



**HAL**  
open science

# RHYTHM TYPOLOGY: ACOUSTIC AND PERCEPTIVE STUDIES

Paolo Mairano

► **To cite this version:**

Paolo Mairano. RHYTHM TYPOLOGY: ACOUSTIC AND PERCEPTIVE STUDIES. Linguistics. Université degli studi di Torino, 2011. English. NNT: . tel-00654261

**HAL Id: tel-00654261**

**<https://theses.hal.science/tel-00654261>**

Submitted on 21 Dec 2011

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Università degli Studi di Torino  
Dipartimento di Scienze Letterarie e Filologiche  
Scuola di Dottorato in Studi euro-asiatici: indologia, linguistica, onomastica  
Curriculum in Linguistica, Linguistica Applicata e Ingegneria Linguistica  
XXIII ciclo

# **RHYTHM TYPOLOGY: ACOUSTIC AND PERCEPTIVE STUDIES**

Tesi presentata da: Paolo Mairano

Tutor: Antonio Romano

Coordinatore della scuola di dottorato: Alda Rossebastiano

Coordinatore del curriculum: Carla Marellò

Anni accademici: 2007-2008 / 2008-2009 / 2009-2010

Settore scientifico-disciplinare di afferenza: L-LIN/01

# Table of contents

Acknowledgments .....	6
<b>1. Introduction.....</b>	<b>8</b>
1.1 Rhythm .....	9
1.2 Speech rhythm.....	10
1.3 Rhythm typology.....	10
1.4 Presentation of the thesis.....	11
<b>2. Nearly 100 years of stress-timing and syllable-timing .....</b>	<b>14</b>
2.1 Introduction .....	15
2.2 Stress-timing vs. syllable-timing.....	15
2.2.1 The pioneers .....	15
2.2.2 The classics: Pike and Abercrombie.....	15
2.2.3. Further studies on syllable-timing and stress-timing.....	17
2.2.4 A comparative study by Roach.....	21
2.3 Departing from isochrony .....	22
2.3.1 The “phonological illusionists” .....	22
2.3.2 Compensation and coarticulation .....	24
2.3.3 Linear regression studies and rhythm modelling.....	26
2.3.4 Rhythm metrics.....	28
2.4 Back to syllable-timing and stress-timing? .....	29
2.4.1 A revisitiation of stress-timing and syllable-timing .....	29
2.4.2 An unplanned experiment.....	30
2.4.3 The results.....	31
2.4.4 A brief discussion of the results .....	33
2.5 Conclusion.....	34
<b>3. 1999-2010: rhythm metrics .....</b>	<b>35</b>
3.1 Introduction .....	36
3.2 An account of rhythm metrics .....	37
3.2.1 The deltas.....	37
3.2.2 The PVI.....	39
3.2.3 Considerations on speech rate and the varcos .....	41

3.2.4 Deltas and varcos versus the PVI, the debate goes on.....	44
3.2.5 Bertinetto & Bertini (2008): the CCI.....	46
3.3 Applying the metrics to the corpus .....	49
3.3.1 Reasons for setting out to compute rhythm measures .....	49
3.3.2 The data .....	50
3.3.3 Labelling the data .....	52
3.3.4 The segmentation.....	53
3.3.5 The deltas and the varcos.....	57
3.3.6 The PVI.....	65
3.3.7 The CCI .....	66
3.4 On the possibility of including pitch and intensity .....	71
3.4.1 Why on earth?.....	71
3.4.2 An attempt .....	72
3.4.3 The results.....	73
3.4.4 So.....	73
3.5 Conclusion.....	74

## **4. Correlatore .....75**

4.1 Introduction .....	76
4.1.1 The reasons for developing <i>Correlatore</i> .....	76
4.2 How to use Correlatore .....	77
4.2.1 The annotation of <i>Praat's TextGrids</i> .....	77
4.2.2 Computing rhythm metrics .....	78
4.2.3 Reports.....	81
4.2.4 Charts.....	83
4.2.5 Preferences and configurations.....	84
4.3 The implementation of Correlatore .....	86
4.3.1 Why Tcl/Tk?.....	86
4.3.2 Overview .....	87
4.3.3 How the segmentation and the transcriptions are dealt with.....	87
4.3.4 How formulae are dealt with .....	89
4.3.5 How reports are dealt with.....	89
4.3.6 How charts are dealt with .....	91
4.3.7 SASASA files .....	91
4.4 Conclusion and future perspectives.....	92

<b>5. Rhythm variation and variability .....</b>	<b>93</b>
5.1 Introduction .....	94
5.2 The study of rhythm variation .....	94
5.3 The study of rhythm variability .....	97
5.4 Inter-operator variability of rhythm metrics.....	98
5.5 Speakers and rhythm variability.....	103
5.5.1 Intra-speaker variability.....	103
5.5.2 Inter-speaker variability.....	106
5.5.3 Geographical variability vs. variation .....	107
5.5.4 Discussion on rhythm variation and variability.....	110
5.6 Conclusion.....	112
<b>6. A perceptive test.....</b>	<b>114</b>
6.1 Introduction .....	115
6.1.1 Why perceptive tests?.....	115
6.1.3 The conception of the test.....	118
6.2 The preliminary phase .....	120
6.2.1 Personal data.....	120
6.2.2 Testing prosodic sensitivity .....	121
6.2.3 The results.....	123
6.3 Testing the perception of discrete rhythm classes.....	125
6.3.1 The format and the interface.....	125
6.3.2 Differences from other tests. ....	127
6.3.3 Perceptive tests within the AMPER project. ....	128
6.3.4 Synthesising audio files with AMPER routines for testing rhythm perception. ....	128
6.3.5 The results.....	131
6.4 Testing the perception of a scalar characterisation of rhythm.....	133
6.4.1 The format and the interface.....	133
6.4.2 The Results .....	135
6.5 Conclusion.....	136
6.5.1 Participants .....	136
6.5.2 The design of the test.....	137
6.5.3 Perception of the rhythm class.....	137
6.5.4 Summary.....	138

<b>7. Final discussion .....</b>	<b>139</b>
7.1 Methodological issues .....	140
7.2 Main findings .....	142
7.3 Future perspectives.....	145
7.4 Conclusion.....	146
<b>8. References.....</b>	<b>148</b>
Webography .....	166
<b>9. Appendices.....</b>	<b>167</b>
Appendix 1: inter-onset values.....	168
Appendix 2: <i>Correlatore</i> 's report .....	169
Appendix 3: Tcl implementations .....	200
Appendix 3a: Tcl implementation of the metrics.....	200
Appendix 3b: reading CV and SAMPA segmentations with a foreach cycle.....	201
Appendix 4: sample log file of a perceptive test.....	203
Appendix 5: Results of the perceptive test by best performers in the preliminary phase.....	205

## **Acknowledgments**

The accomplishment of my PhD and of the drafting of this thesis would have never been possible without the help, assistance and support of many people.

First of all, I am grateful to Antonio Romano (aka Tonino) for following me along my PhD, for his constant precious advice, for the opportunities he granted me, for the confidence he set in me, and, in general, for all the support he gave me during these three years: this work would have been unthinkable without that.

Of course, I am debtor to the PhD school in Linguistics, Applied Linguistics and Linguistic Engineering of the University of Turin, and to much of its staff. In particular, I would like to thank Carla Marengo and Virginia Pulcini for their involvement and for the opportunities they granted me, Mario Squartini and Manuel Barbera for their valuable pieces of advice within my research. A word of mention for my PhD colleagues Lea Glarey, Cristina Colet and Mauro Costantino with whom I shared stimulating discussions, for Adriano Allora, who re-introduced me to computer programming and to Linux and for Maria Chiara Felloni and Daniele Avesani who provided me with precious feedback for the perceptive tests.

I am very thankful to Sigurður Pétursson (University of Iceland) for his outstanding hospitality when I was in Iceland and to the staff of *GIPSA-Lab* and of the University of Grenoble 3 *Stendhal* for receiving me within my PhD. The warm welcome I was reserved at the *Centre de Dialectologie de Grenoble* by Michele Contini, Elisabetta Carpitelli, Jean-Pierre Lai, Giovanni Depau, Maria Goudi and Silvia Gally has no equal. I am full-heartedly thankful to Maria Goudi for her constant scientific and moral support and for feeding me during the drafting of this thesis, to Giovanni Depau and Jean-Pierre Lai for sharing their office with me and for the interesting discussions we had, to Anna Ghimerton for reading and correcting most of my thesis, to Kristina Grabsky for her pieces of “typographic” advice.

Last but certainly not least, I am grateful to my parents for the support they gave me, to Elisabetta, Fabrizio, Marta, Daniela and Patrick for helping me when in need and to my flat mates Arnaud, Morgane and Camille for their joyful and truly rewarding company in the final year of my PhD. In a quite singular sort of way, I would also like to thank Sophie for a much needed turn of the screw.

A Chiara  
*che ha dato una direzione alla mia vita e  
che ha dovuto imparare la pazienza quando a suo papà toccava lavorare per la tesi,*

e a zio  
*che non ha visto la fine del mio dottorato, di cui però andava fiero.*



**1.**

# **Introduction**

# 1. Introduction

## 1.1 Rhythm

Rhythm is generally said to play an important role in human behaviour as it is connected to numerous human activities, such as breathing, walking, dancing, playing, and so on. When referring to rhythm, we usually associate it to auditive phenomena, but this is not always the case: for instance, as reported by Eriksson (1991), repeated visual patterns can also be described as rhythmical. However, in spite of the pervasiveness of rhythm and the apparent straightforwardness of its concept, it is not at all easy to define it.

In its simplest form, rhythm could perhaps be described as a regular succession of events. This definition accounts for the rhythmicity of some human activities mentioned above, such as breathing and walking, and for heartbeat and repeated visual patterns. But it does not fully explain what is meant by rhythm in activities such as dancing and playing. To account for these activities, one needs to add something to the definition given above:

*Rhythm is the structure of intervals in a succession of events.*  
(Allen, 1972:72)

The “structure” refers to a possible hierarchy or a grouping among the events in succession; moreover, the two aspects composing rhythm (structure and succession) tend to go hand in hand. In music<sup>1</sup>, for instance, the succession is given by bars (which recur at regular intervals), while the structure is given by the notes, which can carry the beat in certain positions, but not in others (they are therefore organised hierarchically). Similarly, in poetry, the regular succession of lines is counterbalanced by a hierarchical organisation among syllables, which tend to form groupings composed of one stressed syllable and one or more unstressed syllables.

Moreover, the structure of rhythm is not always given by exclusively acoustic phenomena, but it also has a “subjective” component. For instance, it has been reported (see Allen, 1975, among others) that listeners tend to hear a hierarchy even when it does not exist, *i.e.* they tend to perceive groupings on a succession of identical stimuli. This phenomenon is called *subjective rhythmisation* and emerges, for example, in the onomatopoeia used in various languages to describe the sound of the pendulum: although the acoustic features of its beats are the same, it is usually described as if formed by a grouping of two sounds: *tic-tac* in French and Italian, *ticktack* in Swedish and *tick-tock* in English (this case is reported by Eriksson, 1991). Finally, there seem to be some constraints on the perception of rhythmicity:

*When we hear a sequence of pulses that is neither too rapid nor too slow we hear it as rhythmic [...]. As long as the minimum time between pulses is greater than about 0.1 s, so that successiveness and order are perceivable, and the maximum is less than about 3.0 s, beyond which groupings do not form, we will impose some rhythmic structure on the sequence.*  
(Allen, 1975:76)

---

<sup>1</sup> For sake of simplicity, I shall refer to the music that developed within the Western culture in the last centuries. Other types of music are also rhythmical of course, but their structure can be conceived differently from what is sketched in the text.

## 1. Introduction

In summary, rhythm can be considered to appear in regular successions of events between 0.1 s and 3.0 s and to be composed of a structure (in terms of hierarchy or of groupings) which is either present acoustically, or which is superimposed by the hearer on the succession of events.

### 1.2 Speech rhythm

Although the claim that speech is characterised by a rhythmical component is universally accepted, there is far less consensus as to what should be identified as the rhythmic unit and as the carrier of rhythmic beat in spoken language. In the light of the considerations sketched in the preceding section, if we want to identify rhythm in speech, we have to search for something that is either structured or recurrent within a limited time range:

*Language is produced by humans and is perceived by humans, and it appears to be governed by the same rhythmic constraints as other human motor and perceptual behaviors. These constraints thus set limits on the kinds of rhythms we can expect in languages of the world: they should be simple in structure, confined largely to successions and alternations, depending on the relationship between syllables and stress-accent in the language; the rate of succession of syllables and rhythmic groups should be in or near the range of 0.2 - 1.0/s.*  
(Allen, 1975:82)

This idea, after all, is the same that stands at the base of the conception of metre in poetry. Metres fulfil the requirements in that they are *simple structures* composed of *regular successions* of *stresses* and *syllables*, resulting in *alternations* of stressed and unstressed syllables. Furthermore, stresses and syllables fall precisely within the relevant time range. In effect, the study of speech rhythm has long been associated with metrics and poetry. Aristotle, in *The Art of the Rhetoric*, made an early attempt to describe the rhythm of language, using metric concepts to describe the different speech styles of the people (e.g. the iambic metre was said to be the rhythm of the common people, the trochaic that of rhetoricians, etc.).

Although the question of metre is still of interest today, many phoneticians of the last century have shifted the focus of research on two other issues. One consisted in finding an acoustic correlate of the perception of the rhythm beat: results in this field are not utterly uniform, but there seems to be general agreement on the importance of the vowel onset<sup>2</sup> (see Allen, 1972 and 1975). At any rate, it is the second issue, which concerns the alleged regular occurrence of syllables and/or stresses, that retained most of the attention through the years.

### 1.3 Rhythm typology

The two categories of *stress-timed* and *syllable-timed* languages have been introduced by Pike (1945) referring to the impression that stresses seem to occur at

---

<sup>2</sup> One remarkable alternative theory introduces the so called *p-centre* (perceptive centre), developed initially by Morton et al. (1976).

## 1. Introduction

regular temporal intervals in English, while syllables seem to have similar durations in Spanish. Abercrombie (1967) drew on the distinction, also on the basis that the different rhythmic structure of these languages seems to be reflected by the metrical units adopted in poetry: Germanic languages count the length of verses in feet, while Romance languages use the syllable as the basic metric unit. This view had a great fortune in the following years, but various instrumental experiments failed to give evidence of isochrony at the foot or at the syllable level (some of them are reviewed in chapter 3).

After these failures, some linguists (Bertinetto, 1977, and Dauer, 1983, among others) attributed the impression of stress-timing or syllable-timing to structural properties of languages, such as the absence vs. presence of vocalic reduction and a complex syllabic structure. Relying on these theories, Ramus, Nespore & Mehler (1999)<sup>3</sup> and Grabe & Low (2002) proposed acoustic correlates of these phonological properties based on vocalic and consonantal durations. The authors claimed that their measures allowed for a scalar characterisation of languages on the basis of rhythm properties.

The validity and the stability of the acoustic correlates (soon re-baptised *rhythm metrics*) have soon been tested in other studies (e.g. Schmid, 2001) then gradually introducing new variables, such as different speech rates (e.g. Dellwo & Wagner, 2003), spontaneous conversation (e.g. Barry & Russo, 2003), a larger number of speakers (e.g. Galves *et al.*, 2002, and Rouas & Farinas, 2004) with different and sometimes controversial results. Some authors have proposed modifications of the formulae (e.g. Dellwo, 2006, and Benton, 2010), or have applied the formulae to different durations, such as voiced and unvoiced intervals (e.g. Galves *et al.*, 2002) or feet and syllables (e.g. Wagner & Dellwo, 2004, and Asu & Nolan, 2006), or have proposed new metrics that are based on different rationales (e.g. Bertinetto & Bertini, 2008<sup>4</sup>). Despite some criticisms and a few failures (e.g. Barry & Russo, 2003, and Arvaniti, 2009), these measures and the perspectives they offer have raised (and are still raising) a growing interest within the scientific community: various authors have used them with the aim or in the attempt of categorising different languages, different language varieties (e.g. Schmid, 2004, Romano *et al.*, 2010, Giordano & D'Anna, 2010, within the linguistic area of Italy) and even to detect the interference of the rhythm properties of L1 on productions by L2 speakers (e.g. White & Mattys, 2007). The latest developments in this field include some attempts to merge the two aspects of speech rhythm (namely, the segmental and the accentual levels) into multi-layer models (see Bertinetto & Bertini, 2010, and O'Dell *et al.*, 2010).

### 1.4 Presentation of the thesis

The thesis is collocated within the research frame sketched above. It is mainly concerned with rhythm typology and presents a number of experiments on and with the main rhythm metrics, most of which exploit the same data (a corpus of audio samples that have been gathered and labelled precisely for this purpose). However, the structure of the thesis is conceived in a slightly peculiar way. It is not split into two parts (the first expounding the state of the art, the second presenting the results

---

<sup>3</sup> From now on, Ramus *et al.* (1999).

<sup>4</sup> However, their index is still based on a modification of the rPVI formula (see chapter 3).

## 1. Introduction

of one or more experiments), as it is customary, but mingles the two together. Every chapter deals with one or more aspects of rhythm typology and contains, first, a survey of the theoretical studies that have approached the topic, then, an analysis of the experiments or tests conducted. Of course, some chapters concentrate in outlining previous approaches as well as theoretical issues, whereas others focus on the methodology and the analysis of some experiments. But the two aspects are always present, except for chapter 4.

The following chapter unfolds the research in the field of rhythm typology, from its beginning to present day models, by following the thread of the traditional dichotomy that opposes stress-timing and syllable-timing. In the end, I shall present a minor experiment that was inspired by some recent studies (namely Wagner & Dellwo, 2004, and Asu & Nolan, 2006). Despite using the formulae of rhythm metrics, I have chosen to illustrate it in that chapter because it deals with the inter-onset distance (*i.e.* a certain conception of the syllable), a unit which has been abandoned by mainstream modern approaches, but which has been the focus of past research (together with the inter-stress distance, *i.e.* the foot).

Chapter 3 zooms to rhythm metrics. The main studies on this topic are outlined along an illustration of the formulae of the most frequently used metrics as well as a discussion of practical and methodological issues. Subsequently, I shall present and analyse the results of the metrics for the data I gathered, which presently include 61 speakers of 21 languages. It can be said that this chapter constitutes the core of the thesis.

Chapter 4 is the most technical one and deals exclusively with the illustration of *Correlatore*, a program that I have developed with the aim of accelerating the process of computing rhythm metrics. It is divided into two parts, the first one explaining how to use it, the second one illustrating its implementation and discussing the difficulties encountered.

Chapter 5 treats the theme of variation, which is recurrent in linguistics, from an unusual perspective, *i.e.* rhythm variation and variability. It is not meant to exhaust the topic, rather it is intended to introduce it, as it seems that studies of this type are still rare. After a discussion of the two aspects, it presents the results of the metrics on selected data samples and tries to outline a framework which, I believe, might open interesting perspectives.

Chapter 6 touches on perception, which has been said to stand at the basis of the distinction between stress-timing and syllable-timing. Despite this claim, in fact, only few studies have investigated the ability of naive listeners to discriminate between languages belonging to different “rhythm classes”. After a review of the (mainly modern) studies on the perception of rhythm, I will discuss the format and the controversial results of a perceptive test which has been administered to 43 listeners. A final discussion on different aspects of rhythm metrics and on future perspectives will of course close the thesis.

I shall now raise a terminological point. After numerous studies failed to find evidence of isochrony, many authors working in this field have taken their distance from the terms *stress-timed* and *syllable-timed*, adopting other (more or less original) solutions (see, for instance, the “label inventors” reviewed by Bertinetto, 1989). However, it has to be noted that these two terms do not necessarily evoke isochrony: they simply suggest that the rhythm of the two groups of languages is

## 1. Introduction

based on stresses vs. syllables<sup>5</sup>. For the time being, there seems to be no proof of the opposite: I shall therefore maintain the use of these terms, which have the advantage of being universally understood.

Finally, I would like to warn the reader that, in countercurrent to the present trend, I shall stubbornly follow British English spelling conventions (except, of course, in quotations, where I shall respect authors' choices). This has been consciously brought to exasperation to include forms like *to analyse*, *to visualise*, *to realise* and *to fulfil*, which seem to be long forgotten even in England. I have tried to stick meticulously to my purpose: should any deflection be found in the text, it has to be imputed to Microsoft Word's spellchecker, which apparently did not approve of my choice and attempted to sabotage it in all possible ways.

---

<sup>5</sup> This is valid in English. In Italian, for example, the terms *isosillabico* and *isoaccentuale* do evoke isochrony and, therefore, they have usually been avoided in the publications by Mairano & Romano (2008 and following).

## **2.**

**Nearly 100 years of stress-timing  
and syllable-timing**

## 2. Nearly 100 years of stress-timing and syllable-timing

### 2.1 Introduction

Research on speech rhythm has mainly (though not exclusively) been concerned with the search for a characterisation of languages on the basis of the distinction between stress-timing and syllable-timing. Such a distinction is based on the perceptive impression that stresses should occur at regular intervals in some languages, and that syllable durations should be fairly constant in some other languages (the theory of isochrony). After empirical data disproved the existence of isochrony, many authors set out to provide alternative visions, claiming that the impression of stress-timing and syllable-timing was given by structural properties of the languages (cf. Bertinetto, 1977 and following and Dauer, 1983) that accounted for their classification to one or the other rhythm class. More recently, some authors (cf. Ramus *et al.*, 1999, and Low *et al.*, 2000) proposed some acoustic measures to reflect these structural properties (at segment and/or syllable level), so that empirical data could be tested with a more explicit model. Besides, other authors provided rhythm models that accounted for the accentual level of speech rhythm (see O'Dell & Nieminen, 1999). The latest attempts include the development of multi-layer models meant to account for both levels of speech rhythm.

In the following pages, I shall try to provide a sketch of the research on these topics, from the quest for isochrony that characterised the first studies in this field, to more recent approaches.

### 2.2 Stress-timing vs. syllable-timing

I shall now present a survey of the main studies which have been carried out in the field of stress-timing and syllable-timing. It is not meant to be exhaustive, rather I shall present the most significant steps in a mainly chronological order following the evolution of research in this field.

#### 2.2.1 The pioneers

Although the terms “stress-timed” and “syllable-timed” were introduced by Pike (1945), the existence of two different rhythm groups of languages had already been noticed earlier. Eriksson (1991) reports that the 18<sup>th</sup> century phonetician Joshua Steele had already put forward the idea that stresses in English occurred at fixed temporal intervals. His claim was supported only by intuition as, obviously, no tools were available at that time to provide instrumental evidence. In the 20th century, Classe (1939) tried to provide experimental evidence of the existence of regular inter-stress intervals in English, but he had to conclude that their duration is not independent of the number of syllables composing them: isochrony only emerges under special circumstances. Lloyd James (1940) used a vivid expression to distinguish between languages characterised by a *machine-gun rhythm* (*i.e.* syllable-timed languages) and languages characterised by a *Morse code rhythm* (*i.e.* stress-timed languages).

#### 2.2.2 The classics: Pike and Abercrombie

As has been previously mentioned, Pike (1945) was the first to use the terms *stress-timed* and *syllable-timed* languages, which many authors still use today. He claimed



## 2. Nearly 100 years of stress-timing and syllable-timing

that the duration of inter-stress intervals in stress-timed languages are more or less constant and, therefore, independent of the number of syllables (which are, consequently, compressed in function of the number of syllables contained in one inter-stress interval); conversely, in syllable-timed languages, syllable duration is more or less constant and, therefore, the duration of inter-stress intervals is proportional to the number of syllables. He mentioned English as an example of stress-timed language and Spanish as an example of syllable-timed language. However, in his book *The Intonation of American English*, the aim of which is to teach the American intonation to foreigners, he merely hints at this distinction and provides no empirical tests to support his claims<sup>6</sup>.

Abercrombie (1967) drew on Pike's distinction and terminology. The influential yet controversial passage is entirely reported below:

*It is the way in which the chest-pulses and the stress-pulses recur, their mode of succession and co-ordination, that determines the rhythm of a language. There are two basically different ways in which the chest-pulses and the stress-pulses can be combined, and these give rise to two main kinds of speech-rhythm. As far as is known, every language in the world is spoken with one kind of rhythm or with the other. In the one kind, known as syllable-timed rhythm, the periodic recurrence of movement is supplied by the syllable-producing process: the chest-pulses, and hence the syllables, recur at equal intervals of time – they are isochronous. French, Telugu, Yoruba illustrate this mode of co-ordinating the two pulse systems: they are syllable-timed languages. In the other kind, known as stress-timed rhythm, the periodic recurrence of movement is supplied by the stress-producing process: the stress-pulses, and hence the stressed syllables, are isochronous. English, Russian, Arabic illustrate this other mode: they are stress-timed languages.*  
(Abercrombie, 1967:97)

It can be said that he reformulated Pike's hypotheses, introducing the concept of isochrony and going slightly further in claiming that all languages of the world are either stress-timed or syllable-timed. The claim that a language can either be stress-timed or syllable-timed is supported by the observation that "when one of the two series of pulses is in isochronous succession, the other will not be"<sup>7</sup> (1967:97). This fact is illustrated by Abercrombie with two sentences, in English and French respectively:

Which is the | train for | Crewe, | please?

and

C'est absolument ridicule

---

<sup>6</sup> Eriksson (1991) states that Pike's book is written in the same traditional style as some normative grammars. He adds that "[r]eading his book today, one finds it surprising that his ideas got the attentions they did" (1991:19).

<sup>7</sup> To be precise, this observation only explains why a language cannot be both stress-timed and syllable-timed; it does not prove that all languages belong to one of the two rhythm classes.

## 2. Nearly 100 years of stress-timing and syllable-timing

In the English sentence, stressed syllables (although recurring at regular intervals) are separated by a different number of unstressed syllables (2 in the first unit, 1 in the second, 0 in the third). In contrast, in the French sentence (in which italics indicate stress), the syllable *ment* will supposedly be nearer to the preceding stressed syllable *so* than to the following stressed syllable *cule*, as only 1 unstressed syllable separates *so* and *ment*, while 2 unstressed syllables separate *ment* and *cule*. In fact, in compliance with what had already been stated by Pike (1945), the consequences of the two alleged types of isochrony are, on the one hand, that

*there is considerable variation in syllable length in a language spoken with stress-timed rhythm whereas in a language spoken with a syllable timed rhythm the syllables tend to be equal in length*

and, on the other hand, that

*[...] in French, a language with a syllable-timed rhythm, the constant rate of syllable-succession means that stresses separated by different numbers of unstressed syllables will be separated by different intervals of time.*

*(Abercrombie 1967:98)*

### 2.2.3. Further studies on syllable-timing and stress-timing

As a consequence of Pike's and Abercrombie's claims, and despite the claims of stress-timing for English having already been disproved by Classe's (1939) experiment, many authors set out to verify the supposed stress-timing or syllable-timing of the languages mentioned by Pike and Abercrombie or to attempt a rhythmic categorisation of other languages<sup>8</sup>. Most of the experiments were carried out by measuring syllable durations (either in their phonological notion or as onset-to-onset) and/or inter-stress intervals (feet). Indeed, these studies are so numerous that some authors have even compiled reviews of the research in this subject (such as Bertinetto, 1989 – who introduces ironically labelled partitions between the studies according to their attitude towards isochrony –, Eriksson, 1991, and the most exhaustive and emblematically impressive 54-pages unpublished *Bibliography of Timing and Rhythm in Speech* by P. Roach<sup>9</sup>). I shall now report a selection of relevant studies in this field with no ambition of completeness whatsoever.

Bolinger (1965, reported by Eriksson, 1991:21, and Bertinetto, 1989:102) measures inter-stress intervals in 2 English sentences read by 6 speakers and finds that the number of syllables composing each interval seems to determine its length. Bertinetto (1989) classifies Bolinger (1965) together with Classe (1939), Allen (1975), Lehiste (1977) and others as “perceptual illusionists” since they suggest that isochrony might be evident perceptively, but seems to have no acoustic counterpart.

Allen (1975) stated that the impression of isochrony is given by the listeners' tendency to superimpose a rhythmic structure on the recurrent linguistic patterns (given by syllables and/or stresses). He also reformulated the classical dichotomy in

<sup>8</sup> Abercrombie's statements have been very influential, but scepticism about the stress-timed vs. syllable-timed dichotomy was fairly widespread at any rate: Mitchell writes that the classification of languages on the basis of this dichotomy is “something of an oversimplification” (1969:156).

<sup>9</sup> This document has been last updated in April 2003 but is currently still available online at the following address: <http://www.personal.reading.ac.uk/~llsroach/timing.pdf>

## 2. Nearly 100 years of stress-timing and syllable-timing

terms of rhythms of alternation (between heavy, fully articulated stressed syllables and reduced unstressed syllables) and rhythms of succession (of similar non reduced syllables). It is well-known that the verse of the supposed stress-timed languages is based on the foot as a metrical unit, while the verse of the supposed syllable-timed languages is based on the syllable. Starting from this observation, Allen states that:

*The huge preponderance of English metrical verse has feet that are either two or three syllables long, with accent either beginning or ending the foot. Most of these metres will give rise to an alternating rhythm, since they have one or more unaccented syllables in each foot [...] Romance (and Japanese) poetry requires only a fixed number of syllables per line [...]. That is, since accentuation plays a weaker role in Romance phonology, the poetry of these languages makes little use of differences in syllabic accent, grouping the syllables instead into sequences of equals.*

*Natural language rhythms thus appear to be largely either simple alternations or successions.*

*(Allen, 1975:77).*

This claim is supported by the different status of unstressed syllables in the two categories of languages:

*Unstressed syllables in English [...] are “reduced” in both quality and quantity to the extent that the resulting rhythmic pattern consists of the stressed syllables alternating with all of the intervening unstressed syllables, i.e. a sort of massive off-beat. When the unaccented syllables retain their phonetic shape, however, as in French or Japanese, the resulting rhythmic pattern remains tied as much to syllables as to accents. Stress rhythms are thus rhythms of alternation, whereas syllable rhythms are rhythms of succession.*

*(Allen, 1975:80)*

The author states that listeners have a “centralizing tendency”, a tendency to hear regularity where it does not necessarily exist. In other words, they tend to superimpose a kind of recurrent rhythmic structure on these two underlying rhythmic patterns. Hence the impression of isochrony: “we perceive speech as rhythmic because it is fairly regular in its sequential sound patterns often enough that we can impose upon it simple rhythmic structures” (Allen, 1975:78).

This view is shared by Lehiste (1977), who ran some experiments on the perception of timing and duration: she found that listeners obtained better scores when rating the duration of non-speech stimuli than of intervals of speech. So, she suggests that isochrony is language-bound, “rather than being a feature of the perception of rhythm. At least in terms of a gradient, it is slanted in favor of perception of spoken language” (1977:257). Like Allen (1975), she considers it “quite likely that the listener imposes a rhythmic structure on sequences of inter-stress intervals” (1977:258). However, she then presents a study of the interaction

## 2. Nearly 100 years of stress-timing and syllable-timing

between syntax and isochrony<sup>10</sup> (for this reason she is classified among the “optimists” by Bertinetto, 1989).

Uldall (1971, reported by Eriksson, 1991:25, and by Lehiste, 1977:254) measured inter-stress durations in a speaker reading the narrative *The North Wind and the Sun* and claimed to have found a strong tendency to isochronism<sup>11</sup> as feet made up of less than 4 syllables tended to have durations between 385 and 520 ms. However, Eriksson (1991) wonders “by what standard of comparison Uldall is able to determine that a difference between 385 and 520 ms is ‘small’” and argues that “all one has to say to upset the whole ‘proof’ is that one finds the increase in foot duration as a function of the number of syllables ‘very striking’” (Eriksson, 1991:26).

A series of articles with emblematic titles (*Is Spanish really syllable-timed?*, *Is French really syllable-timed?*, etc.) appeared on the *Journal of Phonetics* in the early 80s. Authors verified the presence of isochrony at the syllable in supposedly syllable-timed languages: these studies include Pointon (1980), Wenk & Wioland (1982) and Borzone de Manrique & Signorini (1983) – the first two ones are classified as “label inventors” by Bertinetto (1989) as they proposed new terms to indicate the different (and often deviating) rhythm tendencies of the languages analysed: segment-timing (Pointon, 1980) as well as trailer-timed vs. leader-timed (Wenk & Wioland, 1982). In the former study, the author reviewed a number of previous studies on Spanish rhythm; the label segment-timing appears in Pointon’s conclusion and is supposed to refer to a language “in which the number and type of segments in each syllable [...] determine the duration of a syllable” (Pointon, 1980:302). In the latter study, the two authors had 12 native French speakers read the following sentence (12 + 6 syllables):

Il a sollicité ma collaboration, car Pierre aime toujours l’art.

If hypotheses about the presence of isochrony at syllable level for French were correct, one would expect the first part (12 syllables) of the sentence to last approximately twice as long as the second part (6 syllables), but the results were far from confirming these hypotheses. Instead, the authors present a series of “phonetically<sup>12</sup> [sic] ambiguous, rhythmically distinct utterances” (such as /sedɔpapa/, *c’est de papa* or *c’est deux, papa*) and argue that “in simplest terms, other things being equal, the greater the number of rhythmic groups<sup>13</sup> in an utterance, the greater the amount of time the utterance will be allotted” (Wenk & Wioland, 1982:194-195). Then, they carried out a perceptual test and found that “rhythm groups” are distinguishable by listeners, possibly thanks to the increase in duration and corresponding decrease in intensity of French final vowels. On the grounds of this explanation, Wenk & Wioland conclude that:

---

<sup>10</sup> She probably refers to the “impression of isochrony” on the part of English speakers and listeners, rather than to “isochrony” proper.

<sup>11</sup> I have decided to include this study for, if one accepts its author’s conclusions, the finding is definitely counter-current.

<sup>12</sup> They possibly mean “phonologically”, since they use slashes for the transcriptions, which also look phonological.

<sup>13</sup> It seems that a definition of “rhythmic group” is not immediately provided by the authors, but their conclusion includes an indication on how to determine them (see below). Judging from their examples, their notion of “rhythm groups” seems to coincide with the interval between two minor prosodic breaks.

## 2. Nearly 100 years of stress-timing and syllable-timing

*what serves to establish rhythmic groups in French is a lengthening of what is perceived as the final syllable in each group [...]. For this reason [...] it is proposed to characterize French as being trailer-timed.*

*As English rhythmic groups, on the other hand, are delimited by the regular occurrence of stronger syllables at the beginning of each group, it is natural to regard English as being leader-timed<sup>14</sup>.*  
(1982:214)

Borzzone de Manrique & Signorini (1983) studied the durations of segments (both consonants and vowels, all divided in several different groups according to stress and positioning), syllables and inter-stress intervals in Argentine Spanish on sentences by 4 speakers. They found that neither segment (cf. Pointon, 1980) nor syllable duration was constant and, on the contrary, they were influenced by several factors, such as stressed or unstressed position. Surprisingly, instead, the measurements of inter-stress intervals seem to cluster around the same area: such findings bring the authors to the odd conclusion that “Spanish has a tendency toward stress-timed rhythm with differentiating characteristics in the way in which this is manifested” (1983:127).

Another “label inventor” (to use Bertinetto’s, 1989, classification) is Hoeqvist (1983a and b), who considered data of English (supposedly stress-timed), Spanish (supposedly syllable-timed) and Japanese (supposedly mora-timed). As for measurements at syllable level, results for Japanese “indicate that lengthening due to an added mora is considerably more than that found for non-phonemic lengthening” (1983b:222), which therefore seems to confirm at least a partial tendency towards mora-timing. Moreover, results for English reveal a “strong shortening effect due to an adjacent accented syllable” (1983b:223), which seems to be absent in data of Spanish<sup>15</sup> and Japanese. Hoeqvist argues that “perceptually, this shortening might serve to highlight the stressed syllable” or it may also be “a consequence of an attempt to maintain some overall durational structure” (1983b:223). On the basis of these observations, the author claims that Japanese could be defined as duration-controlling and English could be defined duration-compensating, while Spanish does not seem to belong to any of these two categories.

Major (1981) examined data involving both real and nonsense “citation” words as well as casual speech<sup>16</sup> of Brazilian Portuguese. The author concludes that Brazilian Portuguese shows a tendency towards stress-timing for several reasons, among others: “inter-stress durations are not directly proportional to the number of syllables; [...] syllable duration is inversely proportional to the number of syllables in a word; [...] in casual speech unstressed syllables delete, which has the effect of

---

<sup>14</sup> This distinction definitely reminds of the distinction between trochee and iambs. Curiously, it can be noticed that French also shows a tendency towards post-modification, whereas English shows the opposite tendency towards pre-modification. However, this has to be taken merely as a remark: I am not implying any sort of correspondence between the two typological categories (leader vs. trailer-timing and pre- vs. post-modification) other than chance.

<sup>15</sup> Hoeqvist reports that this effect is actually visible in Spanish, but only in syllables preceding stress. However, he claims that “the differences between the amounts of shortening shown are so small that it is doubtful that any genuine cross-language difference is showing itself between Spanish and English” (1983b:225-226) The authors, therefore, considers only post-stress shortening (only visible in English) to be characteristic of stress-timing.

<sup>16</sup> By “casual speech” Major refers to speakers reading three sentences rapidly.

## 2. *Nearly 100 years of stress-timing and syllable-timing*

equalizing the number of syllables in each stress group” (1981:350). However, he finds differences as to these phenomena across the citation and casual speech styles (see also chapter 5).

Lehiste (1990, reported by Eriksson, 1991:22) did not find evidence of supposed stress-timing in Icelandic since the duration of feet turned out to be proportional to the number of syllables.

Finally, I shall hint at the numerous phonological contributions on rhythm that mainly attempt to analyse the distribution of stresses and accents<sup>17</sup>. In particular, much of the research in this field concentrated on the study of how languages solve stress clashes, the so-called Rhythm Rule, or stress lapses (see, among others, Liberman & Prince, 1977, Farnetani & Kori, 1990, Nespor, 1993, Arvaniti, 1994). Interestingly, Arvaniti (1994) suggests that languages belonging to different rhythm classes differ in the degree of toleration of these eurhythmic phenomena and in the way they try to solve them. In particular, English (stress-timed) tolerates neither stress clashes nor stress lapses and tends to correct both: the former are corrected with the insertion of an extra stress, whereas the latter are corrected with a stress shift. Instead, Italian and Greek are reported to be more tolerant to stress lapses and to correct stress clashes in a different way from English, namely by either destressing or by inserting extra duration between clashes. So, the behaviour of languages in relation to these phenomena might be indicative of their belonging to one of the two rhythm classes.

Before concluding this summary it is worth to point out the different problems in measuring inter-stress distances noticed by Eriksson (1991) in his review of previous research on stress-timing: apart from the intrinsic difficulties in identifying stresses, inter-stress intervals can be measured from vowel onset to vowel onset, or from syllable onset to syllable onset, or even from and to syllable centres. However, despite the different criteria used to identify inter-stress intervals, the result seems to be the same: inter-stress intervals are not isochronous in stress-timed languages, rather their duration seems to increase as a function of the number of syllables. Similar problems and parallel findings have been reported for syllable-timing: syllable boundaries are not always clear (some authors simply measured vowel-to-vowel distances, other used phonological criteria), but, at any rate, syllable durations do not seem to be constant in syllable-timed languages. As a coronation to the failure of research on isochrony, I shall report in the next paragraph the comparative study by Roach (1982).

### **2.2.4 A comparative study by Roach**

Roach (1982) carried out an experimental test based on Abercrombie’s assumption that syllable length tends to be greatly variable in stress-timed languages and equal in syllable-timed languages. His study involved the six languages quoted by Abercrombie (1967), three of which had been given as examples of stress-timing (English, Russian and Arabic), while the other three had been given as examples of syllable-timing (French, Telugu and Yoruba) – see above.

Firstly, Roach calculated the standard deviation of the durations of the syllables in the four languages assuming that if Abercrombie’s hypothesis (the durations of syllables is constant in syllable-timed languages but greatly variable in

---

<sup>17</sup> Many of these studies actually seem to even take the distinction between stress-timing and syllable-timing for granted.

## 2. Nearly 100 years of stress-timing and syllable-timing

stress-timed languages) was right, its value had to be higher for stress-timed languages and lower for syllable-timed languages. However, Abercrombie's hypotheses were not confirmed by the results: for some syllable-timed languages (French and Yoruba) the value of the standard deviation of syllable lengths was indeed lower than for English, but it was higher for Yoruba than for both Russian and Arabic, which is in contradiction with Abercrombie's statement. At any rate, Roach notes that the differences among the values obtained are too small (ranging from 66 milliseconds in Telugu to 86 in English) to justify the classification of a language as syllable-timed as opposed to stress-timed.

Secondly, he calculated the standard deviation of inter-stress intervals in order to test Abercrombie's second statement (*i.e.* that the length of inter-stress intervals should be constant in stress-timed languages and greatly variable in syllable-timed languages). One would expect the standard deviation of the duration of inter-stress intervals to be lower for stress-timed languages and higher for syllable-timed languages. But again, the results did not confirm expectations: surprisingly enough, the values given by syllable-timed languages (French, Yoruba and Telugu) are all higher than those given by stress-timed languages (English, Arabic and Russian).

However, as Roach says, the results of this experiment may have been influenced by the fact that only one speaker per language was recorded and by the difficulty in establishing which are the prominent stresses and, consequently, where the precise boundaries of inter-stress intervals fall. Yet, there seems to be no doubt about the fact that the differences are all too small to be able to draw any conclusions as to the classification of a language into a rhythmic category. Therefore, Roach suggests that Abercrombie's criteria for the distinction between the two rhythmic groups are inadequate and that stress-timing and syllable-timing may only be a matter of perception: "a language is syllable-timed if it *sounds* syllable-timed" (Roach, 1982:78).

### 2.3 Departing from isochrony

#### 2.3.1 The "phonological illusionists"

In the 80s, many phoneticians abandoned Abercrombie's hypotheses as it was clear that the theory of isochrony was not supported by experimental data; at the very least, it had to be re-interpreted as a continuum spanning from a purely hypothetical stress-timing pole to a likewise hypothetical syllable-timing pole. And even so, acoustic measurements on syllables and inter-stress intervals did not always support these claims. Some authors proposed a set of phonological properties held to be responsible for the classification of a language as stress- or syllable-timed: for this reason, they have been labelled as "phonological illusionists" by Bertinetto (1989).

Bertinetto (1977, expanded in 1981 and 1983) measured syllable and foot durations in 15 sentences read by two native speakers of Italian. The durational increase of the foot seems to be a function of the number of syllables or phones that compose it, just like the durational increase of the syllable seems to be a function of the number of phones that compose it. These results support the interpretation of Italian as tending towards syllable-timing: in all the charts reported by Bertinetto (1983) "l'intercettamento dell'ordinata è sempre localizzato nei pressi dell'origine.

## 2. Nearly 100 years of stress-timing and syllable-timing

Ciò significa, senza possibilità di dubbio, che la tendenza verso l'isocronia sillabica dell'italiano si afferma in maniera costante, indipendentemente dal tipo di unità prescelta per la verifica" (1983:1081-2). Bertinetto (1977 and following) also proposes a list of phonological properties which characterise languages tending to one or the other pole of the rhythm continuum. These properties are then reported in his following publications dealing with this topic. I shall report below the version by Bertinetto (1989)<sup>18 19</sup>:

- a) *Vowel reduction vs. full articulation in unstressed syllables;*
- b) *relative uncertainty vs. certainty in syllable counting, at least in some cases;*
- c) *tempo acceleration obtained (mainly) through compression of unstressed syllables vs. proportional compression;*
- d) *complex syllable structure, with relatively uncertain syllable boundaries, vs. simple structure and well-defined boundaries;*
- e) *tendency of stress to attract segmental material in order to build up heavy syllables vs. no such tendency;*
- f) *relative flexibility in stress placement [...] vs. comparatively stronger rigidity of prominence.*
- g) *relative density of secondary stresses, with the corresponding tendency towards short ISI (inter-stress intervals, my insertion), and (conversely) relative tolerance for large discrepancies in the extent of the ISI. This feature seems to oppose languages like English or German on the one side, to languages like Italian or Spanish on the other.*

(Bertinetto, 1989:108-9)

Bertinetto (1989) recognised a) and d) as the most important ones, a view which is essentially shared by Dauer (1983, see below). The phenomenon of vowel reduction is typically a phonological property of stress-timed languages and contributes to give prominence to stressed vowels (and, consequently, to stressed syllables) by shortening the length of unstressed vowels and making their quality less definite (which usually tends to be in the schwa area). On the contrary, in the languages where this phenomenon does not exist or is not consistent (e.g. syllable-timed languages), unstressed vowels tend to have a comparable length and a similar quality to stressed vowels, thus creating the impression that the duration of stressed and unstressed syllables is nearly alike. As for d), it is normally accepted that the syllabic inventory is larger in stress-timed languages than in syllable-timed languages. As a consequence, we can state that while syllable-timed languages have a simple syllabic structure<sup>20</sup> (*i.e.* only light consonantal clusters), stress-timed languages have a complex syllabic structure (*i.e.* they also have heavy consonantal clusters) particularly in stressed syllables, which are then given further prominence.

---

<sup>18</sup> The reason for choosing the 1989 version is simply that it is in English. However, the first version dates to 1977 and Bertinetto (1981) also includes a similar list.

<sup>19</sup> For each pair of properties, the one on the left is typical of stress-timed languages, while the one on the right is typical of syllable-timed languages.

<sup>20</sup> This assumption relies on the fact that languages presenting complex syllables without possessing simple syllables are not known. So, only languages with a large syllabic inventory can have complex syllables. Indeed, the larger the inventory, the more complex syllables a language can afford.



