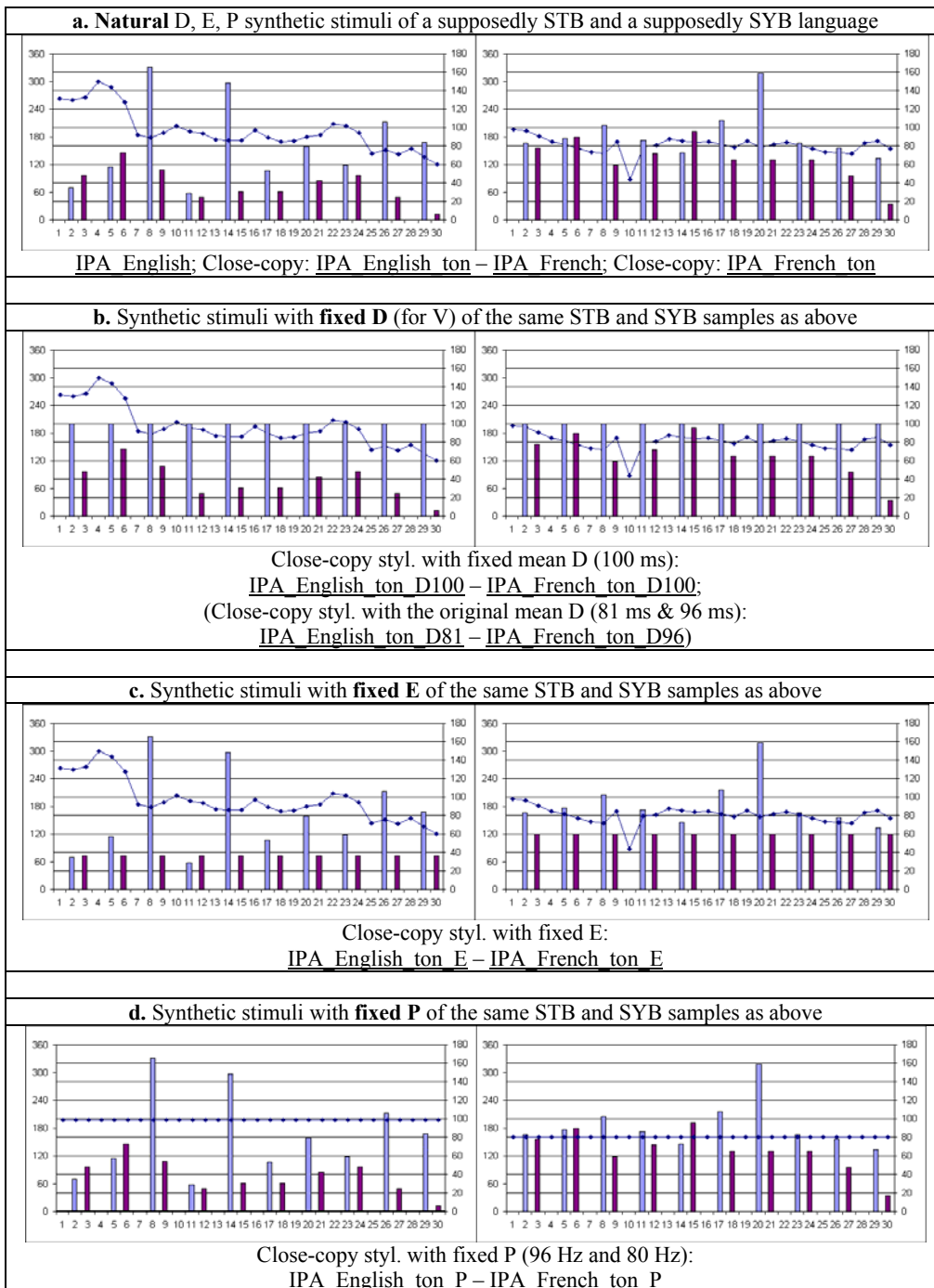
Figure 14: Curves and sounds for a simple listening test giving an insight into the role of the three different parameters D (duration), E (energy) and P (pitch) in rhythm perception

The loss of pitch variations, which can be experience by listening to the synthetic samples in Figure 14d. (with monotonised f_0 respectively at 96 Hz and 80 Hz), seems to prove the relevance of this parameter in rhythmic judgements: the flattening of f_0 information causes a dramatic loss in the different rhythmic characterisation of the two samples.