## 2. Nearly 100 years of stress-timing and syllable-timing

Schmid (2004) added other phonological properties, such as the preference for closed syllables in stress-timed languages vs. the preference for open syllables in syllable-timed languages.

Another “phonological illusionist” is Dauer (1983), who measured the duration of inter-stress intervals on read passages of English, Thai, Greek, Spanish and Italian. She found no significant differences in variances of the values obtained for the five languages: in all cases, inter-stress intervals seemed to increase as a function of the number of syllables contained. Dauer suggested that the perceived difference between the two groups of languages might be an effect of differences in “language structure”:

*I would like to propose that the rhythmic differences we feel to exist between languages such as English and Spanish are more a result of phonological, phonetic, lexical, and syntactic facts about that language than any attempt on the part of the speaker to equalize interstress or intersyllable intervals.*  
(1983:55)

She stated that the areas in which languages differ (in respect to rhythm) are the following: syllable structure, vowel reduction and stress (this view is definitely in compliance with Bertinetto, 1977 and following, see above). On these grounds, she also proposed a change in terminology in favour of the term *stress-based*: furthermore, she claimed that it is not necessary to have a second term in opposition to stress-based (such as syllable-based) as the rhythmic distinction between languages has to be done along a continuum and therefore does not need two dimensions: “[I]anguages can be compared to each other along the dimension as having a more or less stress-based rhythm” (1983:59).

In conclusion, as resumed by Bertinetto (1989), “the original dichotomy has gradually lost much of its dichotomic character, and has more and more acquired the aspect of a scalar orientation” in terms of a continuum. The studies by Bertinetto (1977 and following) and Dauer (1983) suggest that the more phonological properties typical of stress-timing a language possesses, the more it can be placed near the stress-timing pole of the continuum; conversely, the more phonological properties typical of syllable-timing a language has, the more it can be placed near the syllable-timing pole of the continuum<sup>21</sup>.

### 2.3.2 Compensation and coarticulation

The term “compensatory shortening” refers to the phonological phenomenon by which, in certain languages, the stressed syllable of a foot or word tends to be compressed according to the number of the following unstressed syllables in that foot or word. More precisely, this phenomenon is called *inter-syllabic compensation* in opposition to *intra-syllabic compensation*, which refers to the phenomenon by which the phonemes of a syllable tend to be compressed in function of the number

---

<sup>21</sup> It has to be remarked that some languages possess properties typical of both rhythmic groups: Nespore (1990, quoted in Ramus et al., 1999) notes that Catalan has a simple syllabic structure but allows for vowel reduction, while, conversely, Polish has a complex syllabic structure but does not allow for vowel reduction. But this of course does not constitute a disproof of Bertinetto’s and Dauer’s hypotheses. It simply indicates that these languages are somewhere in an intermediate position of the continuum.

## 2. Nearly 100 years of stress-timing and syllable-timing

of the other phonemes present in that syllable. Intuitively, inter-syllabic compensation has been associated with stress-timing, whereas intra-syllabic compensation has been associated with syllable-timing. The tendency of readjusting the length of the syllables of each foot (inter-syllabic compensation) and the phonemes of each syllable (intra-syllabic compensation) are in fact interpreted as an attempt to standardise the length of the feet and the syllables, respectively. Studies of this type were initiated by Lindblom & Rapp (1973) and soon caught on.

Fowler (1981) makes an attempt to explain compensatory shortening and coarticulation as closely related phenomena, both due to co-production: in Fowler's words shortening and coarticulation are "different measures of the same articulatory phenomena" (1981:128). This hypothesis is based on a model that sees consonantal segments superimposed on flanking vowels: according to this view, when a vowel is preceded or followed by one or more consonants, it "is measured to be shorter not necessarily because it is shorter in any articulatory sense, but because most of the durational extent over which it is coproduced with a consonant is conventionally assigned only to the consonants" (1981:128). The author also argues that if inter-syllabic compensatory shortening is explained as co-production, then also inter-syllabic compensation can be explained as the superimposition of unstressed vowels on stressed vowels. These hypotheses have been supported by an experiment which revealed the existence of bidirectional formant co-articulation between trans-consonantal vowels. More in particular, the influence of the stressed vowel seems to be more powerful on following than on preceding unstressed vowels, reproducing what Lindblom & Rapp (1973) found for intra-syllabic and inter-syllabic compensation. However, as stated by Fowler, "to demonstrate that coarticulation and shortening, both at the level of the segments in a syllable and at the level of stressed and unstressed syllables, are symmetric is not to confirm that they are in fact two different measures of the same phenomenon" (1981:131).

As for inter-syllabic compensation, it has been studied by various linguists (see Bertinetto, 1989 and 1990, for a summary of many of these studies) and the results seem to confirm that it is a characteristic of stress-timed languages. However, as for intra-syllabic compensation, the results of some studies did not confirm that it is a characteristic of syllable-timed languages. Vayra, Fowler & Avesani (1987, reported by Bertinetto, 1989) noticed more intra-syllabic compensation in English than in Italian and therefore suggested that English presents "intimations of syllable-timing". This view is not shared by Bertinetto (1989), who claimed that "no (alleged) isosyllabic language examined so far exhibits strong inclinations towards intra-syllabic compensation" (Bertinetto, 1989:122). He argued, instead, that intra-syllabic and inter-syllabic compensation should be considered as the different facets of the same property, which is symptomatic of *duration compensating languages* (i.e. stress-timed languages, re-proposing the term introduced by Hoqvist, 1983). He supports his claims by observing that it would be difficult for a language to obey two opposite tendencies, flexibility at the syllable level and compensation at the foot level (or vice versa): "it seems much more sensible to imagine that both levels obey the same tendency" (1989:123). He then argued that "the ultimate difference between iso-accentual and iso-syllabic languages might lie in the *different degrees of flexibility they exhibit at all relevant levels of structure*" (Bertinetto, 1989:123).

## 2. Nearly 100 years of stress-timing and syllable-timing

Finally, he proposed a new list of features that are supposed to characterise compensating languages<sup>22</sup>:

- i. *more intrasyllabic compensation;*
- ii. *more CS [compensatory shortening] at the foot (and word) level;*
- iii. *more vowel reduction in unstressed syllables;*
- iv. *more tolerance for extreme shortening of unstressed syllables;*
- v. *sharp contrast in the exploitation of prosodic features in stressed vs. unstressed syllables;*
- vi. *in general, less sensitivity to all linguistic and non-linguistic events localized on unstressed syllables.*

(Bertinetto, 1989:124)

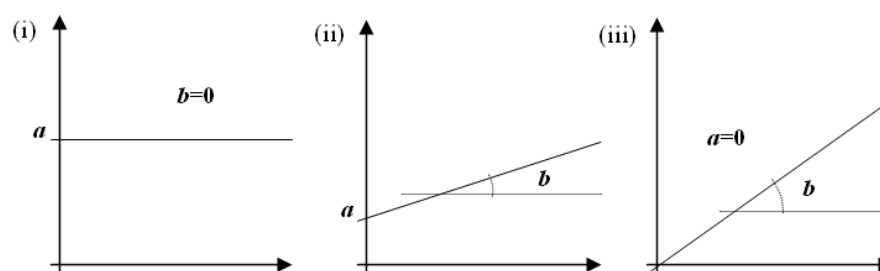
### 2.3.3 Linear regression studies and rhythm modelling<sup>23</sup>

An approach that seems to have yielded interesting perspectives consisted in defining the stress group (inter-stress interval,  $I$ ) as a function of the number of syllables ( $n$ ) according to the following formula:

$$I(n) = a + b * n$$

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:

- a) an absolute stress-timing, when  $b$  is naught and, therefore, the inter-stress interval is a constant ( $b=0 \rightarrow I=a$ ; see figure 2.1);
- b) an absolute syllable-timing, when  $a$  is naught and the inter-stress interval is directly proportional to the number of syllables ( $a=0 \rightarrow I=bn$ ; see figure 2.1);
- c) yet, languages usually tend to show an intermediate form (see figure 2.1).



**Figure 2.1 (from Romano & Mairano, 2010c).** 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).

This approach is fully explained by Eriksson (1991) and Barbosa (2006), but has been used earlier (with some variations) by other authors, such as Bertinetto (1983), Marotta (1985), Farnetani & Kori (1986 and 1990), and later by, for instance,

<sup>22</sup> The author claims that this list has to be intended as an addition to (not a substitution of) the list of phonological features reported above.

<sup>23</sup> This paragraph has been extracted from Romano & Mairano (2010c) and partly re-manipulated.

## 2. Nearly 100 years of stress-timing and syllable-timing

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  $\omega_1$  is the oscillation velocity of the accentual oscillator,  $\omega_2$  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 (Eriksson, 1991, and others 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).

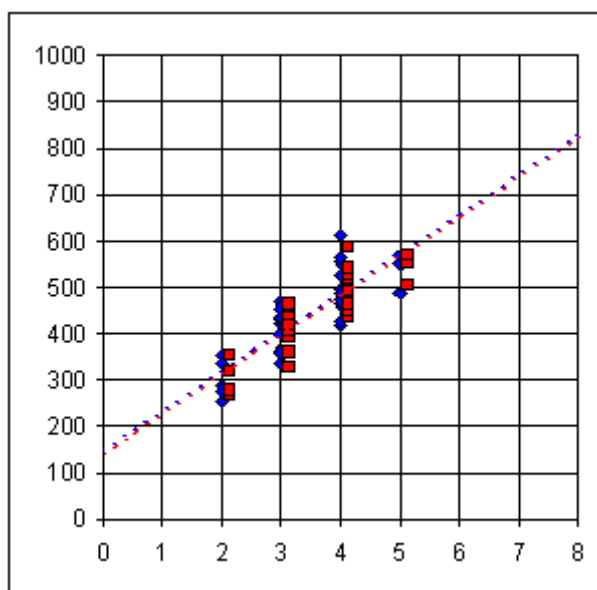
Romano & Mairano (2010) tested this model on a corpus of Italian sentences like the ones analysed by Marotta (1985): *Perciò pésa((me)lo) tûtto di nuovo...* *Perciò pesàte((me)lo) tûtto di nuovo...*<sup>24</sup>. They included similar sentences with different segmental structures (*sposta* instead of *pesa*) and even a nonsense word with growing number of inter-stress syllables as it happens in reiterant speech (*tàta*, *tàtata*, *tàtatata*). Sentences were pronounced by a male speaker with a mean syllable rate of 7.74  $\sigma/s$  (with local minima down to 5.66 and maxima up to 10.25). The five series range from a mean syllable rate of 6.77 to 8.71  $\sigma/s$  defining a fairly homogeneous corpus in terms of tempo. Measurements were taken from the stressed syllable of the word to the first following stressed syllable (*tût*, excluded) thus obtaining  $\sigma$ s-to- $\sigma$ s measures and from the stressed vowel (without the syllable onset) to the first following stressed vowel (*ùt*, excluded) thus obtaining V-to-V measures. The results are summarised in figure 2.2 together with the two regression lines

---

<sup>24</sup> Similar sentences were measured and tested by other authors (Bertinetto, 1977, 1983; Vayra *et al.*, 1984, and Farnetani-Kori, 1986, 1990) bringing evidence on the reduced compression properties of Italian and discussing the discriminant role of stiffness parameters related to syllable and segment durations (similar outcomes are summarised for Spanish and French by Pamies, 1999).

## 2. Nearly 100 years of stress-timing and syllable-timing

giving estimates of a roughly linear growth in both cases: *a* has quite high values (slightly lower for PV: 148 vs. 145) whereas *b* is almost the same for the two measures (rounded to 85).



**Figure 2.2.** The growth of Interstress Intervals for a sample of 27 Italian sentences (◇ for measures at foot level vs. □ for PV measures). (Taken from Romano & Mairano, 2010)

This yields *r* values of 1.74 and 1.70, respectively, as if the accentual oscillator were dominant at phrase level (this is not very surprising according to Bertinetto, 1983, Vayra *et al.*, 1984, Marotta, 1985, and Farnetani-Kori, 1986 and 1990)<sup>25</sup>.

### 2.3.4 Rhythm metrics

A recent approach to speech rhythm involves the so-called *rhythm metrics*. These are treated in detail in the next chapter, so I shall only provide a concise account of the topic. Rhythm metrics are formulae applied to measures of vocalic and consonantal durations giving a representation of the degree of variability of these measures. The rationale behind measuring the variability of consonantal and vocalic intervals rests on the observation by Bertinetto (1977 and following) and Dauer (1983) that the impression of stress-timing or syllable-timing may be given by specific structural properties of the languages (see the list reported above): the most relevant of these properties (mainly vowel reduction and syllable structure) are arguably reflected by the variability of consonantal and vocalic durations.

The first and most used of these metrics include the deltas (see Ramus, Nespor & Mehler, 1999), their normalised versions called varcos (see Dellwo & Wagner, 2003) and the pairwise variability index (see Grabe & Low, 2002). Various authors set out to test the reliability of the metrics, while others used them to classify languages; many of these studies are reported in the next chapter.

<sup>25</sup> The sensitivity of the measures to the segmental content of syllables is evident when analysing separately the fifth series: the low value of the coupling strength (0.14) accounts for very variable results (in contrast with e.g.  $r = 4.72$  of the second series): these variations are greater than the ones induced by changes in speech rate.

## 2. *Nearly 100 years of stress-timing and syllable-timing*

A new and alternative method has recently been proposed by Bertinetto & Bertini (2008 ad following), which has its root in studies of compensation by Fowler (1977 and following) and in the idea expressed by Bertinetto (1989) that stress-timed (or compensatory) languages might be characterised by a higher compressibility at all levels (both intra- and inter-syllabic).

### **2.4 Back to syllable-timing and stress-timing?**

#### **2.4.1 A revisitiation of stress-timing and syllable-timing**

As has been outlined, the failure of the quest for isochrony at the syllable and foot level has discouraged research to persist on using the syllable and the foot as rhythm units: in fact, rhythm metrics are applied to vocalic and consonantal intervals. However, in recent years, some authors went back to look for a way of re-integrating these units into an account of speech rhythm, still of course keeping in mind the newer approaches. I shall report on a few studies in which rhythm metrics were applied to syllable and/or inter-stress durations.

Wagner & Dellwo (2004) proposed a new metric, ironically named YARD (Yet Another Rhythm Determination), which is actually constituted by the formula of the raw pairwise variability index but which is applied to z-transformed syllable durations. They tested it on samples of English, German, French and Italian from the BonnTempo corpus: the results were encouraging, with lower values (indicating lower syllable variability) for Italian and French than for English and German. The authors conclude that “future research should again concentrate on regarding rhythm as a sequence of – roughly – isochronous events within language specific internal structures” (2004:4/4).

Asu and Nolan (2006) applied the PVI to measures of consonantal and vocalic durations of Estonian (which is classified as a syllable-timed language), but also (more innovatively) to syllable and foot durations of Estonian and English. While consonantal and vocalic PVI values mainly reflected Grabe & Low’s findings for Estonian, syllabic and foot PVI values showed interesting results: English, in compliance with the alleged tendency of stress-timed languages to control for the foot and not for the syllable, seems to have a greater variability at the syllable level than Estonian; instead, these two languages exhibit a nearly identical low-value foot PVI, thus indicating a low variability at the foot level. This seems to suggest that Estonian tends to control both at the syllable and at the foot level, thus proving that the results for the PVI calculated at the syllabic or intra-syllabic level are independent of the results for the PVI obtained at the stress level. The authors conclude that “a two-dimensional characterisation using syllable and foot PVIs [...] gives a more appropriate and subtle account of the rhythm of languages, and in the case of Estonian, examined here in detail, explains the intuition that it is both stress and syllable timed” (Asu & Nolan 2006:4/4).

I believe their results present extremely interesting consequences because they suggest a revision of the traditional stress-timed vs. syllable-timed dichotomy into at least a quadriparted rhythmic classification. In other words, languages are no longer classified as either stress-timed or syllable-timed, or rather on a bi-polar continuum; instead, they can be spaced within a quadri-polar area which allows for languages being stress-timed (controlling for feet), syllable-timed (controlling for

## 2. Nearly 100 years of stress-timing and syllable-timing

syllables), stress-and-syllable-timed (controlling for feet and syllables) and, perhaps, a-timed (controlling for neither syllables nor feet). These ideas have to be connected to what has been suggested by Bertinetto & Bertini (2010), who also suggest the possibility of classifying languages on a bi-dimensional paradigm.

### 2.4.2 An unplanned experiment

I shall now present an experiment which I had not planned, but which was inspired by the reading of the two studies reported above. I realised that I disposed of a segmented and CV-labelled multi-language corpus which I had prepared for other purposes (see chapter 3 for the details) but which could easily be exploited to compute rhythm metrics on syllable durations<sup>26</sup>. It has to be noted that by “syllable durations” I do not mean a phonological notion of syllable, rather an onset-to-onset distance (henceforth *inter-onset distance*): this is in compliance with many other authors, such as Farnetani & Kori, who view the rhythm syllable “as the temporal interval extending from the onset of a vowel to the onset of the following one” (1986:27). I shall also remark that vowels in hiatus have been considered as two intervals.

I shall not give details on the data, the segmentation and the methodology because all these aspects are detailed in chapter 3. I shall only briefly say that samples are constituted of read speech (the narrative *The North Wind and the Sun* translated in all the languages analysed – some versions were taken from the *Illustrations of IPA*, others recorded at LFSAG<sup>27</sup> or during fieldwork). Samples were segmented and labelled on *Praat* by two different phoneticians (AR and PM), but only data by PM has been used for this experiment<sup>28</sup>. In order to obtain the values of inter-onset distances, I wrote a *Praat* script that outputs the distances between each successive pair of vocalic labels. The values were finally analysed with a special function of *Correlatore* (see chapter 4).

It is important to note that this experiment is only presented here as a preliminary test, which may or may not bring to any relevant results. Although it exploits the formulae of the rhythm metrics and despite the data is described in detail in chapters 3 and 5, it was decided to present this test here because it is inspired by the studies quoted above and because it includes inter-onset measures: I thought that it would find its most natural collocation in this chapter, which deals more directly with syllable- and stress-timing. Moreover, it should be noticed that the test also recalls Roach (1982), who calculated the standard deviation (the delta) on inter-onset and inter-stress durations. As has been said above, his results were far from confirming Abercrombie’s isochrony hypotheses, but the PVI might (or might not) yield better results (see chapter 3 for an account of the differences between the delta and the PVI).

---

<sup>26</sup> Unfortunately not on inter-stress durations because stresses had not been marked.

<sup>27</sup> Laboratorio di Fonetica Sperimentale “Arturo Genre”, University of Turin.

<sup>28</sup> I have also excluded the 10 Icelandic and the 6 Romanian for compatibility reasons. As explained in chapter 3, the 10 Icelandic samples had been labelled before the development of *Correlatore* and therefore using other criteria. The 6 Romanian samples, instead, have not been labelled with CV transcriptions, but as SAMPA transcriptions (which are also accepted by *Correlatore*); it would have not been difficult to convert SAMPA transcriptions into CV sequences, but for the moment I decided to stick with the data that was readily available.

## 2. *Nearly 100 years of stress-timing and syllable-timing*

### 2.4.3 The results

Figure 2.3 shows the results for the deltas and the PVI<sub>s</sub> on inter-onset durations for the data. All single values for each sample analysed can be consulted in appendix 1, while the charts only show the mean of samples of the same language. As it can be seen, Spanish, Italian and Greek (supposedly syllable-timed languages) show a lower variability of inter-onset distances, while German and, even more, Czech exhibit a high variability. However, one would not expect to find such high values for French (which sticks between Czech and German), Japanese<sup>29</sup> and Estonian. By checking the results obtained on each single sample (reported in appendix 1), it can be seen that variability between speakers of the same language is extremely high (this is clearly visible for the 15 Italian speakers) and, consequently, samples of languages belonging to different rhythm classes greatly overlap.

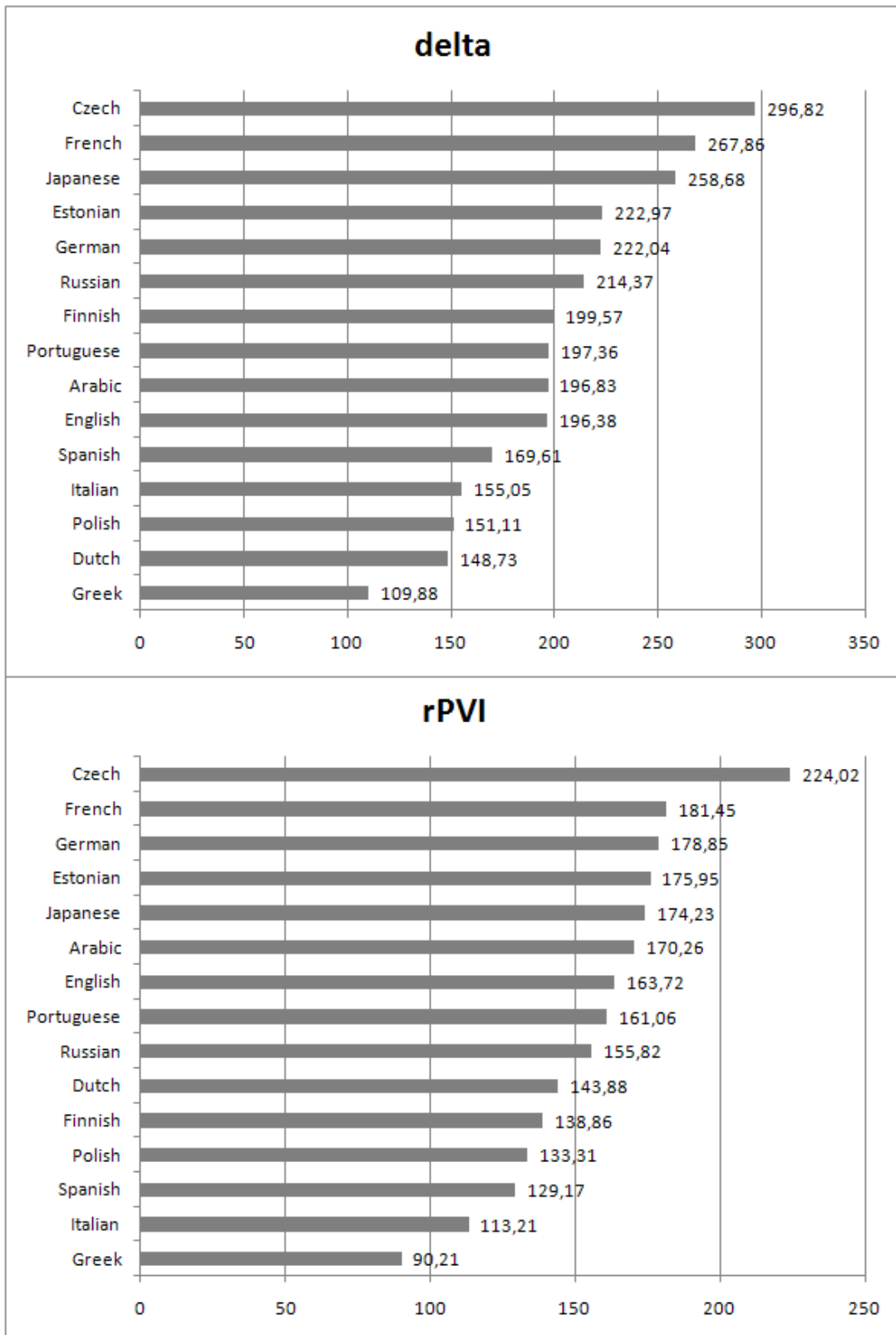
As for the hypothesis that the rPVI might provide a better discrimination, results do seem to improve with such a measure (see for example Dutch and Polish). However, on the whole, I would say that the discriminatory power of these measures applied to inter-onset distances looks dubious.

---

<sup>29</sup> It has to be specified that the Japanese sample has been labeled in two different ways, phonologically (counting devoiced vowels as vocalic segments) and phonetically (counting devoiced vowels as consonantal segments). For practical reasons, in the present experiment I used only the phonetically labelled file (since the script also calculated f<sub>0</sub> values for other experiments which are not included in this thesis). It is of course probable (and, indeed, logical) that a phonological labelling would have resulted in more congruent results: as is cleared in chapter 3, devoiced vowels have in fact a great impact on the variability of consonantal intervals, and consequently of inter-onset intervals as well.



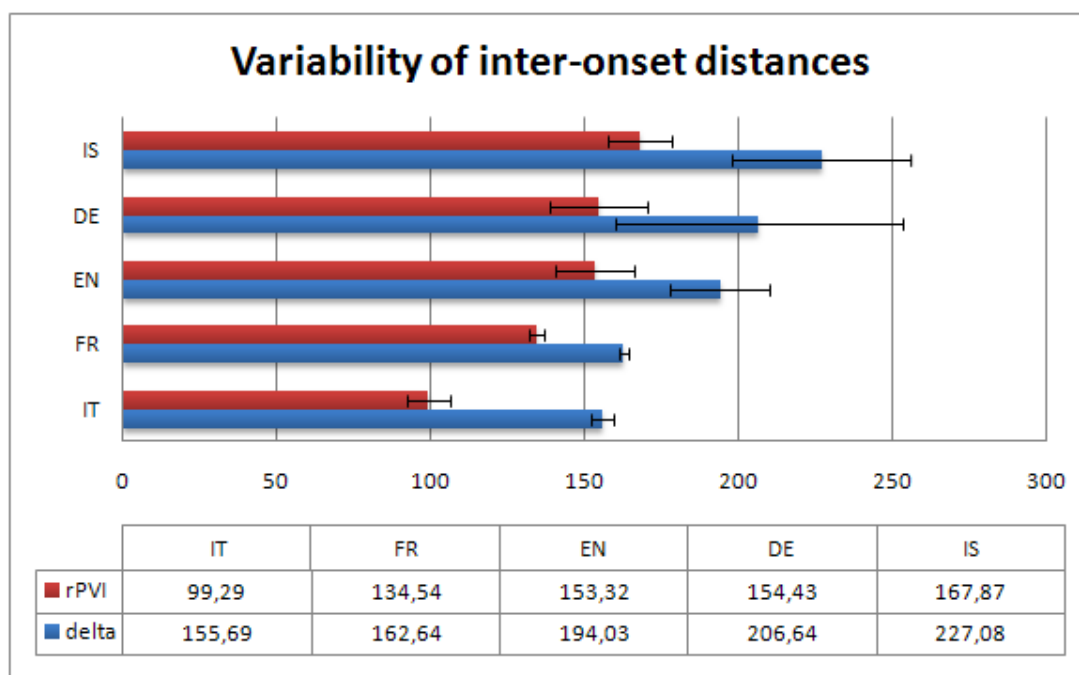
## 2. Nearly 100 years of stress-timing and syllable-timing



**Figure 2.3.** The delta (above) and the PVI (below) applied to inter-onset durations of samples of 15 languages (for more details on the data, see chapter 3 keeping in mind that only the files segmented by PM have been considered for the present experiment).

## 2. Nearly 100 years of stress-timing and syllable-timing

In order to verify this approach on more controlled data, I also applied the metrics to 10 CV-labelled samples of one and the same speaker reading *The North Wind and the Sun* in 5 different languages (Italian, English, French, German and Icelandic, twice for each language). Data of this speaker are presented in more detail in chapter 5, it will suffice here to say that he is a native speaker of Italian and he is fairly fluent in English, French and, to a lesser degree, German. Results of the delta and the rPVI are shown in figure 2.4.



**Figure 2.4.** Values of the delta and the rPVI for inter-stress intervals calculated on a (native Italian) speaker in 5 languages. Results shown constitute the mean of the values obtained for two repetitions of *The North Wind and the Sun*. Standard deviations are shown as error bars.

Even though most of the data comes from an L2 speaker, the results seem to better meet expectations: samples of Italian and French (supposedly syllable-timed) show a low durational variability of inter-stress intervals both with the delta and with the PVI, whereas samples of English and German (supposedly stress-timed) show a high variability with both measures. Icelandic presents very high values even though the classification of this language is controversial (see chapter 3). Also, it has to be considered that the speaker is not proficient in this language, so conclusions are at best not drawn on this particular sample.

### 2.4.4 A brief discussion of the results

In summary, it can be said that these measures (in the particular the PVI) applied to inter-onset intervals do seem to have something to say about speech rhythm. It might be interesting to calculate them on inter-onset intervals as well and put the results on bi-dimensional charts. This would also allow for a unified representation of the two levels of speech rhythm, allowing for an ideally quadriparted rhythm space in the chart (for instance, the syllable level on x axis and the stress level on the y axis). However, the idea is very hazardous and there are no results at the moment that can confirm the reliability of the PVI at higher levels. As put forward by Bertinetto &

## 2. *Nearly 100 years of stress-timing and syllable-timing*

Bertini (2010), in fact, variability at the second level is more difficult to capture as it might be realised in different ways in different languages.

### **2.5 Conclusion**

In these pages I have provided a sketch of the history of research on stress-timing and syllable-timing. Many ‘sceptic’ authors have wondered why a distinction based on no empirical data (actually, often contradicted by empirical data) has had such a fortune. Barry & Andreeva write:

*One of the intriguing and, at the same time frustrating things about the instrumental analysis of spoken-language rhythm is that it has stubbornly survived, without empirical justification it would seem, for a human life-span. In dynastic terms, we should be well into the third generation, traditionally a guarantee of the approaching demise.*  
(2010:27)

However, judging from the present vitality of the studies in this field, one would not say that the demise is getting closer. On the contrary, it seems that the flourishing of new methods and models (be they successful or not) is convincing more and more people to invest in the research of speech rhythm. This is clearly illustrated in the next chapter, which focuses on the studies that have used rhythm metrics.

**3.**

**1999-2010:  
rhythm metrics**

### 3.1 Introduction

In the last decade, research in speech rhythm has focused on rhythm metrics (initially called rhythm correlates), that is to say in variables derived from durational measurements of consonantal and vocalic intervals. Studies following this new approach have at least two things in common, which I shall illustrate by quoting the two initiators:

*We depart [...] from the search for isochrony.*  
(Grabe & Low, 2002:516)

*Instead, [...] we measured the duration of vocalic and consonantal intervals. A vocalic interval is located between the onset and the offset of a vowel, or of a cluster of vowels. Similarly, a consonantal interval is located between the onset and the offset of a consonant, or of a cluster of consonants.*  
(Ramus, Nespor & Mehler, 1999:272)

Research on rhythm has gained new vitality after the publication by Ramus, *et al.* (1999), which has been the turn of the screw in the studies of linguistic rhythm. They proposed a set of three phonetic correlates and claimed that the results managed to discriminate languages belonging to the three traditional rhythm categories (stress-timed, syllable-timed and mora-timed<sup>30</sup>). More or less at the same time, a similar approach had been independently developed by Low and co-workers, started by Low & Grabe (1995, quoted by Grabe & Low, 2002): they proposed another index which differs from the one suggested by Ramus *et al.* (1999) in that it takes in consideration the temporal succession of segments.

Further studies by various authors have been aimed at testing these rhythm metrics on different data and many have shown their instability in relation to some factors (particularly speech rate). Controversial results have been obtained for the same languages across different studies or in different speech styles. For these reasons, some authors have taken their distance from rhythm metrics (e.g. Barry & Russo, 2003), while others have proposed a modification of some metrics in order to reduce their sensitivity to speech rate (e.g. Dellwo & Wagner 2003). More recently, Bertinetto & Bertini (2008) have proposed a new index which is based on a different rationale from the others and which has its roots in previous studies on compensation.

In this chapter, I shall give a detailed account of the research dealing with rhythm metrics from the publications of Ramus *et al.* (1999) to the present day (2010). In 3.2 the most frequently used rhythm metrics will be explained and illustrated one by one in an effort to try and follow the linear development of research in this topic. In 3.3 I shall present the results obtained by calculating the metrics on the corpus which was introduced in chapter 2. Finally, in 3.4 I shall hint at the possibility of applying the metrics on parameters other than duration for the study of speech rhythm. An important remark concerns the terminology: in an attempt to respect the preferences of each author, I shall use the terms *rhythm*

---

<sup>30</sup> Mora-timing is the rhythm class to which Japanese is said to belong.

### 3. 1999-2010 : *rhythm metrics*

*correlates*, *rhythm metrics* and *rhythm measures* as synonyms to refer to these variables calculated from consonantal and vocalic measurements.

## 3.2 *An account of rhythm metrics*

### 3.2.1 The deltas

Ramus, Nespor & Mehler (1999) proposed three phonetic correlates of rhythm relying on the list elaborated by Dauer (1983) and which contains the phonological properties which are believed to be responsible for the perception of a language as either stress-timed or as syllable-timed. As already mentioned in chapter 2, the most important of these phonological properties are: a) the presence vs. absence of vowel reduction and b) a complex vs. simple syllabic structure. The acoustic correlates they proposed are supposed to account for these properties by applying specific mathematical formulae to the durations of vocalic and consonantal intervals.

The authors derive three variables from these measurements, namely  $\Delta V$ ,  $\Delta C$  and %V. They consider the standard deviation of vocalic intervals ( $\Delta V$ ) to be indicative of the presence/absence of a high degree of vowel reduction in unstressed syllables: stress-timed languages, allowing for a high degree of vowel reduction, are supposed to present a higher variability between the length of stressed, fully articulated vowels and unstressed, short, reduced vowels. Therefore, these languages were expected to result in a higher value of  $\Delta V$ .

The standard deviation of consonantal intervals ( $\Delta C$ )<sup>31</sup> was instead supposed to be indicative of syllable complexity: the higher its value, the more complex the syllabic structure. This claim relies on the observation that, as has been previously mentioned, languages with a simple syllabic structure (syllable-timed languages) presumably only allow for simple consonantal clusters, whereas languages with a complex syllabic structure (stress-timed languages) allow for both simple and complex consonantal clusters resulting in a higher value of  $\Delta C$ . The standard deviation measures the degree of variability on a list of values by comparing every possible pair according to the following formula:

$$\sigma_x = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$$

The vocalic percentage (from now on %V) is meant to be an acoustic correlate of both the complexity of the syllabic structure and the presence/absence of a high degree of vowel reduction<sup>32</sup>. The fact that %V will be lower for languages presenting a high degree of vowel reduction (stress-timed languages) is intuitive enough not to need further clarification, while the assumption that %V will also be lower for languages allowing for more complex syllables can be accounted for by

---

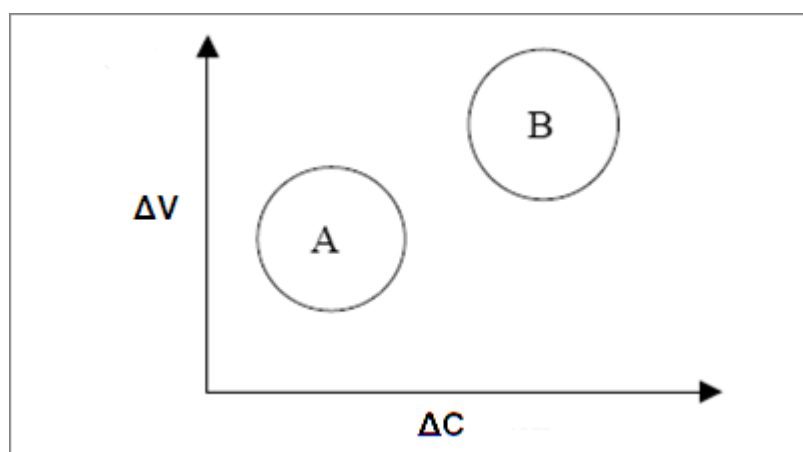
<sup>31</sup> Ramus et al. (1999) refer to  $\Delta C$  as to “standard deviation of inter-vocalic intervals”. I have decided to use the term “standard deviation of consonantal intervals” for its straightforwardness and in order to be consistent with its abbreviated form  $\Delta C$ .

<sup>32</sup> It has to be noted that, as Ramus et al. (1999) point out, the consonantal percentage is isometric to the vocalic percentage and thus needs not be calculated.

### 3. 1999-2010 : rhythm metrics

saying that a complex structure implies a higher consonantal percentage, that is to say a lower vocalic percentage.

The author's expectations are illustrated in figure 3.1 (taken from Barry & Russo, 2003), where the A circle represents languages traditionally classified as syllable-timed (expected to result in lower values of  $\Delta V$  and  $\Delta C$ ), whereas the B circle represents languages traditionally classified as stress-timed (expected to result in higher values of  $\Delta C$  and  $\Delta V$ ).



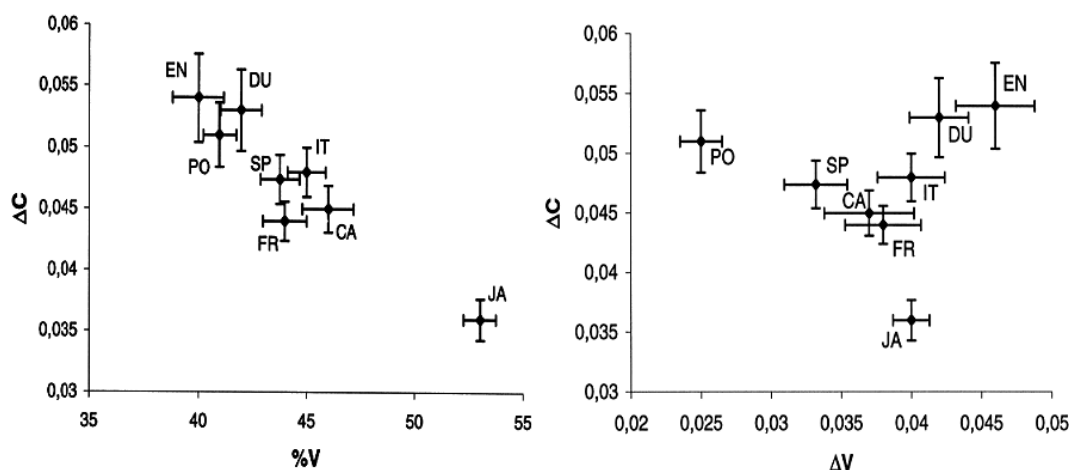
**Figure 3.1.** Chart showing the alleged difference in the values of  $\Delta V$  and  $\Delta C$  for syllable-timed (A) and stress-timed (B) languages. (From Barry & Russo, 2003).

Ramus *et al.* carried out an experiment based on eight languages (three supposedly stress-timed – English, Dutch and Polish –, four supposedly syllable-timed – Italian, French, Catalan and Spanish – and one supposedly mora-timed – Japanese) with 5 sentences uttered by four native speakers per language. The authors describe their data as follows:

*Sentences were short news-like declarative statements, initially written in French, and loosely translated into the target language by one of the speakers. They were matched across languages for the number of syllables (from 15 to 19), and roughly matched for average duration (about 3 s).*

*(Ramus et al., 1999:271)*

They calculated the three acoustic correlates for their data and put the results on three charts combining the values obtained for the correlates in each language. These charts seem to confirm their hypotheses: English and Dutch cluster in a group with high values of  $\Delta C$  and  $\Delta V$  and lower values for %V, whereas Catalan, French, Italian and Spanish cluster in a group with opposite values; Japanese occupies an isolated position, presenting even lower values of  $\Delta C$  and  $\Delta V$  than syllable-timed languages and a higher value of %V (thus looking more syllable-timed than languages traditionally considered as such); Polish, however, sticks with English and Dutch as for %V and  $\Delta C$ , but it exhibits very low values of  $\Delta V$ .



**Figure 3.2.** %V/ΔC and ΔC/ΔV charts, from Ramus *et al.* 1999:273.

In the second part of their article, the authors described a set of experiments about the perception of rhythm which seemed to suggest that both adults and infants can discriminate languages belonging to different rhythm classes. These experiments are dealt with in more detail in chapter 6. The authors conclude that the %V/ΔC chart (see figure 3.2) has the best discriminatory power for languages belonging to different rhythm classes, as ΔV is too heavily influenced by other factors.

Finally, I would like to stress something that seems to pass unnoticed in most modern studies on speech rhythm: *the deltas in se have nothing new*, they are simply the standard deviation, which had already been used in previous studies looking for isochrony in the '70s and '80s (e.g. Roach, 1982). Rather, what is innovative is the fact that they are not applied to syllable or to inter-stress durations, but to other linguistic entities, *i.e.* to vocalic and consonantal durations. This implies that the phonetic reality corresponding to the perceptive distinction between stress-timing and syllable-timing is no longer sought in terms of syllable or foot isochrony, but by measuring acoustic correlates of the phonetic properties associated to stress-timing or syllable-timing.

### 3.2.2 The PVI

Low and co-workers developed a slightly different approach based on the Pairwise Variability Index (PVI), which is meant to give an indication of the variability of vocalic and consonantal intervals.

The PVI, just as the deltas, is applied to the duration of vocalic and consonantal intervals, but its advantage consists in considering the segments in their temporal succession ( $m$  is the number of intervals, while  $d_k$  is the duration of the  $k^{th}$  interval):

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

In other words, the formula of the raw PVI (rPVI) calculates the difference in duration of all pairs of successive intervals and finally calculates the mean of all differences. This is in contrast with the rationale of the standard deviation, whose formula considers all possible pairs (successive and non successive). The normalised

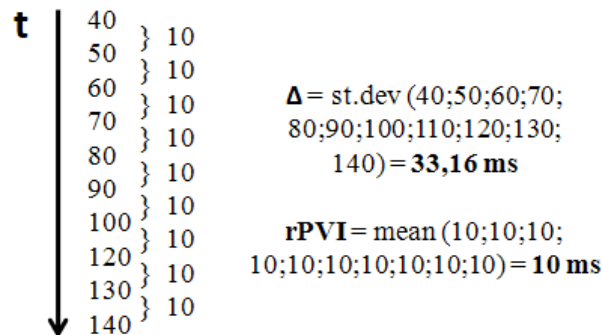


### 3. 1999-2010 : rhythm metrics

PVI (nPVI) is basically the same formula, but it adds a normalisation by dividing the duration of each interval by the mean duration of pairs:

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

The different rationales of the standard deviation and the PVI are illustrated in figure 3.3 with an example: a purely hypothetical and extremely artificial sequence of vocalic durations in decreasing tempo, in which each vowel increases by 10 ms would yield very different results of  $\Delta$  and rPVI: the standard deviation formula considers the differences of all possible pairs, while the rPVI exclusively considers the differences between successive pairs (which in this case is always 10). This example is, of course, absurdly artificial, but it clearly illustrates the difference between the two formulae.

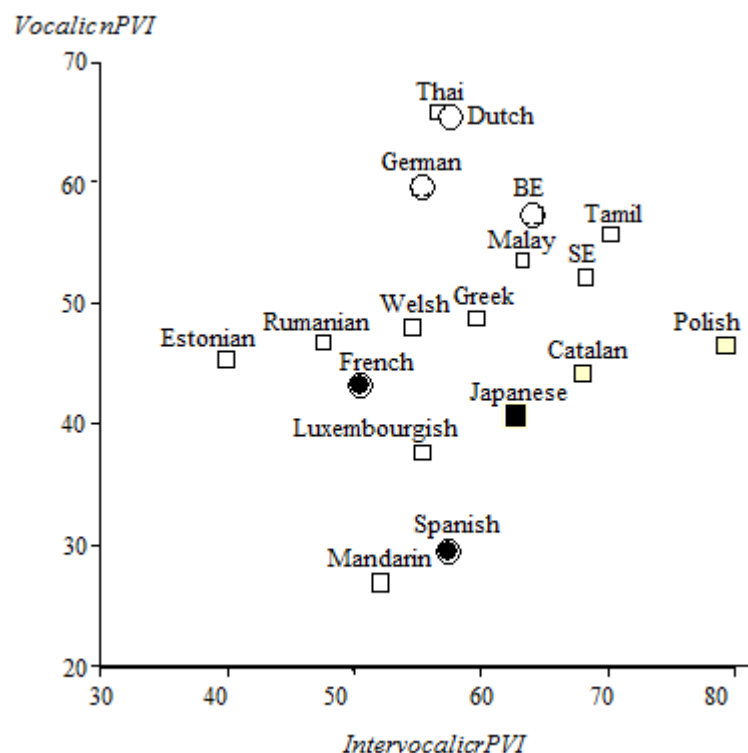


**Figure 3.3.** A comparison of the  $\Delta$  and the rPVI.

Work on the PVI started with Low & Grabe (1995, quoted by Grabe & Low, 2002) and Low, Grabe & Nolan (2000, quoted by Grabe & Low, 2002), who applied the nPVI to vocalic measures and found a higher variability for British English (supposedly stressed-timed) than for Singapore English (supposedly syllable-timed). They also claimed that the nPVI of vocalic interval is a better indication of rhythmicity than  $\Delta C$  and  $\Delta V$ . These ideas were further developed in Grabe & Low (2002), an article which has now become the reference for all researchers using the PVI for the study of linguistic rhythm.

Grabe & Low (2002) conducted an experiment in order to test the deltas against the PVI. They included all the languages studies by Ramus *et al.* (1999) apart from Italian – thus Catalan, Dutch, British English, French, Japanese, Polish, Spanish – and added many more – Estonian, German, Greek, Luxembourgish, Malay, Mandarin Chinese, Rumanian, Singapore English, Tamil, Thai and Welsh. They ended up with four stress-timed languages (English, German, Dutch and Thai), four syllable-timed languages (French, Spanish, Tamil and Singapore English), one mora-timed language (Japanese) and 9 mixed or uncategorised languages. Instead of “news-like declarative statements”, they used translations of the *North Wind and the Sun* and only one speaker per language.

### 3. 1999-2010 : rhythm metrics



**Figure 3.4.** Vocalic nPVI/Consonantal rPVI chart, taken from Grabe & Low (2002:7/16).

Their results of the PVI (see figure 3.4) confirmed expectations fairly well, with stress-timed languages exhibiting high values of vocalic nPVI. Consonantal rPVI, instead, seemed to have a lower discriminatory power, with comparable values for both stress-timed and syllable-timed languages. The chart seems to work well also for mixed languages such as Polish, whose high rPVI value reflects its complex syllable structure (with long consonant clusters, up to 5 segments) and whose low nPVI value reflects the fact that it does not have phonological vowel reduction (these findings are similar to the ones obtained by Ramus *et al.* 1999).

The authors also calculated %V and  $\Delta C$  for the same data in order to compare results with the PVI. They found that the disposition of languages was in some cases similar, but then some languages moved to a different area of the chart, less in compliance with expectations. The authors explained this by claiming that the three correlates proposed by Ramus *et al.* (1999) do not work when some variables come to play a role, especially speech rate.

Grabe & Low's conclusion is that the disposition of languages in rPVI/nPVI chart proves that "a categorical distinction between stress-timing and syllable-timing cannot be defended" (2002:525) as languages tend to scatter within the chart and many of them occupy intermediate positions even with some overlap between the two traditional classes.

#### 3.2.3 Considerations on speech rate and the varcos

Ramus (2002) is a response to Grabe & Low (2002) in which the author comments on their results and calculates the PVI on his own corpus. He finds that the results given by the PVI are similar to those given by the deltas (though nPVI actually provides a clearer differentiation than  $\Delta V$ ) and stresses the fact that, since Grabe &

### 3. 1999-2010 : rhythm metrics

Low used just one speaker per language, their results could reflect speaker's idiosyncrasies. Ramus advocates for a higher number of speakers being studied and admits that the validity of  $\Delta C$  and  $\Delta V$  as correlates of rhythm can easily be influenced by speech rate. He claims:

*It is essential to have a variety of speakers for each language;  
It is essential to control for speech rate, either by constraining the corpus, or by using a normalisation procedure;  
The usefulness of variables such as  $\Delta V$  and  $\Delta C$  may well be limited to corpora where speech rate is strictly controlled.*  
(Ramus, 2002:117)

He then proceeds to a survey of several problems which arise when attempting to control speech rate and, in conclusion, states: "Salvation lies in larger data sets. [...] Automatic speech processing carries hopes of effortless constitution of unlimited corpora, as well as the spectres of imprecision and meaninglessness." (Ramus, 2002:119).

Barry & Russo (2003) calculated both the deltas and the PVIs on semi-spontaneous dialogues from the AVIP corpus for 13 Italian speakers (6 from Naples and 7 from Pisa) and from the Kiel corpus for 4 speakers (the dialogues are longer than in the AVIP). The results they obtained seem to confirm neither the validity of  $\Delta V$  and  $\Delta C$ , nor that of the PVIs as correlates of rhythm since their values, contrary to expectations, are on average higher for speakers from Naples and Pisa than for German speakers. The authors argue that speech rate heavily influences the values of these correlates: they presented a chart suggesting that an increase in speech rate takes the languages towards a position associated with syllable-timing according to the methods proposed by Ramus *et al.* (1999). Russo & Barry (2008) include further considerations from the observation of the same data. In particular, the authors notice that %V is "by far the most tempo-resistant and the most language-distinguishing measure" while the "[...] Ramus delta values and the Grabe and Low PVI values are to a considerable part a function of articulation rate" (Russo & Barry, 2008:4/4).

Dellwo & Wagner (2003) conducted an experiment on English, French and German in order to examine the influence of speech rate on %V and  $\Delta C$ . Their results reflect the language groups found by Ramus *et al.* (1999). Speech rate is found to affect the values of the correlates (above all  $\Delta C$ ), but its influence does not seem to be strong enough to prevent the clustering of stress-timed and syllable-timed languages into separate areas. In order to correct the sensitivity shown by  $\Delta C$ , the authors propose to divide this variable by the mean value of consonantal intervals, thus obtaining varco $\Delta C$ <sup>33</sup>:

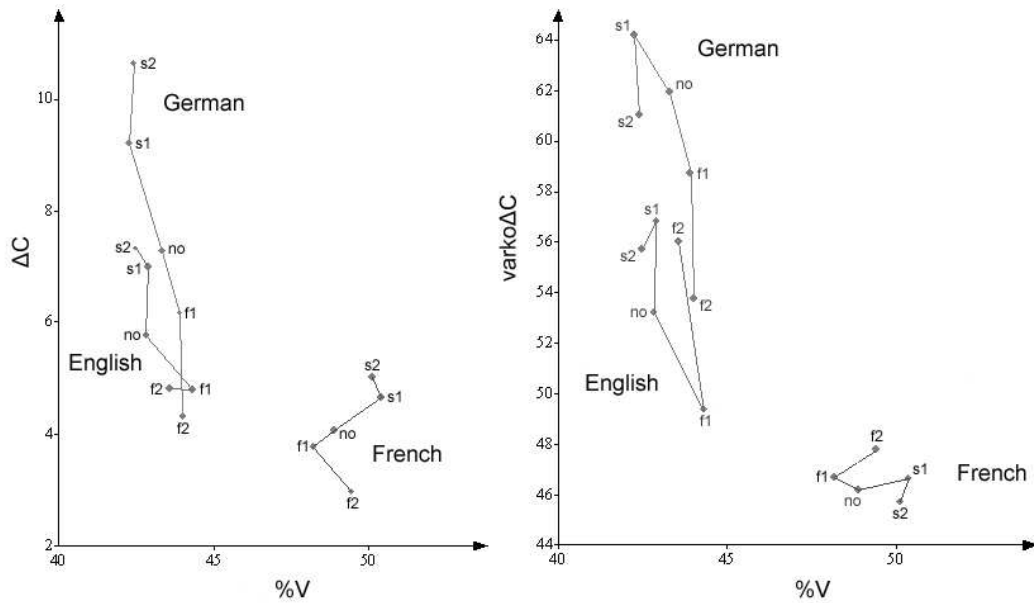
$$\text{varco}\Delta C = \Delta C / \text{meanC} * 100$$

This idea follows the same rationale as the normalisation found in the formula of the nPVI, where interval durations are divided by the mean duration. Varco $\Delta C$  is then tested in Dellwo (2006), where the author calculates this parameter on data drawn from the BonnTempo corpus<sup>34</sup>. The results can be seen in figures 3.5.

<sup>33</sup> Varco stands for variation coefficient.

<sup>34</sup> The Bonntempo corpus is a corpus devised by the V. Dellwo for the combined study of speech rhythm and rate (see Dellwo *et al.*, 2004). The data presented in Dellwo (2006) includes the same

### 3. 1999-2010 : rhythm metrics



**Figure 3.5.**  $\Delta C$ / $\%V$  chart, taken from Dellwo (2006).

The figures show that the results are encouraging, as

*general cluster patterns of stress-timed and syllable-timed languages are clearer with varco $\Delta C$  than with  $\Delta C$  since all F [French] versions lie well below E [English] and G [German] on the varco $\Delta C$  scale which is not the case for  $\Delta C$  [...]. In other words: the use of a variation coefficient for  $\Delta C$  enhances differentiability of rhythm classes for the data presented. (Dellwo, 2006:5/8)*

More recently, Benton (2010) proposes a slight modification of the normalisation applied by Varco $\Delta C$  and by the nPVI. As for Varco $\Delta C$ , the author proposes to divide  $\Delta C$  for the local mean of consonantal durations, instead of the overall mean. However, this approach is not entirely new and a similar remark had already been put forward by Mairano & Romano (2007a and b) referring to the deltas: the two authors calculated the deltas both “globally” (the A method) and “locally” (the B method)<sup>35</sup>. As for the modification of the nPVI, the author suggests that the formula should not only normalise an interval in relation to the preceding interval, rather the “Reverse-normalized PVI would then attempt to normalize the utterance over all past durations from that particular utterance”. The formula he proposes is the following:

$$\text{Rev\_nPVI}_i = 100 \times \left[ \frac{\sum_{k=1}^{m-1} |d_k - d_{k+1}|}{\left( \sum_{i=0}^n d_{k-i} + d_{k+1} \right) / (2+i)} \right]$$

data used in Dellwo & Wagner (2003) plus more (in total, 12 German speakers, 7 English speakers and 7 French speakers).

<sup>35</sup> The two methods are also available in the *Correlatore* software, see chapter 4 and Mairano & Romano (2009).

### 3.2.4 Deltas and varcos versus the PVI, the debate goes on

The studies of Ramus, Nespor & Mehler (1999) and Grabe & Low (2002) have given new impulse to the research in the field of rhythm and several authors have attempted to enlarge the perspectives with the aim of including as many different variables as possible. I shall now report on some of the numerous authors who set out to test the validity of these correlates. However, I shall concentrate here on authors who have dealt with general problems of the rhythm categorisation of languages or who studied characteristics of a specific language. Those who focused on the issue of variation and/or variability in speech rhythm using the rhythm metrics are reported in chapter 5<sup>36</sup>.

Schmid (2001) conducted an experiment on English, German, Swiss German (stress-timed languages), Italian, French, Spanish (syllable-timed languages) and calculated the values of the three parameters proposed by Ramus *et al.* (1999). His results seemed to confirm the validity of  $\Delta C$  and  $\Delta V$  as correlates of rhythm, but not of %V, which was surprisingly high for English and very low for French. He concluded that the results may have been influenced by the fact that only a small corpus was used and he puts forward the need of wider studies including more speakers and more data.

Galves *et al.* (2002) drew on this quest (also stressed by Ramus, 2002, and by most authors who worked on rhythm metrics). They argue that hand labelling is very time-consuming and, furthermore, brings about several inconsistencies in relation to, for example, phonological choices: most emblematically, the decision on whether or not a certain vocalic segment is present has huge implications as it affects greatly the value of  $\Delta C$  (because the surrounding consonants will be considered as one long or two short segments according to the choice). On the basis of the perceptive tests carried by Mehler *et al.* (1996), in which new-borns are found to discriminate between languages belonging to rhythmic classes with a signal filtered at 400Hz, Galves *et al.* argue that

*at this level it is hard to distinguish nasals from vowels and glides from consonants. This strongly suggests that the discrimination of rhythm classes by babies relies not on fine-grained distinctions between vowels and consonants, but on a coarse-grained perception of sonority in opposition to obstruency.*

*Therefore a natural conjecture is that the identification of rhythm classes must be possible using a rough measure of sonority.*

*(Galves et al., 2002:2/4)*

The authors manage to capture the variability of voiced and devoiced intervals by applying some autosegmentation procedures on the same corpus used by Ramus *et al.* (1999). Results mainly reflect those obtained on vocalic and consonantal intervals.

A similar idea was applied by Dellwo, Fourcin & Abberton (2007) who calculated Varco $\Delta$ Unvoiced and %Voiced on English, German, French and Italian: results suggest that it is possible to discriminate between rhythm classes by using

---

<sup>36</sup> The distinction between the two approach is of course not always neat, as some authors deal with both. Still, I have tried to draw a line in order to shape the two chapters along these two different themes.

### 3. 1999-2010 : *rhythm metrics*

these measures as the classification is very similar to the one obtained with  $\Delta C$  and %V.

Lijan (2004) computed the nPVI on vocalic durations (separately for full and reduced vowels) of American English (considered as stress-timed) and Taiwan English (supposedly syllable-timed). The results confirm author's expectations as they show higher nPVI values for American English than for Taiwan English both for full and reduced vowels<sup>37</sup>.

Rouas & Farinas (2004) conducted an experiment on English, German, Mandarin Chinese (stress-timed languages), French, Italian, Spanish (syllable-timed languages) and Japanese. They included a large number of speakers by using an autosegmentation procedure. They calculated both the three correlates proposed by Ramus *et al.* (1999) and those proposed by Grabe & Low (2002) and concluded that the results were not thoroughly satisfactory. So, they proposed new correlates based on the automatic segmentation of the speech signal into "pseudo-syllables"<sup>38</sup>. For each "pseudo-syllable", the authors calculated: a) the total duration of the consonantal segments, b) the total duration of the vocalic segment and c) the number of consonantal segments. Their correlates are shown in charts where stress-timed and syllable-timed languages seem to form two fairly separated groups (though Italian tends to stick in the neighbourhood of stress-timed languages). Nevertheless, no other author seems to have taken their hint and their parameters have not been tested by anyone else.

Dankovicova and Dellwo (2007) added Czech to the data already investigated by Dellwo (2006) expecting that it should fall half-way between the two rhythm classes because (a) it does not have phonological vowel reduction but does have vowel length opposition (b) its syllable structure is fairly complex, but not as complex as in stress-timed languages. The authors, however, found that only %V managed to reflect this, while both Varco $\Delta C$  and consonantal rPVI classified Czech as stress-timed, while vocalic nPVI classified it as syllable-timed.

Benton *et al.* (2007) applied the deltas and the PVI's to large corpora of non-laboratory speech of American English and Mandarin Chinese. They used news broadcasts with more than 50 speakers per language totalling more than 100 minutes of speech that was analysed with the help of automatic natural language processing and other scripts. Mean results for both the deltas and the PVI's confirm the classification of Mandarin Chinese as syllable-timed and of American English as stress-timed, though values for single speakers are reported to vary greatly.

White & Mattys (2007) tested the various correlates on first and second language speakers of English, Dutch, Spanish and French. In particular, they tested native English speakers, native Dutch speakers, native French speakers, native

---

<sup>37</sup> I find these results slightly controversial: the values of vocalic nPVI are meant to be higher in stress-timed languages because of the remarkable difference in these languages between, on the one hand, fully-articulated vowels and, on the other hand, reduced vowels. Obviously, calculating the nPVI separately for fully articulated and reduced vowels fails to capture this difference and one may well wonder what the rationale behind this choice is and on what grounds this particular application of the nPVI was expected to yield higher values in stress-timed than in syllable-timed languages.

<sup>38</sup> Their automatic system is said to recognise vocalic intervals (V) and non-vocalic intervals (C). The segments are consequently grouped into "pseudo-syllables" according to the scheme C...CV. This is to say that any sequence of C segment will be put together with the following V segment in order to form a "pseudo-syllable". In other words, the duration of a "pseudo-syllable" coincides with the inter-onset distance (the distance between two successive vocalic onsets).

### 3. 1999-2010 : rhythm metrics

Spanish speakers, non-native English speakers (Spanish natives), non-native English speakers (Dutch natives), non-native Dutch speakers (English natives) and non-native Spanish speakers (English natives). As expected, the metrics classified Dutch speaker of English and English speakers of Dutch in the stress-timed area (both Dutch and English are stress-timed languages), while Spanish speakers of English and English speakers of Spanish resulted in an intermediate area, a fact that presumably reflects a certain degree of adaptation on the part of L2 speakers to the rhythm of the target language. According to the two authors, VarcoV/%V seems to be the combination that best reflects their hypotheses.

Tortel & Hirst (2008) used the PVI in order to discriminate the rhythmical properties of French learners of English. They carried out an experiment in which three groups of French learners (FR1 non specialist English speakers, FR2 first-year university students of English, FR3 fourth and fifth-year university students of English) and one group of British English native speakers had to listen to some sentences pronounced by a “model” and then repeat it as closely as possible. The vocalic nPVI is found to better reflect the language competence of speakers than the consonantal rPVI. In effect the nPVI values increase from FR1 through FR2 and FR3 and up to BR, whereas the rPVI values are less consistent. Tortel & Hirst (2010) add the deltas and the varcos (which they questionably call cv – standing for coefficient of variation), finding that these measures allow three types of discrimination: (a) learners vs. natives (b) FR1 vs. FR2 (c) a graduation from FR1 to BR (which they call GB in their study in their study in 2010).

Loukina *et al.* (2009) calculated a plethora of rhythm measures (15 different indices, each computed in 3 different ways<sup>39</sup>, which gives a total of 45) on voiced and devoiced intervals retrieved by a procedure of automatic segmentation. The authors then built classifiers<sup>40</sup> and tested how often they could correctly predict the language, based on a combination of one or more rhythm metrics. Results show that some rhythm measures calculated on automatic segmentation are better than others at separating languages, but on the whole “there exists substantial variation within languages which makes it impossible to reliably separate languages based on the rhythm of a single paragraph”. The authors argue that such results reflect human identification of delexicalised speech. They also claim that rhythm seems to be a two or three-dimensional phenomenon as classifiers based on more than 2 rhythm measures do not significantly improve the success rate.

More authors who used the metrics are listed in chapter 5 as their attention was concentrated in showing rhythm variation and variability across different dialectal or regional varieties, or across different speech rates.

#### 3.2.5 Bertinetto & Bertini (2008): the CCI.

A new proposal, the Compensation and Control Index (CCI) was put forward by Bertinetto & Bertini (2008). Its formula is inspired by the rPVI and, like the other

---

<sup>39</sup> (a) calculating them for each sentence and then averaging partial results (cf. the B method in Mairano & Romano, 2009);

(b) calculating them for each sentence leaving out segments preceding pauses (as they are often lengthened) and then averaging partial results;

(c) calculating them on the entire text (cf. the A method in Mairano & Romano, 2009).

<sup>40</sup> A classifier is defined by the authors as “an algorithm that will optimally predict which language was most likely to have produced the observed RMs [rhythm measures]” (Loukina *et al.* 2009:1932).

### 3. 1999-2010 : *rhythm metrics*

metrics, is applied to vocalic and consonantal measures (though with important differences), but it has a different rationale and draws on previous works by Fowler and colleagues (1977 and following) on compensation (see also Farnetani & Kori, 1986 and following). Details about these works are given in chapter 1, I shall therefore only briefly resume the conclusions: data on compensation did not confirm the hypothesis that stress-timed languages should exhibit more inter-syllabic compensation (to equalise feet durations) and syllable-timed languages more intra-syllabic compensation (to equalise syllable durations); instead, stress-timed languages were found to exhibit both intra- and inter-syllabic compensation. In the light of these observations, Bertinetto (1989) revisited the traditional dichotomy of stress-timing vs. syllable-timing in terms of control vs. compensation (although the terms were drawn from Hoqvist, 1983): controlling languages (corresponding to syllable-timed languages) are supposed to show low levels of compensation at all levels, whereas compensating languages are supposed to show higher levels of compensation at all levels (intra and inter-syllabic). As stated by Bertinetto & Bertini, “[t]he CC view aims at describing the intra-syllabic behaviour, which in turn affects (or is possibly affected by) the overarching accentual alternation” (2008:1/4) and will possibly integrate in the future a model for this second level (like the one proposed by O’Dell and Nieminen, 1999).

In order to account for the intra-syllabic behaviour, then, the CCI needs to consider the segments composing each vocalic and consonantal interval: its formula is in fact a modification of the rPVI that divides interval durations by the number of segments<sup>41</sup> that compose it:

$$CCI = \frac{100}{m-1} \sum_{k=1}^{m-1} \left| \frac{d_k}{n_k} - \frac{d_{k+1}}{n_{k+1}} \right|$$

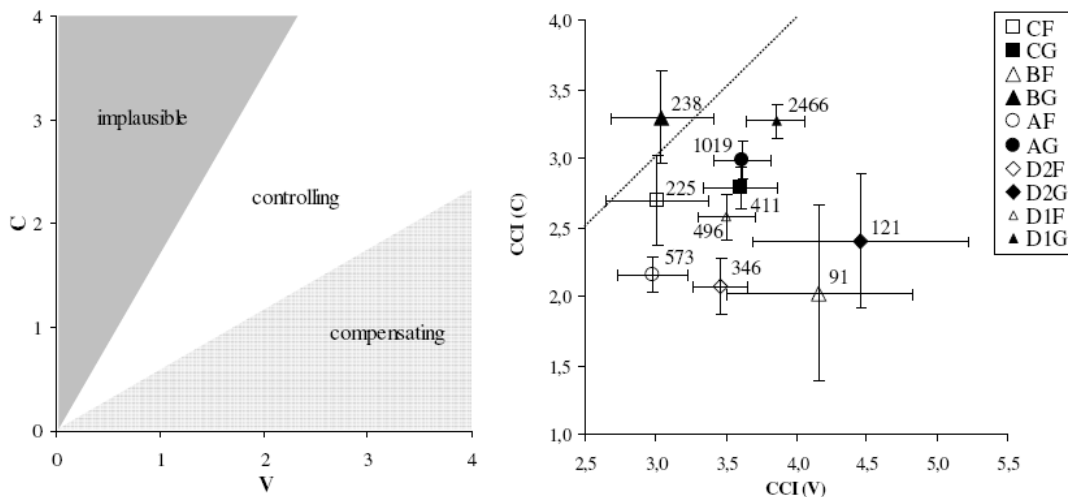
The authors’ prediction was that controlling languages should be located along the bisecting line as vocalic and consonantal fluctuations should tend to be more or less the same in these languages. Instead, compensating languages should tend to be placed below the bisecting line as vocalic fluctuations should be higher in these languages than consonantal fluctuations due to the difference between fully articulated stressed vowels and reduced unstressed vowels. The two authors expected no language to be collocated in the zone far above the bisecting line as this would imply a higher level of consonantal compensation than of vocalic compensation. They provided the results of the CCI computed on dialogues by 10 Italian speakers and compared them with the values obtained with other metrics (%V, deltas and PVIs): results for Italian seemed to reflect expectations for controlling languages, with the ten speakers clustering along the bisecting line (see figure 3.6). Then, the authors divided speakers into three tempo groups and, in line with expectations, they found that with decreasing speech rate, consonants and vowels are compressed more or less at the same level up to a threshold (which is sooner reached for consonants than for vowels).

---

<sup>41</sup> As noted by the authors, the idea of introducing the number of segments composing each interval is not entirely new, as it had already been experimented by Rouas & Farinas (2004), who used the number of consonants composing a C interval as one of many indices.



### 3. 1999-2010 : rhythm metrics



**Figure 3.6.** Results of the CCI for semi-spontaneous productions of 10 Italian speakers (taken from Bertinetto & Bertini, 2008).

Bertini & Bertinetto (2009) gave detailed information on how they computed the CCI: as has been said, each interval has to be divided by the number of phonological segments composing it, but this has the drawback of creating ambiguities as for the phonological interpretation of some segments. For example, the two authors considered both on- and off-glides as consonantal: while considering on-glides as consonantal is a common feature of most other studies on rhythm metrics<sup>42</sup>, the fact of considering off-glides as consonants is at odds with most other authors' choices<sup>43</sup>. Italian geminated consonants were considered as intervals composed by 2 phonological segments and it was prescribed that phonologically long vowels (in languages like Finnish) should also be considered as double intervals. Vocalic intervals were mostly composed of one segment, apart from cases of synaloepha, which were treated as double intervals in contrast to cases of hiatus, which were assigned to two different nuclei. Furthermore, the two authors decided on the following criteria to select utterances for inclusion in the study:

- I) sono stati scelti enunciati privi di esitazioni, pause, forme di assenso, esclamazioni, false partenze, fenomeni vocali non verbali, sequenze inintelligibili, routines discorsive (come le frasi fatte ricorrenti);
- II) sono state ulteriormente eliminate le parti terminali di enunciato di tipo asseverativo (ad esempio: ..., no?), così come quelle introduttive (ad esempio: cioè,...);
- III) sono state selezionate sequenze che, in trascrizione ortografica, avessero almeno 9 sillabe (> 9 sillabe) e, foneticamente, almeno 8 sillabe (> 8 sillabe). La differenza è motivata dal fatto che non tutte le sillabe potenziali si realizzano effettivamente nel parlato, a causa

<sup>42</sup> Actually, Grabe & Low (2002) considered even on-glides as vocalic when it was not possible to distinguish them on the spectrogram: “[w]e excluded initial glides from vocalic portions if their presence was indicated by clearly observable changes in formant structure or in the amplitude of the signal. Otherwise, glides were included in the vocalic portion.” (Grabe & Low 2002:5)

<sup>43</sup> The estimate applies to those authors who have declared the criteria followed in the segmentation.

### 3. 1999-2010 : rhythm metrics

*di possibili fenomeni di fusione tra vocali adiacenti, riduzione di iato, ipoarticolazione etc.*  
(Bertini & Bertinetto, 2009:4/17)

The two authors also state that they “discarded the final portion of each utterance, from the last stressed syllable (inclusive) onward. This portion has an entirely different rhythmic behavior, that should best be analyzed on its own” (2008:2/4).

The CCI is younger than the other rhythm metrics and, for this reason, it has not yet been applied to many data; so far, apart of course from its authors, it has only been used by Mairano & Romano (2008 and following) as far as I am aware. Results of the CCI for various languages are also reported below.

## 3.3 Applying the metrics to the corpus

### 3.3.1 Reasons for setting out to compute rhythm measures

The brief description sketched above has shown that many authors have tested the metrics in a fair amount of conditions, checking their (in)stability over different speech styles and speech rates. In most cases, while the hypothesis of two (or three if we include mora-timing) clear-cut rhythm classes does not seem to be plausible, rhythm measures do seem to yield a more or less approximate scalar characterisation of the rhythm typology of languages. However, with a few exceptions, most studies on the field of rhythm metrics have included limited data sets (which is understandable as the computing of these measures is a long a time-consuming procedure). Furthermore, data from different studies are often not directly comparable because of different choices in the segmentation or in the interpretation of some specific phonological segments. For example, as has been stressed by Mairano & Romano (2007a and b) and Loukina *et al.* (2009), there does not seem to be a general agreement as to the procedure with which rhythm metrics should be calculated: some authors compute them on all vocalic and consonantal (or voiced and unvoiced) interval durations, while others exclude the last segments, and others compute the measures separately for each sentence or inter-pausal unit finally averaging results.

As a consequence, what was claimed nearly 10 years ago (e.g. Schmid, 2001, and Ramus, 2002) is still relevant: these studies have proven the necessity of enlarging the perspectives and conducting experiments on a wider range of data including more languages, more speakers, more sentences per speaker, different registers and speech styles. In particular, the number of languages for which correlates have been calculated is still fairly small, the wider spectrum probably still being the one presented by Grabe & Low (2002)<sup>44</sup>.

So, around 2007 I set out to gradually gather data, which has already been described in chapter 2. At first, the corpus was very limited, but the number of languages and speakers grew gradually and currently totals 21 national languages and 61 speakers. For these data, the most commonly used rhythm metrics were calculated with *Correlatore* 2.2 (see chapter 4) in two different ways, namely globally (the A method) and locally (the B method). Loukina & al. (2009) have reported that such different methods do not yield significant differences, yet at the

---

<sup>44</sup> Although it was corroborated by only one speaker per language.

beginning of my research in this field (cf. Mairano & Romano, 2007a and b), I found that the B method provided a slightly better classification of rhythm classes. The results are presented below.

### 3.3.2 The data

I shall now introduce the data, which are the same throughout all tests on rhythm presented in this chapter. I have been gathering recordings of the narrative *The North Wind and the Sun* by different speakers in different languages for some years as a constantly growing corpus. At the time of writing, the corpus was composed of 61 speakers of 21 national languages (in alphabetical order: Arabic, Chinese, Czech, Danish, Dutch, English, Estonian, Finnish, French, German, Greek, Icelandic, Italian, Japanese, Polish, Portuguese, Romanian, Russian, Spanish, Swedish and Turkish).

The corpus is thus homogeneous as for text type (a narrative), speech style (read speech), text length (translations of the same text, ranging from 23.30s to 49.78s and averaging at 32.50s  $\pm$ 5.29s) and number of segments uttered by speakers; speech rate, unfortunately, varies a little according to speakers, ranging from 4.13 syll/s to 6.84 syll/s and averaging at 5.84 syll/s ( $\pm$ 0.72). All recordings involved native speakers of the languages analysed<sup>45</sup>.

The main drawbacks of the corpus at its present state are (a) its strong bias towards the Indo-European family and (b) the under-representation of some languages in comparison to others, with only 1 speaker for Czech, Danish, Estonian, Greek, Japanese, Polish, Swedish and Turkish as opposed to the 15 speakers of Italian. Furthermore, the samples have different provenances as some were recorded in a sound-proof booth at LFSAG (*Laboratorio di Fonetica Sperimentale 'Arturo Genre'*, University of Turin), whereas a few others were recorded during fieldwork and others were taken from the *Illustrations of the International Phonetic Association* either published in the *Handbook of the IPA* (1999) or in various issues of the *Journal of the IPA*. Still, the corpus is fairly wide and certainly includes a higher number of languages than most other comparative studies on rhythm. A list of all samples (and their source) in alphabetical order by national language follows.

- Arabic - 2 speakers: 1 male speaker of Standard Arabic taken from Thelwall & Akram Sa'adeddin (1999) and 1 female speaker of Lebanese Arabic recorded at LFSAG (CELI, 2009).
- Chinese - 2 speakers: 1 male speaker of Mandarin Chinese from the province of Chao Yang and 1 female speaker of Mandarin Chinese from Hongkong. Both samples were recorded at LFSAG (S. Pittoni, 2008).
- Czech - 1 speaker: a male speaker of standard Czech recorded at LFSAG (D. Brdičko, 2007).
- Danish - 1 speaker: a female speaker of standard Danish taken from Grønnum (1999).
- Dutch - 1 speaker: a male speaker of standard Dutch, taken from Gussenhoven (1999).

---

<sup>45</sup> With the possible exception of the Indian English speaker: even though this speaker might have learnt English in the family and have been educated in English, there are clear auditive cues that suggest that he might be assimilated to a speaker of English as an L2. At any rate, the issue seems to be questionable.

### 3. 1999-2010 : *rhythm metrics*

- English - 5 speakers: a female RP speaker (taken from Roach, 2004), a female GA speaker taken from Ladefoged (1999), a male speaker of Australian English (recorded at LFSAG, *CELI* 2009), a female speaker of New Zealand English (taken from Bauer *et al.*, 2007) and a male speaker of Indian English (recorded at LFSAG).
- Estonian - 1 speaker: a female speaker of standard Estonian, taken from Asu & Teras (2009).
- Finnish - 2 speakers: 2 female speakers of standard Finnish recorded at LFSAG (L. Capovilla, 2007).
- French - 2 speakers: 1 female speaker of standard French (taken from Fougeron & Smith, 1999) and 1 female speaker of Canadian French recorded by P.L. Salza (2006), *Loquendo*, (who kindly granted me permission to use the recording for research purposes).
- German - 2 speakers: 2 female speakers of Standard German recorded at LFSAG (L. Capovilla, 2007).
- Greek - 1 speaker: 1 female speaker of standard Greek taken from Arvaniti (1999).
- Icelandic - 10 speakers: 8 speakers (7 M and 1 F) of Icelandic recorded in 2007 by me in Reykjavik during fieldwork and 2 (1 M and 1 F) more speakers recorded at LFSAG.
- Italian - 15 speakers: 4 speakers of supposedly Standard Italian: Italian01 and Italian04 are males and were recorded at LFSAG in 2009 respectively by A. Romano and L. Calabrò, Italian02 is a male speaker taken from Canepari (2004), Italian05 is a female speaker taken from Rogers & D'Arcangeli (2004); one female speaker of Rome regional Italian (Italian03, taken from Costamagna, 2000), 4 (2 M and 2 F) speakers of Piedmontese regional Italian (Italian06, Italian07, Italian08, Italian15, all from Turin and recorded at LFSAG), 1 female speaker of Sicilian regional Italian (Italian09 from Mazara del Vallo, recorded at LFSAG), 1 female speaker of Northern Apulias regional Italian (Italian10 from Bitonto recorded at LFSAG), one male speaker of Calabrian regional Italian (Italian11, from Vazzano, recorded at LFSAG), 1 female speaker of Sardinian regional Italian (Italian12, from Nuragus, recorded at LFSAG), 1 female speaker of Neapolitan regional Italian (Italian 13, from Tramonti, recorded at LFSAG) and 1 female speaker of Venetian regional Italian (Italian14, from Treviso, recorded at LFSAG). Note that all 15 recordings are samples of standard or regional varieties, no dialectal varieties of Italy were included in the corpus<sup>46</sup>.
- Japanese - 1 speaker: 1 female speaker of standard Japanese recorded at LFSAG.
- Polish - 1 speaker: 1 female speaker of Standard Polish taken from Jassem (2003).
- Portuguese - 3 speakers: 1 female speaker of Standard European Portuguese taken from Cruz-Ferreira (1999), 2 female speakers of Brazilian Portuguese (1 from Manaus recorded at LFSAG and 1 from Sao Paulo taken from Barbosa & Albano, 2004).

---

<sup>46</sup> I did record 6 samples of Piedmontese, but they have been kept separated from the rest of the corpus and are presented further on (see chapter 5).

### 3. 1999-2010 : rhythm metrics

- Romanian - 6 speakers: 1 female speaker from Bucharest, 1 female speaker from Brasov, 1 male speaker from Bucovina, 1 male speaker from Moldavia, 1 male speaker from Muntenia and 1 male speaker from Oltenia. The first 3 speakers were recorded at LFSAG, while the last three speakers were recorded, respectively, by H. Bendea, D. Jitaru and G. Stan during fieldwork.
- Russian - 2 speakers: 1 male and 1 female speaker both recorded at LFSAG (Romano, 2010).
- Spanish - 5 speakers: 1 female speaker of Castilian Spanish taken from Martín Celdrán *et al.* (2003), 1 female speaker from Granada (Spain, recorded by I. Giacoletto), 1 female speaker from Bogotá, 1 male speaker from Caracas and 1 male speaker from Lima (all three recorded by S. Amorosini).
- Swedish - 1 speaker: 1 female speaker of Standard Swedish taken from IPA (1999).
- Turkish - 1 speaker: 1 speaker of standard Turkish recorded at LFSAG.

#### 3.3.3 Labelling the data

Initially, all samples were analysed with *Praat* and the durations of each vowel and consonant were saved in a spreadsheet for analysis. This procedure was very time-consuming and so, when *Correlatore* became available (see chapter 4), all data were labelled as CV (consonantal and vocalic intervals) following the conventions outlined in chapter 4 (or, in very few cases, as SAMPA transcriptions) and were saved as *TextGrids*<sup>47</sup>. However, the transition was long as all old recordings had to be re-labelled and sometimes the results were slightly different<sup>48</sup>. This is also the reason why the values of the correlates presented here are not exactly the same as in old publications by Mairano & Romano (e.g. 2007).

As a rule, all samples were labelled manually by two trained phoneticians (PM and AR) in order to minimise fluctuations in the results due to the author's choices or bias. However, this does not apply to the 10 Icelandic speakers and to 8 of the 15 Italian speakers (who were labelled exclusively by PM) and to the Danish, Swedish, Turkish and the 2 Chinese speakers (who were labelled exclusively by AR). Unless otherwise specified, the results presented in the charts of this thesis represent the average of the results obtained by each phonetician and error bars illustrate (unless otherwise specified) the standard error of the mean<sup>49</sup> of the values obtained by PM and AR.

---

<sup>47</sup> *TextGrids* are *Praat* annotation files in text format.

<sup>48</sup> The 10 Icelandic speakers still have not been relabelled. Results presented for these samples are still the ones obtained with a spreadsheet. Yet, they have been inserted in *Correlatore*'s reports using an extra functionality and this is why some metrics are missing in the report (such as vocalic rPVI and consonantal nPVI, which are usually not used and which therefore were not calculated).

<sup>49</sup> Please note that error bars in this thesis differ sensibly from error bars published in preceding works by Mairano & Romano and Romano & Mairano. This is because in previous publications, we did not use the standard error of the mean, but the standard deviation (which is also used here in some cases but for other purposes, see chapter 5). In fact, the standard error of the mean is the standard deviation divided by the square root of the number of samples (in this case 2 - PM and AR). The standard error of an estimation is defined as the standard deviation of the sampling distribution associated with the estimation method. I am aware that the standard error (as well as the standard deviation) tends to underestimate the real population variability, still it was impossible to have the data labelled by more than 2 phoneticians for evident practical reasons. At any rate,

### 3. 1999-2010 : rhythm metrics

Of course, both phoneticians were not familiar with all the 21 languages taken in consideration. In particular PM is fluent in Italian (native speaker), English, French, German and has some knowledge of Icelandic, Spanish and Portuguese, while AR is fluent in Italian (native speaker), English, French and has some knowledge of Greek, Spanish and Portuguese. For the languages which both PM and AR are familiar with (Italian, English, French, Spanish, Portuguese), the segmentation was carried out by the two phoneticians in a completely independent way and relying on no external resource. For the languages which neither was familiar with, both authors made use of the phonetic transcriptions provided by the authors of the illustrations of the IPA (either in IPA, 1999, or in different contributions in the *Journal of the International Phonetic Association*, see the bibliography) if available. If no such transcription was available and for languages which either phonetician was familiar with, PM or AR set out to provide a transcription of the narrative for the language in question with the help of handbooks or other phonetics manuals: this transcription was then used by both phoneticians for help in the segmentation and in order to fulfill the conditions set by the *CCI*, which requires the specification of the number of phonological segments for each vocalic or consonantal interval (see chapter 3 and 4). More details about the work done by the two phoneticians will be given in chapter 5.

#### 3.3.4 The segmentation

As has been said, audio data needs to be segmented into consonantal and vocalic intervals for a computation of rhythm metrics. As has already been stated in chapter 2, the segmentation of data has been carried out manually. Automatic segmentation has not been taken in consideration because its results are sometimes questionable<sup>50</sup> and I preferred to have control over decisions on segments' classification. The drawbacks noted by Galves *et al.* (2002) and Loukina *et al.* (2009) concerning the intrinsic subjectivity of manual segmentation have been mitigated by a double segmentation and labelling, carried out independently by two trained phoneticians, PM and AR, using *Praat*. The segmentation was then saved in text files containing interval durations called *TextGrids* and analysed with *Correlatore* (see chapter 4).

As already mentioned, work by PM and AR has been carried out completely independently for languages which both authors were familiar with (Italian, English, French, Spanish, Portuguese) and semi-independently for languages which neither or only either of the two authors was familiar with (see chapter 2 for further clarifications). The variability of rhythm metrics according to the measurements by each phonetician will be dealt with in more detail in chapter 5.

In many cases, above all for languages in which the segmentation was carried out in a completely independent way, authors' choices diverged either for segmentation criteria, or for the phonological interpretation of specific segments. I have compiled an explanatory and absolutely non-exhaustive list of ambiguous or potentially problematic segments or phenomena found in the languages taken in consideration.

---

most comparable studies rely on one only segmentator, so the fact of using 2 segmentators already places these results at a higher level of objectivity.

<sup>50</sup> This, is particularly true for heavily co-articulated passages or anomalous phonetic realisations, such as voiceless vowels and approximants.

### 3. 1999-2010 : *rhythm metrics*

#### ***Semi-consonants ([j] and [w])***

They are classified as approximants by the IPA because they have some vocalic characteristics but they are intrinsically dynamic (formants move) and are distinguished from corresponding fricatives. They are present in most languages (e.g. Italian *viaggiatore* [vjaddʒa'to:re], *più* [pju]) and it is very difficult to delimit them from the following vowel precisely because of their dynamicity. Furthermore, in some languages (including English) they tend to be at least partially devoiced after voiceless plosives (e.g. *disputing* [dɪs'pju:tɪŋ]).

In compliance with most studies in which segmentation choices are explicitly declared, on-glides have been considered as consonantal segments by both PM and AR. In some cases, high unstressed vowels in particular contexts are realised as approximants by some speakers (e.g. in Italian *innanzi avvolto* pronounced as [in:antsja'vɔlto] by one of the speakers from Turin): in such cases, the two phoneticians adopted different solutions, following their impressions. Clearly, differences are also found in many specific cases as to where each phonetician set their boundaries.

It has to be noted that Finnish has raising diphthongs (which are perhaps best described as vowel clusters like *ie* in *pieni* 'piccolo', *ia* in *pian* 'presto'), which were considered totally as vocalic intervals since, from a phonetic point of view, they are not approximants<sup>51</sup>.

#### ***Off-glides***

The final part of falling diphthongs was considered as vocalic in accordance with the bibliography (see Ramus *et al.*, 1999, among others) and with real/observed phonetic realisations.

#### ***Postvocalic r in many languages***

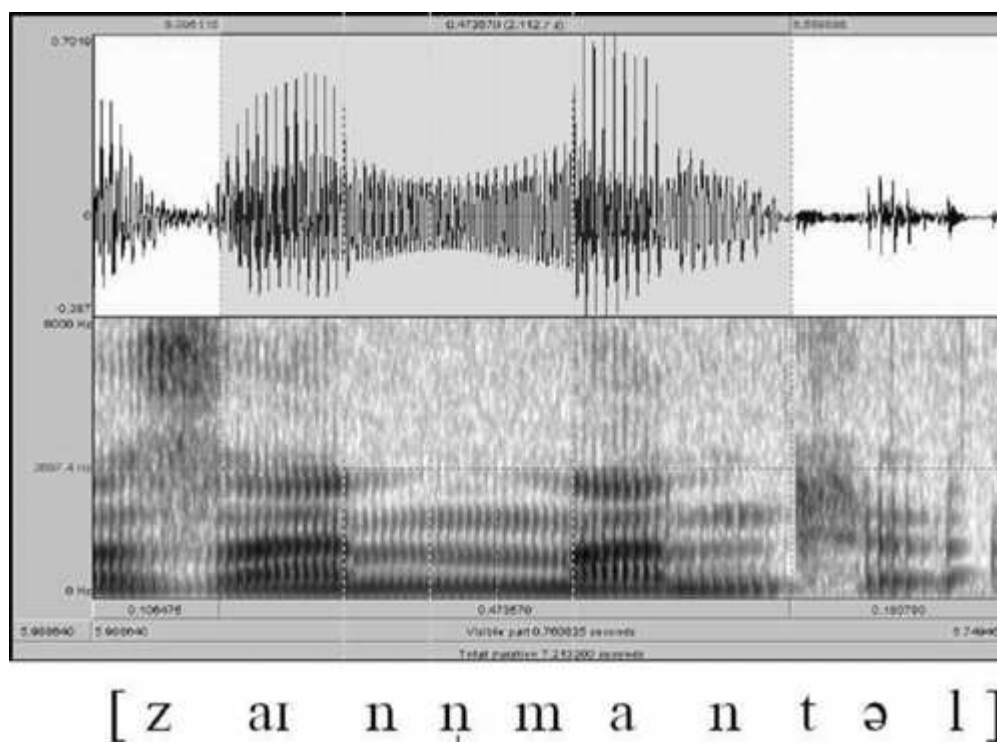
Postvocalic r was indeed fairly difficult to categorise as C or V in many languages. PM and AR's choices diverged, for instance, in American English for words such as *stronger*, which could be interpreted as either ['stɹɑ:ŋgə] or as ['stɹɑ:ŋgə.] or even as ['stɹɑ:ŋgɪ]. The cases of American English are perhaps the most emblematic, but similar problems are found in German (e.g. *Nordwind*, with pronunciations oscillating between ['nɔɐtvɪnt] and ['nɔɐtvɪnt]), Swedish and Danish.

#### ***Syllabic consonants***

Syllabic consonants were found in many languages. Phonologically, they are present in English but do not occur in the English version of *The North Wind and the Sun*. The German version had several occurrences of syllabic n and l (e.g. *stritten sich* ['ʃtɪtɪŋ zɪç] and *Mantel* [mantl]), sometimes making segmentation choices extremely challenging (e.g. *einen Mantel*: [zainən 'mantl] vs. [zainŋ 'mantl] vs. [zain: 'mantl]), see figure 3.7).

---

<sup>51</sup> Yet, they were classified as consonants in those cases in which phonetic cues suggested that they were consonantal. Choices of PM and AR sometimes diverged in this respect.



**Figure 3.7.** Spectrogram for German *seinen Mantel*.

Syllabic consonants were also found in other languages, in some cases even at a purely phonetic level, as realisations of an unstressed vowel followed by a nasal or a lateral. In all cases, syllabic consonants were considered as vocalic segments (again, in agreement with the bibliography) on the grounds that they occupy a syllabic nucleus, and therefore constitute a prominence peak within the syllable and within surrounding consonants. However, some doubts arose about this choice because syllabic consonants, in spite of occupying a nucleus, maintain the acoustic characteristics of consonants; further considerations will be discussed when analysing the results.

#### ***Sentence-initial voiceless plosives***

Sentence-initial voiceless plosives are of course only partially visible in the spectrogram: only the outburst can be observed, as the hold phase is of course represented by silence. So, it is impossible to determine the exact moment in which the plosive begins, and consequently its duration cannot be measured exactly. Both phoneticians, however, agreed to arbitrarily attribute a duration of 70 ms to all sentence-initial plosives<sup>52</sup>.

#### ***Sentence-final vowels***

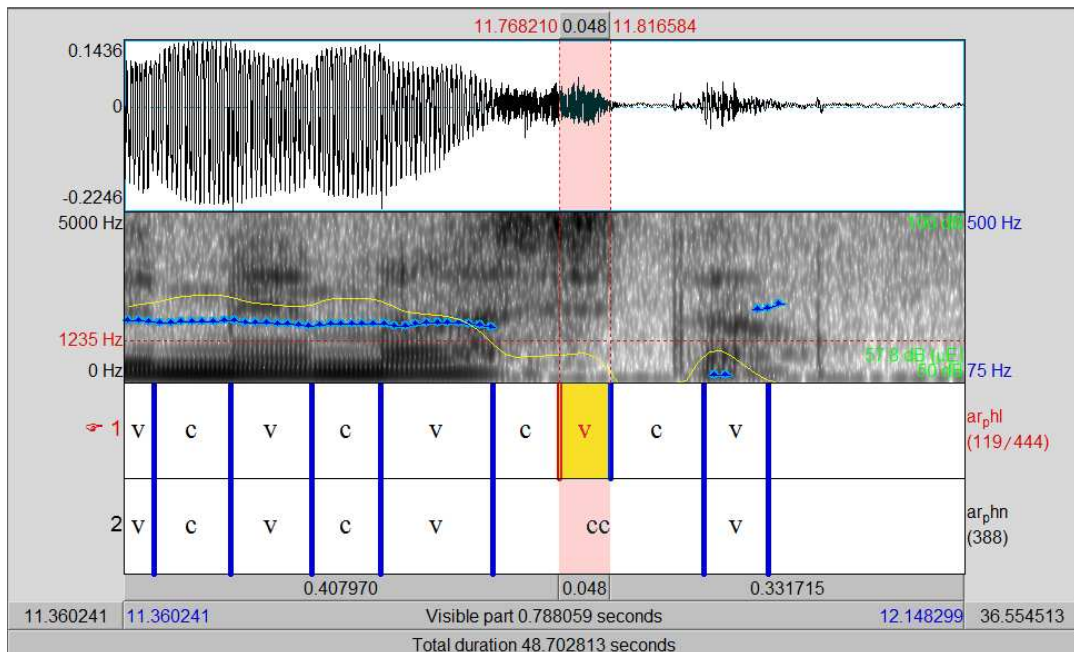
Sentence final vowels are hard to delimit as well, because they are usually lengthened and, furthermore, echo might make it impossible to take a decision. Following widespread conventions, both PM and AR considered the end of the second formant as the end of the vowel.