A simple experiment like this allowed us to start thinking in a different way with regard to speech rhythm:

(1) it demonstrates the inadequacy of metrics based on durations only;

(2) the reduced importance of vocalic durations (observed by listening to the stimuli in Figure 14b.) suggests the possibility that the distance in time between f_0 peaks or specific movements could be one of the main cues in listening discrimination of different rhythmic types (cp. *P-centre* methods).

a. Synthetic stimuli with **fixed D** (for V and C) of the STB and SYB samples as in *Table IV*

Close-copy styl. with the original mean D for V & C (81 ms & 96 ms):
IPA_English_ton_D81_VC – IPA_French_ton_D96_VC

b. Synthetic stimuli with **fixed D** (for V and C clusters) of the two samples (cp. above)

Close-copy styl. with the original mean D for V & C clusters (81 ms & 96 ms):
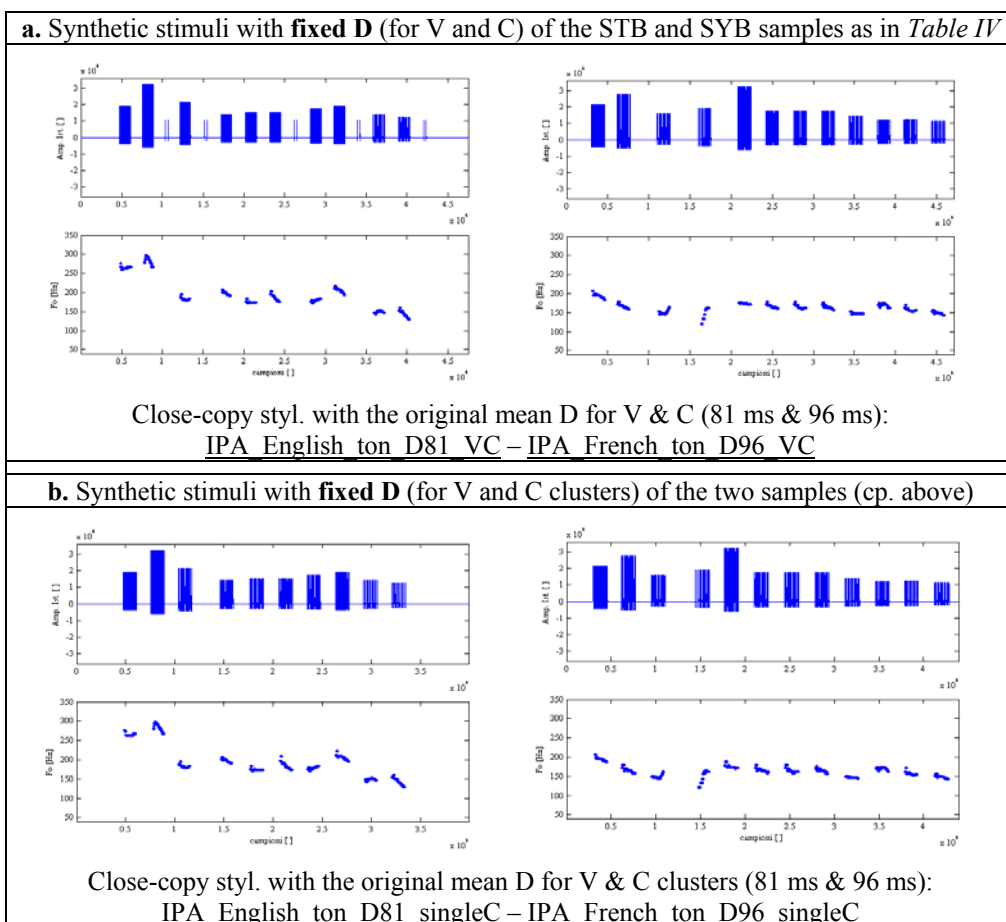IPA_English_ton_D81_singleC – IPA_French_ton_D96_singleC

Figure 15: Sounds for a further listening test giving an insight
into the role of pitch and duration of V and C.