<sup>52</sup> In compliance with CLIPS annotation criteria.



**Devoiced or voiceless vowels**

Devoiced or voiceless vowels systematically occur in Japanese (see figure 3.8) and in Romanian (although the Romanian speakers recorded for the corpus often pronounced voiceless vowels as only partially devoiced), but also sporadically in certain speakers of other languages, including Italian. It is hard to decide whether devoiced and voiceless vowels have to be classified as vocalic or consonantal segments: acoustically, they are consonants (see figure 3.8); phonologically, they are vowels.



**Figure 3.8.** Spectrogram for a voiced Japanese vowel (in the selection): as it can be seen, the /i/ can hardly be distinguished from the preceding fricative (perceptively, the result is [ʃ]). The two transcription tiers show the possible interpretations as a consonantal or as a vocalic segment.

PM and AR sometimes adopted different solutions as to their classification. In Japanese, the consonantal aspect of this type of segments was so strong and so clear that both authors considered them as consonants. However, the results yielded by the metrics were quite in disagreement with those of other studies, so we tried to classify them as vowels and got results comparable to the ones found in the literature. More details will be discussed in the analysis.

**Glottal stops**

Phonological glottal stops were considered as consonantal segments if they were visible in the spectrogram (e.g. German *wurden sie einig* [ˈvʊədŋ zi ˈʔaɪnɪç]), otherwise they were ignored as, in such a case, accounting for them would imply measuring silence. Glottal stops in languages for which they are not described or are described as extralinguistic or paralinguistic sounds were ignored (e.g. in Italian, Spanish, French etc.).

### ***Epentheses***

Various cases of epenthesis (both vocalic and consonantal) were found in many languages. Perhaps the most common one was schwa epenthesis within complex consonantal clusters, which was usually considered as an independent vocalic interval by both PM and AR. In Italian, some schwa epentheses were also found in word-final position at the end of a sentence in words ending in a consonant (e.g. *nord* pronounced by some speakers as [ˈnɔrdə]).

### ***Other conventions***

In some studies (e.g. Bertinetto & Bertini, 2008, Grabe & Low 2002), the final parts of each sentence have been excluded as they are considered to contain spurious data that does not contribute and is not beneficial to speech rhythm. In previous trials, Mairano & Romano (unpublished poster) tried adopting this approach and got no significant differences in the results<sup>53</sup>. So, for sake of simplicity, all segments uttered by the speakers were labelled and classified as either consonantal or vocalic.

Unlike Grabe & Low (2002), PM and AR did not include consonantal spikes as part of the vowel.

### **3.3.5 The deltas and the varcos**

I shall now discuss the values of the deltas and the varcos obtained for the 61 samples in the corpus. These results were obtained by saving the segmentations as text files in *TextGrid* format and by using *Correlatore* for computing the metrics and building the charts (see chapter 4).

Numerical results and other information (number of consonantal intervals, number of vocalic intervals, number of pauses, mean duration of vocalic intervals, mean duration of consonantal intervals) for all data in the corpus can be consulted in appendix 2, which contains *Correlatore*'s entire report for the corpus. However, only the charts will be used in the discussion, as they provide a clearer visualisation (although they are of course merely a graphic representation of numerical data). Samples of different languages were averaged altogether even in those cases in which speakers came from different areas (e.g. GA English, RP English, AUS English and so on)<sup>54</sup>. I am aware that this is risky, but it responds to the practical need of not loading too many samples on the charts, which are already overcrowded. Moreover, the variation and variability of rhythm metrics are treated in detail in chapter 5, while appendix 2 contains the results for each single speaker.

Figure 3.9 shows data for the rhythm metrics proposed by Ramus *et al.* (1999), that is to say  $\Delta C/\Delta V$  and  $\Delta C/\%V$ . Expectations<sup>55</sup> seem to be confirmed for

---

<sup>53</sup> The validity of this observation may well be limited to the speech style considered, *i.e.* read speech.

<sup>54</sup> It has to be noticed that the two samples of Mandarin Chinese (one from the province of Chao Yang, the other from Hong Kong – I shall specify that the speaker of the sample from Hong Kong really spoke Mandarin, not Cantonese) have been kept separated as the results yielded by most metrics placed those two varieties fairly apart, often in different rhythm groups (also cf. the discussion in Romano, 2010:51).

<sup>55</sup> I am aware of the problems concerning expectations for rhythm metrics stressed by Bertinetto & Bertini (2010) and discussed in various parts of the thesis. Rhythm metrics do not constitute a fully predictive model and expectations are based on perceptive impressions (which most of the times are not even verified with perceptive tests) and/or on an estimate based on the phonological properties of the language(s) analysed. Using either criterion, it is not possible to quantify exactly the values of the metrics and thus the position they will occupy in the chart, so that only relative distances can be commented.

### 3. 1999-2010 : *rhythm metrics*

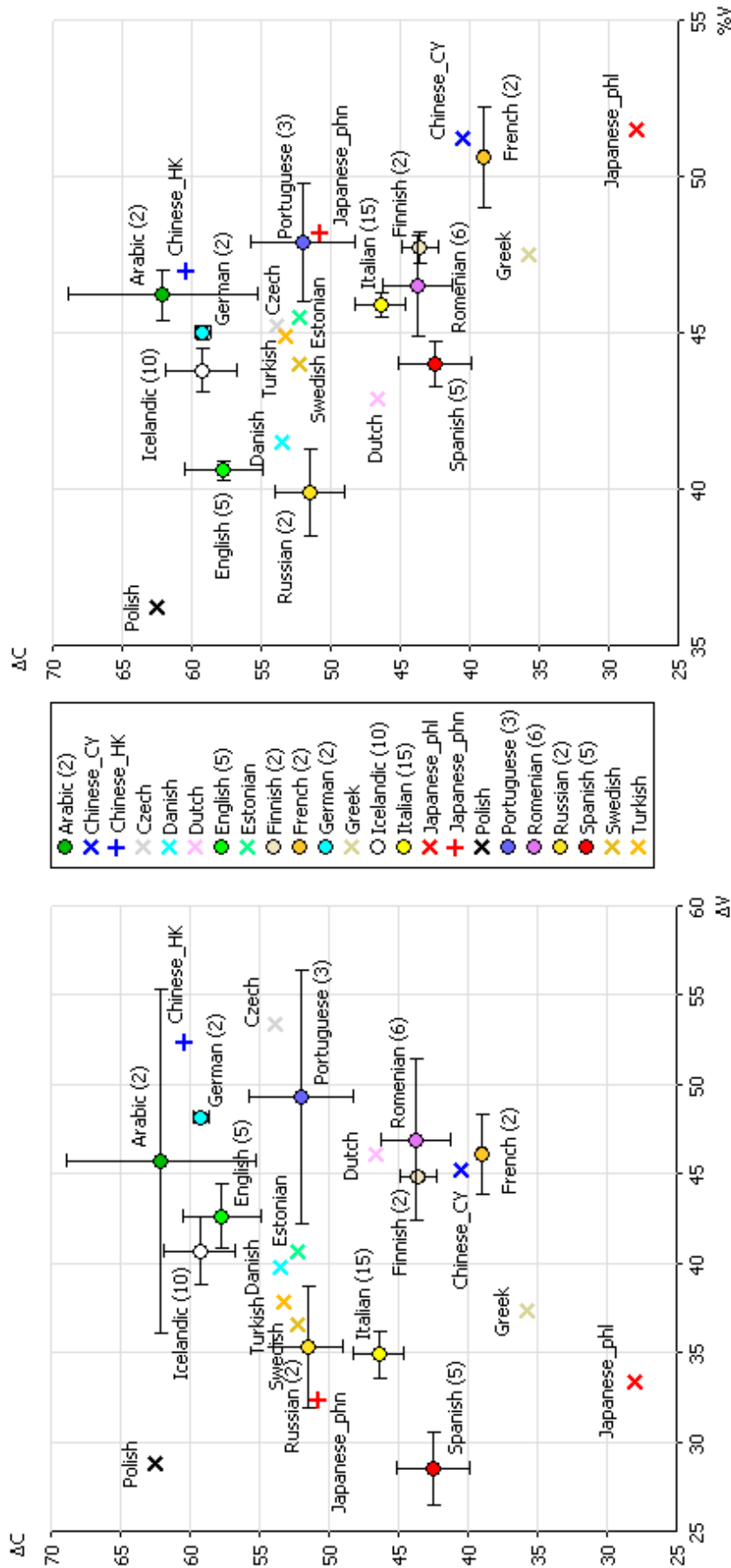
most languages traditionally considered as stress-timed or as syllable-timed; English, German and Arabic (usually considered as stress-timed) show high values of both vocalic and consonantal deltas and are therefore situated in the north-east corner of the first chart; they also show low values of %V and thus occupy the north-west corner of the second chart. Conversely, Italian, Spanish and Greek (traditionally considered as syllable-timed languages) exhibit low values of both vocalic and consonantal deltas and therefore occupy the south-west corner of the first chart; however, their values of %V are not extremely low, which accounts for their southern position in the second chart (not far from the expected south-eastern position, occupied instead by French, also usually classified as syllable-timed, Chinese\_CY and to a lesser extent Greek and Finnish).

Indeed,  $\Delta C$  values are in general more consistent with expectations than %V and  $\Delta V$  values; in effect, languages traditionally considered as syllable-timed, instead of showing low values of  $\Delta V$  and low values %V tend to exhibit either low or high values of both metrics: French, Finnish, Chinese\_CY and Romanian have fairly high %V and  $\Delta V$ , whereas Italian and Spanish, as already mentioned, have low values of  $\Delta V$  and medium values for %V. In the same way, Russian (traditionally considered as stress-timed) exhibits low values of both %V and  $\Delta V$  (the latter was expected to be greater as Russian has phonological vocalic reduction).  $\Delta V$  shows several inconsistencies as to expectations, such as English and Russian having lower values than French. The case of Portuguese is also worth mentioning: even if it has been controversially classified as syllable-timed by others (Fikkert *et al.*, 2004) or as stress-timed (e.g. Major, 1981, and more recently Mairano & Romano, 2010a), it certainly shows those phonological features typical of stress-timed languages and is thus expected to cluster with other stress-timed languages: in effect, it does exhibit high values of  $\Delta C$  and  $\Delta V$ , but it also bizarrely shows high values of %V, which do not seem to account for its evident phenomena of vocalic reduction (which can go as far as vowel deletion).

On the whole,  $\Delta C$  values seem to better separate supposedly stress-timed languages from supposedly syllable-timed languages. However, one very noticeable exception is represented by Dutch, which shows very low  $\Delta C$  values and which consequently clusters amid syllable-timed languages. Also, Turkish and Estonian values for  $\Delta C$  are fairly high and place these two languages in the stress-timed area, although perceptively they rather sound syllable-timed. Danish and Swedish have not so far been included in any other studies using rhythm metrics as far as I know, so no particular expectation or reference was available, apart from the observation that they both allow for a fairly complex syllable structure (though not as complex as the German one) and that vocalic reduction is present but not particularly evident in these languages (similar considerations can be made for Icelandic, vocalic reduction being in this case even more limited). The second chart (and the first one to a lesser extent) seems to suggest that these languages are mildly stress-timed.



3. 1999-2010 : rhythm metrics



**Figure 3.10.**  $\Delta C/\Delta V$  and  $\Delta C/\%V$  chart with calculated locally (at sentence level, cp. the B method, chapter 4) for all data in the corpus, finally averaging local results. Dots represent mean values of different speakers for the same language, error bars indicate the standard error. Numbers in parentheses indicate on how many samples the mean was calculated.

### 3. 1999-2010 : rhythm metrics

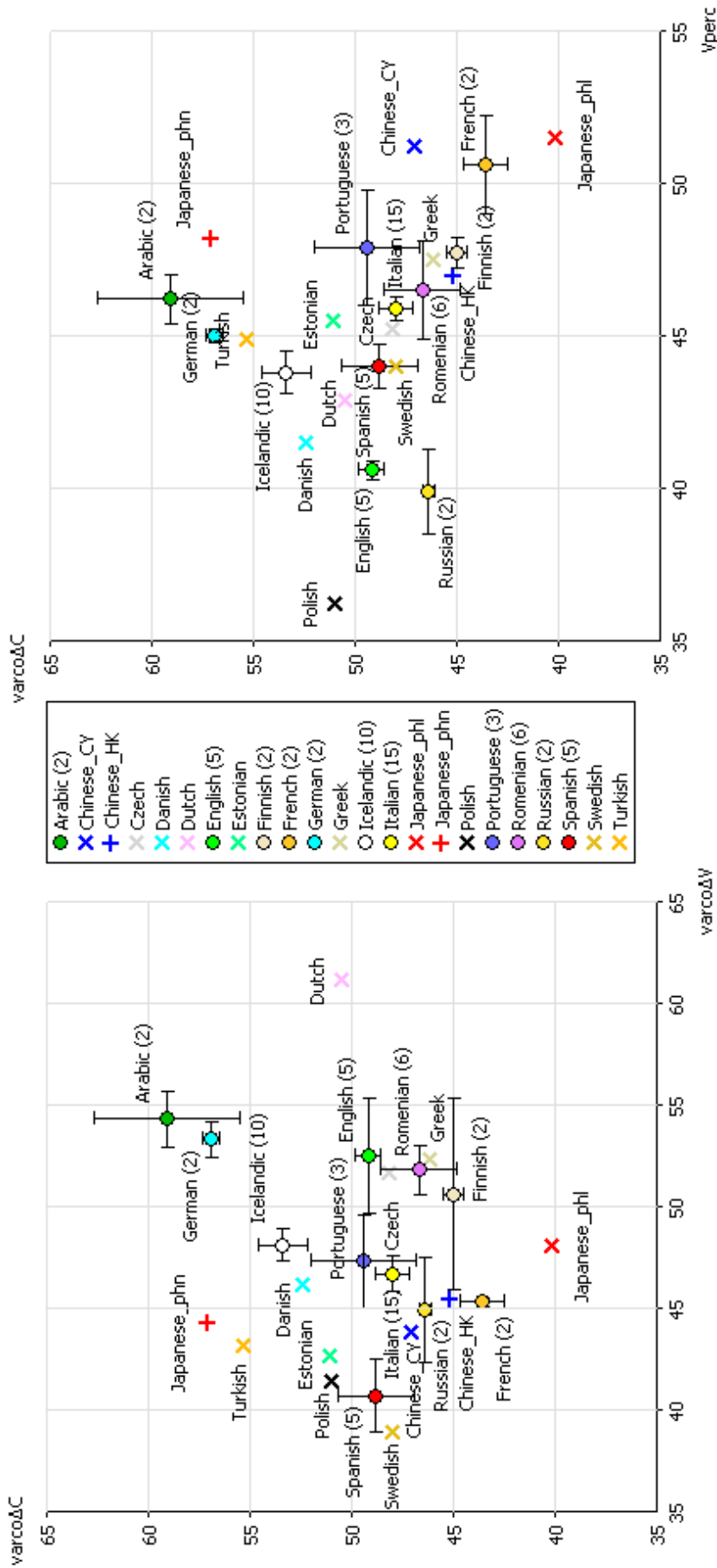
One thing that strikes out is the departing of Polish and Japanese from the other languages, though in opposite directions: Polish has the highest  $\Delta C$  and the lowest %V and  $\Delta V$  of all languages (apart from the Spanish  $\Delta V$  value), whereas Japanese\_phl has the lowest  $\Delta C$  value and the highest %V value,  $\Delta V$  values being medium. Visibly, Polish is confined to the far north-western corner of both charts, whereas Japanese\_phl is confined to the far south or south-western corner. This is in compliance with previous studies (cf. Ramus *et al.*, 1999) and with expectations, since Polish allows for complex consonant clusters (hence a high  $\Delta C$ ) but does not have macroscopic phenomena of vowel reduction (hence a low  $\Delta V$ ); Japanese, conversely, has a very simple syllable structure (hence a low  $\Delta C$  and a high %V) and does not have evident vowel reduction (but it has mono and bi-moraic vowels – hence perhaps a medium  $\Delta V$ ). However, as it can be seen in the legend, two Japanese values have been included in the chart, Japanese\_phl and Japanese\_phn; both correspond to the same language sample, but the former has been segmented phonologically (*i.e.* considering voiceless vowels as vocalic segments), while the latter has been segmented phonetically (*i.e.* considering voiceless vowels as consonantal segments). The results are immensely different, as Japanese\_phn comes to be placed in the stress-timed area: this confirms the hypothesis that segmentation choices have an influence on the results and, in particular, that the phonological interpretation of voiceless vowels has a huge and direct effect on  $\Delta C$  (the effect on %V is of course present but far less evident, while  $\Delta V$  is virtually the same for Japanese\_phn and Japanese\_phl). Finally, one should remark that these charts seem to classify Chinese\_CY as syllable-timed and Chinese\_HK as stress-timed (most previous studies of these varieties classified them as syllable-timed, see Grabe & Low, 2002, and Mok & Dellwo, 2008).

Results presented so far have been calculated globally, that is to say by applying the formulae to all consonantal and vocalic intervals in the narrative for each speaker (obviously separately for consonants and vowels). I shall now present the values of the deltas calculated locally for each inter-pausal unit and finally averaged, which are shown in figure 3.10. As it can be seen in the two charts, differences are minimal. The only remarkable discrepancy concerns the Arabic  $\Delta V$  value, which moves to a less peripheral region and which shows an even greater error bar. To a minor extent, Chinese\_HK is also affected (its  $\Delta C$  rises slightly, but it remains next to languages traditionally classified as stress-timed) and Japanese\_phl exhibits a lower vocalic variability (more in compliance with Ramus *et alia's* results). Since the two methods of computing rhythm metrics do not seem to cause significant differences, the two approaches will not be reviewed for the other metrics and the analysis will only<sup>56</sup> include results computed globally, which seems to be the most used approach in the literature (although many authors do not clarify this aspect).

---

<sup>56</sup> The results of the B method can still be consulted in appendix 2, with results for each language sample of the corpus.

3. 1999-2010 : rhythm metrics



**Figure 3.11.** VarcoΔC/VarcoΔV and VarcoΔC/%V chart for all data in the corpus. Dots represent mean values of different speakers for the same language, error bars indicate the standard error. Numbers in parentheses indicate on how many samples the mean was calculated.

### 3. 1999-2010 : *rhythm metrics*

Figure 3.11 presents the results obtained with the varcos for the same data. As has been said above, the idea behind the varco is to introduce a variation coefficient with the aim to normalise in respect to speech rate (see Dellwo & Wagner, 2003 for reference). Again, it is possible to recognise a group of languages traditionally classified as stress-timed (such as German and Arabic<sup>57</sup>) in the north-eastern corner of the first chart and in the north of the second chart; conversely, a group of languages traditionally classified as syllable-timed (such as French and Italian) is situated in the south-western corner of the first chart and the south-eastern corner of the second chart. However, some differences distinguish the results of the varcos from those of the deltas.

Firstly, it catches the eye that error bars are far slighter: the  $\Delta C/\Delta V$  chart shows that these two metrics have great fluctuations within different samples of some languages (noticeably in Arabic and Portuguese), which is notably reduced by the use of  $\text{varco}\Delta C$  and  $\text{varco}\Delta V$ . Therefore, the variation coefficient seems to effectively neutralise between different characteristics of speakers of the same language. These characteristics may well include speech rate, which varies remarkably between the two Arabic speakers, but could also include other parameters (the two Arabic speakers and the three Portuguese speakers have a different geographical provenance).

Moreover, it should be noted that the varcos place Dutch and French in a position that is more consistent with expectations, French exhibiting a low  $\text{varco}\Delta V$  (which was not the case with  $\Delta V$ ) and Dutch exhibiting an extremely high  $\text{varco}\Delta V$  and a high  $\text{varco}\Delta C$  (while it had a fairly low  $\Delta C$  and merely medium  $\Delta V$ ). On the other hand, Russian has moved to an even less comfortable position, amid syllable-timed languages with low values of  $\text{varco}\Delta C$ , which is difficult to account for (given its phonological features). Spanish and Greek lose a bit of their marked characterisation as syllable-timed, moving towards more central regions,  $\text{varco}\Delta C$  being greater than  $\Delta C$  for these languages. Polish lowers slightly, suggesting a less exceptional consonantal variability, still conserving low levels of  $\text{varco}\Delta V$  which indicates a very limited vocalic variability. Interestingly,  $\text{varco}\Delta C$  values for Chinese\_HK drop dramatically and move this language from the stress-timed group to the syllable-timed group, while Japanese\_phl conserves its confinement to the south (first chart) and south-east (second chart) corners.

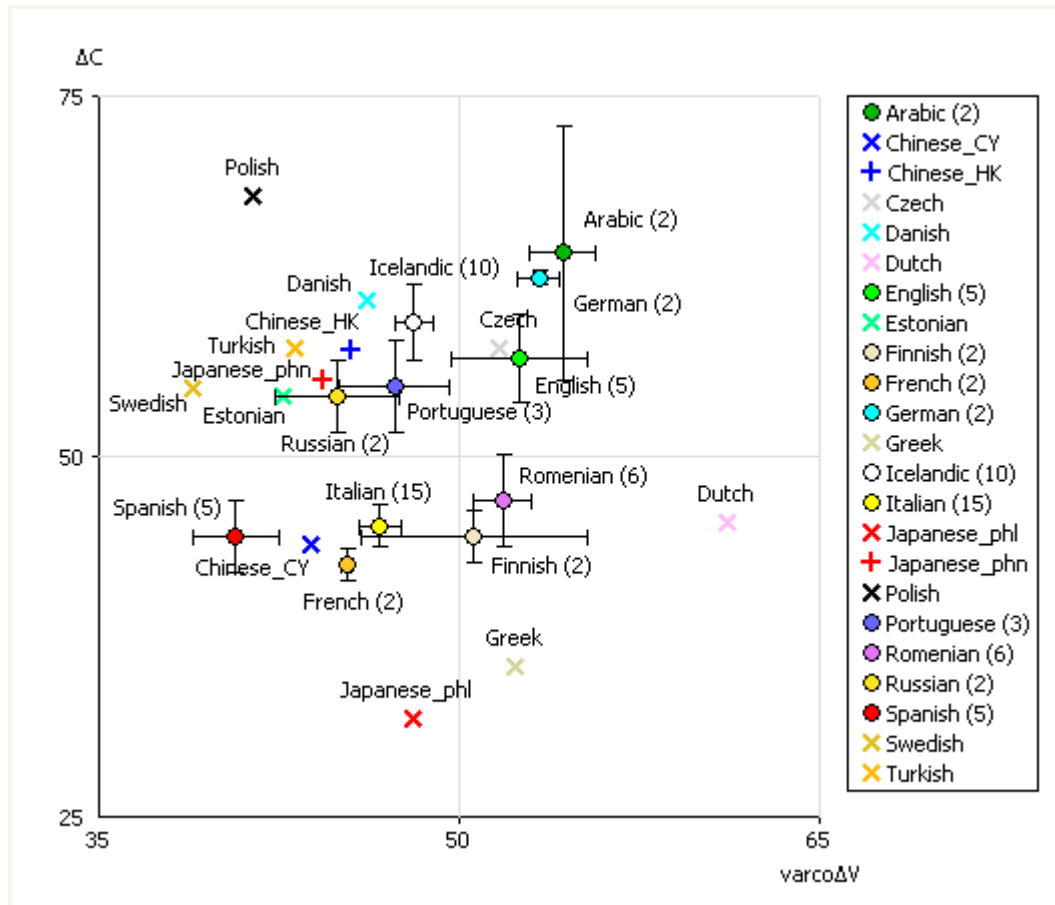
So, applying a variation coefficient seems to be a better solution for  $\Delta V$  than for  $\Delta C$ , as is shown by the chart in figure 3.12, with  $\Delta C$  on the y axis and  $\text{varco}\Delta V$  on the x axis. This solution seems to yield a better scenario for a rhythmic categorisation of languages than with only the deltas or only the varcos. Moreover, it reflects the usual visualisation of the PVI, which, following Grabe & Low (2002), applies a normalisation to vocalic measures (using the formula of the nPVI), but not to consonantal measures (for which the rPVI formula is used).

---

<sup>57</sup> For the classification of Arabic as stress-timed, see for instance Miller (1984) and Ghazali et al. (2002, although  $\Delta C$  shows sometimes aberrant values).



### 3. 1999-2010 : rhythm metrics



**Figure 3.12.**  $\Delta C$ /Varco $\Delta V$  chart for all data in the corpus. Dots represent mean values of different speakers for the same language, error bars indicate the standard error. Numbers in parentheses indicate on how many samples the mean was calculated.

The chart in figure 3.12 offers an interesting scenario:  $\Delta C$  clearly separates two groups of languages. Languages traditionally classified as syllable-timed (Italian, French and Spanish) are situated below 50ms, while languages traditionally classified as stress-timed (English, German, Arabic, Russian and Portuguese) as situated above this line. However, the chart also presents a vertical partition at roughly 50ms on the x axis (values for varco $\Delta V$ ), which again separates French, Spanish and Italian from English, German and Arabic. Interestingly, the two lines create a quadripartition with four zones in the chart, none of which is empty. Actually, each of the four zones seems to have a cluster of languages: French, Italian, Spanish and Chinese\_CY occupy the south-western slot, traditionally allocated to syllable-timed languages; English, German, Arabic and Czech occupy the north-eastern slot, traditionally allocated to stress-timed languages; Finnish and Romanian occupy the south-eastern slot, indicating a simple syllabic structure combined with high durational variability at the vocalic level (explained in Finnish with the phonological opposition of vowel length, in Romanian with devoiced vowels); Danish, Swedish, Icelandic, Estonian, Turkish, Chinese\_HK, and rather surprisingly, Russian and Portuguese occupy the north-western slot, indicating limited vocalic variability (only explainable for Danish, Swedish, Icelandic and to a lesser extent for Estonian, Turkish and Mandarin) combined with a complex syllabic structure (fairly unexpected for Estonian and Turkish).

### 3. 1999-2010 : *rhythm metrics*

Furthermore, it is interesting to remark that languages tend to cluster in a focal area extending from north-east to south-west (see the red lines). Only four languages are scattered away from the focal area: Polish to the far north-west (as expected), Japanese\_phl to the far south-south-west (again, as expected), Greek to the far south-south-west, Dutch to the far south-east-east. However, these four languages are all represented by only one speaker and it may well be possible that, averaging results from more speakers, their isolation would be less or not at all remarkable.

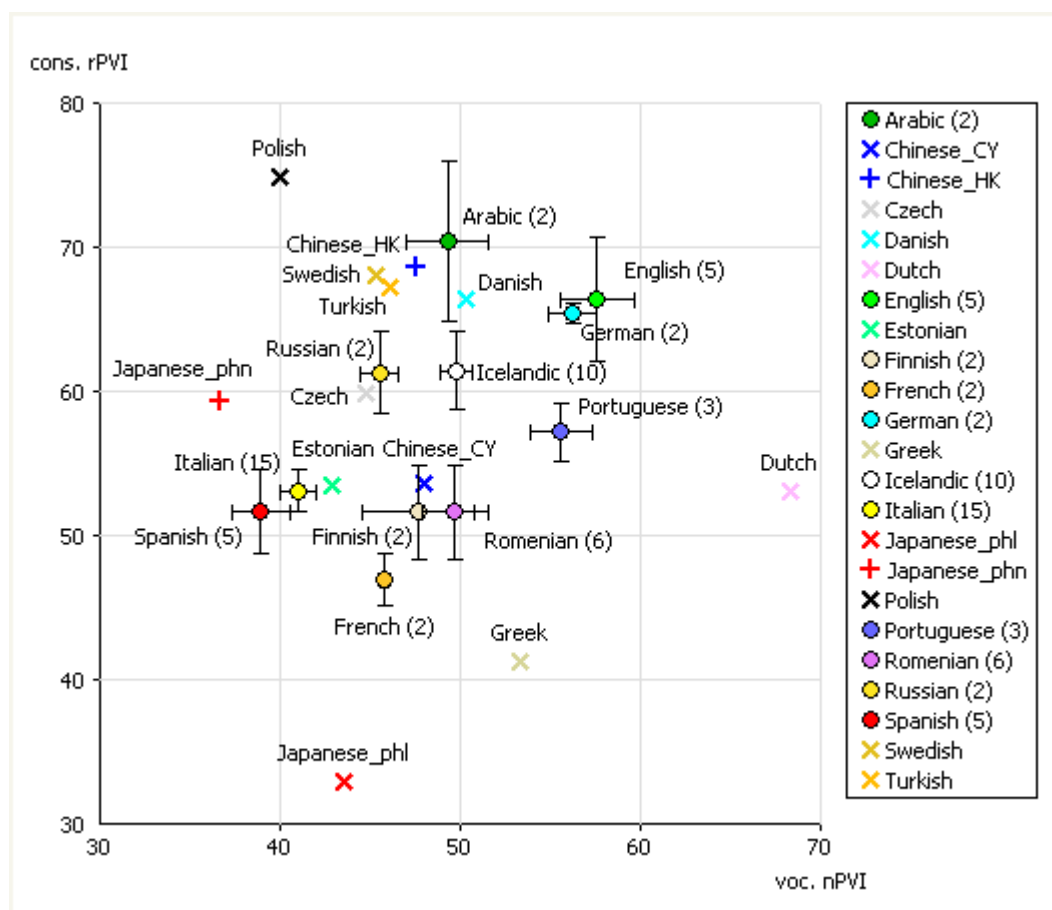
#### 3.3.6 The PVI

As has been previously mentioned, the PVI differs from the deltas in that it takes in consideration the sequential order of the segments by calculating the difference between successive vocalic or consonantal intervals (see above). Figure 3.13 shows the results of the consonantal rPVI and the vocalic nPVI calculated on the corpus.

Probably, this chart provides the best representation for a rhythmic categorisation of languages according to expectations coming from perceptual evidence and phonological properties. English, German, Arabic, Danish, Swedish and to a lesser extent Russian, Icelandic, Czech and Portuguese (supposedly stress-timed languages) cluster in the north-eastern area, exhibiting high values of both rPVI and nPVI, indicating great durational variability both at the vocalic and at the consonantal level. On the contrary, Italian, Spanish, French, Romanian, Greek, Finnish and Estonian (supposedly syllable-timed languages) cluster in the south-west area, with low values of both rPVI and nPVI, which in turn indicate a limited consonantal and vocalic durational variability. Again, the two Chinese samples occupy different portions of the chart: Chinese\_HK amid stress-timed languages, Chinese\_CY not far from syllable-timed languages.

Similarly to what happened with the deltas and the varcos, Japanese\_phl (segmented phonologically) and Polish occupy isolated positions: both exhibit a limited vocalic variability (neither of them has evident phenomena of vowel reduction) though the values for Japanese\_phl (which does have mono and bi-moraic vowels) are slightly higher. Polish has the highest value for the consonantal rPVI, while Japanese\_phl has by far the lowest: as already said, this is in compliance with expectations based on the syllabic structure of these two languages and it reflects results obtained by other authors. Dutch, as well, occupies an isolated position with very high values of vocalic nPVI and medium values of consonantal rPVI.

Results for consonantal rPVI and vocalic nPVI roughly reflect those for  $\Delta C$  and  $\text{varco}\Delta V$ . Still, the PVIs provide a scalar characterisation of languages that is more in line with expectations coming from perceptual evidence and structural properties (e.g. for Greek as syllable-timed and for Portuguese as stress-timed). A quadripartition is in this case far less evident and, as expected, only Polish, Japanese and to a lesser extent Dutch are isolated from the central area. Remarkably, the diamond-shaped focal area confirms the scenario found by Grabe & Low, 2002, for a similar though smaller sample of languages. This may suggest that languages showing a very high variability on one axis and a very low variability on the other axis, should be considered as marked.



**Figure 3.13.** Consonantal rPVI / vocalic nPVI computed for all data in the corpus. Dots represent mean values of different speakers for the same language, error bars indicate the standard error. Numbers in parentheses indicate on how many samples the mean was calculated.

### 3.3.7 The CCI

As has been said above, the Control and Compensation Index (CCI, devised by Bertinetto & Bertini, 2008) is a modification of the rPVI formula by which the duration of each vocalic and consonantal interval is divided by the number of phonological segments that compose it and aims at measuring the level of segmental compensation allowed by different languages. This of course implies that the segmentation cannot simply be carried out as C-V, but requires a more sophisticated analysis. Determining the number of phonological segments of vocalic and consonantal intervals might seem a trivial task, but it has brought about several interpretation dilemmas. I shall review the most relevant ones below.

#### *Gemination*

As suggested by Bertinetto & Bertini (2009), Italian, Icelandic, Finnish and Estonian geminate consonants were considered as one interval composed of two segments: this included Italian intrinsically geminate consonants  $\int$   $\lambda$   $\eta$   $ts$   $dz$  in inter-vocalic position (e.g. Italian *riuscito*). However, for speakers of Northern regional varieties of Italian, intervocalic  $\int$  was considered as having only one segment as its short

### 3. 1999-2010 : rhythm metrics

realisation does not depend on compensation, but on the fact that this consonant is not intrinsically geminated in these varieties. Moreover, Icelandic preaspirated consonants were also considered as intervals composed of two segments as they are the phonetic realisation of phonologically geminated voiceless plosives (see Thrainsson, 1978, and Helgason, 2002, for reference).

#### ***Diphthongs***

In line with most other studies on linguistic rhythm, on-glides were considered as consonantal, off-glides were considered as vocalic (in contrast to Bertinetto & Bertini, 2008 and following, who considered both as consonantal), so rising diphthongs were labelled as |c|v|, while falling diphthongs were labelled as |vv| (a vocalic interval consisting of two segments). This is valid for most languages included in the corpus, but with some exceptions: for example, Finnish and Estonian diphthongs (rising and falling) were always considered as vocalic segments with two intervals |vv| on the grounds of their acoustic characteristics. German diphthongs resulting from postvocalic r were considered as |vv| if the r segment presented evident vocalic properties as shown in the spectrogram.

#### ***Insertions***

Realised but non expected segments (epentheses) were included in the segments count if they were clearly visible in the spectrogram and perceptively evident; that is to say that Italian *nord* pronounced as [ˈnɔrdə] was labelled as |c|v|cc|v| and measured accordingly.

#### ***Deletions***

Expected but non realised segments were counted in the total amount of segments composing one interval: for instance, German *einst stritten* expected as [amst ˈʃtɪtɪŋ] but realised as [am ˈʃtɪtɪŋ] or as [ains ˈʃtɪtɪŋ] was labelled as |vv|cccc|v|c|v|; similarly French *reconnaitre que* theoretically expected as [ʀəkɔnɛʀ(ə) kə] was labelled as |c|v|c|v|ccc|v| even when it was realised as [ʀəkɔnɛt kə]. The same applies to nasals realised only through nasalisation of the preceding vowel, for example, Italian *un mantello* realised as [ũ mǎnˈtɛllo] was labelled as |v|cc|v|cc|v|cc|v| (in which the second interval – cc – also includes the last third of the [ũ] in order to account for the cancelled nasal). The reason for including cancelled segments in the count is that the cancellation of a segment corresponds to the highest possible level of compensation and, therefore, they have to be accounted for in a measure that aims at describing compensation. Some particular cases of deletion were excluded from this reasoning as they were not caused by compensation, for example, h-droppings in English weak forms (e.g. *his* pronounced as [ɪs] or [ɪz] in unstressed position) and others (such as *and* pronounced as [ən]).

#### ***Hiati and synaloephae***

According to the indications given by Bertinetto & Bertini (2009), synaloephae were considered as one vocalic interval composed of one segment |vv|, while hiati were considered as two separate vocalic intervals |v|v|. Cases such as Italian *tolse il mantello*, <se il> was considered as |c|v|v|c| if it was realised as a hiatus [se.il]

### 3. 1999-2010 : *rhythm metrics*

(rarely), as |c|vv|c| if it was realised as a diphthong [seil] (still rather rarely), as |c|v|c| if the second part was deleted [sel] (most often). In this case, the cancelled segment was not accounted for in the segmentation because its deletion is not due to a compensation phenomenon. The same considerations applied in cases in which the first and the second vowel had the same quality, e.g. in Italian *togliersi il*, in which realisations as [si.il] were labelled as |c|v|v|c|, realisations as [si:l] were labelled as |c|vv|c|, realisations as [sil] were labelled as |c|v|c|.

#### **Other**

Sentence-final parts were not excluded from the segmentation both for practical reasons and because previous trials showed little difference in the results.

The values of vocalic and consonantal CCI obtained for all samples in the corpus are reported in figure 3.15. As it can be seen, some compensating languages such as German, Czech and Portuguese are collocated below the bisecting line, while some controlling languages are collocated along the bisecting line, such as French, Chinese\_CY, Finnish, Spanish and Japanese. On the one hand, the disposition of some languages given by the CCI is more convincing than for others, such as Turkish (which is classified as a controlling language by the CCI but as a stress-timed language by the deltas and PVIs against perceptive impressions).

On the other hand, some supposedly compensating languages (namely Arabic and Polish) occupy the position above the bisecting line: this was thought to be impossible (or at least improbable) by Bertinetto & Bertini (2008) as such a result suggests that these languages compensate more for consonants than for vowels. However, it is possible that this prediction failed to take in consideration languages like Polish, which do have very complex consonantal clusters (and thus presumably need a certain amount of consonantal compensation), but which do not exhibit evident phenomena of vowel reduction (and thus do not compensate much at the vocalic level).

As claimed by Bertinetto & Bertini (2010), the CCI is a “phonologically-driven” index, as it needs a careful phonological evaluation of each segment as, for instance, deleted segments have to be accounted for (but only if their deletion has to do with compensation) and numerous other decisions have to be made. This task, as stated by their authors, requires a deep knowledge of the phonetic and phonological characteristics of the language(s) analysed. However, this is of course difficult to achieve for 21 languages and, as has been stated in the presentation of the data, the two phoneticians have mostly relied on transcriptions and descriptions published as illustrations of the IPA<sup>58</sup>. Problems often arise when choices have to be made for ambiguous segments, such as on- and off-glides, syllabic consonants, etc. As can be seen from the segmentation criteria specified above, PM and AR made some different choices from Bertini & Bertinetto (2009). PM and AR generally explicitly adopted a more phonetically-oriented segmentation, such as the interpretation of off-glides as consonantal segments (despite their phonological consonantal value, as stressed by Bertini & Bertinetto, 2009). Indeed, it is well possible that changing criteria even just for these very frequent segments would yield completely different

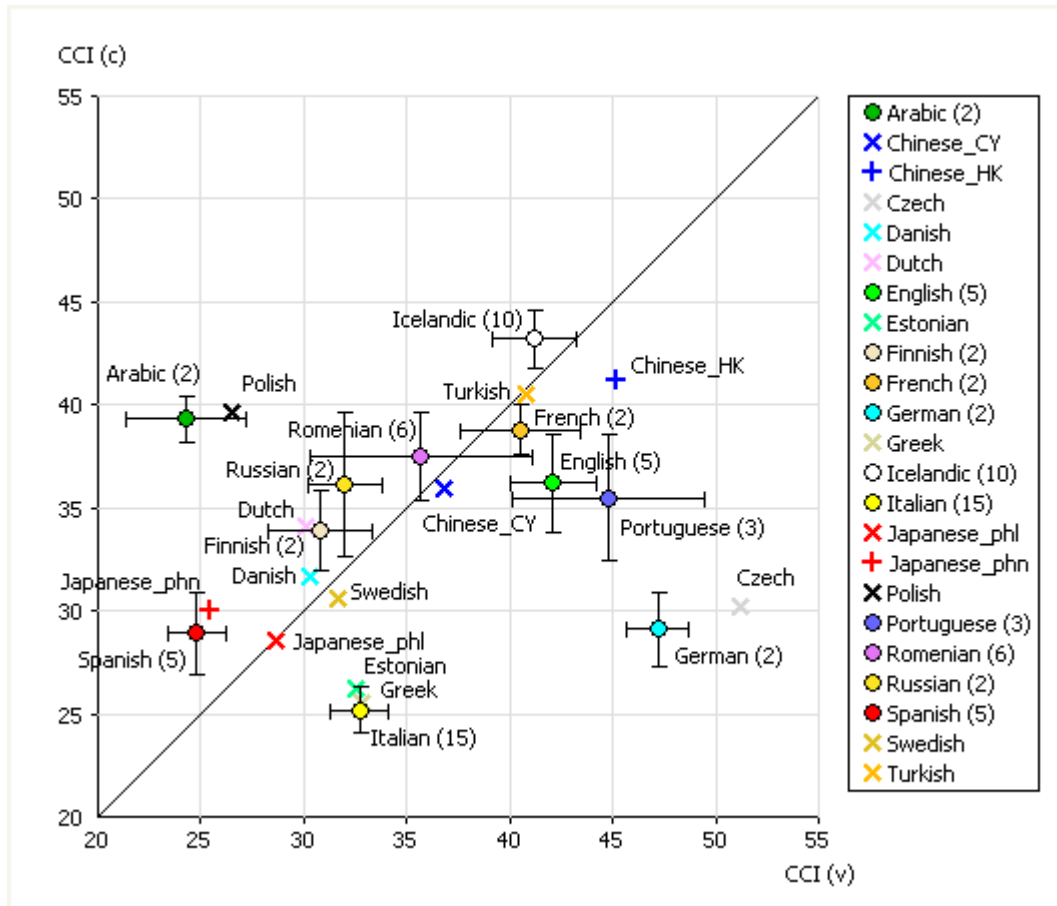
---

<sup>58</sup> PM gave up the segmentation of the Chinese samples discouraged by the difficulties he encountered.

### 3. 1999-2010 : rhythm metrics

results and would perhaps fix the incoherencies presented by the CCI chart, such as the placement of Arabic and Russian.

Still, these results clearly ask for further analysis and interpretation of the disposition of languages on the CCI chart, which can only be done with more data for each speaker. Other considerations on these new metrics follow in the next lines.



**Figure 3.14.** Consonantal CCI / Vocalic CCI computed for all data in the corpus. Dots represent mean values of different speakers for the same language, error bars indicate the standard error. Numbers in parentheses indicate on how many samples the mean was calculated.

The formula of the CCI tries to capture segments' variability which is supposed to be higher in compensating languages (precisely as an effect of compensation) than in controlling languages. In order to do this, intervals are divided by the number of segments that compose them. In practice, similar results could presumably be obtained by directly applying the PVI formula to segments' durations<sup>59</sup>. The comparison of the CCI and this approach would certainly be of

<sup>59</sup> I shall explain myself more clearly. Grabe & Low (2002) apply the PVI formula to vocalic and consonantal intervals in order to capture fluctuations in intervals' duration. Bertinetto & Bertini (2008) change the focus from intervals to segments and aim at measuring segments' variability (which is thought to be greater in compensating languages as they have fewer constraints and are freer as to fluctuations at all levels). So, if we want to capture fluctuations in segments' length, the most straight-forward way of doing it is to apply the same formula to segments' durations. Predictions would of course be unchanged, that is to say that controlling languages should tend to align along the bisecting line, while compensating languages should tend to cluster below it.

### 3. 1999-2010 : *rhythm metrics*

some interest; however, I suppose that such a choice would probably bring to biased results as the final value would be heavily influenced by the intrinsic length of different categories of sounds (e.g. voiceless fricatives being far longer than taps and so no). Indeed, the CCI does have the same problem, although it is somehow mitigated by considering each vocalic or consonantal interval and dividing it by the number of segments that compose it: this way, for intervals of more than one segment, the value obtained does not correspond to any actual duration, but is rather an abstract representation of the level of compensation within that interval. Yet, for mono-segmental intervals, the problem is not solved.

Another consideration regards the fact that the CCI does not preserve each segment's weight on the final value: in fact, each segment's weight on the final value is inversely proportional to the number of segments composing the interval<sup>60</sup>. I shall not deal with this matter in any further detail now, except for stating that if ever this is found to be a problem, a possible solution would perhaps simply imply the introduction of a coefficient for increasing the weight of each segment in plurisegmental intervals.

More importantly, compensation should become more evident in a language at increasing speech rate. Therefore, the CCI ought to be computed at different speech rates for each language, like its authors did (see Bertinetto & Bertini, 2008). This is even more true given the fact that the CCI does not normalise for speech rate (as a specific choice by its authors on the basis of the tight relationship between rhythm and speech tempo): such a choice certainly has rooted grounds, but it implies that for cross-language comparative studies the CCI must be applied to samples rigorously comparable as for speech rate. Otherwise, the results might reflect the effects of speech tempo as well as of compensation properties and other rhythmic cues. And, of course, it would be difficult to identify those characteristics that have to be attributed to speech tempo and those that that can be attributed to real cross-language differences.

At the light of these considerations and also keeping in mind that the CCI is younger than the other rhythm metrics, it can be claimed that it still needs to be tested across a combination of different languages and, crucially, different speech rates. Unfortunately, my corpus is not adequate (at least at the moment) for testing both different languages and different speech rates and, so, results present some incoherencies that might otherwise be fixed: the disposition of the samples within the CCI chart meets expectations for some languages, while others occupy an area which was predicted to be empty. For the time being, the interpretation is unclear and represents a challenging perspective for future studies in this field.

---

<sup>60</sup> As it has already been said and repeated, the CCI divides each interval by the number of segments composing it. Consequently, each interval issues one value, whether it is composed of one, two or more segments. So, a segment included in an interval with 3 segments will weight as 1/3 of segment on the final value, while a segment composing an interval in itself will weight as 1 segment. Furthermore, since it is usually stressed syllables that tend to attract segmental material and thus present consonantal clusters, and that stressed syllables are less numerous than unstressed syllables, it is well possible that reduced (compensated) segments will be underrepresented in the final computation.

### **3.4 On the possibility of including pitch and intensity**

#### **3.4.1 Why on earth?**

Even though only few studies on speech rhythm have attempted to include parameters other than duration, the importance of pitch and intensity for an account of speech rhythm has been stressed by several authors, mainly on perceptive grounds. In the 70s, some experiments were conducted by different authors, among which Lehiste and Allen (some of them are reported by Allen, 1975, and, Lehiste, 1977). They have shown that sequences of stimuli can be perceived as having an alternating<sup>61</sup> rhythmic structure by manipulating not only durations, but also intensity or pitch alone. So, a difference in pitch and/or intensity is enough for our mind to create rhythm: this seems to me a fairly good reason to try and integrate these two parameters into a speech rhythm account.

More recently, other authors have stressed the importance of pitch and intensity on the perception of rhythm and some of them have put forward some proposals. Lee & Todd (2004, but see also Lee, 2010) hypothesised that stress-timed languages displayed a higher variability in the prominence of syllable nuclei than syllable-timed languages. Since prominence is not only given by duration, they measured the delta and the rPVI on values of intensity and pitch on 2 different sets of data (one made up of English and French sentences, the other containing samples of Dutch, English, French and Italian drawn from the corpus used by Ramus *et al.*, 1999). All measures proposed in their study seemed to reflect a rhythm categorisation of languages.

Cumming (2009) carried out a perceptive experiment on French and Swiss German speakers on the spur of the ones made by Lehiste (see above). She found that  $f_0$  has an impact on the perceived duration of speech stimuli ([si] syllables in this case) for both groups of listeners:

*Since length judgments do not differ significantly between the two (prosodically different) language groups, this tends to suggest that the perceived lengthening effect of dynamic  $f_0$  is not dependent on language background [...].*

*If  $f_0$  changes affect the subjective duration of successive intervals in several languages, the rhythm of a language which tends to use  $f_0$  dynamism within the syllable may be perceived differently from that of one in which  $f_0$  changes minimally within the syllable. However, durational rhythm metrics may not accurately reflect this difference; therefore, in answer to the title, rhythm metrics should take account of  $f_0$ . Finding a suitable means of integrating duration and  $f_0$  change into metrics such as the PVI could be the next challenge.*

*(Cumming, 2009:14)*

On the basis of informal tests carried out by manipulating the prosodic parameters of the incipit of *The North Wind and the Sun* in English and French, Romano (2010) suggested that much rhythm information is lost by equalising the

---

<sup>61</sup> Allen (1975) makes a distinction between rhythms of alternations rhythm of successions. An alternating rhythm is given by an alternation of more prominent and less prominent beats. See 2.2.3 for further details.



### 3. 1999-2010 : rhythm metrics

pitch contour and, similarly, that some rhythm information is kept when equalising vocalic durations and leaving the original pitch values. The author concludes as follows:

*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 [...] 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.*

*(Romano, 2010:66)*

#### 3.4.2 An attempt

I made an attempt to introduce pitch and intensity in an account of speech rhythm. It was merely a preliminary test, whose results did not confirm expectations and which was therefore discontinued. However, since I have not yet found any other solution for the integration of prosodic parameters in the study of speech rhythm, I shall briefly present the experiment.

The test was based on the very simple idea already attempted by Lee & Todd (2004) of applying rhythm metrics to measures not only of duration, but also of pitch and intensity. This may seem weird, but there is a rationale: as noted by Bertinetto, in stress-timed languages, there is a “tendency of stress to attract segmental material in order to build up heavy syllables” (1989:108). This certainly is one of the reasons<sup>62</sup> why stressed syllables are more salient in some languages (such as English) than in others, but prominence is usually achieved through other prosodic parameters as well, such as pitch and intensity. So, it may be that the higher degree of prominence showed by stressed syllables in stress-timed languages (if compared to stressed syllables in syllable-timed languages) is contributed by a bigger difference between the pitch and intensity values of stressed and unstressed syllables in stress-timed languages than one would find in syllable-timed languages. A way to test this difference in pitch and intensity variability consists then in applying the metrics to these measures.

However, I should stress two important problems concerning the comparison of intensity and pitch values. First of all, raw  $f_0$  values (in Hz) are obviously of no use as it is nonsense to calculate variability on values distributed on a non-linear scale: I therefore need to use semitones. Secondly, it is extremely risky to compare intensity values of different recordings as they are heavily influenced by several factors (such as, most importantly, the distance of the speaker from the microphone). For this reason, it was not possible to use the same as for the experiments presented above (some of the samples were recorded at LFSAG, but many others were drawn from the *Illustrations of the IPA*, each produced by different authors). So, I used the most uniform and comparable data I had, the same that have been used in chapter 2: 10 samples by the same speaker, recorded the same day and in the same conditions,

---

<sup>62</sup> The other being that unstressed vowels are often reduced and centralised.

### 3. 1999-2010 : rhythm metrics

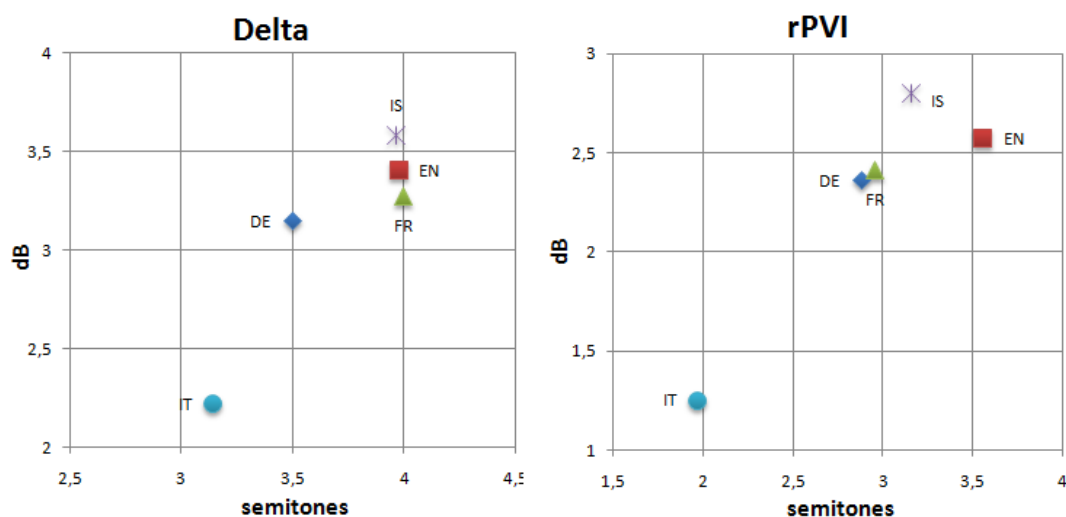
reading *The North Wind and the Sun* twice in 5 languages (English, French, German, Icelandic and his mother tongue Italian).

As for the methodology, I prepared a *Praat* script that saved in a text file the mean values of intensity and  $f_0$  (the latter were converted in semitones) for each vocalic interval (meaning that the values of the metrics would then be calculated with *Correlatore* – see chapter 4). Then, the 10 recordings were labelled manually according to the criteria already specified. I shall linger no longer on the description of the data and the segmentation in detail, as they are presented, respectively, in chapter 5 and above. Instead, I shall pass directly to the results

#### 3.4.3 The results

I computed the delta and rPVI on semitones and intensity values (obtained through the *Praat* script) with a special function of *Correlatore* and put the results on the charts shown in figure 3.15 (each dot represents the mean obtained for each pair of samples of the same language).

As it can be seen, the delta and the rPVI reflect basically the same scenario, with low variability of pitch and intensity values for the Italian samples and high values for English, French, German and Icelandic. The results do not seem to reflect a standard rhythm representation of languages, since French sticks next to English (on the left) and German (on the right), with high values of variability for both intensity and semitones. Instead, for some reason, they seem to distinguish between the speaker's L1 (Italian) and all L2s, so that the interpretation is challenging. These values may result from insecurity on the part of the speaker as regards the L2s, but this should somehow be reflected by speech rate, as well (which does not seem to be the case, see chapter 5).



**Figure 3.15.** Charts showing the values obtained by computing the deltas (on the left) and the rPVI (on the right) on semitones and intensity values of vocalic intervals. Data consist of 10 recordings of *The North Wind and the Sun* in 5 languages (twice per language) by a native speaker of Italian.

#### 3.4.4 So...

Whatever the interpretation of these results, the integration of pitch and intensity is certainly no easy task and has to be pondered more. Besides, this method does not

### 3. 1999-2010 : *rhythm metrics*

appear to be suitable for a direct integration in a speech rhythm account for a number of reasons. Firstly, it is very problematic to acquire comparable data (as has been said, intensity values differ across studies according to recording methods and conditions). Furthermore, results run the risk of being influenced by various phenomena and are not easily interpretable. Arguably, the integration of prosodic parameters could more sensibly integrate a speech rhythm account at a higher level.

#### **3.5 Conclusion**

By applying the formulae of various rhythm metrics on the corpus, it has been shown that all of them provide an acceptable representation of speech rhythm that is mainly consistent with expectations based on perceptive impressions. Moreover, other studies have shown that it is possible to obtain a rhythm categorisation even by applying the deltas and the PVIs to voiced and devoiced intervals. This somehow suggests that the different formulae are interesting and challenging more for the theoretical perspectives and the rationales standing at their base, than for the practice. It has also been shown that each of them has advantages and drawbacks, both at the theoretical and at the practical level. For example, metrics that normalise for speech rate run the risk of ignoring relevant phenomena; yet, they usually provide a more solid representation of different language samples precisely because they neutralise differences in speech rate.

One thing to keep in mind when observing the results for rhythm metrics is that they reflect exclusively the first level of rhythmicity, that is to say the segmental one. They do not measure anything at the second level. Therefore, researchers should be careful at classifying languages on the basis of what they see on these charts. It can be inferred that low deltas, varcos or PVIs characterise languages tending towards syllable-timing or segmental control. However, languages that show high values of deltas, varcos or PVIs cannot be said to tend towards stress-timing, because nothing has been measured at that level: simply, these languages do not tend to syllable-timing, but there is no proof at all that they tend towards stress-timing. In fact, in line with other studies (see the scheme presented in Bertinetto & Bertini, 2010), I suggest that the two levels allow for a quadripartition of languages based on control/compensation at each level.

It also has to be noted that the four slots of the quadripartition need not necessarily be all represented by languages: time will tell. The CCI are clearer as the other metrics as for this, as they only intend to describe the intra-syllabic behaviour of languages; therefore, languages aligning along the bisecting line are only said to show segmental control, while languages clustering below the bisecting line are only considered to compensate at the segmental level.

As for the inclusion of prosodic parameters (namely pitch and intensity), the preliminary test has not provided a clear representation of language groups. This may well be because the values have been computed on a non-native speaker, but procedural caveats make it difficult to obtain comparable data by several speakers. However, it is desirable that intensity and pitch be somehow integrated in a fulfilled speech rhythm account.

**4.**

**Correlatore**

### 4.1 Introduction

This chapter is the most technical one of the thesis as it introduces *Correlatore*, a programme that I have developed in Tcl/Tk in order to automatically compute the most commonly used rhythm metrics (%V,  $\Delta C$ ,  $\Delta V$ , Varcos, PVIs, CCIs) from *Praat's* annotation files. Initially, *Correlatore* was merely conceived as Perl script applying the formulae of rhythm metrics to raw data, but it then developed into a full framework that can be used for the study of linguistic rhythm with the metrics. It includes a tool for saving and organizing data into multiple reports, a module for building customizable charts, and various extras, such as the visualisation of the segmentation (into vocalic and inter-vocalic segments), the consultation of the formulae for each rhythm metric, the customisation of how SAMPA transcriptions should be treated and the creation of simple SASASA files for perceptive tests.

The version of *Correlatore* that will be presented here is 2.2<sup>63</sup>. At first, I shall briefly review the reasons for setting out to develop such a programme, and then I shall illustrate how it works and how it has been implemented.

#### 4.1.1 The reasons for developing *Correlatore*

Since the computing of rhythm correlates is an extremely time-consuming task, most authors have used (above all at first) a restricted corpus for their research. Some authors (e.g. Rouas & Farinas, 2004) have thus used automatic segmentation tools in order to speed up the procedure and be able to use greater amounts of data. However, the results are marred by the fact that the tools that are presently available are not yet able to produce a high-quality segmentation. Most authors assume though that the drawback of low-quality segmentation is somehow compensated by the use of great amounts of data.

On the other side, instead of automating speech segmentation, it is possible to speed up the processing of formulae on data. Everyone working with rhythm metrics has experienced that computing them in spreadsheets is not only time-consuming, but also fairly uncomfortable and complicated. This is even more so when one sets out to enlarge their corpus and/or decides to use other metrics.

So, after annotating speech samples of English, French, German, Italian, Icelandic and Finnish with *Praat* and computing the deltas, the varcos, the PVIs and the CCI, I experienced growing difficulties in the organisation and maintenance of data on spreadsheets. Furthermore, since I had used different spreadsheets for different metrics, making modifications and/or corrections on data was extremely complicated and redundant, which, in turn, made it difficult to make cross-comparisons between specific sets of data and metrics.

For these reasons, I decided that I needed at least a faster and simpler way of calculating rhythm metrics, so at first I wrote a Perl script which processed the formulae on *Praat's* annotation files (*TextGrids*). Then, I thought I might make it publicly available and built a rudimentary graphic user interface with Perl/Tk. At that point, I realised that I could develop a more sophisticated program, so I turned to Tcl/Tk and gradually added extra features, which finally merged into *Correlatore*

---

<sup>63</sup>At the moment of writing, the last version downloadable from the website was 2.1, but the 2.2 version should be soon available. It can be retrieved at the following Internet address [http://www.lfsag.unito.it/correlatore/index\\_en.html](http://www.lfsag.unito.it/correlatore/index_en.html) (it has a GPL license).

## 4. *Correlatore*

1.0. A series of bug fixes and modifications in the interface and report system were then integrated in *Correlatore* 2.0.

So, if one wishes to carry on research on rhythm metrics, one only needs to annotate wave files with *Praat* and, then, to open the *TextGrids* in *Correlatore*: one will obtain the values for the correlates quoted above and will be able to build charts with the results displaying all possible combinations of metrics.

Finally I shall explain the name *Correlatore*: it refers to how the metrics were first called by Ramus, Nespore & Mehler (1999), that is to say *rhythmic correlates*, *correlati del ritmo* in Italian. I added the Italian suffix *-tore* (which can be compared to the English suffix *-er* and which denotes the agent or the instrument) to the stem *correlato* to form *Correlatore* (that is to say, “somebody or something that produces correlates”); but it is also a pun, as the word *correlatore* actually exists in Italian and indicates the co-director of a thesis.

### 4.2 How to use *Correlatore*

#### 4.2.1 The annotation of *Praat*'s *TextGrids*

*Correlatore* works on *Praat*'s *TextGrid* files, so the first step is of course to annotate sound files in *Praat*. As of version 2.2, annotations with two types of transcriptions are supported: in CV and SAMPA (CV transcriptions are implemented with a higher degree of correction). However, whatever the choice of using CV or SAMPA, one should carefully follow the conventions reported below in order to avoid the risk of *Correlatore* interpreting the transcriptions badly. Although these conventions may seem odd and unnecessary, there are reasons why I had to introduce them: they are explained in chapter 3 with the illustration of how the segmentation is carried out.

For CV transcriptions, every label should correspond to a vocalic/consonantal interval and be annotated with as many “c” or “v” as the number of segments composing the interval. For instance, <campus> should be annotated as |c|v|cc|v|c| and Italian <palla> as |c|v|cc|v|. Pauses should be left empty or labelled as |#|. This leaves the user free to decide how to treat segments whose phonological status is debated (such as syllabic consonants) and to be in full control of the subdivision of intervals; so, it is possible to follow the instructions set by Bertini & Bertinetto (2009) for the calculation of the CCI: for example, *hyati* can be labelled as 2 distinct intervals: e.g. Italian 'suo' |c|v|v|.

Alternately, it is also possible to use a simpler segmentation that does not take into consideration the number of segments composing each interval, ex. <campus> |c|v|c|v|, Italian 'palla' |c|v|c|v|, but keep in mind that this will result in faulty CCI values (their formula requires each interval to be divided by the number of phonological segments composing it, which in this case would be interpreted as 1, thus giving the same results as the rPVIs).

For SAMPA transcriptions:

- a) every label should correspond to one and only one phone (that is to say a vowel or a consonant, not a vocalic or consonantal interval);
- b) phonologically long phonemes (like long vowels in Finnish and geminate consonants in Italian) should be annotated with two distinct labels (even though the boundary between the two is of course fictitious); for instance,

## 4. Correlatore

- the Finnish word 'saami' should be annotated as: |s|a|a|m|i|, not as |s|a:|m|i| and nor as |s|aa|m|i|;
- c) it is normally possible to use standard SAMPA diacritics, but if you use any non standard diacritic or annotation convention, this may interfere with the substitution variable (see below). For instance, if you use t\_u (instead of t\_w) to indicate a labialised voiceless alveolar plosive, this label will be interpreted as a vowel because of the 'u'.
  - d) Pauses should be labelled as '#' or left empty.
  - e) *Correlatore* uses a substitution variable to transform SAMPA transcription into CV sequences that contains all the symbols that should be considered as vocalic: if a label contains one of these symbols, it will count as a vowel, otherwise as a consonant (except for '#', which indicates pauses). The variable's value is shown in the statusbar in the main window and it is possible to change it by clicking on it.
  - f) During the process of segmentation of a tier labelled as SAMPA, *Correlatore* builds vocalic and consonantal intervals by summing the duration of adjacent consonants/vowels. This means that *hyati* will be considered as one interval: although, this does not have any effects on the results of the deltas, varcos and PVIs, it does have some consequences on the values of CCI. So, if one wishes to obtain more precise results for the CCI, it is advisable to opt for a CV segmentation.

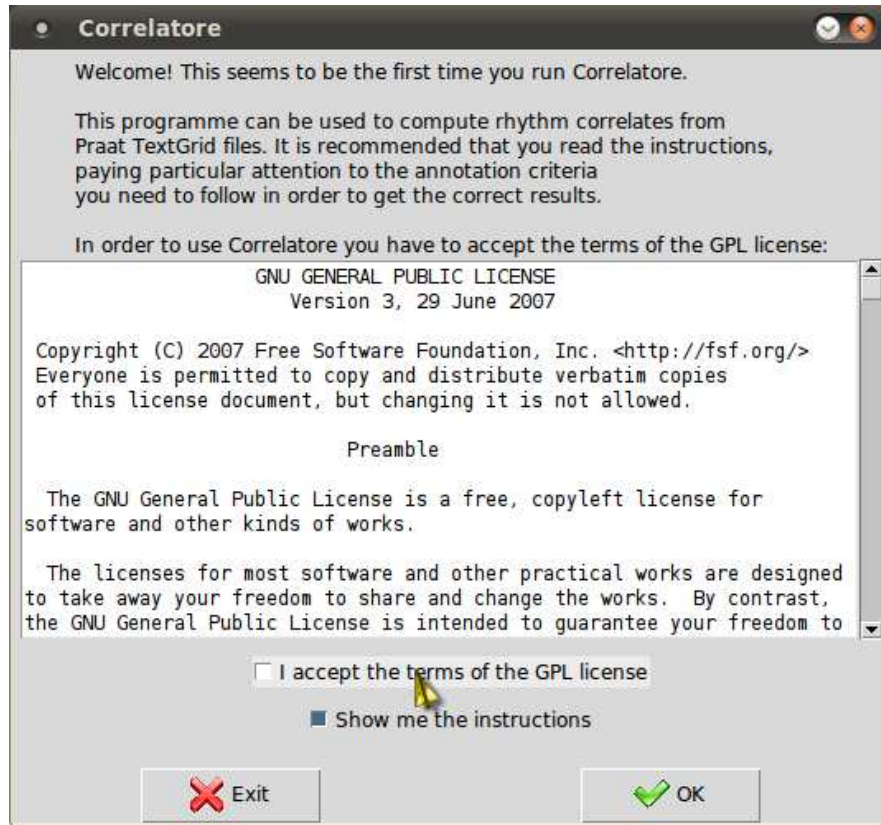
### 4.2.2 Computing rhythm metrics

Once the *TextGrids* have been annotated, it is possible to open them in *Correlatore*. As already mentioned, *Correlatore* does not need installation, it is started by simply double-clicking on the executable. The first time it is run, the user will be prompted with a window asking to specify the language (English or Italian, see figure 4.1), and then to accept the terms of the GPL license and whether he/she would rather see the instructions (see figure 4.2).

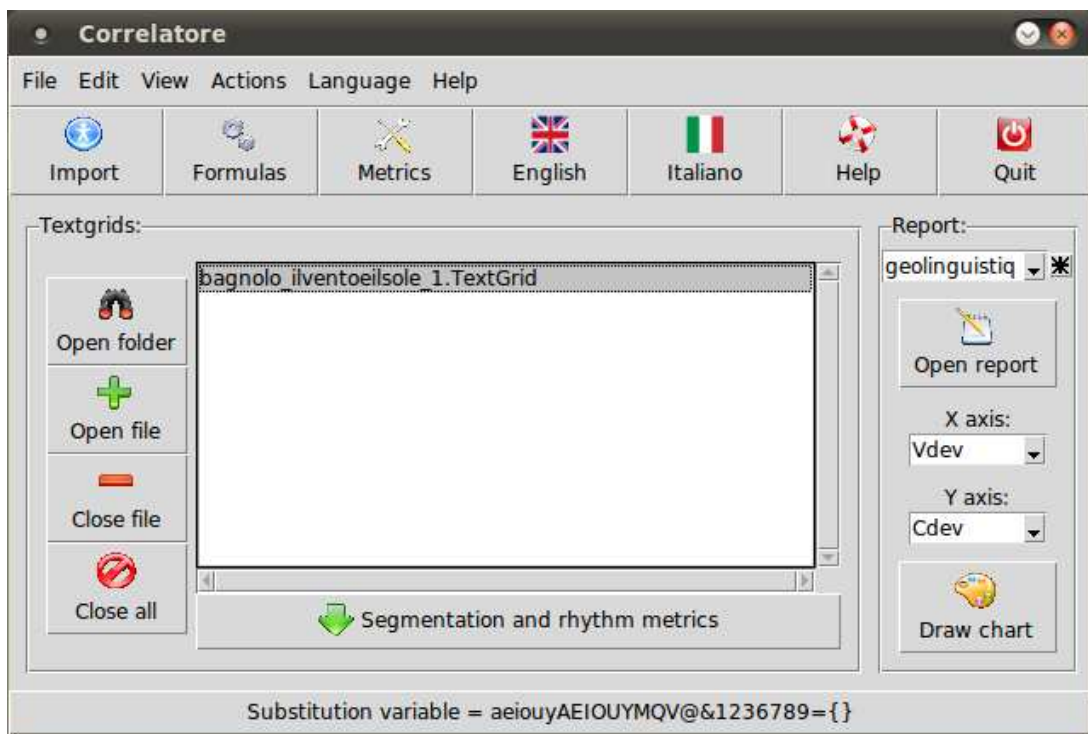


**Figure 4.1.** *Correlatore*'s window asking to choose the language. This window is only shown the first time the executable is run.

## 4. Correlatore



**Figure 4.2.** *Correlatore*'s window asking to accept the GPL license. This window is only shown the first time the executable is run.



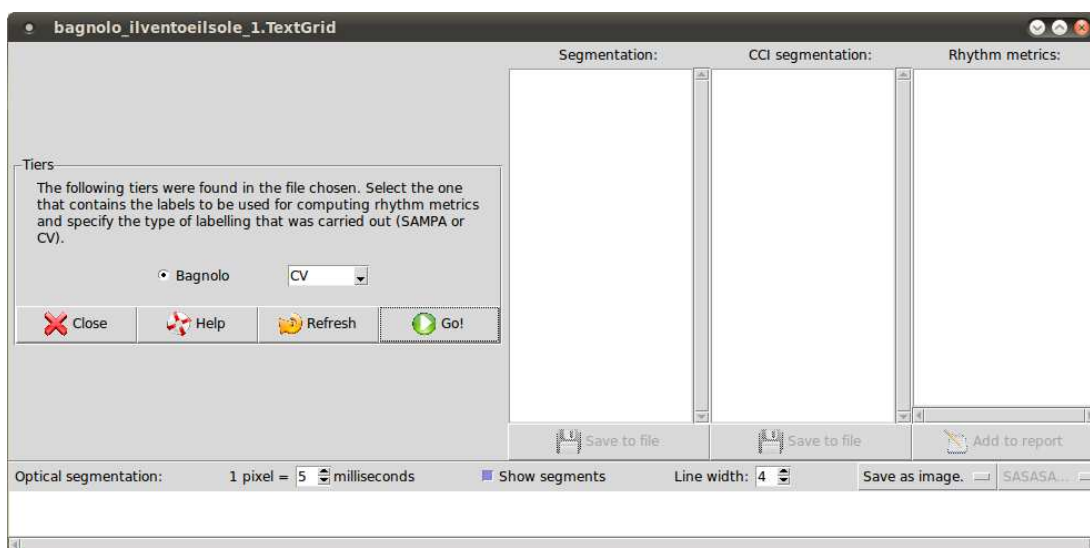
**Figure 4.3.** *Correlatore*'s main window.



## 4. Correlatore

After making these choices, the user will be presented with *Correlatore*'s main window (figure 4.3), which includes a menu, a toolbar, some buttons and an empty box. The left part of the window deals with *TextGrids*, while the right part deals with reports and charts. The statusbar indicates the current value of the SAMPA substitution variable. Once *Correlatore* starts, it automatically scans in its folder for files with a *TextGrid* extension; if it finds any, they will appear in the box of the main window; otherwise, it will be possible to click "Open file" or "Open folder" and browse to the folder containing the *TextGrid*(s). Once open, its/their name(s) will be shown in the box. They can be closed by clicking "Close file" and "Close all".

In order to compute the rhythm metrics, one has to select one and only one *TextGrid* and click on "Segmentation and rhythm correlates". A new window will pop up (see figure 4.4) showing on the left the names of the tier(s) found in the *TextGrid*. The user is asked to select the tier (if there is more than one) he/she intends to work with and to specify the type of annotation used<sup>64</sup> (aka SAMPA or CV).



**Figure 4.4.** *Correlatore*'s segmentation window.

The user should make his/her choices and press "Go!". The three boxes on the right will be filled (see figure 4.5) and, in case of problems, a log window will pop up (for example, if *Correlatore* finds unexpected labels, such as a |b| in a *TextGrid* which has been annotated as CV). In the first box on the left, it is possible to see how the data were segmented (*i.e.* the consonantal and vocalic intervals with the corresponding durations in ms) for the deltas, the varcos and the PVIs. In the second box you will see how the data has been segmented for the CCI: note that there will only be a difference between the contents of those two boxes if the *TextGrid* has been annotated with a CV transcription following the conventions specified above. Both segmentations can be saved in TXT format by pressing "Save to file". In the third box you will see some information about the file (n° of V and C intervals, n° of pauses, mean duration of V and C segments) and the values of rhythm correlates. It is possible to save these results to a report by clicking on "Add to report" and then

<sup>64</sup>*Correlatore* does try to detect whether every tier has been labelled as SAMPA or CV but the algorithm used is extremely simple and should not be trusted blindly.

## 4. Correlatore

by selecting the report that should contain them (by default they will be saved in the active report).

At the bottom of the window, it is possible to see a graphic representation of the vocalic and consonantal segments (the colours, size and scale of the lines can be customised using the appropriate controls). Using the controls on the right at the bottom, it is also possible to listen to a SASASA file which *Correlatore* builds automatically during the segmentation<sup>65</sup>: it consists of a rudimentary synthesis of C intervals as [s] and of V intervals as [a] that can be used for perceptive tests. It can also be saved in WAV format.

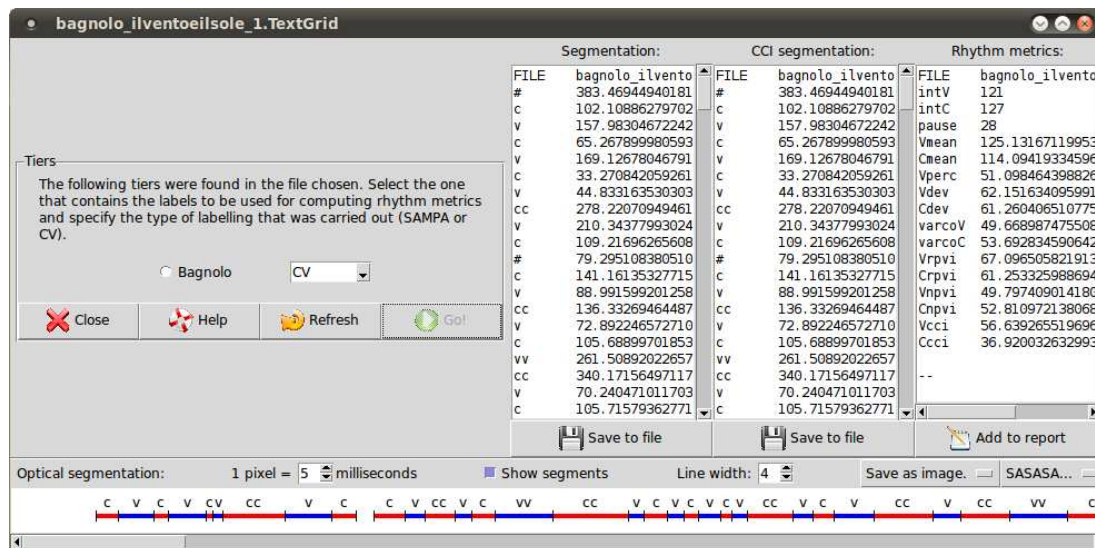


Figure 4.5. *Correlatore*'s segmentation window after the computation of rhythm metrics.

Finally, it is possible to compute the metrics on other tiers by clicking on “Refresh” or to go back to the main window by clicking on “Close”.

### 4.2.3 Reports

Reports contain the results of correlates computed on one or more files. They can be viewed, modified and exploited from the right frame of the main window: a pop-down menu allows the user to select one report among the existing ones. Pressing “Open report” will open a new window which allows to see the results of the metrics by clicking on the name of each item (see figure 4.6). The user can view and edit the items stored in the active report: it is possible to rename one or more items, to delete them or to calculate the mean of their values. In the last case, a new item will be created containing the means and the standard deviation, which will be used as the value of error bars when building charts. So, for instance, it is possible to have a sound file annotated by 2 different people, to calculate the correlates on both resulting TextGrids, to save data in the report and to calculate the mean: this way, when charts will be created with these data, a circle will be shown to indicate the value of the mean, while error bars will reflect inter-operator variability. Also, one can annotate say 10 sound files from different speakers of the same language and save the results of the metrics in the report, then calculate the mean and draw the

<sup>65</sup>If *Correlatore* is executed from the sources, the *Snack Sound Toolkit* needs to be installed. If it is not, this feature is disabled.

#### 4. Correlatore

chart in order to visualise the mean value with error bars indicating inter-speaker variability within the same language<sup>66</sup>.

Items	Metrics	Values A	Values B	Stdev A	Stdev B
bagnolo_ilventor	FILE	bagnolo_ilventoeilsole_1			
briga	intV	121			
campertogno	intC	127			
capanne	pause	28			
exilles	Vmean	125.13167119953573			
roccaforteligure	Cmean	114.09419334596413			
	Vperc	51.09846439882685			
	Vdev	62.15163409599165	59.50381732829607		
	Cdev	61.260406510775795	52.3177788696173		
	varcoV	49.66898747550832	44.74819690107703		
	varcoC	53.692834590642	44.59884765040941		
	Vrpvi	67.0965058219137	68.64128126747033		
	Crpvi	61.25332598869428	66.59048056570023		
	Vnpvi	49.79740901418009	50.223122388213184		
	Cnpvi	52.81097213806834	56.23021637950512		
	Vcci	56.63926551969613	56.50819246474357		
	Ccci	36.92003263299336	36.605567507615476		
	colour	#CC4466			
	border	black			
	symbol	c			
		--			

Figure 4.6. Correlatore's report window.

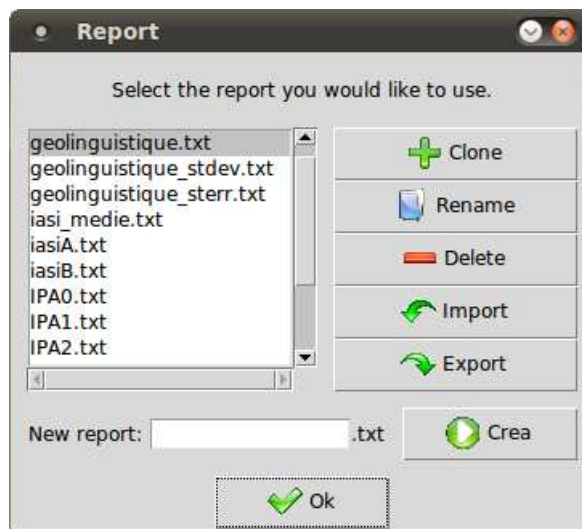


Figure 4.7. Correlatore's window from which it is possible to clone, rename, delete, export and import reports.

<sup>66</sup>One may be surprised when visualising *Correlatore*'s error bars, as they are greater than error bars in many similar studies. This is because these works use the standard error, whereas *Correlatore* uses the standard deviation. The standard error is the standard deviation of the sampling distribution associated with the estimation method. So, in order to get the value of the standard error, one has to divide the value of the standard deviation by the square root of the number of samples used to calculate the mean: the standard error is therefore smaller than the standard deviation (unless the mean is calculated on one only sample, which definitely is neither useful nor advisable).

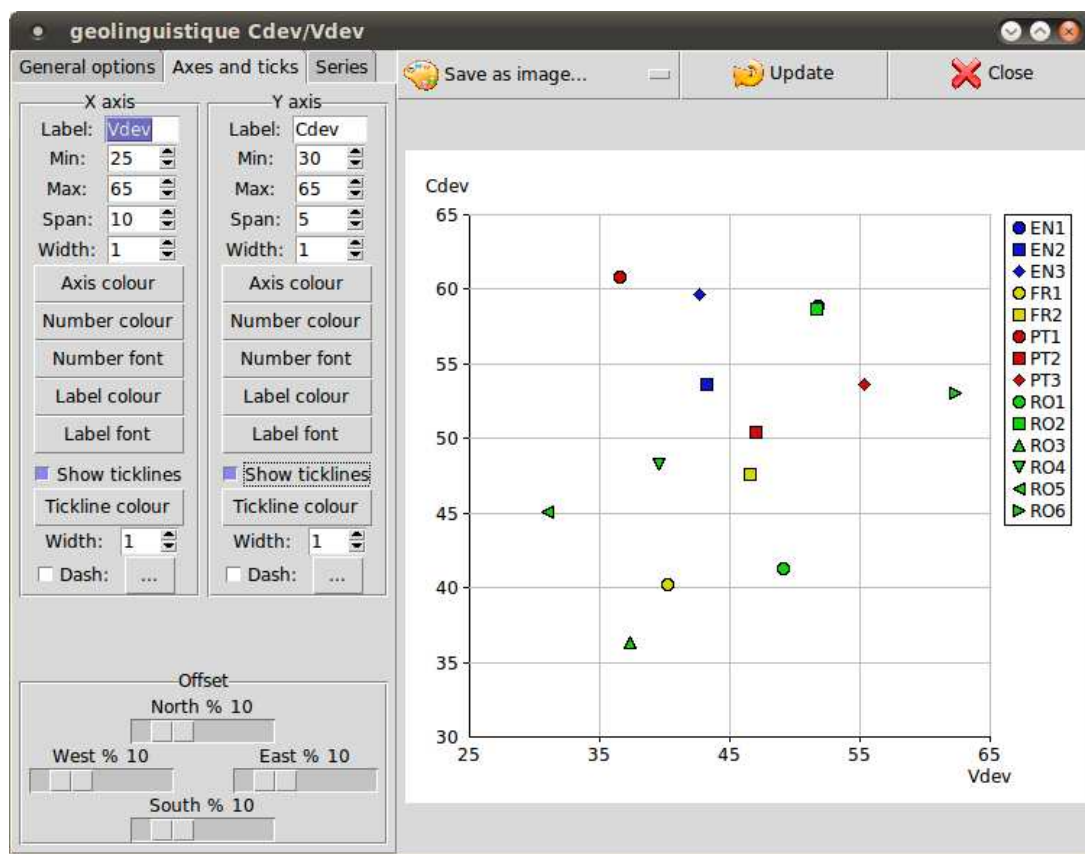
## 4. Correlatore

Of course it is possible to create new reports, to rename them, to export and import them. These operations can be carried out by clicking on the asterisk-button beside the pop-down menu of the main window, which will pop up a new window (see figure 4.7).

Although the import/export facility allows the user to easily exchange data among different computers and/or users, one should be careful as *Correlatore* does not check the validity of imported reports (it only checks that they are in TXT format) and even a small change may make them unusable or faulty. Importing valid reports is considered to be the user's responsibility.

### 4.2.4 Charts

Charts can be built from data in the report. In the main window (see figure 4.3) it is possible to choose which metrics have to be represented in the x and y axes using the two pop-down menus at the right. Then, by pressing “Draw chart”, a window will open (see figure 4.8) with a chart and several controls for customisation: one can specify preferences as for the size of the chart, the indicators' shapes and colours, the legend, the labels, the title, the axes, etc.



**Figure 4.8.** *Correlatore*'s chart windows. The chart can be modified using the controls on the left (not all are visible because they are distributed on three tabs).

Charts can also be exported to several image formats (JPEG, PNG, GIF, BMP, GIF, etc.<sup>67</sup>) by clicking on “Save as image” in order to be inserted into

<sup>67</sup>If the user is running *Correlatore* from the sources, the extension *Tking* needs to be installed. Otherwise, it will only be possible to export charts in *PostScript* format.

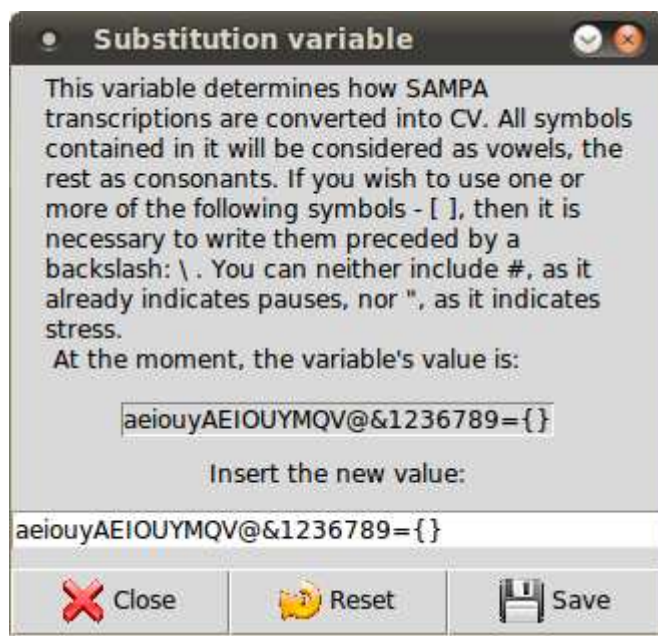
#### 4. Correlatore

publications. Most charts presented in the previous chapter have been created this way.

##### 4.2.5 Preferences and configurations

Configuration options are stored in a configuration file in order to make them persistent. This means that they will be remembered if you close and restart *Correlatore*.

The SAMPA **substitution variable** is used in the transformation of SAMPA transcriptions into CV sequences. It contains all the symbols which are to be considered as vocalic: that is to say, when a *TextGrid* labelled with SAMPA is opened, every label containing one of the symbols in the substitution variable is replaced with V, in all other cases with C (except for # which indicates pauses). Its default value is `aeiouyAEIOUY@MQV&1236789={}` (so syllabic consonants are included, while glides are considered as consonantal), but you can modify it by clicking on “Edit variable” or through the menu “Edit” (see figure 4.9).



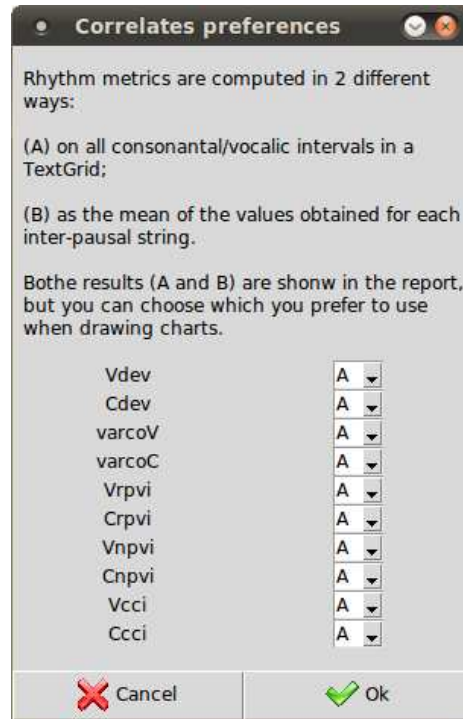
**Figure 4.9.** *Correlatore*'s window for the modification of the substitution variable.

**Rhythm metrics preferences** control how the metrics are computed. There are two possibilities:

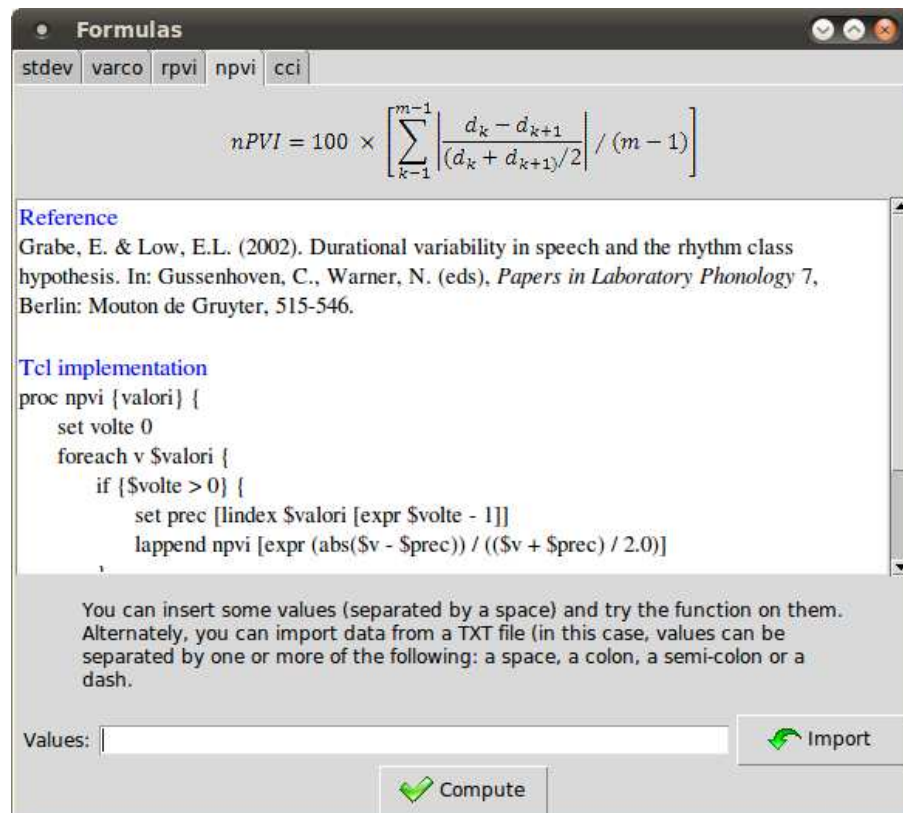
- A) they can be calculated by applying the formulae (delta, varco, rPVI, nPCI or CCI) to all the vocalic and consonantal intervals found in a tier;
- B) they can be calculated by applying the formulae (delta, varco, rPVI, nPCI or CCI) to the vocalic and consonantal intervals of every single inter-pausal segment and then calculating the mean of the values obtained.

Starting from *Correlatore* 2.0, all correlates are computed both ways (and both results are saved in the report); however, it is necessary to specify which type of results you wish to use when building charts: *Correlatore* uses method A by default, but it is possible to modify this behaviour by clicking on “Metrics” in the toolbar, or through the menu “Edit” (see figure 4.10).

#### 4. Correlatore



**Figure 4.10.** *Correlatore*'s window for setting the preference on the A/B methods of computing rhythm metrics.



**Figure 4.11.** *Correlatore*'s window showing the formulae and their Tcl implementation. The user can try the formulae by inserting values in the text field at the bottom and clicking on “Compute”.

## 4. Correlatore

If one wishes to see how rhythm metrics are computed, it is possible to click on **Formulae** or on the menu “Edit” and then on “View TCL implementation of rhythm metrics”. A new window will pop up (see figure 4.11), showing the formulae of rhythm metrics and their TCL implementation. It is possible to insert numerical values or to import them from a TXT file in order to try the formulae.

Moreover, *Correlatore* is available in English and Italian, it is possible to switch language simply by clicking on the corresponding button. Moreover, it is possible to control other more futile preferences, such as hiding/showing the statusbar, the toolbar and the tooltips; on Unix machines (other than MacOSX) one can also choose three different themes for the interface.