This hypothesis could be partially invalidated if, even after normalising C durations to a fixed value and thus altering the distances between vocalic pulses, a distinct rhythmic characterisation still persists (listen to the stimuli in Figure 15a.). The possibility of perceptively distinguishing between the two rhythm types is definitely reduced only when the durational information of consonant clusters is limited (listen to stimuli in Figure 15b).

8. CONCLUSIONS

In this intervention I summarised early contributions to rhythm measurement and assessment carried out by the laboratories of Turin and Pisa. I also tried to briefly illustrate the recent attention reserved in Turin to rhythm perception (a listening experimental protocol is being defined by P. Mairano for his *PhD*) in order to include the apparently relevant role of f_0 in speech rhythm evaluation ("otherwise we have a duration model, but not a rhythm model" as stated by Gibbon & Gut, 2001: 96).

Preliminary testing showed that duration could be a good correlate, giving a physical estimate of rhythmic (maybe derived) properties, but we believe that most direct correlates should take into account (distance, extension and shape of) peaks and general melodic profiles.

Linguists' intuitions about speech rhythm are mainly based on perceived or expected properties (often perhaps heavily influenced by the knowledge of phonological features; see e.g. Canepari, 2006, and Ghazali *et al.*, 2002, for examples on how impressionistic evaluations could match experimental/instrumental research).

Even though common people perhaps associate perceived rhythmic properties to stress occurrence conditions or relate it to an impressionistic evaluation of faster vs. slower (or rapid changes in) tempo, we believe that the general intuitions could help to understand the basis of a well assessed dichotomy between two basic rhythm types.

The need to better investigate the relations between stress, intonation and rhythm – which also arises from the simple experiment we proposed in §6 – is also expressed by Giordano (2008). While various authors are addressing their interests in extending the assessment dimensions to other variables (as Lee & McAngus Todd, 2004, with intensity and rhythmograms), other promising results should perhaps derive from different ways to calculates metrics (as it is proposed by Mok & Dellwo, 2008, with DeltaS, or by Gibbon & Gut, 2001, with the rhythm ratio, both accounting for syllable durations).

One of the perspectives for research in this field is to improve rhythm metrics in order to integrate them into a rhythm model (cp. Barbosa, 2006), possibly a multi-layer one, where durational properties could be associated with strong-weak measures or other stress/syllable properties at different levels (as proposed by Bertinetto & Bertini, 2008 and forthcoming; or on the wake of Gibbon & Gut, 2001, distinguishing focal and non-focal components) merge high level (linguistic) information with measures of more than one parameter (namely pitch, duration and perhaps intensity).

9. REFERENCES

AMPER (*Atlas Multimédia Prosodique de l'Espace Roman*), http://www.lfsag.unito.it/amper.html

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.

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

Barbosa, P.A. & Albano, E.C. (2004), Brazilian Portuguese, *Journal of the International Phonetic Association*, 34, 227-232 (sound samples available on-line for the *IPA* members: http://web.uvic.ca/ling/resources/ipa/members).

Barry, W. & Russo, M. (2004), Isocronia soggettiva o oggettiva? Relazioni tra tempo articolatorio e quantificazione ritmica, in *Il parlato italiano* (F. Albano Leoni, F. Cutugno, M. Pettorino & R. Savy, editors), Atti del Convegno nazionale di Napoli, 13-15 febbraio, 2003, Napoli: D'Auria (CD-ROM).

Barry, W.J., Andreeva, B., Russo, M., Dimitrova, S. & Kostadinova, T. (2003), Do rhythm measures tell us anything about language type?, in *Proceedings of the 15th International Congress of Phonetic Sciences*, Barcelona, Spain, 2693-2696.

Bertinetto, P.M. (1977), *Syllabic Blood* ovvero l'italiano come lingua ad isocronismo sillabico, *Studi di Grammatica Italiana*, 6, 69-96.