### 4.3 The implementation of Correlatore

*Correlatore* was developed in Tcl/Tk and consists of 4663 lines of code in total (plus all the documentation in English and Italian). All procedures<sup>68</sup> (including the ones for the creation and customisation of charts) were written by me in order to make *Correlatore* completely independent from other projects and to better suit to my needs. I only used two external libraries (*Tking* and *Snack*), for two extra features: the first controls the conversion of charts into several image formats (jpeg, gif, png, tiff, etc.), whereas the second is used for the manipulation of sounds for the synthesis of SASASA files. However, these two libraries are optional, meaning that if for some reason they are not provided, *Correlatore* will run normally, except for the fact that these two extra features will be disabled.

I shall now present how it was implemented, first explaining the reasons for choosing Tcl/Tk, then providing an overview of its structure and then focusing on specific issues.

#### 4.3.1 Why Tcl/Tk?

Tcl (Tool Command Language) is a scripting language created by John Ousterhout and first appeared in 1988, while Tk (ToolKit) is an extension to Tcl that makes it possible to build graphic user interfaces (GUIs). The combination of the two is generally referred to as Tcl/Tk and has been used widely, also in several academic projects<sup>69</sup>. The reasons for choosing it over other languages are the following:

- 1) It is open-source and licensed under very liberal terms.
- 2) It is a multi-platform interpreted procedural language, which allows the sources to run unmodified on many operating systems.
- 3) It provides several useful features such as regular expressions (which have been essential for the analysis of Sampa-labelled TextGrids) and a canvas widget (which has made possible the development of a module for the creation of charts).
- 4) Tk has sometimes been criticised because the graphic interfaces created with this toolkit are said to look ugly and do not integrate with the operating system. If ever this has been true and of any relevance, it is no longer so starting from version 8.5, which uses native widgets on Windows and

---

<sup>68</sup>Subroutines are called procedures in Tcl.

<sup>69</sup>For instance, *Wavesurfer*, a well-known sound visualisation and manipulation tool, is written in Tcl/Tk.

## 4. *Correlatore*

MacOsX (not yet on other Unix platforms – but the appearance has been improved here as well).

- 5) With Tclkit, it is possible and easy to create executables that do not need an installation of Tcl/Tk and that are extremely light in terms of system requirements (*Correlatore* for Windows requires less than 2 MB) and that do not even need any installation at all.
- 6) Several extensions exist for the simplification of some tasks: in particular I used Tking (for the conversion of charts into several image formats for use in articles and other publications) and *Snack* (which provides facilities for acoustic analysis and is used for the synthesis of simple SASASA files).

It has to be remarked that I have used many features of Tk 8.5 which are not available in the old but far more popular 8.4 version. However, it is possible to run *Correlatore* under Tcl/Tk 8.4 by installing an extension called *tile*, (which provides the missing widgets) as I made an effort to write Tcl code that is completely compatible with Tcl 8.4.

Finally, it is worth to mention that although the sources have been designed with portability in mind and should run unmodified on all platforms, *Correlatore* has been developed since the beginning on Ubuntu Linux and is only tested on this platform and on Windows XP.

### 4.3.2 Overview

*Correlatore* consists of eight source files and a folder containing a library (*plib*) which I developed specifically and which contains the most commonly used procedures. `correlatore.tcl` is the main file (the one that has to be executed to start *Correlatore* from the sources): firstly, it does a couple of checks (e.g. it controls Tcl/Tk's version), then it creates an interface to deal with any runtime error, finds its path, loads *plib*, all images for the GUI and the procedures contained in the other files, sets some preferences, checks if there has been a version upgrade (if so, it loads `versioni.tcl`, which fixes compatibility issues between versions) and finally loads `start.tcl`. This file prompts the user to accept the GPL license at the first execution, that it reads (or creates at the first execution) the configuration file, sets global variables and builds the graphic user interface (GUI) of the initial window. As in normal GUIs, all items in the window are associated to procedures that are executed at specific events: for instance, the button *Draw chart* is associated to the procedure `disegna`, which is executed when this button is pressed by the user.

I shall now discuss a number of issues that have posed problems or difficulties in the implementation.

### 4.3.3 How the segmentation and the transcriptions are dealt with

As has been said above in this chapter, *Correlatore* accepts CV and SAMPA transcriptions, as SAMPA transcriptions are internally converted to CV transcriptions before the segmentation. The first thing to be done is to fix SAMPA characters that may potentially interfere with Tcl, namely inverted commas, backslashes, as well as square and curly brackets. Then, SAMPA transcriptions are converted into CV transcriptions using the so called substitution variable: this variable (whose content is customisable and persistent as it is saved in the



#### 4. *Correlatore*

configuration file) contains a string with all SAMPA symbols that have to be considered as vowels. *Correlatore* compares each SAMPA annotation with the substitution variable and considers it to be a vowel/consonant if it matches/does not match any character of the substitution variable. This type of conversion is far from being sophisticated and should therefore be improved in the future. Instead, CV transcriptions are only checked: should one segment not contain any of the following five symbols "c", "C", "v", "V" or "#", it is discarded from the list and is appended to the error variable, which is then used for a final log.

As far as the segmentation is concerned, it is very complex mainly because of the different needs of the deltas, varcos, and PVIIs on the one hand, and the CCI on the other hand, which requires a partially separate treatment. In fact, this is also the reason why annotation criteria for CV files are so complex (see above, 4.2.1). I shall now analyse the difficulties in the implementation and solutions adopted while illustrating the reasons for the constraints on transcription criteria.

The “older” metrics merely require a segmentation into consonantal and vocalic intervals so that the implementation would be simple both for CV transcriptions and for SAMPA annotations. As for CV transcriptions, the situation would be extremely straightforward as a simple |c| annotation for consonantal intervals and a |v| annotation for vowels would be enough: the script should just create two lists (C and V) and then apply the formulae to the durations. As for SAMPA transcriptions, each annotation only contains one phoneme: so, the implementation would simply need to construct consonantal and vocalic intervals by summing up the durations of, respectively, consonantal clusters and adjacent vowels (be they diphthongs or hiati).

Instead, the formula of the CCI divides each interval duration by the number of phonological segments that compose it. For SAMPA transcriptions, this does not imply any particular problem: as has been said, vocalic and consonantal intervals have to be constructed, so one simply has to keep track of how many segments are united to compose each interval and then divide the duration of the interval by that number. On the contrary, this does have some repercussions on CV annotations, as simple |c| and |v| labels for whole vocalic and consonantal intervals are no longer adequate. In fact, since *Correlatore* does not access sound files<sup>70</sup>, there is no way it could distinguish a |c| indicating a simple consonant from a |c| indicating a cluster of two, three or more consonants; in short, *Correlatore* would not know how many segments compose the interval and, therefore, could not make a division by that number. So, each label needs to be annotated so that the number of segments is explicit: |c| for a simple consonant, |cc| for a cluster of two consonants, |ccc| for a cluster of three consonants, and so forth. In the same way, |v| for a single vowel, |vv| for a diphthong, |vvv| for a triphthong. This results in an acceptable implementation of all correlates.

However, if one wants to stick to the indications on how to compute the CCI given by the two authors (Bertini & Bertinetto, 2009), there are further complications: hiati have to be considered as two separate segments. Again, the problem is that *Correlatore* does not access sound files and therefore cannot distinguish between a “vv” label indicating a diphthong from a “vv” label indicating a hiatus. Its implementation is possible by slightly changing transcription criteria

---

<sup>70</sup>And, even if it did, it would need to do some speech recognition in order to establish how many segments compose an interval, which is well beyond my purposes.

## 4. *Correlatore*

and conventionally establishing that diphthongs must be annotated in one single label, whereas hyati have to be annotated in two separate labels: so, |vv| (or, for instance, |aI| in SAMPA) should indicate a diphthong and |v|v| (or |a|I| in SAMPA) should indicate a hyatus. This has been implemented exclusively for CV transcriptions<sup>71</sup> and is also the reason while *Correlatore* (from version 2.2 onwards) presents two segmentations: the first is the one that is used for the computing of deltas, varcos and PVIs (where hyatus labels such as |v|v| have to be united and considered as a single vocalic interval); the second is the one that is used for the computing of the CCI (where such hyatus labels are kept separated).

In conclusion, the analysis of the transcription and the process of segmentation are fairly complicated and reserve a separate treatment to SAMPA and CV transcriptions as well as to deltas, varcos, and PVIs on the one hand, and to the CCI on the other hand. The procedures dealing with them are all contained in the file `calcoli.tcl`: while reading the *TextGrid*, the content of each tier is stored in a Tcl list where the annotation of each segment and its duration in ms are appended one after the other. Once the user chooses the tier that contains the CV or SAMPA annotation, each element of the list is analysed in a `foreach` cycle<sup>72</sup> in which SAMPA transcriptions are converted to CV transcriptions and where the two groups of metrics are dealt with separately.

Future improvements of the process of segmentation should include a new implementation of the SAMPA to CV conversion and the possibility of using IPA transcriptions.

### 4.3.4 How formulae are dealt with

All formulae of rhythm metrics are implemented as single and independent procedures<sup>73</sup> and it is even possible to visualise them within *Correlatore* itself (by pressing the *Metrics* button in the toolbar). Their implementation is reported in appendix 3a.

Deltas, varcos and the PVIs require just one argument, that is to say a list of numeric values (on which the formulae are applied). Instead, the CCI is slightly more complex and it requires the number of phonological segments for each vocalic or consonantal interval to be passed as an argument as well. So, the CCI procedure has to be called with two parallel lists as arguments, the first being a the list of the durations of each interval, the second being a list of the number of segments contained in the corresponding interval.

It has to be noted that these procedures simply apply the formula on the numeric values that are passed as arguments; the job of make a difference between consonantal and vocalic segments is done previously and separately.

### 4.3.5 How reports are dealt with

As it has been said, results of the metrics can be saved in reports. Of course, reports have to be persistent through different instances of *Correlatore*, that is to say that

---

<sup>71</sup>The reason for not implementing this for SAMPA transcriptions is that it would create further problems because of the method used for the SAMPA to CV conversion (that is to say, the substitution variable). Since this method is not sophisticated and at any rate will be improved in the future, I decided not to implement it for the moment.

<sup>72</sup>The `foreach` cycle is reported in appendix 3b for inspection.

<sup>73</sup>Apart from %V, which does not have a procedure for itself as it is simply implemented in-line.

#### 4. *Correlatore*

they should not be lost when *Correlatore* is closed or when the computer reboots; for this reason, they cannot simply be stored in variables. The choice was that of either using a small database (maybe using Metakit, which is readily available as it is embedded in Tclkit), or to save them in text files. I finally opted for the latter possibility and decided that report files should be saved in *Correlatore*'s hidden configuration folder (whose path is `~/.correlatore`<sup>74</sup>). Each report is a file contained in the `report` sub-folder, and includes all data saved to that report in text format (with a UTF-8 encoding). The structure of each entry in the report is as follows:

```
FILE      Eempio
intV      216
intC      212
pause     15
Vmean     72.71495663716493
Cmean     90.27291956026146
Vperc     45.076053559321686
Vdev      32.11878816751892    33.03864836735656
Cdev      39.43809243378562    39.67276512149737
varcov    44.17081389154383    44.439752814033504
varcovC   43.687622629130566    42.58782999749963
Vrpvi     32.05840505039406    34.38483639549482
Crpvi     49.90414965937686    50.76376358755145
Vnpvi     40.39055981686158    41.689759668844076
Cnpvi     54.545009862312966    53.72203260830348
Vcci      33.552866292942625    36.812405375916605
Ccci      21.073056103497294    20.830852059951457
colour    #662211
border    black
symbol    c
```

The first line (FILE) reports the name of the entry and is followed by the number of vocalic and consonantal segments, the number of pauses, the mean duration of vocalic and consonantal intervals (which are useful for calculating speech rate<sup>75</sup>), then the values of %V and all the other metrics (the first column reports the results obtained with the A method, the second column reports the results obtained with the B method, see above for details). The last three lines include the fill and border colours as well as the symbol to be used for chart indicators: the first

---

<sup>74</sup>According to Unix conventions, the symbol `~` indicates the path of the home folder. So, `~/.correlatore` translates differently on different platforms: `/home/paolo/.correlatore/` on all Unix platforms (including MacOSX), `c:\Documents and Settings\paolo mairano\.correlatore\` on Windows XP, `c:\users\paolo mairano\.correlatore\` on Windows Vista.

<sup>75</sup>If CV annotation criteria have been attended, it can be inferred that every vocalic interval corresponds to a nucleus because hyati are assigned to two separate labels and syllabic consonants are considered as V (by default, at least). Of course, the number of nuclei is also equal to the number of syllables. So, it is possible to compute the number of syllables by simply  $1000/V_{\text{mean}}$ . The rate of syll/s is one of the possible indicators of speech rate.

#### 4. *Correlatore*

is a random number generated on the fly at the time of saving, while the other two are set by default to “black” and “c” (for circle) respectively.

During the visualisation of reports (which is dealt with by a set of procedures contained in `report.tcl`), all items in the report are stored in a complex data structure (based on a set of arrays<sup>76</sup> containing lists) for analysis and modification. Only when/if the user decides to save the modifications, the arrays are saved back in the report.

##### 4.3.6 How charts are dealt with

Reports can be used to simply save data (which can then be put in a spreadsheet or in other applications for further treatment), or they can be used to build charts. The creation of charts is rather complex (as I did not rely on any external libraries) and includes several procedures. The two main ones are called `disegna` (685 lines) and `grafico` (81 lines): to put it shortly, they read all items in the report and get the values of the two metrics under consideration (for the x and y axes) as well as the values of “colour”, “border” and “symbol”; then, most of the job consists in building the GUI components that interact with the chart: in fact, the values of all chart elements are attached to a global variable whose value is controlled by a GUI widget and can thus be customised by the user (e.g. the distance from the border and the axes, the scales, width, font and colours of the axes, etc.). However, nearly all of these parameters are set to default each time the chart module is initiated; the only parameters that are persistent are those saved in the report, that is to say the fill and border colour and the symbol used for each indicator (this is essential in order to have the same items constantly represented in the same way); so they are saved to the report every time the user modifies them.

Finally, the actual drawing of the chart into the canvas widget is done by a set of several procedures contained in `plib`, which include the conversion of pixels to chart scale and vice versa, the drawing of the axes, labels, grids, legend, indicators, etc. Drag-and-drop is implemented within the chart exclusively for the title, the legend, the axes labels and indicators labels. For the rest, charts are redrawn from scratch every time the user changes something<sup>77</sup>.

##### 4.3.7 SASASA files

As has already been said, one of the extra features of *Correlatore* is a module that synthesizes SASASA files: these are sound files in which an [s] sound replaces all consonantal intervals of the original file, whereas an [a] sound replaces all vocalic intervals of the original file. Such files have been previously used for perceptive tests (see chapter 6) and can have numerous variants. In particular, the original pitch and intensity contours can either be preserved or leveled.

*Correlatore* only produces flat SASASA files (with leveled pitch and intensity contours). During the process of segmentation and metrics calculation, the durations of vocalic and intervals are passed to a procedure that builds a visual segmentation (which is visible in the lower part of the window). The same values are also used for

---

<sup>76</sup>Arrays are a common data structure found in most programming languages. In short, they are composed of a set of values, each of which corresponds to a unequivocal key.

<sup>77</sup>This solution may not be very economical in terms of system resources, but certainly it is so in terms of ease of implementation.

#### 4. *Correlatore*

the synthesis of SASASA files with the help of *Snack*<sup>78</sup> through a very rudimentary method: the sources of *Correlatore* (only from version 2.2 onwards) include three WAV files, *c.wav*, *v.wav* and *silence.wav*, which contain the recordings of a stable [a] and a stable [s] respectively, both for a duration of 10 seconds. In order to create the synthesis, the procedure takes the necessary duration of [a] for vocalic intervals, of [s] for consonantal intervals and of silence for pauses and concatenates them one after the other. So, for instance, a sample utterance [wɒtʃɪt] (<watch it!>) with [#]=100ms [w]=80ms [ɒ]=90ms [tʃ]=110ms [ɪ]=50ms [t]=70ms [#]=200ms would be synthesised taking the first 100ms of *silence.wav*, then the first 80ms of [s] from *c.wav*, then the first 90 ms of [a] from *v.wav*, then the first 110 ms of [s] from *c.wav*, then the first 50 ms of [a] from *v.wav*, then first 70 ms of [s] from *c.wav* and, finally, the first 200 ms from *silence.wav*. Should a vocalic or consonantal interval be longer than 10 s (which is highly improbable in normal situations), that interval will be set to 10 seconds (the recordings last 10 s, so evidently I cannot provide syntheses longer than that).

This procedure is far from being sophisticated as of course shifts from [s] to [a] and vice-versa are abrupt and without formant transitions, but the result is audibly acceptable. I did the recordings personally at the Laboratory of Experimental Phonetics *Arturo Genre* of Turin in a sound-proof booth in order to get the best possible quality. Future perspectives include of course an improvement of this feature.

#### **4.4 Conclusion and future perspectives**

The functioning and implementation of *Correlatore* has been shown and it has been explained that it is possible to use for a quick computing of the most used rhythm metrics. The data presented in chapter 3, based on the calculation of these measures for the 61 samples and on the segmentation carried out by 2 different phoneticians (totalling nearly 110 *TextGrids*), would not have been possible without a tool like this.

As of now, its main default seems to be the fact that it is not flexible as to choices in how to label data. Future perspectives should include improvements in this sense, at least enhancing a better interpretation of SAMPA transcriptions, and of the technique for creating *flat SASASA* syntheses.

---

<sup>78</sup>The *Snack Sound Toolkit*, which has already been quoted above, is a Tcl and Tk extension that allows for the analysis and manipulation of sound files. It has been written by Kåre Sjölander at the Department of Speech and Hearing, University of Stockholm. It is also used as the base for *Wavesurfer*, a programme for the phonetic analysis of speech by the same author,

**5.**

**Rhythm variation and  
variability**

## 5.1 Introduction

Variation is a recurrent theme throughout linguistics, at all levels of analysis, and it even stands at the base of specific disciplines, such as dialectology, geolinguistics and, above all, socio-linguistics<sup>79</sup>. For this reason, I have decided to devote it an entire chapter of the thesis, focusing, of course, on rhythm variation and variability.

Firstly, I shall specify what I mean by *rhythm variation* and *rhythm variability*. The term *rhythm variation* will be used to refer to linguistic variation that pertains to the *langue* of a linguistic community (to put it in Saussurian words) or to a subset of a linguistic community; instead, the term *rhythm variability* will be used to refer to individual fluctuations which pertain to the *parole*, which are given by the context or by other extra-linguistic factors or even by chance.

After a brief summary of the most important authors who worked on rhythm variation (in 5.2) and rhythm variability (in 5.3), I shall deal with specific aspects of rhythm variation and variability by presenting some data and analysing the most relevant results. Namely, I shall touch upon the variability of rhythm metrics when computed by different phoneticians (in 5.4). Then, I shall analyse intra-speaker (in 5.5.1) and inter-speaker (in 5.5.2) variability of rhythm metrics, finally discussing variation between speakers of regional or dialectal varieties (5.5.3). Finally, I shall discuss the possibility of distinguishing between different degrees of rhythm variability and rhythm variation through the use of rhythm correlates.

## 5.2 The study of rhythm variation

Rhythm variation across different dialectal or regional varieties has been firmly present in the literature along the evolution of rhythm speech theories. Many authors have tried to capture rhythm differences between different dialectal (or, in some cases, regional<sup>80</sup>) varieties attempting to categorise them as stress-timed or syllable-timed. The raising of rhythm metrics has of course given a new impulse to this type of research. I shall now briefly review (with no pretension of exhaustiveness) some

---

<sup>79</sup>In particular, William Labov has devoted his career to the study of variation (see for instance Labov, 1972) and is now regarded as the founder of variationist sociolinguistics. In Italy, this discipline has also been influenced by Coseriu's ideas (e.g. 1958) and is most authoritatively represented by Sornicola (e.g. 1982). However, the bibliography on these themes is endless and a discussion on this topic is well beyond the scope of this thesis.

<sup>80</sup>The distinction between dialectal and regional varieties has been drawn from Italian dialectology (between *varietà regionali* vs. *varietà dialettali*). Dialectal varieties of Italy (which notoriously do not correspond to English *dialects*) are the language varieties which developed from vulgar Latin; they form a continuum in Italy, which is usually broken into smaller areas according to isoglosses by scholars attempting to provide a classification. Dialectal varieties differ one another (and from Standard Italian, which, actually, could itself be considered as a dialectal variety) greatly and on all levels - phonological, morphological, syntactic and phonetico-prosodic. Regional varieties of Italian are different versions of Italian spoken locally and which often reflect at least some characteristics of corresponding dialectal varieties. Regional varieties, as well, can differ one another on all levels, but variation is on a smaller scale, phonetico-prosodic differences being perhaps the most perceptively salient. The bibliography on these subjects is simply huge: for sake of brevity, I shall only mention Maiden & Parry (1997) and Loporcaro (2009), both of which contain many further references.

## 5. Rhythm variation and variability

of the authors who attempted a rhythm categorisation of dialectal or regional varieties.

Most of the works that studied rhythm variation across different dialectal (or, less often, regional) varieties are fairly recent and therefore use recent approaches, such as the metrics. However, the rhythm variation of the dialects of Italy had already been the object of investigation by some scholars, mainly on the basis of vowel durations and of syllable structure. In particular, Mendicino & Romito (1991) as well as Romito & Trumper (1993) attempted a first rhythm classification of some dialects based on the duration of stressed vowels: the authors found several differences across dialectal areas (Venetian and Tuscan dialects being syllable-timed, Apulian dialects being stress-timed, Calabrian dialects being half-way), finally confirming the existence of a continuum and introducing the notion of “fuzzy poles”. Mayerthaler (1996) proposes a different classification based on diachronic processes (which alternatively caused a simplification or a complexification of the syllable structure of dialects): from the dialects of extreme Southern Italy (characterised by a simple syllable structure), through two intermediate areas (central and Southern dialects), to the dialects of Northern Italy (mainly of Piedmont and Romagna), which are characterised by a more complex syllable structure.

More recently, Schmid (2004) conducted an experiment which combines a study of the syllabic structures of different dialectal varieties of Italy (Piedmontese, Milanese, Bitontino, Neapolitan, Venetian and Pisan) and their rhythm properties. Firstly, he checked which and how many different syllable-types each of these varieties allows for (the breadth of the syllabic inventory is in fact one of those phonological properties which are believed to play an important role in the classification of a language as more or less syllable-timed or stress-timed) and, subsequently, ordered them as follows: Pisan (18 different possible syllable-types), Neapolitan (21), Venetian (24), Bitontino (26), Milanese (28), Piedmontese (35). He then calculated the three rhythmic correlates proposed by Ramus *et al.* (1999) and put the results on charts. As expected, Pisan, Neapolitan and Venetian occupy a position of syllable-timing, while Bitontino, Milanese and Piedmontese occupy a position which could easily be associated with stress-timing, thus confirming the relation between the breadth of the syllabic inventory and Ramus’ parameters.

Deterding (2001) investigated rhythm differences between Singapore and British English on the spur of previous studies by Low and co-workers (also see chapter 3) by applying a variability index to syllable durations. Despite the difficulties encountered in establishing clear criteria as to how to measure syllable durations, results (lower values of syllable variability) confirm Low’s previous findings (obtained on vocalic intervals) that Singaporean English shows a tendency towards syllable-timing, at least if compared to British English.

Ghazali, Hamdi & Barkat (2002) studied rhythm variation in 6 Arabic dialects: Moroccan, Algerian, Tunisian (Western dialects), Jordanian, Syrian and Egyptian (Eastern dialects). The three authors computed %V and  $\Delta C$  on versions of *The North Wind and the Sun* and found that Western dialects resulted in lower values of %V and higher values of  $\Delta C$ . The authors consider this to be in compliance with the general impression that Western Arabic dialects tend to have a stronger-tendency towards stress-timing than Eastern dialects, which has been reported by previous studies on the basis of perceptive experiments (e.g. Barkat 2000, quoted by Ghazali, Hamdi & Barkat, 2002).



## 5. Rhythm variation and variability

Mok & Dellwo (2008) computed a variety of rhythm measures (%V,  $\Delta V$ ,  $\Delta C$ ,  $\Delta S$ <sup>81</sup>, VarcoV, VarcoC, VarcoS, rPVI\_C, rPVI\_S, nPVI\_V, nPVI\_S) on Cantonese, Beijing Mandarin, Cantonese English and Mandarin English. Speakers read *The North Wind and the Sun* in Cantonese or Mandarin, then retold it without having access to the script (semi-spontaneous speech) and finally read the English version of the story. Cantonese and Beijing Mandarin are found to be syllable-timed by all rhythm measures despite a high degree of variation of %V values for the two different speaking styles (reading and retelling the story). Results are less clear for Cantonese English and Mandarin English, which are perceptively syllable-timed, but whose categorisation varies according to different rhythm measures. The authors state that these results “pose a challenge to the acoustic measures” (Mok & Dellwo, 2008:4/4).

Mairano & Romano (2008b) calculated the deltas, the varcos, the PVI and the CCI on strictly controlled comparable dialectal samples from linguistic areas of Italy and Romania (data came from fieldwork and belongs to the AMPER database). It was found that dialectal varieties (even of the same linguistic area) could present notably different results and that the categorisation was slightly different using different rhythm metrics.

O'Rourke (2008) computed the deltas and the PVI for 3 groups of Peruvian Spanish (3 native speakers from Lima, 3 native speakers from Cuzco, 3 native bilingual speakers of Spanish and Quechua - 9 speakers in total) and compared the results with data from other languages published by White and Mattys (2008). Despite the author's claims, results look controversial as VarcoC shows a very different scenario for Peruvian Spanish data from that offered by VarcoV, rPVI and nPVI. Furthermore, it seems hazardous to combine results from different studies in one chart because of potential differences in the segmentation or in the treatment of data. However, sticking to Peruvian Spanish, the author remarks a statistically significant difference between Lima and Cuzco speakers, while (quite unsurprisingly) there does not seem to be any significant difference between monolingual and bilingual speakers from Cuzco.

Romano, Mairano & Pollifrone (2009) calculated several rhythm measures on 6 dialectal varieties of Piedmont (speakers translated *The North Wind and the Sun* in their dialect and re-read it aloud). Final results showed great differences among the 6 samples, but confirmed an overall tendency of Piedmontese dialects towards stress-timing/compensation.

White, Payne & Mattys (2009) calculated various metrics on samples of regional Italian (Venetan and Sicilian): they chose, on the basis of their previous studies, the varcoV/%V chart as the most representative and found that samples of both regional varieties clustered in the syllable-timed area. The authors claimed that such results are “perhaps” surprising as “Southern Italian (e.g. Sicilian) has been frequently described as more ‘stress timed’ than northern Italian (e.g. Venetan)” (2009:151). However, I believe that some considerations are necessary here. First of all, I have the impression that the authors do not clearly distinguish between *regional* and *dialectal* varieties: in fact, since they did not give any specification, one may well wonder who made such claims; it is, in effect, the case that much has been said about rhythm variation in Italy, but it concerned dialectal variations (see below) and I am not aware of many reports on the rhythm of regional varieties of Italian

---

<sup>81</sup>S refers to phonological syllables.

## 5. Rhythm variation and variability

(apart from Giordano & D'Anna, who published their contribution in 2010). Furthermore, if the authors referred to dialectal varieties, only some of the northern varieties are classified as syllable-timed (Venetan among them, but certainly not Piedmontese, for example) and only some southern varieties are classified as stress-timed (not Sicilian, at least according to Mayertaler, 1996).

Romano & Mairano (2010) attempted to provide a rhythm categorisation of 6 regional varieties of Romanian by testing various metrics on speech samples of *The North Wind and the Sun*. Differences were remarkable and went from syllable-timing (Brasov, Bucovina, Moldavia and Muntenia) to mild stress-timing (Bucharest and Oltenia).

Giordano & D'Anna (2010) computed %V, the deltas, the varcos and the PVIs on samples by 34 speakers of 15 regional varieties of Italian (Bari, Bergamo, Cagliari, Catanzaro, Florence, Genua, Lecce, Milan, Naples, Palermo, Parma, Perugia, Rome, Turin) and on 3 speech styles (pre-planned monologic speech, spontaneous dialogic speech and read speech) using data from the CLIPS corpus. The  $\Delta C$  value is found to increase from read through dialogic and to pre-planned speech, while, conversely,  $\Delta V$  is found to decrease; consonantal rPVI values are found to reproduce  $\Delta C$ 's behaviour, while, once again, nPVI is found to be stable. The variation of rhythm metrics across the 15 regional Italian varieties was calculated on read passages only: results for different Italian varieties vary greatly from each other, ranging from delta and PVI values associated with syllable-timing to values usually associated to stress-timing.

In conclusion, most studies in the field show that dialectal varieties can differ one another in terms of rhythm properties as much as different languages. Yet, there seems to be less agreement on the degree of rhythm variability given by different regional varieties of the same language. This could be motivated by the fact that regional varieties do not usually show great differences in phonotactics and syllable structure, which therefore do not have remarkable repercussions on rhythm properties. However, testing these different types of rhythm variation provides interesting perspectives and will be dealt with in 5.5.

### 5.3 The study of rhythm variability

The study of rhythm variability has been far more neglected than the study of rhythm variation, apart of course from the fluctuations caused by speech rate. The fact that the effects of speech rate on speech rhythm have been studied extensively is of course perfectly understandable as they are related phenomena. Yet, other aspects of rhythm variability have often been shunned and, indeed, in many cases, treated as spurious data. However, before the "rhythm metrics era", Major (1981, already reported in chapter 2) found rhythm differences across different speech styles for Brazilian Portuguese studying durational properties and some phonological characteristics of this language variety in citation and casual speech: formal Portuguese (represented by his citation data) seems to show the properties typical of syllable-timed languages, whereas informal Portuguese (represented by his casual speech data) seems to show the properties typical of stress-timed languages (most emblematically the shortening or even the deletion of unstressed syllables). On the basis of these observations, he puts forward a singular hypothesis: since historical change in languages generally occurs in the direction of casual speech, "Portuguese

## 5. Rhythm variation and variability

is in the process of changing from a syllable-timed language to a stress-timed language” (1982:350).

Again, the rhythm metrics have offered a framework against which it is possible to test rhythm variability. This certainly explains the flourishing of studies on speech rhythm in relation to speech rate, which are not reported here as they have already been reviewed in chapter 3. Among the few who set out to analyse rhythm variability other than in relation to speech rate, Mairano & Romano (2007a) tested inter-subject variability on the values of the metrics proposed by Ramus *et al.* (1999) on 7 speakers of 4 languages (English, French, German and Italian). The segmentation and the measurements were carried out by the two authors independently and it was found that the values were fairly stable across different segmentators in spite of remarkable differences in the segmentation and in classification choices of phonologically ambiguous segments. General results were in compliance with expectations, with high values of  $\Delta C$  and  $\Delta V$  for German and English, and high values of %V for French and Italian. The authors also presented measures of inter-operator agreement rate.

A similar work appeared in a poster by White *et al.* (unpublished) presented at EASR 2008, in which the authors had independently segmented data from a few languages and presented similar measures of inter-operator agreement rate.

Widget *et al.* (2010) tested the robustness of several rhythm metrics (though only focusing on %V, varcoV and vocalic nPVI) in relation to various factors of variation, namely across 5 different measurers, 6 different speakers (of Standard Southern British English) and 5 different sentences. Results show that fluctuations due to different speakers and different measurers are smaller than those caused by different sentences. They conclude their study by giving pieces of advice to researchers intending to use rhythm metrics.

Yoon (2010) computed the varcos and the PVIs on conversational speech by ten American speakers from Columbus in order to check intra- and inter-speaker variability. Several minutes per speaker were taken in consideration as data was drawn from an annotated corpus just needing a script to convert SAMPA into CV. The author finds higher intra- and inter-speaker variability with the varcos (varcoV, in particular) than with the PVIs and more with consonantal rPVI than with vocalic nPVI; he therefore observes that data from the ten speakers tend to cluster tightly into a vocalic/consonantal nPVI chart thus minimising inter-speaker variability. He claims that “nPVI-V and nPVI-C both make a very compact cloud, suggesting that the normalised variability indices are the best rhythmic metrics that capture the speaker’s dialect similarity in this study” (2010:4/4). However, the author does not seem to be aware that the rPVI and the nPVI are not directly comparable on the same scale. Furthermore, measures that are able to capture similarity are not necessarily able to capture difference: since the author did not include any other language or dialect, it seems impossible to determine whether the combination of consonantal and vocalic nPVI is effectively able to “capture similarities” between dialects or whether it simply levels differences of any kind.

### 5.4 Inter-operator variability of rhythm metrics

As has been said, nearly all data in the corpus presented in chapter 2 and 3 have been segmented and labelled separately by two phoneticians (PM and AR). This has of course been done in an effort to make measurements less bound to subjective

## 5. Rhythm variation and variability

evaluations and therefore more reliable. In effect, apart from discrepancies in the boundaries set for each segment by the two phoneticians<sup>82</sup>, many differences can also be found in phonological choices as to, for instance, whether certain segments exist or not, or about the vocalic or consonantal status of some sounds. This is because segmentation, to a certain extent, is a phonetic transcription.

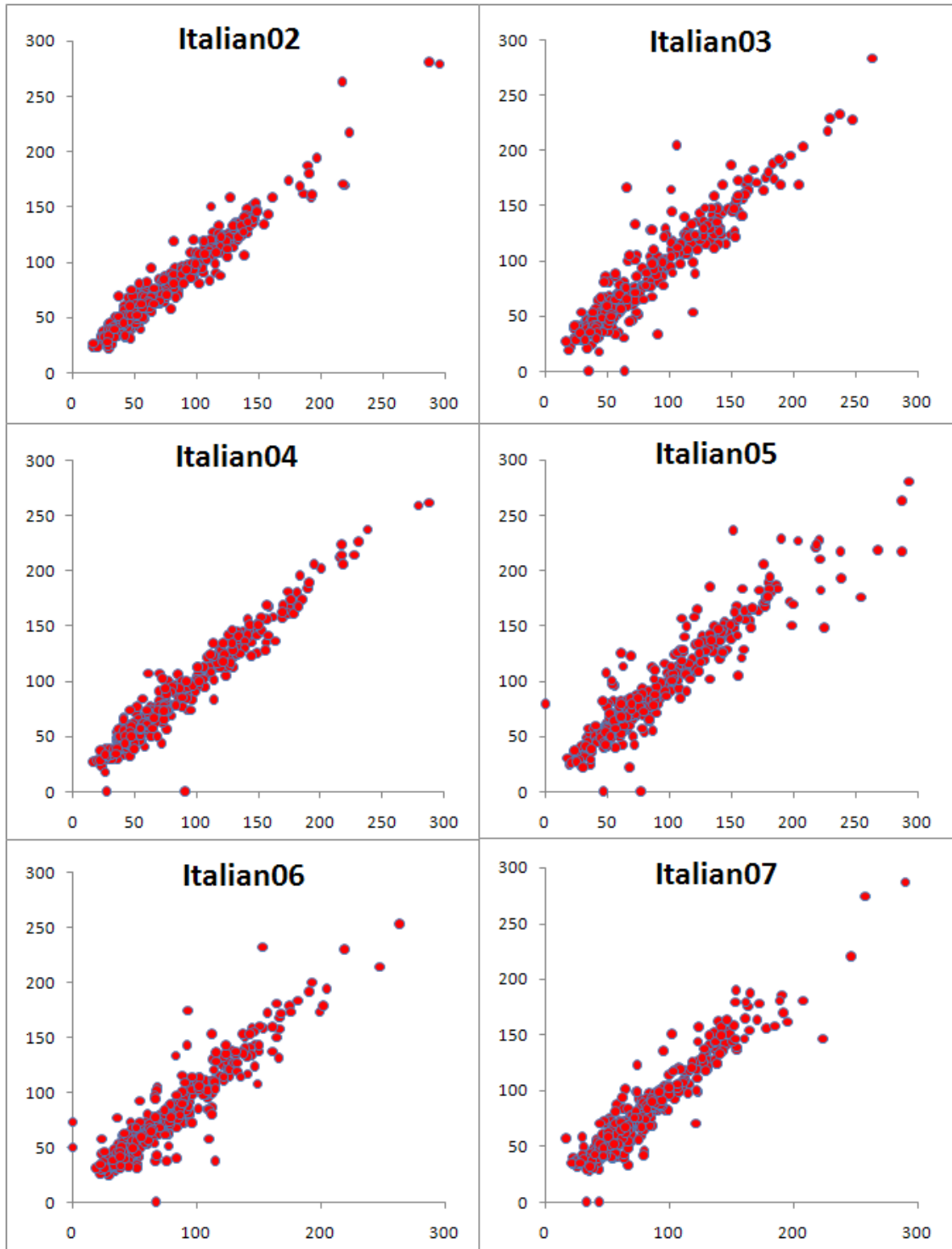
Figure 5.1 represents segmentation choices by the two phoneticians (AR on the x axis, PM on the y axis) for the Italian samples 2-7. The values on both axes represent the duration in ms of each vocalic or consonantal interval calculated on the boundaries set respectively by AR and PM. As it can be seen, measurements align more or less along the bisecting line, sometimes with more remarkable deviations. In fact, the more the values on the chart tighten along the bisector, the higher is the correlation between the two segmentations; in a purely hypothetical condition, two perfectly corresponding segmentations would draw a bisecting line. It also has to be noted that dots resting on the x or on the y axis indicate that one of the two phoneticians considered the corresponding segment as non-existent (hence 0 ms), while the other labelled it (e.g. in cases of epenthetic schwas).

However, what is relevant for this study is the influence given by differences in the segmentation/labelling on the final values of the metrics. In order to get an idea of this, the values of the different metrics were also computed on the basis of the segmentation carried out by PM and AR separately: the results can be seen in figure 5.2, where the final values for the PVI (above) and the CCI (below) for each sample are shown separately for AR and PM. Only the samples segmented by both PM and AR were included in the chart, which is already overloaded and difficult to read. Variability exists but is not impressive: PVI variability goes from the very low scores of Romanian Muntenian, Finnish2, Estonian, Australian English and Lebanese Arabic, to higher scores for Czech, Caracas and Bogota Spanish, Bucharest Romanian and Italian07. CCI variability goes from the very low scores of Lima Spanish (the two samples are practically superimposed), Italian02, Sao Paulo Portuguese, Brasov and Muntenian Romanian, to higher scores of the 2 Finnish samples and of Moldavian Romanian.

---

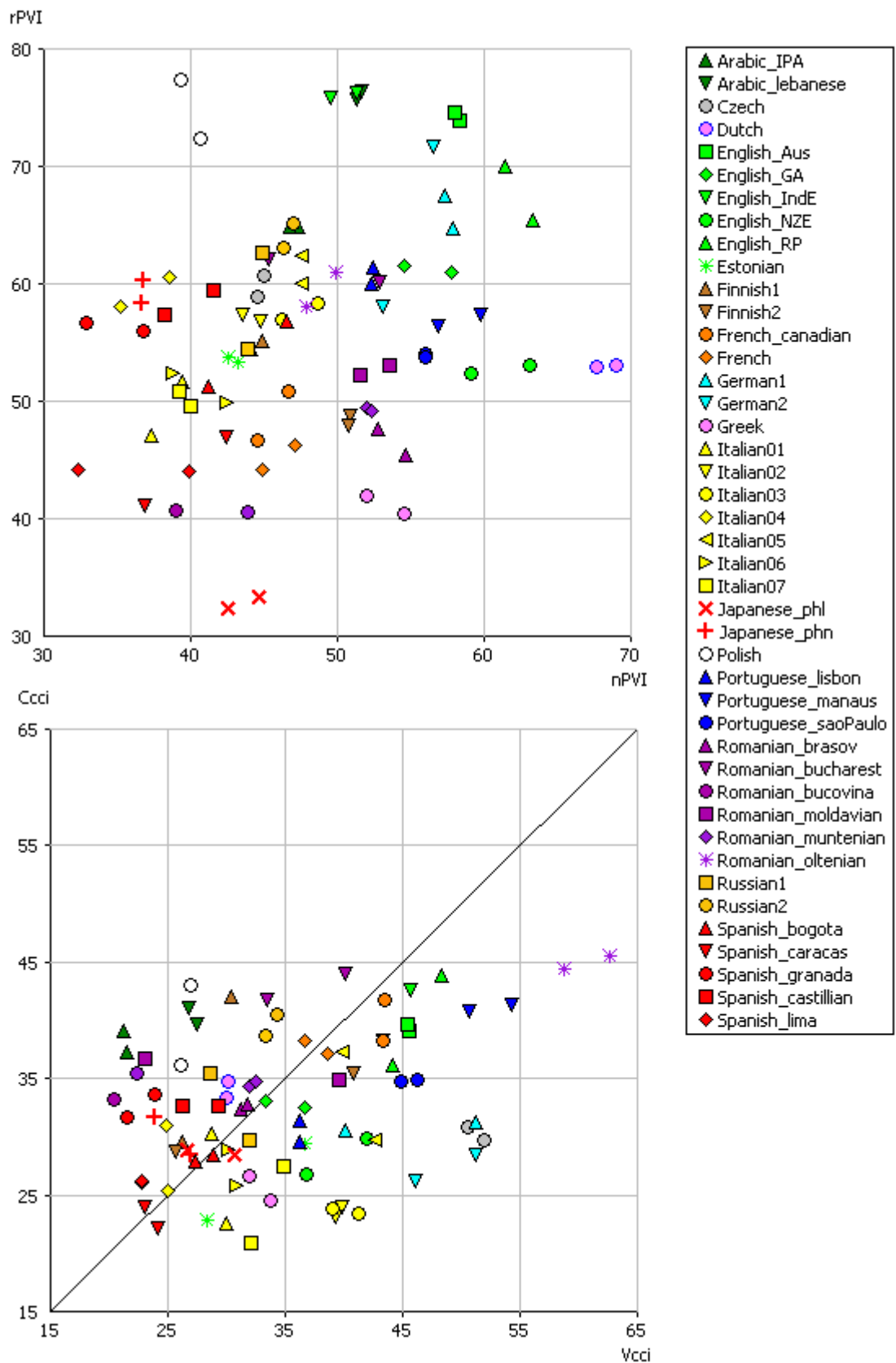
<sup>82</sup>Such discrepancies are of course inevitable in manual segmentation and they would likewise exist even across two segmentations of the same sample done by the same person.

5. Rhythm variation and variability



**Figure 5.1.** Durations in ms of consonantal and vocalic intervals measured on the boundaries set by AR (x axis) and PM (y axis). Perfectly matching segmentations would result in a bisecting line.

## 5. Rhythm variation and variability



**Figure 5.2.** Values of the PVI and the CCI for samples in the corpus. Each sample is represented twice, reflecting the segmentation by the AR and PM.

## 5. Rhythm variation and variability

It is also remarkable that some samples present very limited variability with one metric and high variability with other metrics: it is the case of Lima Spanish, which has nearly the same values of vocalic and consonantal CCI as well as rPVI for PM and AR, but which has very different values of nPVI. This suggests that inter-operator variability is not directly proportional to inter-subject agreement. That is to say that a speech sample might be segmented very differently by two phoneticians and still yield similar results or, *vice versa*, the same sample segmented very similarly by two phoneticians might yield considerably different results of the metrics. Moreover, it is also possible that some metrics are more stable than others in respect to segmentation differences: in order to check that, for each rhythm metric, I calculated the mean of the standard deviation<sup>83</sup> between the values obtained on PM and AR *TextGrids* (see appendix 2 for all single values of inter-operator standard deviation); results are shown in Table 5.1 and show that the CCI seems to be more sensitive to this parameter. This is in effect understandable as it demands a phonological interpretation of each segment, instead of a (comparatively) simpler segmentation into vocalic and consonantal intervals *tout court*. It has also to be noted that the varcos and the nPVI are normalised values and therefore represent completely different entities from the other metrics and cannot be directly compared to them. However, it has to be noted that a higher sensitivity is in itself not necessarily a drawback: as suggested by Bertinetto & Bertini (2008 and following), normalising might mean losing some relevant information. So, high sensitivity to segmentation choices might also mean high sensitivity to other rhythm-related phenomena. Normalisation brings of course more stability but the issue of whether to normalise or not is probably solved case by case depending on data and according to the aim of the study: for a cross-language rhythm categorisation it is probably best to normalise vocalic durations even if this has a risk in terms of a possible loss of relevant information. Instead, for a sophisticated study on specific rhythm properties of a group of speakers or within different styles in the same language, it is probably better not to normalise in order to be able to capture all possible nuances (keeping in mind that, at the state of the art, it is difficult to confidently attribute changes in the values of rhythm metrics to specific phenomena).

%V	Vdev	Cdev	varcoV	varcoC	nPVI	rPVI	CCI(v)	CCI(c)
1,33	1,36	1,21	1,33	0,90	1,12	1,13	3,55	1,47

**Table 5.1.** Inter-operator variability expressed as the mean of the standard deviations between the values obtained on PM and AR segmentation for each rhythm metric.

<sup>83</sup>It was chosen to calculate the mean of the standard deviations for each sample and not the standard deviation *tout court*. This is because a global standard deviation would include the differences between each possible pair of values, which does not make sense in this case: instead, I computed exclusively the differences between each pair of values obtained on the same sample by PM and AR. Therefore, the difference between the  $\Delta C$  value as calculated on, say, Estonian by PM and on Finnish by AR was not considered (while it would be taken in consideration by calculating the overall standard deviation for each metric). This implies that the standard deviation was calculated at each time only on two values and thus it would have been possible to use simply the difference; however, it was chosen to use the standard deviation for comparative reasons.

## 5.5 Speakers and rhythm variability

### 5.5.1 Intra-speaker variability

Apart from the studies on rhythm and speech rate, the other aspects of intra-speaker variability have been completely neglected by the literature and, as far as I am aware, none of them has been extensively treated in any study making use of the metrics. A partial exception consists in some studies that considered productions by bilingual subjects or L2 learners in more than one language (such as the already quoted studies by Mok and Dellwo, 2008, and White & Mattys, 2007). Computing rhythm metrics on multilingual speakers in order to study their rhythm characteristics and, perhaps, to be able to evaluate their level of adaptation to a certain model is no doubt an interesting perspective (see below). However, it by no means constitutes the only aspect of intra-speaker variability: in fact, differences between styles and registers also pertain to this domain. Moreover, it is possible to check variability within equivalent productions by the same speaker in order to uncover pure intra-speaker variability at comparable speech rate, style, register and context. I believe there are good reasons for testing this type of “pure” variability. Most importantly, it has to be done as a first step in order to have a term of comparison in order to evaluate other types of variation and variability. In other words, it is not possible rate intra-speaker variability if we do not rate inter-speaker variability first.

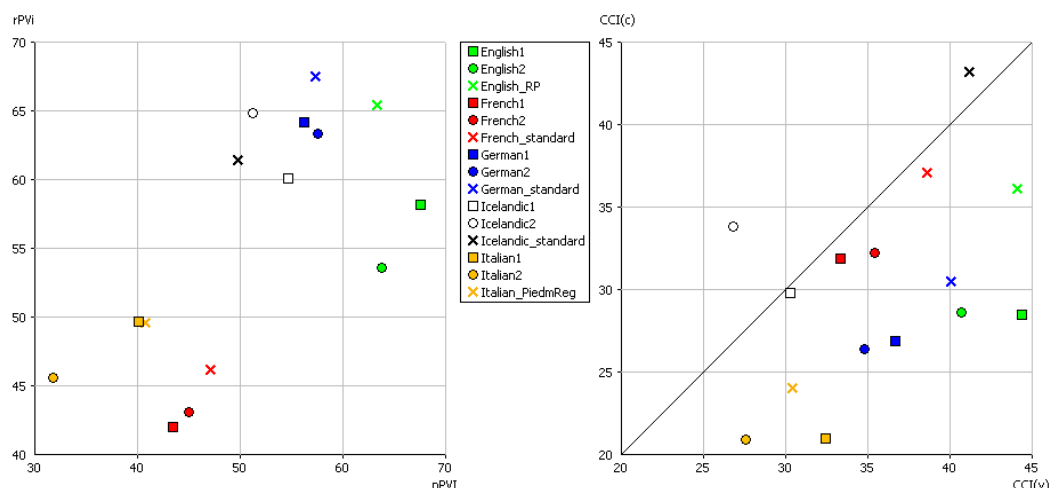
I decided to set up data in order to test at the same time “pure” intra-speaker variability as well as “multilingual” variability. I recorded one single speaker in five different languages, segmented his productions<sup>84</sup> and, as usual, computed the metrics with *Correlatore*. The subject is a male speaker in his twenties with university education, a native speaker of Italian, proficient L2 speaker of English and French, also possessing a fairly good competence of German and a limited competence of Icelandic. He was recorded while reading versions of *The North Wind and the Sun* twice in each language (therefore a total of 10 productions). Results are shown in figure 5.3 for the PVIs and the CCI. Samples of native speakers for each language have been included for comparison: RP English (because the speaker model is definitely British English), Standard French, Standard German and the mean of the 10 Icelandic speakers as well as the mean of the other 3 Piedmontese regional Italian speakers taken from the corpus.

---

<sup>84</sup>The segmentation and labelling of his productions were carried out by PM only as they are not part of the corpus (since they are mostly L2 productions). To be precise, the first Italian production of the speaker is actually part of the corpus (Italian06) and has therefore also been labelled by AR. However, for a better comparability with the rest of his productions, only the segmentation by PM was kept in consideration for the present study on intra-speaker variability.



## 5. Rhythm variation and variability

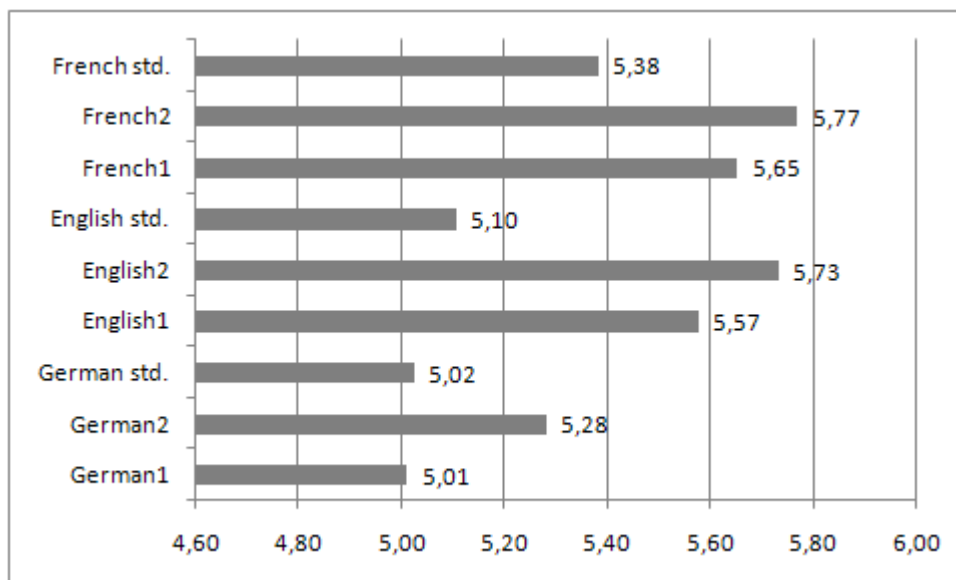


**Figure 5.3.** Values of the PVIs (above) and the CCI (below) for 10 productions of Italian, English, French, German and Icelandic of one (native Italian) speaker. Values of an RP speaker, a standard French speaker, a standard German speaker and the mean of 10 Icelandic speakers and 3 Piedmontese regional Italian speakers (all taken from the corpus presented in chapter 2 and 3) have been added for comparison.

It can be seen that the rhythm categorisation suggested by the PVIs is very coherent within the values of this speaker, although variability within each pair of samples of the same language is remarkable. Productions in French show low PVI values and are disposed along the bisecting line in the CCI chart, whereas productions in English and German exhibit high values of the PVIs and are disposed below the bisecting line in the CCI chart. This seems to confirm that this speaker is fairly proficient in these languages and that he masters control and compensation phenomena. As for his productions of (Piedmontese regional) Italian, they come to be categorised as syllable-timed by the PVIs and as more or less compensating by the CCI, while the opposite happens to his Icelandic productions, which come to be categorised as stress-timed by the PVI and as controlling by the CCI. As for this variety of regional Italian, it is difficult to comment as Italian should be a syllable-timed/controlling language, but the Piedmontese dialect is usually categorised as stress-timed/compensating (see for instance Schmid, 2004 and Romano *et al.*, 2010): it is therefore possible that the CCI value reflects the influence of the Piedmontese dialect on the speech rhythm of the speaker (regional varieties of Italian will be treated in more detail below). As for Icelandic, as well, the situation is complex; first of all, this language should perhaps be regarded as a mixed language, allowing for fairly complex consonantal clusters without having macroscopic phenomena of vowel reduction. It is thus expected to yield low vocalic nPVI and high consonantal rPVI values. Given these features, then, it is perfectly understandable that, in contrast to compensating languages, it falls around the bisecting line in the CCI chart: vocalic CCI is not expected to be low enough to place it below the bisector because the difference between stressed and unstressed vowels is not supposed to be as remarkable as, for instance, in English or German. Also, a further complexity is given by the low competence of this language on the part of the speaker: in this case, it is of course well possible that he does not master all segmental compensation phenomena of Icelandic.

## 5. Rhythm variation and variability

It is also interesting to observe that values for all metrics are slightly lower for this speaker than for the corresponding native speakers. This probably does not have to be attributed to differences in segmental control, but rather to a slightly faster speech rate, as is demonstrated in figure 5.4:



**Figure 5.4.** Speech rate for samples of English, French and German by the L2 speaker and native speakers from the corpus.

Speech rate has been calculated by dividing the total duration (excluding pauses, of course) by the number of vocalic intervals. This procedure might seem arbitrary but is grounded on the fact that each vocalic interval labelled corresponds to a syllable nucleus<sup>85</sup> and therefore gives an indication of the number of syllables. It also has to be noticed that English and German present lower values of speech rate than French, which is compliance with what suggested by Dellwo (2008).

As for “pure” intra-speaker variability, the difference between each pair of productions in the same language is perhaps greater than expected. Table 5.2 shows the mean of the standard deviations between the values obtained for each pair of samples by the speaker. Indeed, the mean values of variability are comparable to the ones obtained for inter-operator variability (given by different phoneticians working on the segmentation of the same sample) apart from the CCI.

Finally, it is interesting to remark that this test provides a sound and original confirmation of the validity of these metrics: the scenario offered is consistent with expectations and, since the 10 samples are perfectly comparable as for speech style, register, speech rate and context and since they all pertain to only one speaker, they cannot be attributed to idiosyncrasies of different speakers representing each language. It seems therefore natural to conclude that the metrics can provide a reliable representation of the segmental properties related to rhythm in controlled productions of this type.

<sup>85</sup> In fact, hiati were labelled as two separate intervals, as suggested by Bertini & Bertinetto (2009).

## 5. Rhythm variation and variability

	%V	Vdev	Cdev	varcoV	varcoC	nPVI	rPVI	CCI(v)	CCI(c)
Engl.	0,86	1,10	3,16	0,49	3,11	1,87	2,27	1,86	0,04
French	0,35	0,11	0,66	0,59	0,72	0,81	0,54	1,06	1,15
Germ.	0,51	2,04	5,40	1,12	4,43	0,64	0,44	0,96	0,24
Icel.	0,61	2,09	0,21	2,28	0,93	1,74	2,33	1,76	2,01
Italian	1,21	3,19	1,25	4,37	0,62	4,14	2,05	2,39	0,08
mean	0,71	1,71	2,14	1,77	1,96	1,84	1,53	1,61	0,70

**Table 5.2.** Intra-speaker variability expressed as the mean of the standard deviations between the values for each pair of productions in the same language for this speaker.

### 5.5.2 Inter-speaker variability

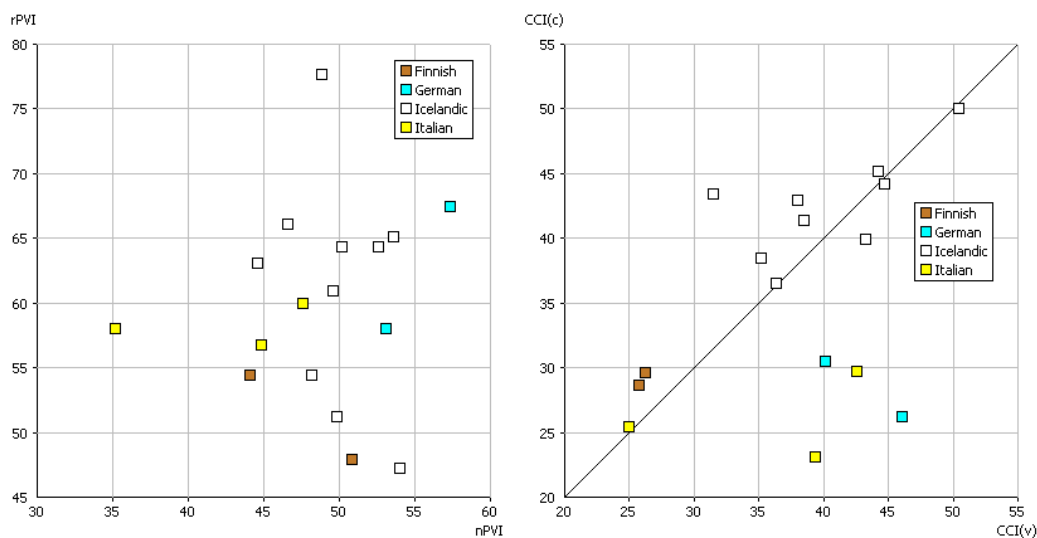
Inter-speaker variability refers of course to variability within different speakers of the same language or linguistic variety. A number of authors have already proved that the metrics yield different results for such data. So, I decided to test inter-speaker variability with the data available in my corpus and compare it with intra-speaker and inter-operator variability. Unfortunately, the only exploitable samples from the corpus are the 10 Icelandic speakers, 3 (out of 15) Italian speakers, the 2 German and the 2 Finnish ones: speakers for the other national languages in the corpus, in fact, present some dialectal/regional differentiation (see below) and thus cannot be considered as representative of simple inter-speaker variability. All 10 Icelandic speakers are from Reykjavík and claimed to have no dialectal/regional accent<sup>86</sup>; the 4 Italian speakers are number 6, 7, 8 and 15 and all live in Turin and speak Piedmontese regional Italian; the Finnish and German speakers all speak the standard variety of their language. Since the ten Icelandic speakers as well as Italian14 and Italian15 have only been segmented and labelled by PM, the values obtained by AR on Italian06, Italian07 and on the Finnish and German samples were not considered for the present analysis. Results for the PVIs and the CCI are shown in figure 5.5.

As can be seen, PVI variability is definitely high, above all for the ten Icelandic speakers and for consonantal intervals more than for vocalic intervals (which is easily explained as an effect of normalisation<sup>87</sup>). The scenario offered by the CCI seems to be more stable: variability is lower - in effect the 10 Icelandic speakers (despite showing a fairly high variability) do not mix with those of other languages. In both charts there is some overlapping between speakers of different languages.

<sup>86</sup>The geolinguistic differentiation within Iceland is, however, very limited (though it does exist, in contrast to what is usually claimed).

<sup>87</sup>In fact, the  $\Delta C/\Delta V$  chart (in which normalisation is neither applied at consonantal level nor at vocalic level) shows that variability is equally important on both axes.

## 5. Rhythm variation and variability



**Figure 5.5.** Values of the PVIs (above) and the CCI (below) for the 10 Icelandic speakers, 3 standard Italian speakers, 2 German and 2 Finnish speakers.

Similarly to what has been done with intra-speaker variability, inter-speaker variability has been evaluated by computing the mean of the standard deviation among samples of the same language. Results are shown in table 5.3.

	%V	Vdev	Cdev	varcoV	varcoC	nPVI	rPVI	CCI(v)	CCI(c)
Finnish	0,05	2,09	1,76	4,64	0,43	3,34	3,23	0,27	0,48
Germ.	1,31	4,03	2,86	3,48	0,72	2,06	4,75	2,99	2,17
Icel.	0,67	1,91	2,62	0,80	1,24	0,95	2,74	2,02	1,41
Italian	0,92	5,10	1,64	3,62	0,70	3,76	0,93	5,40	1,95
Mean	0,74	3,28	2,22	3,14	0,77	2,53	2,91	2,67	1,50

**Table 5.3.** Inter-speaker variability expressed as the mean of the standard deviations between the values obtained for different speakers of the same language.

### 5.5.3 Geographical variability vs. variation

For the present study on rhythm metrics, geographical variation and variability refer to fluctuations in the results of rhythm metrics given by a different geographical provenance of speakers. It therefore includes dialectal and regional variation (see above for the difference between dialectal varieties and regional varieties).

In order to test dialectal variation, I have exploited data of 6 speakers of 6 different dialectal varieties of Piedmont, namely those of the following villages: Bagnolo Piemonte (CN), Briga Alta (CN), Campertogno (VC), Capanne di Marcarolo (AL), Exilles (TO), Roccaforte Ligure (AL). Data are perfectly comparable with the corpus as speakers read *The North Wind and the Sun* after

## 5. Rhythm variation and variability

translating the Italian version into their own dialectal variety. These 6 dialectal varieties belong to the same group of dialects, but each of them has specific characteristics at all linguistic levels: this means that the 6 texts are all different from one another (see Romano et al., 2010, for more details on the speakers and on these dialects and for an orthographical transcription of the texts).

In order to test regional variation, I exploited data from the corpus, namely 10 of the 15 Italian speakers (each speaking a different regional variety of Italian)<sup>88</sup>, the 5 English speakers (RP, American, Australian, New Zealand, Indian), the 3 Portuguese speakers (from Lisbon, Sao Paulo and Manaus), 6 Romanian speakers (from Braşov, Bucharest, Buchovina, Moldavia, Muntenia and Oltenia), 5 Spanish speakers (Castillian, Granada, Bogotá, Caracas and Lima). In contrast to dialectal varieties, the texts of different regional varieties of the same language were the same for all speakers: so, all Italian speakers read the same text (as well as all English, Portuguese, Romanian and Spanish speakers). Since not all 12 speakers had been labelled by AR, measurements by PM only were taken in consideration for consistency across all samples. The results for the PVIs and the CCI obtained on these data are shown in figure 5.6.

Observing the charts, it can be noticed that regional varieties show a moderately high variability, which is greater among dialectal varieties. In effect, despite remarkable differences, samples pertaining to different regional varieties of a language still cluster in the same area of the chart and there seems to be very limited overlapping between stress-timed/compensating languages and syllable-timed/controlling languages. On the other hand, the 6 samples of Piedmontese show a higher degree of fluctuation, particularly for the rPVI values, which range from typical stress-timed values for most samples to the very low values of Capanne di Marcarolo (which are even lower than those for Spanish varieties). Results for Piedmontese varieties seem to be more consistent using the CCI, suggesting that they all allow for a remarkable amount of compensation.

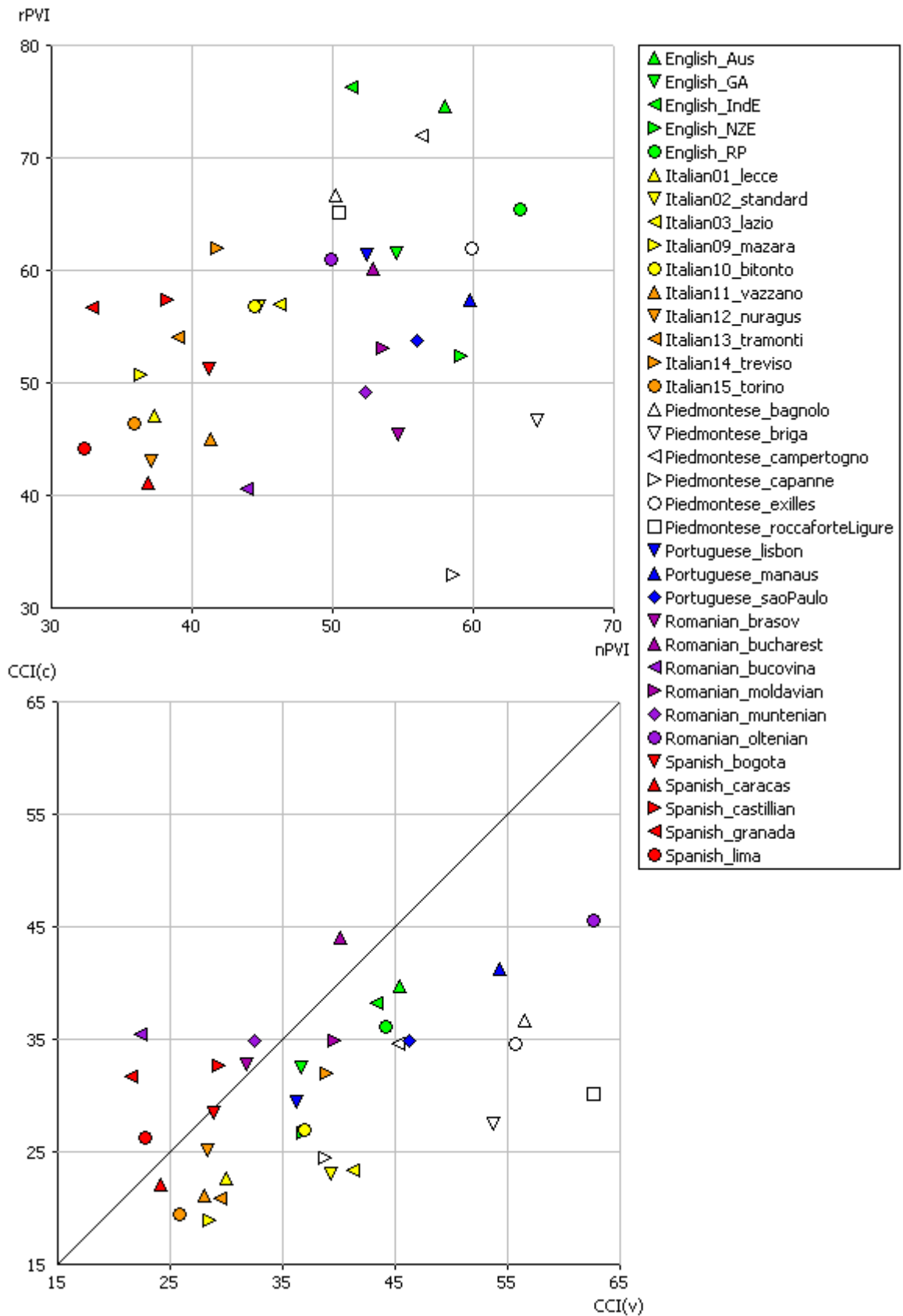
As in the previous cases, I computed the mean of the standard deviations among the values obtained for regional varieties of the same language (see table 5.4); separately, I also computed the standard deviation<sup>89</sup> of the values obtained for Piedmontese varieties (see table 5.5).

---

<sup>88</sup>The different provenance of the 15 Italian speakers are reported in chapter 2 in the presentation of the corpus. Italian speakers 6, 7 and 8 were excluded because they all spoke Piedmontese regional Italian, which is already represented by speaker 15. Speakers 4 and 5 were excluded because they spoke a standard variety, which is already represented by speaker 2.

<sup>89</sup>Obviously, as Piedmontese varieties are taken as the only representation of dialectal (vs. regional) variability, there is no need to compute the mean of the standard deviations (there is just one standard deviation).

### 5. Rhythm variation and variability



**Figure 5.6.** Values for the PVI and the CCI on dialectal varieties of Piedmontese and on regional varieties of Italian, English, Romanian, Portuguese and Spanish.

## 5. Rhythm variation and variability

	%	Vdev	Cdev	varcoV	varcoC	nPVI	rPVI	CCI(v)	CCI(c)
Engl.	0,31	2,37	2,84	3,08	0,70	2,06	4,38	1,86	2,33
Italian	0,44	1,87	1,69	1,05	1,14	1,21	2,01	1,82	1,24
Port.	2,11	7,16	3,28	2,64	2,53	2,13	2,19	5,21	3,41
Rom.	1,71	5,23	3,07	2,05	1,99	1,59	3,32	5,58	2,22
Span.	0,45	1,91	2,94	1,94	2,39	1,66	3,26	1,60	1,91
MEAN	1,00	3,71	2,76	2,15	1,75	1,73	3,03	3,21	2,22

**Table 5.4.** Inter-regional variation expressed as the mean of the standard deviations among different regional samples of the same language.

	%V	Vdev	Cdev	varcoV	varcoC	nPVI	rPVI	CCI(v)	CCI(c)
Piedm.	1,94	3,52	4,26	2,58	0,51	2,29	6,03	3,51	1,94

**Table 5.5.** Inter-dialectal variation expressed as the standard deviation among speakers of different dialectal varieties of Piedmont.

### 5.5.4 Discussion on rhythm variation and variability

After testing variability and variation across different axes and dimensions, it is natural that one may want to compare results. The most obvious approach is to try and rate the degree of variability given by the different factors. In other words, I shall try and establish a “scale of variability”, determining which factors condition a smaller or a greater degree of variability on the metrics. The hypothesis is that variability causes smaller fluctuations than variation and, more particularly, inter-operator and intra-speaker variability should be somehow smaller than inter-speaker variability, which in turn should be smaller than regional variation (which is simply a certain type of inter-speaker variability but which introduces one more factor – geographic provenance), which yet should be smaller than dialectal variation (thus reflecting what happens at other linguistic levels). In other words, I expect the following scale:

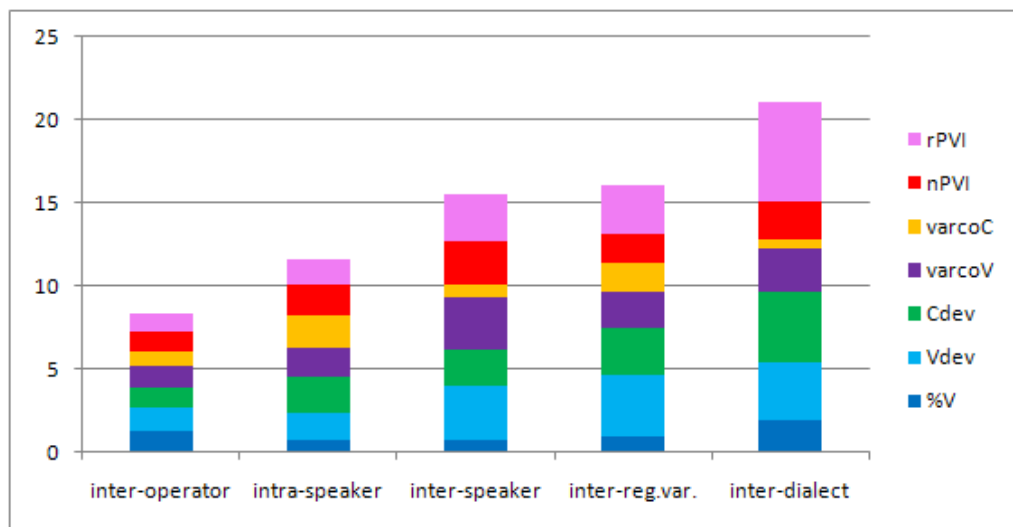
$$\textit{intra-speaker/inter-operator} < \textit{inter-speaker} < \textit{inter-regional variation} < \textit{inter-dialect}$$

This hypothesis can be easily verified by observing the values reported in tables 5.1-5.5 and resumed here in table 5.6 and in figures 5.7 and 5.8:

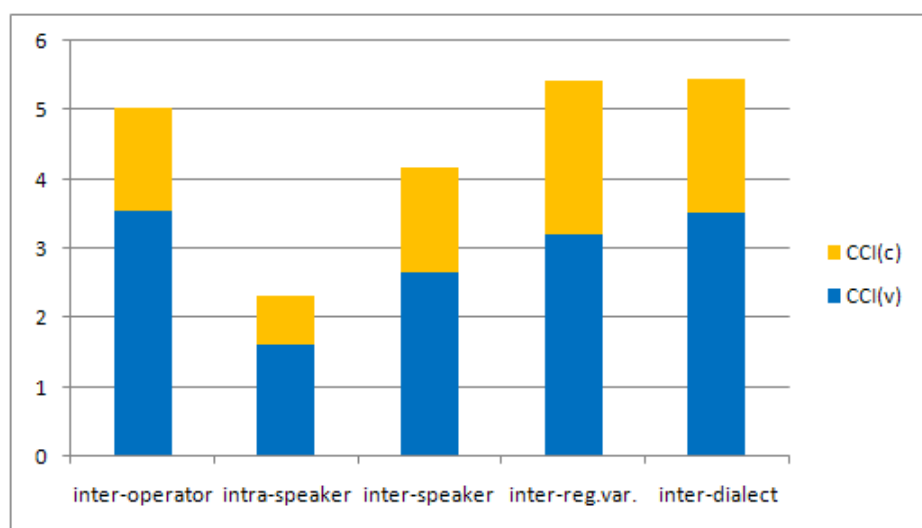
### 5. Rhythm variation and variability

	% V	Vdev	Cdev	varcoV	varcoC	nPVI	rPVI	CCI(v)	CCI(c)
inter-oper.	1,33	1,36	1,21	1,33	0,90	1,12	1,13	3,55	1,47
intra-spkr.	0,71	1,71	2,14	1,77	1,96	1,84	1,53	1,61	0,70
inter-spkr	0,74	3,28	2,22	3,14	0,77	2,53	2,91	2,67	1,50
inter-regvar.	1,00	3,71	2,76	2,15	1,75	1,73	3,03	3,21	2,22
inter-dialect	1,94	3,52	4,26	2,58	0,51	2,29	6,03	3,51	1,94

**Table 5.6.** A comparison of the variability of metrics in function of different parameters (inter-operator, intra-speaker, inter-speaker, inter-regional varieties, inter-dialectal varieties – in order of expected crescendo).



**Figure 5.7.** Fluctuations of %V,  $\Delta V$ ,  $\Delta C$ , varcoV, varcoC, rPVI and nPVI for the different types of variation and variability.



**Figure 5.8.** Fluctuations of vocalic and consonantal CCI for the different types of variation and variability.



## 5. Rhythm variation and variability

First of all, it can be observed that the CCIs seem to follow a different logic from the other metrics (for this reason they have been separated from the others in the charts). In fact, inter-operator variability is, as expected, the lowest type of variability for all metrics except for the CCI, for which it seems to be very high: this can be easily explained by the fact that, as already mentioned, the CCI requires not only a segmentation into vocalic and consonantal intervals, but also a phonological interpretation of each segment, which introduces a further element of subjectivity in the work and which is therefore likely to be the cause of more evident fluctuations in the results. As for the rest, both charts seem to agree on the following:

$$\textit{intra-speaker} < \textit{inter-speaker} < \textit{inter-dialect}$$

This was of course expected and is perfectly understandable: variability between one speaker is lower than variability among speakers of the same linguistic community which in turn is lower than variability among speakers of different (though related) communities. However, the status of inter-regional variability is not clear: fluctuations of the CCI seem to distinguish between inter-speaker variability and inter-regional variation, but not between inter-regional variation and inter-dialect variation; conversely, the other metrics do not seem to distinguish between inter-speaker variability and inter-regional variation, but they do distinguish between inter-regional and inter-dialect variation. In effect, as has been explained in chapter 3, the CCI aims to capture different phenomena from the other metrics, namely the degree of compensation allowed by a language. It is therefore plausible that the regional varieties considered in this study allow for very different degrees of compensation resulting in high fluctuations of the final CCI values (indeed so high that they are comparable to fluctuations for inter-dialect variation). On the other hand, the failure of the “classic” metrics in distinguishing between inter-speaker variability and inter-dialect variation would seem to suggest that rhythm variation across different regional varieties of the same language is perfectly comparable to differences given by the idiosyncrasies of different speakers.

At any rate, we should consider the following: the CCI tries to measure the degree of compensation, the other metrics attempt to measure the syllabic structure given by phonotactics. This could suggest that the degree of compensation of a language (reflected by the CCI) is more variable across different regional varieties and is not directly linked to its phonotactics (reflected by the other metrics), which is of course different across dialectal varieties. However, a word of caution is needed for the number of samples used to rate variability is small. Therefore, further investigation is needed to confirm these results.

### 5.6 Conclusion

Different types of rhythm variability and variation have been tested. On the whole, variability can be said to be high across several dimensions and I have tried to build a “variability scale”. The initial hypothesis resumed as:

$$\textit{intra-speaker/inter-operator} < \textit{inter-speaker} < \textit{inter-regional variation} < \textit{inter-dialect}$$

has only been confirmed for the following:

## *5. Rhythm variation and variability*

*intra-speaker < inter-speaker < inter-dialect*

However, results have provided a coherent paradigm for the study of speech rhythm and interesting perspectives for the future, which certainly include the necessity of enlarging the samples analysed.

**6.**

**A perceptive test**

## 6. A perceptive test

### 6.1 Introduction

#### 6.1.1 Why perceptive tests?

In the past chapters, it has been made clear that the alleged difference between stress-timed and syllable-timed languages was issued out of perception (actually, according to some authors, on perception and nothing else<sup>90</sup>). The traditional dichotomy has often enclosed perceptual evaluations even in its numerous denominations, such as *languages with machine-gun rhythm* vs. *languages with Morse code rhythm* (Lloyd James, 1940). Also the terms introduced by Pike (1945) and then adopted by the community reflect the impression that languages might sound as stress-timed or as syllable-timed. The existence of this impression was usually confirmed even by those authors who set out to look for isochrony and who did not find it, as is evident by the already quoted sentence by Roach: “a language is syllable-timed if it *sounds* syllable-timed” (1982:78).

So, given such a widespread consensus that syllable-timed languages sound syllable-timed and that stress-timed languages sound stress-timed, one would expect that a number of experimental tests have confirmed these claims. However, it seems that this is not the case: some perceptive tests have indeed been conducted on related issues (see for instance those by Allen, 1975, and Lehiste, 1977, reported below), but very few authors set out to verify precisely to what degree languages are perceived as belonging to different rhythm categories (see for instance Miller, 1984). Only recently, a number of authors have started to investigate this field with sophisticated test formats.

After testing the correlates on various languages, it has been difficult to give an evaluation of the results obtained since there is no proper theoretical framework against which one can compare results. Therefore, researchers are forced to comment on their charts by simply stating where each language is factually situated and where they “expected” it to be situated. Unfortunately, the expectations by researchers are usually based on impressions described by other researchers in previous studies (most typically Pike, 1945, and Abercrombie, 1967) or, sometimes, on impressions of their own; rarely are these “impressions” based on data (that is to say on perceptive tests). So, whenever there is a discrepancy between the researcher’s prediction and the actual results, it is not clear whether this has to be attributed to a malfunctioning of the metrics or, rather, to the fact that the language in question belongs to a different rhythm class from what had been inferred (which implies that the impressions were wrong). In short, it is not clear whether we should trust the metrics to be better indicators (or “correlates”, as they were initially called) of rhythm than impressions. For the same reason, it is risky to use metrics to give a rhythm assessment of previously unstudied languages. These characteristics indicate that the model is not entirely “predictive”, to use the words by Bertinetto & Bertini (2010).

A partial solution to this problem is to carry out perceptive tests and put them in relation with the values obtained with metrics. However, it has to be stated that the non-predictiveness is inherent in the metrics (deltas, varcos and PVIs) and perceptive tests do not make them any more predictive; it is nonetheless true that a

---

<sup>90</sup> See the “perception illusionists” reported by Bertinetto (1989).

## 6. A perceptive test

correlation might be found between the results of perceptive tests and the values of metrics: this would certainly imply that the metrics do give an acceptable account of the perception of language rhythm (if not an account of language rhythm tout court) and might thus improve their trustworthiness.

So, I decided to carry out perceptive tests with the initial aim of comparing results obtained with the various rhythm metrics and to check if there really was a correspondence between the two. After all, Ramus *et al.* intended the metrics as “correlates of the perception of rhythm” and specified that their study was “meant to be an implementation of the phonological account of rhythm perception” (1999:274). The authors themselves presented the results of a series of tests carried out on adults and infants on the discrimination of languages on the basis of rhythm. It seems therefore natural that a study on rhythm metrics that does not take perception in any account could be considered as wanting in some respects.

This chapter introduces the test that has been conducted, describes its format and discusses its (controversial) results.

### 6.1.2 Previous tests on rhythm

As has been stated above, very few authors cared to test the perception of speech rhythm until recently. One of the few who did was Miller (1984), who carried out a test on four groups of participants (a- English phoneticians, b- English non-phoneticians, c- French phoneticians, d- French non-phoneticians). He used recordings of read speech (*The North Wind and the Sun*) and spontaneous speech in Yoruba, Japanese, Argentinean Spanish, Indonesian, Arabic, Polish and Finnish. He extracted balanced samples of these data and asked the four groups of participants to rate them as either syllable-timed or as stress-timed (non-phoneticians were first given an explanation of these terms that involved clapping hands in synchrony with stress and syllables). Audio samples were not manipulated in any way. Results showed that phoneticians’ ratings were unsurprisingly more consistent with expectations and that Arabic was nearly universally perceived as stress-timed, while Indonesian, Yoruba and Japanese tended to be classified as syllable-timed; a high level of indecision was found for Finnish, Polish and Spanish.

More recently, the perception of speech rhythm has raised a new interest, above all since the publications of the research by Ramus and co-workers. They conducted numerous tests, proving that languages belonging to the two alleged rhythm classes are distinguished by adults and even by new-borns. Their experimental protocol departs completely from Miller’s and involves delexicalising speech in order to prevent listeners from rating stimuli on the basis of lexical information. This is achieved through a re-synthesis of the original speech samples in a degraded signal. Ramus & Mehler (1999) present four possible re-syntheses: (1) *SALTANAJ* synthesis is achieved by substituting all plosives with [t], all fricatives with [s], all liquids with [l], all nasals with [n], all glides with [j], all vowels with [a] and preserving the original pitch and intensity; (2) *SASASA* synthesis substituting all consonants as [s] and all vowels as [a] preserving the original pitch; (3) *AAAA* synthesis is obtained by substituting all phones with [a] preserving only the original pitch; (4) *flat SASASA* synthesis is obtained like (2) but levelling the fundamental frequency at 230 Hz. They tested 64 students on the 4 types of re-synthesis on samples of English and Japanese and it was found that participants could discriminate the two languages with *SALTANAJ*, *SASASA* and *flat SASASA*

## 6. A perceptive test

syntheses, but not with AAAA synthesis. The authors consider the *flat SASASA* synthesis the most adequate to test the perception of speech rhythm.

Ramus, Dupoux & Mehler (2003) used the *flat SASASA* synthesis for discrimination tests of English vs. Spanish, English vs. Dutch, Polish vs. English, Polish vs. Spanish, Catalan vs. English, Catalan vs. Spanish and Polish vs. Catalan. Results confirmed expectations, with listeners discriminating languages belonging to different rhythm classes (English vs. Spanish), but not being able to distinguish between languages of the same rhythm class (English vs. Dutch and Spanish vs. Catalan). Interestingly, Polish was moderately discriminated from English, Catalan and Spanish, suggesting either that it is a mixed language or that it belongs to a different rhythm class.

White *et al.* (2007b) carried out *SASASA* tests to check whether the results would correlate with the values of the varcos and %V for the same samples (SSBE having high varcoC and low %V, Orkney Islands and Welsh Valleys English having medium values of both measures, Castilian Spanish having low varcoC and high %V). Data included heavily controlled sentences by three English speakers (Welsh Valleys, Orkney Islands and SSBE) and 4 Castilian Spanish speakers. The authors used an MBROLA synthesis to convert vocalic and consonantal intervals into *SASASA* at a constant fundamental frequency of 230 Hz and applied a series of normalisations to samples<sup>91</sup>. Final results confirmed that listeners could discriminate Castilian Spanish vs. the three types of English, but that they could not distinguish between Orkney Islands and Welsh Valleys English.

Dellwo (2008) carried out a test to verify whether speech rate plays a role in the perception of rhythm classes. Participants had to listen to de-lexicalised stimuli of “syllable-timed German and stress-timed French”<sup>92</sup> (consonantal intervals were re-synthesised as white noise, vocalic intervals were re-synthesised as complex periodic waveforms with a constant  $f_0$  at 230 Hz + 2<sup>nd</sup> and 3<sup>rd</sup> harmonics) and to rate them on a scale of regularity. They were unaware that they were listening to manipulated speech samples. Results showed that listeners generally rated the stress-timed French samples as being more regular than the syllable-timed German samples: this proves that they did not use the variability of vocalic and consonantal intervals as cue of regularity. Instead, the author suggests that they used CV rate (the number of vocalic and consonantal intervals per second), which is confirmed by the linear regression in cross-plots of listener ratings of regularity in function of CV-rate.

Arvaniti & Ross (2010) drafted a critical summary of perceptive tests carried out on the matter claiming that

*new protocols may be needed to test the idea of distinct rhythm classes. Such protocols should go beyond simple discrimination (which could be due to a variety of confounding factors) and should*

---

<sup>91</sup> In particular: they truncated sentences after the last stress to avoid final lengthening; they eliminated the first syllable and other following syllables in order to obtain 10 syllables per sample; they stretched or compressed each utterance uniformly in order to obtain a total duration of 1900 ms, thus preventing speech rate effects.

<sup>92</sup> By syllable-timed German and stress-timed French, the author means, respectively, German sentences that showed high %V and low varcoC, and French sentences that showed low %V and high varcoC.

## 6. A perceptive test

*be neither too indirect [...] nor too explicit, like the categorization task of Miller (1984).  
(2010:2/4)*

They built a test in which stimuli were obtained by low-pass filtering sentences of English, German, Greek, Italian, Korean and Spanish at 450 Hz. Sentences of each language were divided into 3 types (syllable-timed, stress-timed, uncontrolled). Listeners (of three different mother tongues, namely English, Greek and Korean) listened to a synthetic trochee series<sup>93</sup> and to a sentence, repeating this task for each sentence. They were asked to rate the similarity of each stimulus to the trochee series on a 7-step scale. Final results show that the native language of speakers did not significantly affect the ratings, that stimuli of English were rated less similar to the trochee series. The three sentence types were rated as more similar to the trochee series along the following scale: syllable-timed - stress-timed - uncontrolled. This is at odds with expectations and the authors conclude that “language classification by means of rhythmic classes cannot be achieved on the basis of listener impressions anymore than it can rely on measuring consonantal and vocalic variability in production” (2010:4/4).

This summary of perceptive tests for the discrimination of rhythm classes has shown that results are controversial among different studies, each of which uses a different protocol. I shall now pass on to describe how the test for the present study was conceived.

### 6.1.3 The conception of the test

When I set out to build perceptive tests (in 2008), I had no general framework<sup>94</sup> on which to base myself and not many previous studies to learn from. The so-called SASASA tests have been used in most recent studies of this type (see above), in several slightly different variations (for instance including/excluding the original pitch contour and/or intensity values). Such tests normally follow the so-called ABX or the AAX formats<sup>95</sup> and share what I consider to be a basic limitation: they ask for a clear-cut decision on the part of subjects<sup>96</sup> as to whether the stimuli are either A or B. There is no way to let them say, for instance, that stimuli n° 4 and n° 5 are both more similar to A than to B, but that n° 4 is even more similar to A than n° 5 is. In other words, they ask for a clear-cut categorisation of languages, reproducing the stress-timed vs. syllable-timed dichotomy, leaving no chance to the subjects to spread the different stimuli along a continuum. Leaving that choice to the subjects certainly would be no easy task as it would introduce variability and add many complications, but it could bring to interesting results and it would connect to the

---

<sup>93</sup> The authors report that the trochee series has been created with the MacOSX “frog” sound.

<sup>94</sup> such as the AMPER project, which not only includes precise prescriptions on how to collect data and how to analyse them, but also offers a standard model for perceptive tests.

<sup>95</sup> ABX tests are conceived as follows: the subject hears a first auditory stimulus (A), then a second (presumably different) auditory stimulus (B) and finally a third one (X): he/she is then asked whether X is more similar to A or to B. This test format is also widely used in other fields and for other purposes, such as to evaluate digital audio data compression methods. Intuitively, AAX tests provides listeners with two identical or similar stimuli plus a third one which has to be categorised as “similar” or as “different”.

<sup>96</sup> I am of course talking about adults subjects. Testing the rhythm perception of infants, though certainly a fascinating ground, was definitely not a possibility for me. So, I shall make no more reference to tests devised for such a purpose.

## 6. *A perceptive test*

theories proposed by Dauer (1987) and Bertinetto (1983) and dealt with in detail in chapter 2. One more difficulty consisted in the fact that I needed participants not to be aware of what was being tested and (in some tasks) of what the language in question was. This, of course, because I did not want the results to be influenced by any ideas or preconceptions on the part of participants.

Moreover, I have a feeling that not everybody is equally sensitive to linguistic/phonetic phenomena in general and to prosodic phenomena in particular, so I wanted a means of verifying the level of participants' prosodic sensibility and, perhaps, of identifying those subjects that were likely not to perform well: this is not to say that I intended to discard their answers, but simply that it may be relevant to see whether their answers were comparable to those given by other subjects.

As for the practical implementation of the test, I had the choice of using specific software or to build it myself with some programming language. The advantage of the former possibility consisted of course mainly in speed and easiness. Nevertheless, I chose to build the test from scratch, using HTML and JavaScript<sup>97</sup> in order to have an absolute control and no limitations as to the format and configuration of each task. Moreover, the choice of HTML and JavaScript over other programming languages or programming toolkits allowed me to build a test which is easily executed on any computer (one just needs a browser) and which could, one day, reside on a website. I have not exploited the possibility of publishing the perceptive test on the world wide web within my PhD: of course, this would bring remarkable advantages in terms of numbers (all Internet users are potential subjects for the test); yet, it does have the very negative side-effect that I would completely lose control over the selection of subjects and their trustworthiness. In fact, there is not even an effective way to make sure that each person takes the test only once<sup>98</sup>.

In the end, the output was a perceptive test written in HTML and JavaScript which consisted of three parts: after filling in a form with personal data, the first part was meant as a preliminary phase to check the prosodic sensibility of subjects (see 6.2.1), the second part consisted of auditory stimuli to be rated on a limited set of possibilities not unlike ABX tests (see 6.2.2), the third part consisted of auditory stimuli to be rated on a continuum scale (see 6.2.3). The results of each participant were contained in a log file which was saved in txt format at the end (a sample log file can be seen in appendix 4).

I shall now describe each part of the test separately and in detail: note that the results are not presented at the end, but within the discussion of each single part. It should be noted that such a presentation was chosen merely for the reader's comfort, but subjects carried out all the parts of the test thoroughly and without any interruption.

---

<sup>97</sup> I also exploited Macromedia® Flash technology in order to play sound in HTML pages.

<sup>98</sup> Checking the IP address is not safe, because one person might use more than one computers from different IP addresses and, conversely, more than one user might share the same IP address. Similarly, cookies are not safe because they can easily be deleted or modified by the user.