Bertinetto, P.M. (1983), Ancora sull'italiano come lingua ad isocronia sillabica, in *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*, 91-92-93, 99-130.

Bertinetto, P.M. (1990), Coarticolazione e ritmo nelle lingue naturali, *Rivista Italiana di Acustica*, XIV, 69-74.

Bertinetto, P.M. & Bertini, C. (2008), On modeling the rhythm of natural languages, in *Proceedings of the 4th International Conference on Speech Prosody*, Campinas, Brazil, May 6-9, 2008, 427-430.

Bertinetto, P.M. & Bertini, C. (forthcoming), Towards a unified predictive model of Speech Rhythm, Manuscript.

Bertinetto, P.M. & Magno Caldognetto, E. (1993), Ritmo e intonazione, in *Introduzione all'italiano contemporaneo. Le strutture* (A.A. Sobrero, editor), Roma-Bari: Laterza, 141-192.

Bertinetto, P.M. & Vékás, D. (1991), Controllo vs. compensazione sui due tipi di isocronia, in *L'interfaccia tra fonologia e fonetica* (E. Magno Caldognetto & P. Benincà, editors), Padova: Unipress, 155-162.

Bertini, C. & Bertinetto, P.M. (2009), Prospezioni sulla struttura ritmica dell'italiano basate sul corpus semispontaneo AVIP/API, in *La fonetica sperimentale. Metodo e applicazioni* (L. Romito, V. Galatà & R. Lio, editors), Atti del 4° Convegno Nazionale dell'Associazione Italiana di Scienze della Voce, Arcavacata di Rende (CS), 3-5 Dicembre 2007, Torriana: EDK Editore, 3-21.

Canepari, L. (2006), *Avviamento alla fonetica*. Torino: Einaudi.

Canepari, L. (2004-), *Fonetica Naturale – Natural Phonetics* (on-line sound samples on Italian pronunciation: http://www.unive.it/canepari [last accessed 30/06/2009]).

Costamagna, L. (2000), *Insegnare e imparare la fonetica*. Torino: Paravia scriptorium.

Crystal, D. (1994), Documenting rhythmical change, in *Studies in general and English phonetics* (J. Windsor Lewis, editor), London: Routledge, 174-179.

Dankovicová, J. & Dellwo, V. (2007), Czech speech rhythm and the rhythm class hypothesis, in *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrücken, Germany, August 6-10, 2007, 1241-1244.

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

Dellwo, V. (2008), The role of speech rate in perceiving speech rhythm, in *Proceedings of the 4th International Conference on Speech Prosody*, Campinas, Brazil, May 6-9, 2008, 155-158.

Dellwo, V. & Wagner, P. (2003), Relations between language rhythm and speech rate, in *Proceedings of the 15th International Congress of Phonetics Sciences*, Barcelona, Spain, 471-474.

Dellwo, V., Fourcin, A. & Abberton, E. (2007), Rhythmical classification of languages based on voice parameters, in *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrücken, Germany, August 6-10, 2007, 1129-1132.

Demuth, K. (1996), The prosodic structure of early words, in *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition* (J. Morgan & K. Demuth, editors), Mahwah, N.J.: Lawrence Erlbaum Associates, 171-184.

Demuth, K. (2003), The status of feet in early acquisition, in *Proceedings of the 15th International Congress of Phonetic Sciences*, Barcelona, Spain, 151-154.

Eriksson, A. (1991), *Aspects of Swedish Speech Rhythm*, Göteborg: University of Göteborg (Monographs in Linguistics, 9).

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.

Fikkert, P. (2007), Acquiring phonology, in *Handbook of phonological theory* (P. de Lacy, editor), Cambridge: Cambridge University Press, 537-554.

Fikkert, P., Freitas, M.J., Grijzenhout, J., Levelt, Cl. & Wauquier S. (2004), Syllabic Markedness, Segmental Markedness, Rhythm and Acquisition, Talk presented at GLOW 2004 (Thessaloniki, Greece, 2004), http://www.del.auth.gr/GLOW2004/abstracts/ Fikkert_etal.pdf [last accessed 30/06/2009].

Fleischer J. & Schmid S. (2006), Zurich German, *Journal of the International Phonetic Association*, 36, 243-253 [sound samples available on-line for the *IPA* members: http://web.uvic.ca/ling/resources/ipa/members].

Fowler, C. (1977), *Timing control in speech production*, Bloomington: Indiana University Linguistic Club.

Frota, S., Vigário, M. & Martins, F. (2002), Language Discrimination and Rhythm Classes: Evidence from Portuguese, in *Proceedings of Speech Prosody 2002* (B. Bel & I. Marlien, editors), Aix-en-Provence, France, 315-318.

Frota, S., Vigário, M. & Freitas, M.J. (2003), From Signal to Grammar: Rhythm and the Acquisition of Syllable Structure, in *Proceedings of the 27th annual Boston University Conference on Language Development* (B. Beachley, A. Brown & F. Conlin, editors), Somerville, MA: Cascadilla Press, 809-821.

Fougeron, C. & Smith, C.L. (1999), French, in *Handbook of the International Phonetic Association,* Cambridge: Cambridge University Press, 78-81.

Ghazali, S., Hamdi, R. & Barkat M. (2002), Speech Rhythm Variation in Arabic Dialects, in *Proceedings of Speech Prosody 2002* (B. Bel & I. Marlien, editors), Aix-en-Provence, France, 331-334.

Gibbon, D. & Gut, U. (2001), Measuring speech rhythm, in *Proceedings of Eurospeech 2001*, Aalborg, Denmark, 95-98.

Giordano, R. (2008), On the phonetics of rhythm of Italian: patterns of duration in pre-planned and spontaneous speech, in *Proceedings of the 4th International Conference on Speech Prosody*, Campinas, Brazil, May 6-9, 2008, 74-77.

Goldstein, L., Chitoran, I. & Selkirk, E. (2007), Syllable structure as coupled oscillator modes: Evidence from Georgian vs. Tashlhiyt Berber, in *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrücken, Germany, August 6-10, 2007, 241-244.

Grabe, E. (2002), Variation Adds to prosodic Typology, in *Proceedings of Speech Prosody 2002* (B. Bel & I. Marlien, editors), Aix-en-Provence, France, 127-132.

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

't Hart, J., Collier, R. & Cohen, A. (1990), *A perceptual study of intonation*, Cambridge: Cambridge University Press.

Hayes, B. (1984), The Phonology of Rhythm in English, *Linguistic Inquiry*, 15, 33-74.

IPA (1999), *Handbook of the International Phonetic Association*, Cambridge: Cambridge University Press (sound samples available on-line: http://web.uvic.ca/ling/resources/ipa/handbook.htm).

Jian, H. (2004), On the syllable timing in Taiwan English, in *Proceedings of Speech Prosody 2004*, Nara, Japan, 247-250.

Keller, E. & Port, R. (2007), Speech timing: Approaches to speech rhythm, *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrücken, Germany, August 6-10, 2007, 327-329.

Krull, D. & Engstrand, O. (2003), Speech rhythm – intention or consequence? Cross-language observations on the hyper/hypo dimension, *Phonum*, 9, 133-136, http://www.ling.su.se/fon/perilus/2003_10.pdf (last accessed 30/06/2009).

Lee, C.S. & McAngus Todd, N. (2004), Towards an auditory account of speech rhythm: application of a model of the auditory 'primal sketch' to two multi-language corpora, *Cognition*, 93, 225-254.

Lenneberg, E. (1967), *Biological foundations of language,* New York: Wiley.

Liberman M. & Prince A. (1977), On Stress and Linguistic Rhythm, *Linguistic Inquiry*, 8, 249-336 [now also in *Phonology: Critical concepts* (Ch.W. Kreidler, editor), London-New York: Routledge, 2001, 152-244].

Lindblom, Bj. & Rapp, K. (1973), Some temporal regularities of spoken Swedish, *Papers in Linguistics from the University of Stockholm*, 21, 1-59.

Lloyd James, A. (1940), *Speech signal in telephony*, London: Pitman & Sons.

MacNeilage, P.F. & Davis, B.L. (1990), Acquisition of speech production: Frames, then content, in *Attention and Performance XIII - Motor Representation and Control* (M. Jeannerod, editor), Hillsdale, NJ: LEA, 452-468.

MacNeilage, P.F. (1998), The frame/content theory of evolution of speech production, *Behavioral and Brain Sciences*, 21, 499-546.

Mairano, P. & Romano, A. (2007a), Lingue isosillabiche e isoaccentuali: misurazioni strumentali su campioni di italiano, francese, inglese e tedesco, in *Scienze vocali e del linguaggio. Metodologie di valutazione e risorse linguistiche* (V. Giordani, V. Bruseghini & P. Cosi, editors)*,* Atti del 3° Convegno Nazionale dell'Associazione Italiana di Scienze della Voce, 29 novembre – 1 dicembre 2006, Povo (Trento), Torriana (RN): EDK editore, 119-134.

Mairano, P. & Romano, A. (2007b), Inter-Subject Agreement in Rhythm Evaluation for Four Languages (English, French, German, Italian), in *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrücken, Germany, August 6-10, 2007, 1149-1152.

Mairano, P. & Romano, A. (2008a), *A comparison of four rhythm metrics for six languages,* Poster presented at the workshop 'Empirical Approaches to Speech Rhythm' (EASR), University College London, 2008.

Mairano, P. & Romano, A. (2008b), Distances rythmiques entre variétés romanes, in *La variation diatopique de l'intonation dans le domaine roumain et roman* (A. Turculeţ, editor), Proceedings of the international symposium, Iaşi, Romania, 2008, Iaşi: Univ. A.I. Cuza, 251-272.

Mairano, P. & Romano, A. (2010), Un confronto tra diverse metriche ritmiche usando Correlatore, in *La dimensione temporale del parlato* (S. Schmid, M. Schwarzenbach & D. Studer, editors), Atti del 5° Convegno Nazionale dell'Associazione Italiana di Scienze della Voce, Zurigo, Svizzera, 4-6 febbraio 2009 (this volume).

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

Martinéz Celdrán, E., Fernández Planas, A.M. & Carrera Sabaté, J. (2003), Castilian Spanish, *Journal of International Phonetic Association*, 33, 255-259 (sound samples available on-line for the *IPA* members: http://web.uvic.ca/ling/resources/ipa/members).

Mehler, J., Dommergues, J., Frauenfelder, U. & Segui, J. (1981), The syllable's role in speech segmentation, *Journal of Verbal Learning and Verbal Behavior*, 20, 298-305.

Mendicino, A. & Romito, L. (1991), «Isocronia» e «base di articolazione»: uno studio su alcune varietà meridionali, *Quaderni del Dipartimento di Linguistica dell'Università della Calabria*, S. L. 3, 49-67.

Miller, M. (1984), On the perception of rhythm, *Journal of Phonetics*, 12, 75-83.

Mok, P.P.K. & Dellwo, V. (2008), Comparing native and non-native speech rhythm using acoustic rhythmic measures: Cantonese, Beijing Mandarin and English, in *Proceedings of the 4th International Conference on Speech Prosody*, Campinas, Brazil, May 6-9, 2008, 63-66.

Molinu, L. & Romano, A. (1999), La syllabe dans un parler roman de l'Italie du Sud (variété salentine de Parabita – Lecce), in *Actes du Colloque des Journées d'Etudes Linguistiques:'SyllabeS',* Nantes, 25-27 mars 1999*, Nantes, France, 148-153.

Nazzi, T., Bertoncini, J. & Mehler, J. (1998), Language Discrimination by Newborns: towards an understanding of the role of rhythm, *Journal of Experimental Psychology: Human Perception and Performance*, 24, 756-766.

Nespor, M. (1993), *Fonologia*, Bologna: Il Mulino.

Nooteboom, S. (1997), The prosody of speech: melody and rhythm, in *The Handbook of Phonetic Sciences* (W.J. Hardcastle & J. Laver, editors), Oxford: Blackwell, 640-673.

O'Dell, M. & Nieminen, T. (1999), Coupled oscillators model of speech rhythm, in *Proceedings of the 14th International Congress of Phonetic Sciences*, San Francisco, USA, August 1-7, 1999, 1075-1078.

Pamies Bertrán, A. (1999), Prosodic Typology: On the Dichotomy between *Stress-Timed* and *Syllable-Timed* Languages, *Language Design*, 2, 103-130.

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

Ramus, F. (2002), Acoustic Correlates of Linguistic Rhythm: Perspectives, in *Proceedings of the International Conference Speech Prosody 2002* (B. Bel & I. Marlien, editors), Aix-en-Provence, France, 115-120.

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

Rhardisse, N. & Abry, C. (1995), Mandible as syllable organizer, in *Proceedings of the 13th International Congress of Phonetic Sciences,* Stockholm, Sweden, 3, 556-559.

Roach, P. (1982), On the distinction between 'stress-timed' and 'syllable-timed' languages, in *Linguistic Controversies* (D. Crystal, editor), London: Arnold, 73-79.

Roach, P. (editor) (2003), *A Bibliography of Timing and Rhythm in Speech* (last updated 2nd April 2003), University of Reading, http://www.personal.rdg.ac.uk/~llsroach/timing.pdf [last accessed 30/06/2009].

Roach, P. (2004), British English: Received Pronunciation, *Journal of the International Phonetic Association*, 34, 239-245 (sound samples available on-line for the *IPA* members: http://web.uvic.ca/ling/resources/ipa/members).

Romano, A. (1999), *Analyse des structures prosodiques des dialectes et de l'italien régional parlés dans le Salento: approche linguistique et instrumentale*, Thèse de Doctorat de l'Université Stendhal de Grenoble (superv. Michel Contini) (part. published in 2001, Lille: Presses Univ. du Septentrion).

Romano, A. (2003), Accento e intonazione in un'area di transizione del Salento centro-meridionale, in *Storia politica e storia linguistica dell'Italia meridionale* (P. Radici Colace, G. Falcone & A. Zumbo, editors), *Atti del convegno internazionale di studi parlangeliani*, Messina, Italy, 2000, Messina-Napoli: Edizioni Scientifiche Italiane, 169-181.

Romano, A., Mairano, P. & Pollifrone, B. (2010), Variabilità ritmica di varietà dialettali del Piemonte in *La dimensione temporale del parlato* (S. Schmid, M. Schwarzenbach & D. Studer, editors), Atti del 5° Convegno Nazionale dell'Associazione Italiana di Scienze della Voce, Zurigo, Svizzera, 4-6 febbraio 2009 (this volume).

Romito, L. & Trumper, J. (1993), Problemi teorici e sperimentali posti dall'isocronia, *Quaderni del Dipartimento di Linguistica dell'Università della Calabria*, S. L. 4, 10, 89-118.

Russo, M. & Barry, W.J. (2008a), Measuring rhythm. A quantified analysis of Southern Italian Dialects Stress Time Parameters, in *Experimental Prosody* (A. Pamies, M.C. Amorós & J.M. Pazos, editors), Actas del IV Congreso de Fonética Experimental, Granada, Spain (= *Language Design*, special issue 2), 315-322.

Russo, M. & Barry, W.J. (2008b), Isochrony reconsidered. Objectifying relations between rhythm measures and speech tempo, in *Proceedings of the 4th International Conference on Speech Prosody*, Campinas, Brazil, May 6-9, 2008, 52-55.

Sachs, C. (1953), *Rhythm and tempo: a study in music history*, London: Dent.

Schmid, S. (1996), A typological view of syllable structure in some Italian dialects, in *Certamen Phonologicum III* (P.M. Bertinetto, L. Gaeta, G. Jetchev & D. Michaels, editors),

Papers from the Third Cortona Phonology Meeting, April 1996, Torino: Rosenberg & Sellier, 247-265.

Schmid, S. (2001), Un nouveau fondement phonétique pour la typologie rythmique des langues, *Poster présenté au 10ème anniversaire du laboratoire d'analyse informatique de la parole (LAIP)*, Université de Lausanne.

Schmid, S. (2004), Une approche phonétique de l'isochronie dans quelques dialectes italo-romans, in *Nouveaux départs en phonologie* (T. Meisenburg & M. Selig, editors), Tübingen: Narr, 109-124.

Schmid, S. (2008), *Measuring the rhythm of Italian dialects*, Poster presented at the workshop 'Empirical Approaches to Speech Rhythm' (EASR), University College London, 2008.

Scott, D., Isard St.D. & de Boysson-Bardies, B. (1986), On the measurement of rhythmic irregularity: a reply to Benguerel, *Journal of Phonetics*, 14, 327-330.

Sock, R., Löfqvist, A. & Perrier, P. (1996), Kinematic and acoustic correlates of quantity in Swedish and Wolof: a cross-language study, in *Proceedings of the 1st ETRW on Speech Production Modeling 'From Control Strategies to Acoustics'*, Autrans, France, 81-84.

Vayra, M., Avesani, C. & Fowler, C. (1984), Patterns of temporal compression in spoken Italian, *Proceedings of the 10th International Congress of Phonetic Sciences*, Utrecht, The Netherlands (1983), Vol. 2, 541-546.

Volín, J. (2005), Rhythmical properties of polysyllabic words in British and Czech English, in *Patterns* (J. Cermák, A. Klégr, M. Malá & P. Šaldová, editors), Praha: Kruh moderních filologu, 279-292.

White, L., Mattys, S.L., Series, L. & Gage, S. (2007), Rhythm Metrics Predict Rhythmic Discrimination, in *Proceedings of the 16th International Congress of Phonetic Sciences,* Saarbrücken, Germany, August 6-10, 2007, 1009-1012.

11. APPENDIX. SOURCES OF THE SPEECH SAMPLES ANALYSED IN 6.1

Original data for several languages have been recorded at the *LFSAG* within the framework of various research projects or by individual dissertations on different topics (in parentheses we summarise the reference to the *project name* or to the author of the recording).

French_european	*IPA* (1999)
Romanian_muntenian	*LFSAG* (H. Bandea, 2006)
Italian_1	*LFSAG* (A. Romano - P. Mairano, 2006)
Finnish_2	*LFSAG* (L. Capovilla, 2007)
Chinese_Mandarin	*LFSAG* (S. Pittoni, 2008)
Icelandic_10	*LFSAG* (P. Mairano, 2006)
Icelandic_4	*LFSAG* (P. Mairano, 2006)
Italian_2	Canepari (2004-)
Finnish_1	*LFSAG* (L. Capovilla, 2007)
Romanian_moldavian	*LFSAG* (A. Romano, 2007)
French_canadian	*LFSAG* (*Loquendo*, 2006)
Italian_3	Costamagna (2000)
Portuguese_brazilian (SP)	Barbosa Almeida & Albano Cavalcanti (2004)
Italian_4	*LFSAG* (L. Calabrò, 2009)
Spanish_castilian	Martinéz Celdrán, Fernández Planas & Carrera Sabaté (2003)
Japanese_fast	*LFSAG* (A. Romano - P. Mairano, 2006)
English_GA	*IPA* (1999)
Japanese_slow	*LFSAG* (A. Romano – P. Mairano, 2006)
Chinese_Cantonese	*LFSAG* (S. Pittoni, 2008)
Turkish	*LFSAG* (A. Romano – P. Mairano, 2006)
Russian	*LFSAG* (*CELI*, 2008)
Czech	*LFSAG* (D. Brdičko, 2007)
English_RP	Roach (2004)
English_Aus	*LFSAG* (*CELI*, 2009)
Portuguese_european	*IPA* (1999)
German_IPA	*IPA* (1999)
Zurich_German	Fleischer & Schmid (2006)
German_1	*LFSAG* (L. Capovilla, 2007)
German_2	*LFSAG* (L. Capovilla, 2007)
Arabic_lebanese	*LFSAG* (*CELI*, 2009)