## 6. A perceptive test

### 6.2 The preliminary phase

#### 6.2.1 Personal data

The first part was not properly meant to test anything. It was a simple questionnaire<sup>99</sup> with personal data (reproduced in figure 6.1): participants entering the test were asked their name, age, level of studies, whether they had taken any exams in linguistics and phonetics (and, if applicable, how many exams they had taken and/or how many university credits<sup>100</sup> they were attributed in these subjects)<sup>101</sup>, their mother-tongue and their level of proficiency in any foreign languages<sup>102</sup>. The age range was between 19 and 60 years averaging 25.25. 36 participants had Italian as a mother tongue, 2 French, 1 English, 1 German, 1 Romanian, 1 Arabic and 1 claimed that the dialect of Verona was his mother tongue. The mean number of university credits acquired in phonetics was 3.72 (spanning between 0 and 10), while the mean number of university credits acquired in linguistics was 8.04 (spanning between 0 and 10).

The image shows a digital form with a light blue background. At the top, a blue-bordered box contains the instruction: "Per favore, riempi i campi con i tuoi dati." Below this, the form consists of several rows of input fields and dropdown menus. The labels for the fields are: "Nome", "Cognome", "Età (anni)", "Titolo di studio", "CFU sostenuti in fonetica e/o fonologia", "CFU sostenuti in linguistica (esclusi fonetica e fonologia)", "Madrelingua", "Conoscenza dell'inglese", "Conoscenza del francese", "Conoscenza del tedesco", "Conoscenza dello spagnolo", and two "altro(specificare)" fields. The "Titolo di studio" field is a dropdown menu currently showing "Maturità". The language proficiency fields are dropdown menus currently showing "no".

**Figure 6.1.** The form which had to be completed by participants entering the test.

<sup>99</sup> As it can be seen in question 6.1, it was not a paper questionnaire, but a digital one. It was in HTML and was attached to the rest of the test without interruption: participants were already sitting for the test when they completed it.

<sup>100</sup> For those who completed university studies before university credits were introduced, I established the equivalence of 1 exam with 10 university credits, in line with the choice taken by the University of Turin.

<sup>101</sup> The rationale behind this question is of course that students who had taken exams in linguistics and/or phonetics are more likely to be aware of the traditional stress-timed vs. syllable-timed dichotomy and to be sensitive to prosodic phenomena.

<sup>102</sup> Participants should self-evaluate their competence using the European common framework (A1 to C2). Participants who did not know the meaning of these levels usually asked information about it, so the data should be fairly accurate (keeping in mind the obvious limitations of self-evaluations, of course).

## 6. *A perceptive test*

The 43 participants did not gain any compensation (neither pecuniary, nor in terms of university credits or else) for taking the test and all of them gave me permission to use the results for research purposes. They were mainly (though not exclusively) recruited among students of the Faculty of Foreign Languages of the University of Turin. Some students came to the Laboratory of Experimental Phonetics Arturo Genre of Turin and took the test inside the sound-proof booth. However, most took the test straight after passing their first exam of general linguistics because that was the moment in which they were most easily available. This exam is usually taken in the first year, which means that most candidates were at the beginning of their university career. It has been decided that it was best to let students take and register the exam first, in order to avoid stress and to prevent them from thinking that their performance in the test might somehow have an influence on their final mark of the exam. Students who did not pass the exam were not asked to take part in the test<sup>103</sup>.

I supervised personally all participants taking the test, so some of them felt free to ask me indications if they did not understand some tasks or else. The test was not designed to keep track of the time taken by each participant to reach the end, and not even of the time taken to complete each task. For this reason, I have no precise data as to time variables, but since (as already said) I was present during all the tests, I can give an estimate of the time generally needed by participants in order to get through the end, spanning roughly from 20 to 40 minutes but probably averaging a little more than 30.

### 6.2.2 Testing prosodic sensitivity

In the second part, participants started the real test and had to respond to auditory samples. Yet, it was a kind of introductory task in which they were presented with the orthographic transcription of an Italian sentence and a recording by a professional speaker of that very sentence. They had to mark stresses on the transcription according to how the sentence was actually pronounced by the speaker (less prominent stresses on the first level, more prominent ones on the second level)<sup>104</sup>. Instructions were given at the top of the page, while in the middle there was an example sentence and its recording. Participants could listen to both the example sentence and the target sentence as many times as they wished by clicking on the corresponding icon<sup>105</sup>. The task was controlled by a JavaScript routine so that it was not possible to proceed with the test until at least one primary stress had been marked. This task was repeated five times, each time with one of the following target sentences:

---

<sup>103</sup> This choice was not taken on any “discriminatory” grounds. Simply, it was felt that asking students who had just failed our exam to kindly take a test for our research purposes was not precisely courteous. It is true, however, that students might also have been upset by their negative performance and, consequently, they might not have been able to concentrate properly for the test.

<sup>104</sup> I am aware that this distinction poses theoretical and practical problems. By “more and less prominent” stresses, I actually meant the distinction between stresses and accents (the latter are usually defined as those stresses that carry prosodic relevance). I initially included the distinction between first-level and second-level stresses in order to provide participants with an extra difficulty, I realised only too late that this complicated things enormously.

<sup>105</sup> The audio player used is an open-source and is called XSPF Web Music Player. It has been retrieved at the following webpage: <http://musicplayer.sourceforge.net/>

## 6. A perceptive test

1. In **diversi** paesi **africani**, il **tenore** di **vita** si sta **innalzando** **rapidamente**.
2. **L'uomo** ha riconosciuto **sin** da **tempi** **antichissimi** l'**importanza** dell'**acqua** per la **vita**.
3. Allo **stato** **solido** è **nota** come **ghiaccio**, allo **stato** **aeriforme** è **nota** come **vapore** **acqueo**.
4. Sono **note** **anche** altre due **forme** **solide**, **quella** del **ghiaccio** **vetroso** e **quella** del **solido** **amorfo**.
5. Per**ché** non **proviamo** a **risolvere** il **problema** **insieme**, **invece** di **litigare**?

Sentences were of course given without the stress indications reported above. Here they have been marked with first-level stresses (in bold) and second level stresses (in bold and underlined)<sup>106</sup>. Such an evaluation was of course necessary in order to have a term of comparison against which to compare the answers given by participants. I then calculated the correlation between her answers and those given by each participant, which ranged between 3.08% and 80,62% with a mean of 49.57% (quartiles at 37.67% , 54.26%<sup>107</sup> and 63.13%).

The instructions given and the format of the test can be seen in figure 6.2 (English translation in the footnote).

The screenshot shows a test interface with three main sections:

- Istruzioni:** Ascolta i campioni sonori (cliccando sull'altoparlante) e indica dove senti gli accenti; ti viene chiesto di segnare 2 tipi di accento: sul primo livello (in basso) devi segnare tutti gli accenti che senti, sul secondo livello (in alto) solo gli accenti più prominenti.
- Esempio:** Gli affreschi dell'antica parrocchiale di San Marco sono ora al museo Borgogna di Vercelli. Below the text is a row of 24 syllable markers (squares) with checkmarks indicating stresses: Gli af fres chi del l'an ti ca par roc chia le di San Mar co so no o ral mu se o Bor go gna di Ver cel li.
- Campione 1:** In diversi paesi africani, il tenore di vita si sta innalzando rapidamente. Below the text is a row of 24 syllable markers.

A play button icon is visible in the bottom right corner of the interface.

**Figure 6.2.** A screenshot of part 1 of the test (the first sentence). It is possible to see the instructions at the top<sup>108</sup>, the example in the middle and the task at the bottom of the page. Participants could mark stresses by simply clicking on check-buttons corresponding to each syllable.

The rationale behind this task is to have an evaluation of the prosodic sensitivity of participants to prosodic phenomena and, if needed, to be able to

<sup>106</sup> It can be seen that, although recordings were made by a professional speaker of Standard Italian, the accentuation is sometimes atypical, above all at the second level.

<sup>107</sup> The second quartile corresponds, of course, to the median.

<sup>108</sup> English translation of the instructions: *Listen to the audio samples (by pressing on the loudspeaker) and mark where you hear stresses; you are asked to mark two different types of stresses: on the first level (below) you have to mark all stresses that you hear, on the second level (above) you have to mark only the more prominent stresses.*

## 6. A perceptive test

identify those participants who appear to have a limited prosodic sensitivity. It is true that we do not have any proof that a limited sensitivity to stresses necessarily implies a limited sensitivity to rhythm phenomena; however, at least intuitively, this is likely to be so, as stresses and rhythm are both realised through prominences (see chapter 1 for a discussion on what rhythm is and for more details about the relation between rhythm and prominence).

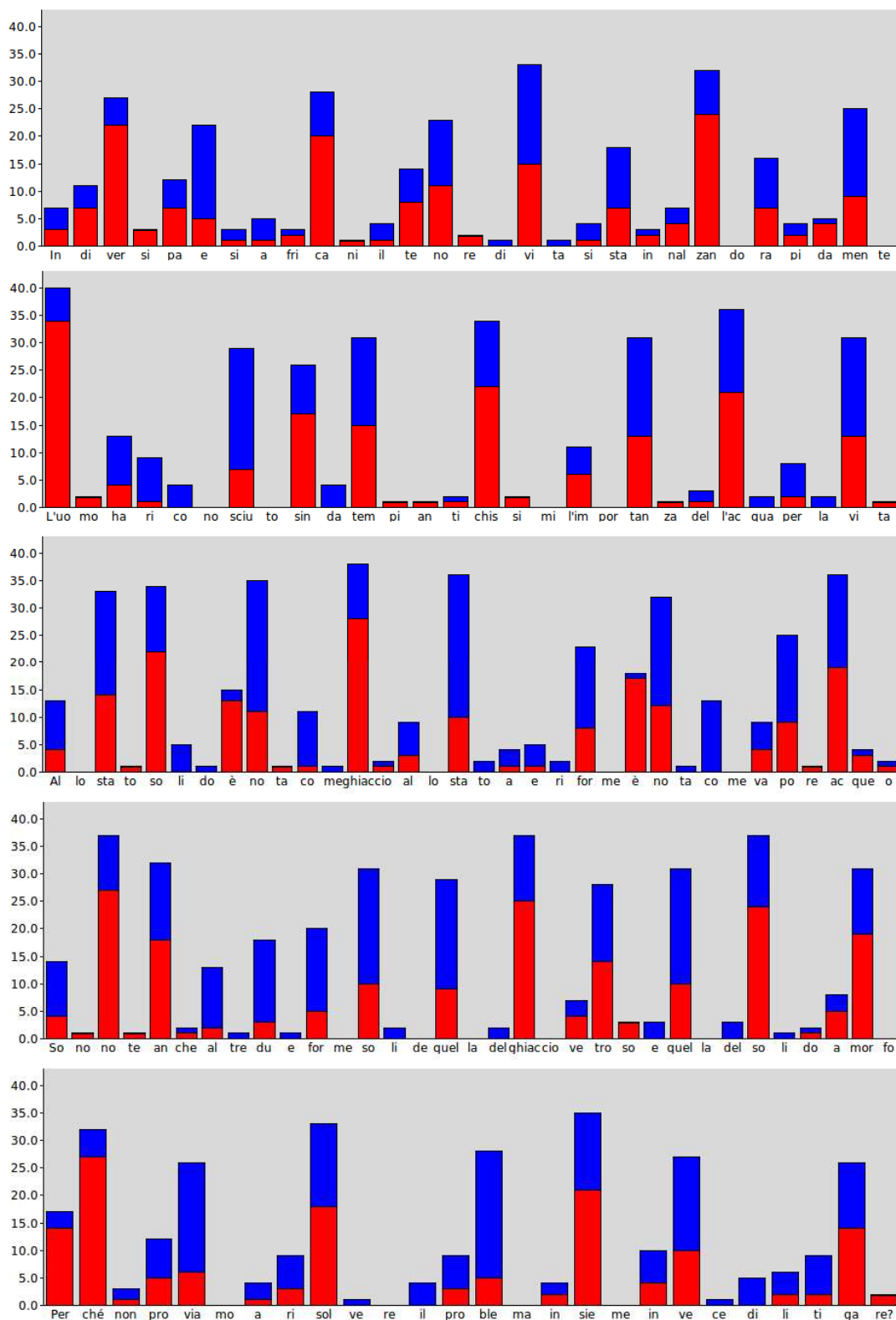
### 6.2.3 The results

I shall now illustrate the answers given by participants for the five sentences. They are shown by the histograms in figure 6.3, which represent the amount of participants that marked a first-level stress (in blue) and a second-level stress (in red) on each syllable. On the whole, it can be claimed that most participants found it difficult and did not perform splendidly. When supervising participants who took the test, I was surprised at seeing that, even though it was meant as a preliminary task, people took often long time to complete each sentence, listening various times to the recordings and taking long to mark (often very few) stresses. It should also be noted that this cannot be attributed to a misunderstanding of the instructions, because those who were not sure about what the task involved were free to ask me clarifications (and some did).

Despite the difficulties encountered by single participants, the overall results reflect quite well expectations set by the model and phonological evaluations *a priori*. In particular, effectively stressed syllables have been marked with either a first-level or a second-level stress by usually a high number of participants, whereas completely unstressed syllables have been usually marked by no or few participants (most emblematically, sentence final syllables). Interestingly, the most frequent mistakes involve marking a stress on syllables immediately preceding stress (such as ‘pa’ in *paesi*, ‘te’ in *tenore* and ‘nal’ in *innalzando* in sentence 1, as well as ‘è’ in *è nota* twice in sentence 3).

I shall not discuss these results in more detail here, as their role within the test was simply that of providing an idea of the prosodic sensitivity of participants (a detailed analysis is in preparation and is thought to appear soon). So, for this purpose, what matters are the scores of correlation (see above) between the answers given by each participant and the answers considered as the model. It has to be made clear that the correlation is not meant to be a sophisticated and faultless evaluation of listeners’ prosodic sensitivity: needless to say, this would not be feasible by simply calculating the correlation with answers by one model because, clearly, such a task has more than one acceptable answer. However, the values of correlation should be enough to identify those participants whose performance was extremely low. The problem consists in establishing the exact threshold under which participants might be considered as “prosodically insensitive”: the distribution is fairly homogenous apart from 2 participants who scored 3.08% and 6.89% (the rest of the population ranging from 18.54% upwards). I decided not to exclude any sample, at least for a first analysis, which means that the results presented above refer to the entire population of 43 participants. The results by segments of the population will be discussed at the end.

## 6. A perceptive test



**Figure 6.3.** Histograms representing the answers given to the preliminary phase of the test. Syllables are shown in the x-axis, while the y-axis represents the number of people who marked a stress on each syllable (first level stresses in blue, second level stresses in red).

## 6. A perceptive test

### 6.3 Testing the perception of discrete rhythm classes

#### 6.3.1 The format and the interface

After the preliminary phase, participants started the real test, in which I meant to verify the discrimination of rhythm classes in discrete terms, that is to say in the traditional way, asking people to decide whether a sample is stress-timed or syllable-timed, without any gradient. This part was divided into two similar tasks, each consisting of 15 samples.

Participants were first asked to listen to masked synthetic audio samples and decide which language was being spoken: they had the possibility of choosing between: (1) *Spanish, French or similar* (2) *English, German or similar* (3) *Other* (4) *I don't know* (the two possible choices obviously reflected the two traditional rhythm classes). Samples were synthesised using the AMPER routines (see below for details) and preserved the duration, pitch and intensity of the original samples: in short, they reproduced the prosodical but hid the lexical information of speech. Participants could listen to each sample only once. Instructions were given immediately before starting the task and read as follows (see footnote for an English translation):

*Ora ascolterai dei campioni audio e dovrai cercare di capire se si tratta di inglese/tedesco o francese/spagnolo. Il parlato è stato modificato in modo che tu non possa riconoscere le parole delle due lingue, quindi dovrai giudicare solo dall'impressione che ne ricevi. Per ogni frase clicca sul tasto "francese spagnolo o simile" o "inglese tedesco o simile"; se il campione audio non ti sembra simile a nessuna di queste lingue, clicca su "altro"; se invece non hai idea, piuttosto che premere un bottone a caso premi "non so", ma ti preghiamo di cercare di utilizzare il meno possibile questa soluzione. Nota che i campioni sonori vengono estratti da un database ampio in ordine del tutto aleatorio. Ora premi il tasto avanti per iniziare.*<sup>109</sup>

The interface was extremely simple and intuitive, consisting exclusively of the stimulus label and the four buttons, as can be seen in figure 6.4. Participants only had to press the button corresponding to their answer and they were immediately put forward to the next audio sample. So, the procedure was very strict: it was not possible to listen to the audio sample more than once nor to go back and correct the answer once it was given<sup>110</sup>. Each stimulus lasted between 4 and 8 seconds.

---

<sup>109</sup> *You will now listen to some audio samples and you will have to understand if they are English/German or French/Spanish. Speech has been modified so that you will not be able to recognise the words of these languages, so you will have to judge exclusively on the impressions you get. For each sentence, press "French, Spanish or similar" or "English, German or similar"; if you think that the audio sample does not sound like any of these languages, press on "other"; if you really have no idea, press on "I don't know" rather than taking a wild guess, but we kindly ask you to try and avoid this solution. Please note that audio samples are randomly retrieved from a large database. Press "forward" to start.*

<sup>110</sup> One may think that it was possible to navigate back and forward with browser controls. Actually, participants could not do that because browser controls were disabled by Javascript and, at any rate, the test was executed in full-screen mode. Furthermore, all 15 samples resided in one HTML page, so, even succeeding in navigating back would have led the participant to the instruction page.

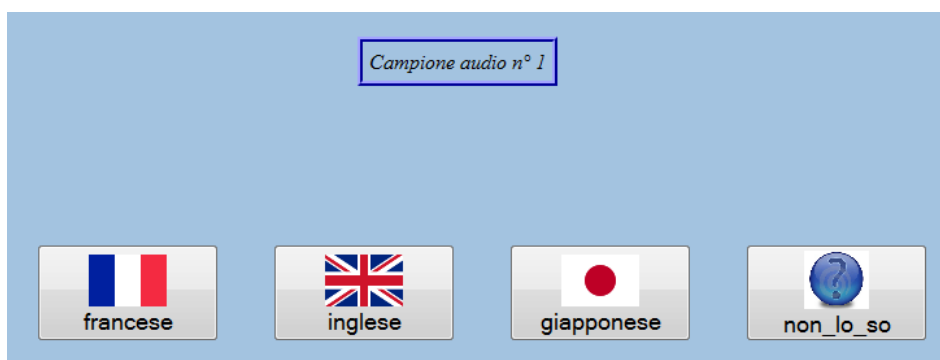
## 6. A perceptive test

The synthesised stimuli included 15 different languages. Audio samples were taken from the Illustrations of the IPA (with the exception of the Italian, Icelandic, Romanian, Finnish and Japanese samples, which were recorded at LFSAG), either included as an attachment to IPA (1999) or in various articles appeared in the *Journal of the IPA*, and synthesised as described below. They corresponded to some of the data used in chapter 3 to test rhythm metrics (which can be consulted for further details and for the bibliographic reference corresponding to each sample).

The hypothesis was that participants would classify tendentially stress-timed languages as “English, German or similar”, tendentially syllable-timed languages as “Spanish, French or similar” and mixed languages as “other” or “I don’t know”. Of course, I did not expect all participants to agree on each sample, rather it was presumable that the alleged rhythm classification would emerge as a general trend.



**Figure 6.4.** Screenshot of the interface. It consists merely of a label indicating the stimulus number and the four buttons. There is no button to listen to audio samples nor to proceed, as the succession of events is completely controlled by JavaScript.



**Figure 6.5.** Screenshot of the interface. As for the former task, there is only the stimulus number and the four choice buttons as participants have no control over the succession of events.

After the first 15 stimuli, the test proceeded with a similar task made up of another series of 15 stimuli, in which participants were, again, asked to listen to differently synthesised audio samples and had to decide which language was being spoken. In this case, the possible choices included: (1) *English* (2) *French* (3) *Japanese* (4) *I don’t know* (representing respectively stress-timing, syllable-timing, mora-timing). The audio samples of these three languages were similar to the ones presented in the former part of the test, but were manipulated in order to normalise

## 6. A perceptive test

at each turn one or two of the three prosodic parameters (pitch, intensity and duration, see below for details).

The interface (see figure 6.5) is very similar to the one in part 3 and the instructions given read as follows (English translation in footnote):

*Ora ascolterai altri campioni audio e il tuo compito è simile a prima, dovrai cercare di capire se si tratta di inglese, francese o giapponese. Tuttavia, questa volta il parlato è stato mascherato in maniera diversa.  
Ora premi il tasto avanti per iniziare.<sup>111</sup>*

### 6.3.2 Differences from other tests.

The format is not so different from SASASA tests. However, in this case, participants did not hear three stimuli (A, B and X), but only one (X), and had to classify it on the basis of categories which they presumably already knew<sup>112</sup>. This has been done in order to avoid that the results be marred by the choice of A-B sentences. In fact, the choice of A and B in ABX tests is a very delicate matter: I feared that the classification of X in ABX tests might be influenced by the choice of A and B, that is to say that using one particular sentence in language A might bring to (slightly) different results from those that one would get by using some other sentence in language A (and the same could be said about B). So, I preferred to rely on participants' knowledge of languages A and B. Furthermore, in this way, the participants' task is far simpler and this contributes to decrease the impact of errors due to misunderstandings.

An argument that has been raised concerns the fact that participants who speak a certain language might tend to classify audio samples as that language because they are more familiar with it<sup>113</sup>. Of course, this cannot be excluded *a priori*; yet, on the other hand, one could also hypothesise the opposite, *i.e.* that participants who speak a certain language might tend NOT to classify audio samples as that language precisely because they know it and they find that artificial stimuli sound different. Be it as it might, if need be, it would be possible to verify these hypotheses since questions about foreign language proficiency were included in part I of the test.

Another fundamental difference from other similar tests consists in the way the audio samples were synthesised. In fact, as in most perceptive studies on rhythm categorisation, I needed to find a way of hiding lexical information and the segmental characteristics of speech. Obviously, asking participants to categorise an overt speech sample as either Italian/French or as English/German would not allow to discover anything about rhythm: participants would most likely recognise speech

---

<sup>111</sup> *You will now listen to other audio samples and your task is similar, you will have to understand if English, French or Japanese are being spoken. However, this time speech has been masked in a different way. Press "forward" to start.*

<sup>112</sup> It could be argued that not everybody has ever heard Italian, French, German and English. The answer to this is that, first of all, it is not necessary to know all four of them: it is enough to have heard at least Italian or French and English or German. Secondly, most participants were university students at the faculty of foreign languages, who should then be fairly knowledgeable about languages. Finally, nobody complained that they did not have a sufficient knowledge of these languages in order to complete the task.

<sup>113</sup> I am grateful to Lea Glarey for raising this observation.



## 6. A perceptive test

samples on the basis of lexical information as well as of segmental clues (and even if they did not, it would be difficult - if not impossible - to make sure of that).

So, long time was spent in deciding how to mask speech. Most studies of this type (see above) make use of the already quoted SASASA method, by which consonantal intervals are re-synthesised as [s], while vocalic intervals are re-synthesised as [a]. However, for the purpose of this test I decided to use a different procedure for masking speech, the one already used within the AMPER project, where it is exploited for different purposes. I shall spend a few words about it.

### 6.3.3 Perceptive tests within the AMPER project.

The AMPER (*Atlas Multimédia Prosodique de l'Espace Roman*) project (which saw the light at the *Centre de Dialectologie* of the *Université Stendhal Grenoble 3*) enhances the creation of a multimedia prosodic atlas of the Romance area through a cooperation of different research teams belonging to each Romance linguistic area. Members of each research team collect data of their own linguistic domain according to a shared protocol, which includes various utterance types, various syllable and stress patterns as well as three repetitions for each sentence. Data are then labelled by members of the research teams and first processed with a *Praat* script and then with a programme called *Interface*, which has been developed by Albert Rilliard (currently LIMSI - *Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur*, Orsay, formerly *Université Stendhal Grenoble 3*) on the basis of Matlab procedures written by Antonio Romano (currently University of Turin, formerly *Université Stendhal Grenoble 3*) for his PhD thesis. The *Praat* script outputs a text file with the durations of each vocalic interval and their relative values of pitch and intensity at the onset, middle and offset. Taking these text files as input, *Interface* computes the values of prosodic cues (pitch, intensity and durations) for each set of three repetitions, creates charts for visualising the results and produces sound files which are then used for perceptive tests with the aim of validating the data collected. The sound files are synthesised by averaging the values of the three prosodic parameters at the onset, middle and offset of each vowel. Currently, consonants are not considered and consequently not re-synthesised (they result in silence), as their relevance at the prosodic level is thought to be limited. However, the possibility of including them in future analyses is being discussed.

The values obtained are meant as the standard prosodic contour for that utterance type in the linguistic variety taken in consideration and sound files are then used for perceptive tests for validation (see, for instance, Felloni, 2010, and Interlandi, 2003 and 2004). The usefulness of this synthesis has already been proved by several studies conducted by the various members of the numerous AMPER teams. For further details on the AMPER project, see Contini *et al.* (2002).

### 6.3.4 Synthesising audio files with AMPER routines for testing rhythm perception.

The 15 re-synthesised sentences which have been used as stimuli for the first part of the test are the incipit of the narrative *The North Wind and the Sun* in 15 different languages (in the order, German, Italian, Russian, Icelandic, Brazilian Portuguese, Romanian, European Portuguese, Finnish, Turkish, Japanese, Spanish, RP English, GA English, French, Czech). Only the first sentence was selected, because of course the stimulus should not be too long in order to avoid that the subject gets confused

## 6. A perceptive test

or tired: for instance, as far as the English version is concerned, I only used “The North Wind and the Sun were disputing on which was the stronger, when a traveller came along wrapped in a warm cloak”. I then labelled it following AMPER conventions (which involve marking the boundaries of vowels<sup>114</sup>) and processed data with *Interface* in order to get the synthesised sound files.

A relevant issue concerns the choice of preserving vs. levelling the original pitch and intensity contours. Leaving these parameters lets the participants free to use all prosodic cues to complete the task, while removing them forces the subjects to concentrate on durations. The decision has very important implications because, of course, samples for which  $f_0$  is not levelled could be recognised by their pitch contour; on the other hand, if  $f_0$  plays a role in rhythm perception, levelling it might bring to inconsistent results. Probably, the choice of what to do depends on the conception of rhythm, and I am fairly convinced that duration (although undoubtedly being an important factor – probably the most important) is not the only parameter on which rhythm is perceived (see also Allen, 1975, Cumming, 2009, and Romano, 2010). I could find no satisfactory solution to the problem as it seems to be impossible to separate the rhythmic aspect of  $f_0$  from its pitch aspect. So, after a long hesitancy and several perplexities, I finally decided to leave the original pitch and intensity contours of the samples (however, see part 2 of this task). When interpreting results, one will have to keep in mind this decision.

In the first task, participants had to give their judgment as to whether each synthetic stimulus was more similar to French/Italian or to English/German. As has been said, the 15 sound samples corresponded to 15 different languages, so only 4 of them were truly French, Italian, English or German (5 to be precise, because I included both RP and GA English). However, this has no importance at all, as I was not testing people’s ability to recognise a language on the basis of synthetic stimuli and I did not expect them to; rather, I was prompting them for a categorisation of languages based on prosodic cues (even though, of course, they were not aware of this and they were bound to think they were taking a discrimination task).

In the second task, participants had to judge whether they were listening to English, French or Japanese. Effectively, they were always listening to samples of these three languages, but in this case they were manipulated as for  $f_0$ , intensity and duration. The audio samples were similar to the ones presented in the former task as they were also created by applying the AMPER routines on the initial sentence of *The North Wind and the Sun*. However, in turn, the following manipulations were applied to each of the three samples:

- 1) *ergconst*: intensity levelled to 80 dB ( $f_0$  and durations untouched);
- 2) *f0const*:  $f_0$  levelled to 200 hz (intensity and durations untouched);
- 3) *dconst*: vocalic durations levelled to 80 ms (intensity and  $f_0$  untouched, each vocalic onset is placed at the original distance from the preceding vocalic onset);

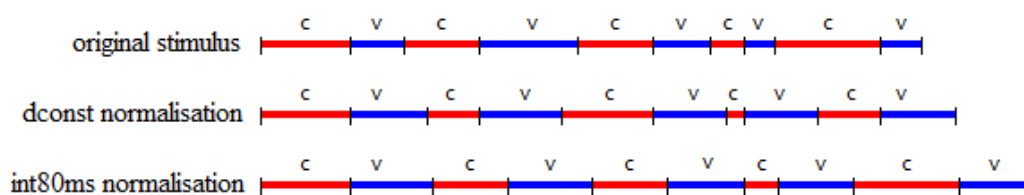
---

<sup>114</sup> While AMPER does not currently grant any importance to consonants, one of its creators and coordinators (Antonio Romano) felt that they also play a relevant role and advised that I should insert the spikes of consonantal clusters involving more than one plosive in the synthesised sound files. The rationale behind this choice is that spikes constitute a prominence within the speech chain which might have some relevance as far as the perception of rhythm is concerned. Only clusters were considered, anyway, as pre-vocalic plosives are of course followed by a vocalic onset and therefore their presence needs not be emphasised.

## 6. A perceptive test

- 4) *int80ms*: vocalic durations levelled to 80 ms (intensity and  $f_0$  untouched, each vocalic onset is placed at the original distance from the preceding consonantal onset);
- 5) *dconstconst*: consonantal and vocalic durations levelled to 80 ms (intensity and  $f_0$  untouched).

So, a total of  $3 \times 5 = 15$  manipulated samples were obtained and inserted in the test in random order. I shall briefly clarify the difference between manipulation (3) and (4): in both cases, vowels are normalised to a duration of 80 ms; however, in (3) the consonantal interval (which is synthesised with silence) is either enlarged or shrunk in order to let the following vocalic onset be at the original distance from the preceding vocalic onset as illustrated below.



In practise, the manipulation has been done by manually modifying the values in the text files (which are produced by the *Praat* script quoted above and which are given as inputs to *Interface*). As has been said above, these text files contain the durations of all vocalic and consonantal segments and the values of  $f_0$  and intensity at their onset, middle and offset. So, the manipulation is easily done by changing the numeric values of the text files: for example, figure 6.6 shows a manipulated text file (for French) with levelled intensity.

The screenshot shows a text file named 'c12\_int80ms.txt' with the following content:

```

File  Modifica  Formato  Visualizza  ?
C:\MATLAB6p5p1\work\oggi\c12.txt      size: 114624
22-May-2009

          duration [ms]          energy [dB]          fo1          fo2          fo3 [Hz]
1              27              80              220          218          215
2             118              80              273          324          341
3              64              80              239          229          218
4              59              80              181          177          167
5             103              80              162          138          135
6              34              80              182          186          192
7              49              80              228          230          218
8              53              80              273          269          253
9              57              80              185          164          160
10             27              80              112          96          92
11             28              80              192          180          167
12             30              80              190          181          180
13             83              80              188          166          159
14             75              80              175          178          182
15             40              80              243          248          254
16             51              80              276          293          293
17             84              80              320          308          286
18             51              80              190          183          169
19             98              80              167          150          138
20             44              80              137          136          134
21            120              80              130          127          131
22             84              80              146          148          154
23             38              80              178          175          173
24             59              80              181          179          170
25            122              80              141          125          131
26            142              80              233          240          316

values at:
12269 12909 13549 14657 15297 15937 18712 19352 19992 22513 23153 23793 28251 28891 29531 32379
33019 33659 34577 35217 35857 39866 40506 41146 42823 43463 44103 45622 46262 46902 49433 50073
50713 52548 53188 53828 56619 57259 57899 60547 61187 61827 72205 72845 73485 74128 74768 75408
77803 78443 79083 80824 81464 82104 84083 84723 85363 86144 86784 87424 88637 89277 89917 95189
95829 96469 99127 99767 100407 101008 101648 102288 103365 104005 104645 108422 109062 109702
  
```

**Figure 6.6.** A text file issued as output by a *Praat* script and used as input for *Interface* (see above). This file has been manipulated by setting all intensity values to 80 ms. A similar operation has been done to level pitch and durations. More details about this file format can be found in Romano (2010:62-64).

## 6. A perceptive test

### 6.3.5 The results

The results are reported in figure 6.7; histograms show the answers given by the 43 participants for each sample. For the first task (see the 15 samples reported in the chart above), red bars indicate ratings for “English, German or similar”, green bars indicate ratings for “French, Spanish or similar”, blue bars indicate ratings for “Other”, while yellow bars indicate ratings for “I don’t know”. For the second task (see the 15 samples reported in the chart below), red bars indicate ratings for “English”, green bars indicate ratings for “French”, blue bars indicate rating for “Japanese”, while yellow bars indicate ratings for “I don’t know”.

Results are quite surprising and do not confirm expectations: the German, Brazilian and European Portuguese, Romanian, Japanese and French samples display the highest ratings as “*English, German or similar*”, while the Russian, Finnish and English (both GA and RP) samples display the highest rating as “*Spanish, French or similar*”. High levels of indecision seem to affect the classification of Italian, Icelandic, Turkish, Spanish and Czech. It goes without saying that this scenario does not reflect the traditionally accepted rhythm classes. It is particularly surprising to notice that Japanese (an alleged mora-timed language, which should be at the other end of the supposed rhythm continuum) has been rated together with German (an alleged stress-timed language): these two languages present opposite values of deltas and PVIs for these very samples (see chapter 3).

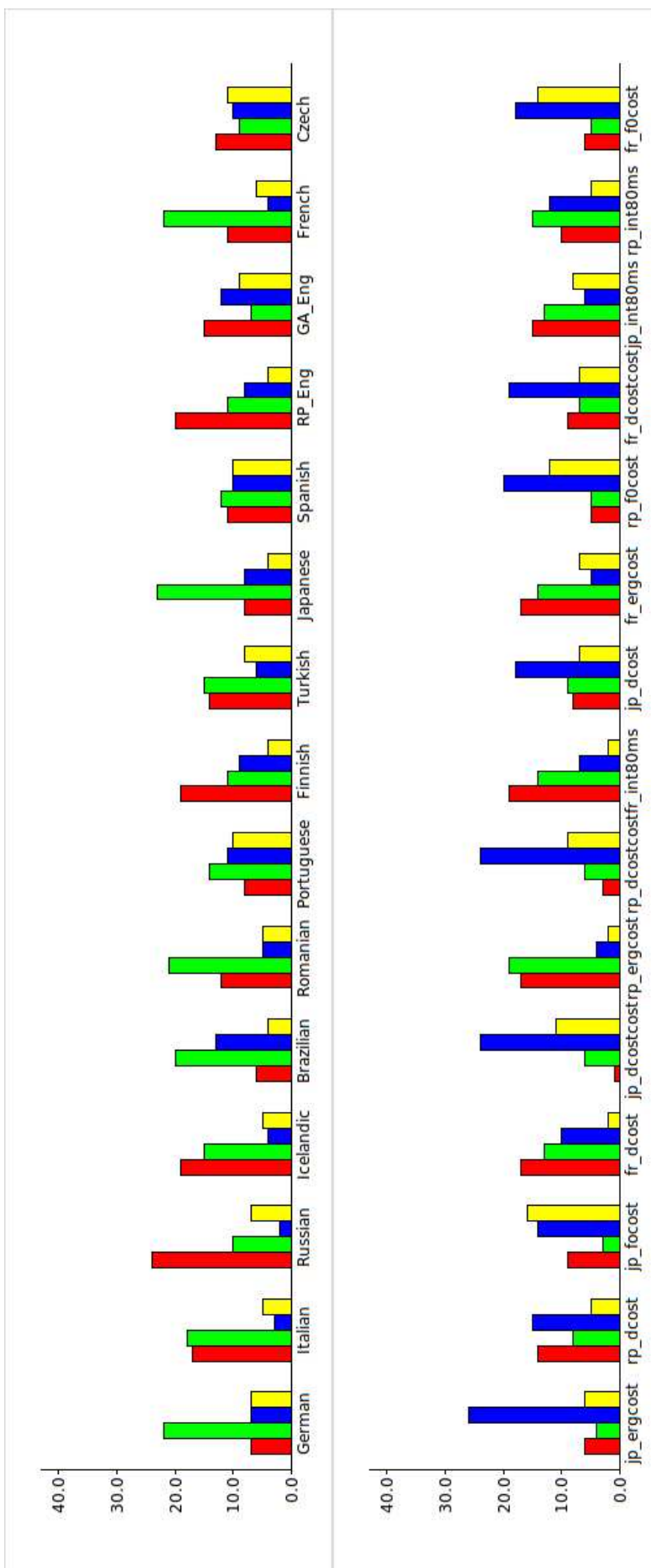
Examining the answers given to the second task, three main observations emerge. Firstly, it can be seen that all three samples normalised as *dcostcost* (by levelling both vocalic and consonantal durations) have been most frequently classified as Japanese. This, in a singular and unexpected way, seems to confirm that participants have a mental representation of this language as having constant intervals: whenever they heard very regular sequences, they classified them as Japanese.

Secondly, it can be seen that indecision seems to be greater in correspondence of *f0cost* normalisations: in these 3 cases the yellow bars are in fact the highest of all 15 samples. This clearly suggests that participants were using  $f_0$  for the classification of samples and that its levelling disorients them.

Thirdly, in 4 out of 5 cases, Japanese has been correctly identified (the only exception being the *int80ms* normalisation). Interestingly, it seems to be the only one of these three languages that can be recognised, since samples of French and English are either classified as Japanese (in *dcostcost* normalisations, as remarked above) or are classified as French and English by roughly the same amount of participants. What is even more surprising is that the Japanese sample was classified as “English, German or similar” in the preceding task. However, if participants were able to recognise Japanese even when one of the three prosodic parameters was neutralised, there is no reason to believe that they could not recognise it when all prosodic parameters were original: this implies that speakers probably recognised that sample as “Japanese or similar”, but, in lack of such a label, they associated it with “English, German or similar”. However, as already stated, there is no way of telling whether this association has been made on the base of purely rhythmic properties or only on the basis of the pitch contour.

In general, it can be said that the results of this part of the test were far from confirming expectations, but an analysis of the answers given by participants revealed some aspects of coherence and opened interesting and challenging perspectives.

## 6. A perceptive test



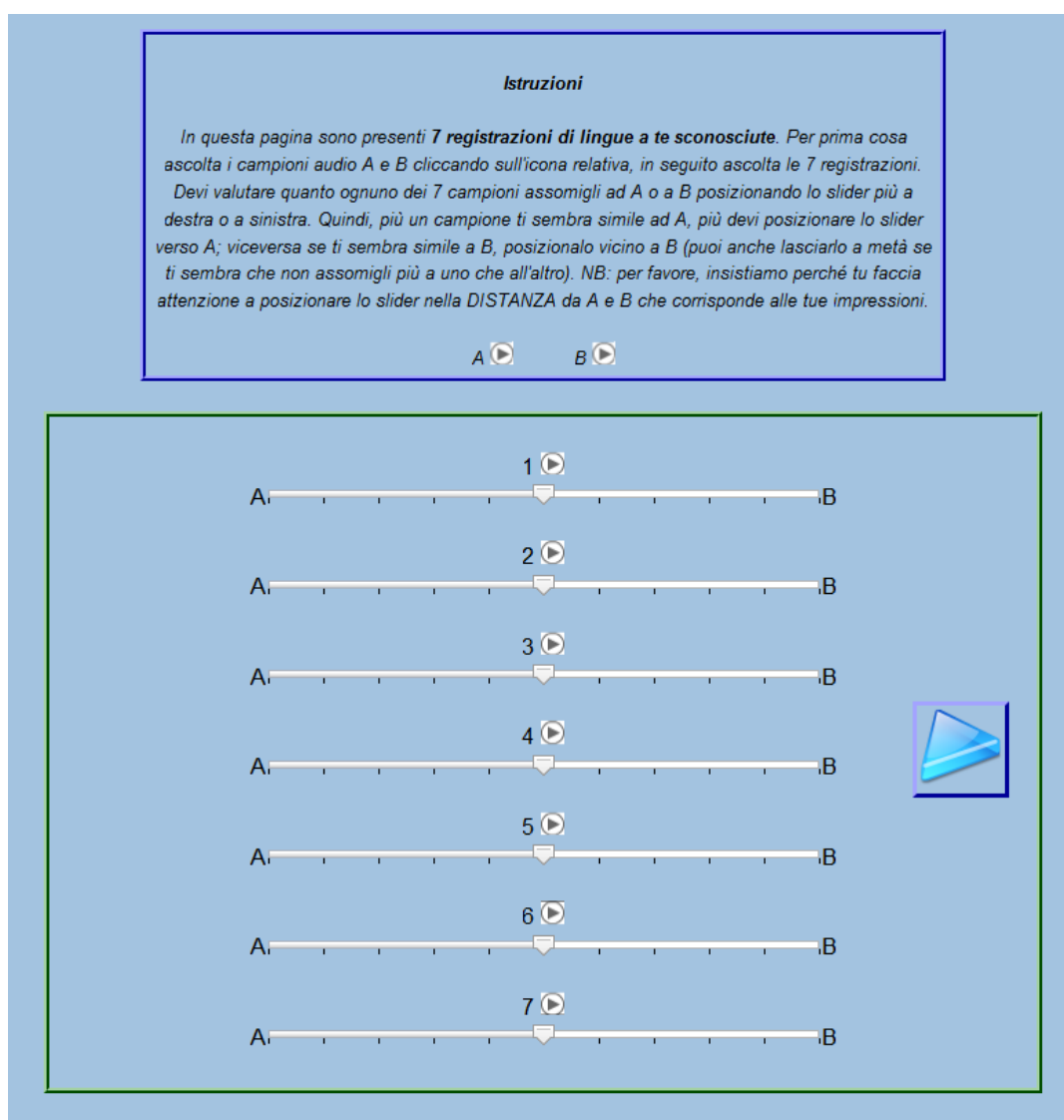
**Figure 6.7.** Answers given by the 43 participants for each audio sample. For the first task (above), red bars indicate ratings for “English, German or similar”; green bars indicate ratings for “French, Spanish or similar”; blue bars indicate rating for “Other”; while yellow bars indicate ratings for “I don’t know”. For the second task (below), red bars indicate ratings for “English”; green bars indicate ratings for “French”; blue bars indicate rating for “Japanese”; while yellow bars indicate ratings for “I don’t know”.

## 6. A perceptive test

### 6.4 Testing the perception of a scalar characterisation of rhythm

#### 6.4.1 The format and the interface

The final part of the test consisted in a “scalar” implementation of a traditional ABX format. Participants had to listen to two synthesised versions of the first sentence of the *North Wind and the Sun* in RP English (A) and Standard French (B), though of course they were not told it was French and English. The 2 samples were synthesised using the AMPER routines by neither modifying nor levelling the values of duration, pitch and intensity (in short, the stimuli were only masked, not manipulated). Then, they had to listen to 7 different stimuli of 7 presumably unknown languages (see below) and to decide whether they resembled more to A or to B. They had to express their judgment with the help of a slider, which went from A to B (see figure 6.8).



**Figure 6.8.** The interface of the final part of the test. Participants could listened to A, B and the 7 samples by clicking on the corresponding icons. Sliders could be dragged left or right to reflect each sample’s resemblance to A and B.

## 6. A perceptive test

They had to place the slider at the distance from A and B which corresponded to their impression: for instance, if they found that sample 1 was very similar to A, they had to drag the slider next to A; if they found that it was slightly more similar to B than to A, they had to drag the slider a bit more than half-way towards B, and so on; if they found that it was neither similar to A nor to B, they could leave it in the initial position half-way between the two. So, languages were rated on a continuum, the task did not prompt for a clear-cut bi-polarisation. This also allowed participants to create an order among the seven unknown languages as they were let free to adapt their judgments after one or more ratings. In other words, they could rate the first language as 80% closer to A, then listen to the second language and rate it as 85% closer to A, then modify the first at 70% closer to A because they felt that there still was a remarkable difference between the two. Even though the test was not conceived to keep track of how many times participants changed their mind, I noticed that the majority of listeners did modify their first ratings (at least once) after listening to the following ones.

This procedure was repeated twice: the first time with 7 supposedly unknown languages (in order of appearance: Amharic, Czech, Finnish, Standard Belgian Dutch, Icelandic, Indonesian and Turkish<sup>115</sup>), the second time with 7 regional varieties of English (in order of appearance: RP English, Tyneside English, New Zealand English, GA English, Australian English, Liverpool English, Southern Michigan English<sup>116</sup>). Like before, data samples were drawn from the illustrations of the IPA (apart from Icelandic, see chapter 3 for more details). In both tasks, the A and B stimuli remained unchanged and participants were free to listen to them as many times as they wished; likewise, they could listen to the 7 samples as many times as they wished and in the order they preferred.

The instructions for task 1 can be seen in figure 6.8<sup>117</sup>, while the instructions for task 2 read as follows:

*In questa pagina sono presenti 7 registrazioni di varietà regionali di inglese (cioè di inglese parlato con accenti diversi). Il tuo compito è lo stesso di prima; se vuoi puoi riascoltare A e B una o più volte, poi ascolta i sette campioni audio e dai una valutazione di quanto ognuno di essi assomiglia ad A e a B posizionando lo slider.<sup>118</sup>*

---

<sup>115</sup> All samples were taken from the corpus (Czech, Finnish, Icelandic and Turkish) or from the Illustrations of the IPA (Amharic from Hayward & Hayward, 1999, Standard Belgian from Verhoeven, 2005, and Indonesian from Soderberg & Olson, 2008). It was chosen to refer to Standard Belgian Dutch as *Flemish* for practical reasons despite the ambiguity of this term (the label *Standard Belgian Dutch* is too long to fit in the charts).

<sup>116</sup> They were all taken from the Illustrations of the IPA (see respectively Roach, 2004, Watt & Allen, 2003, Bauer et al., 2007, Ladefoged, 1999, Cox & Palethorpe, 2007, Watson, 2007, Hillenbrand, 2003).

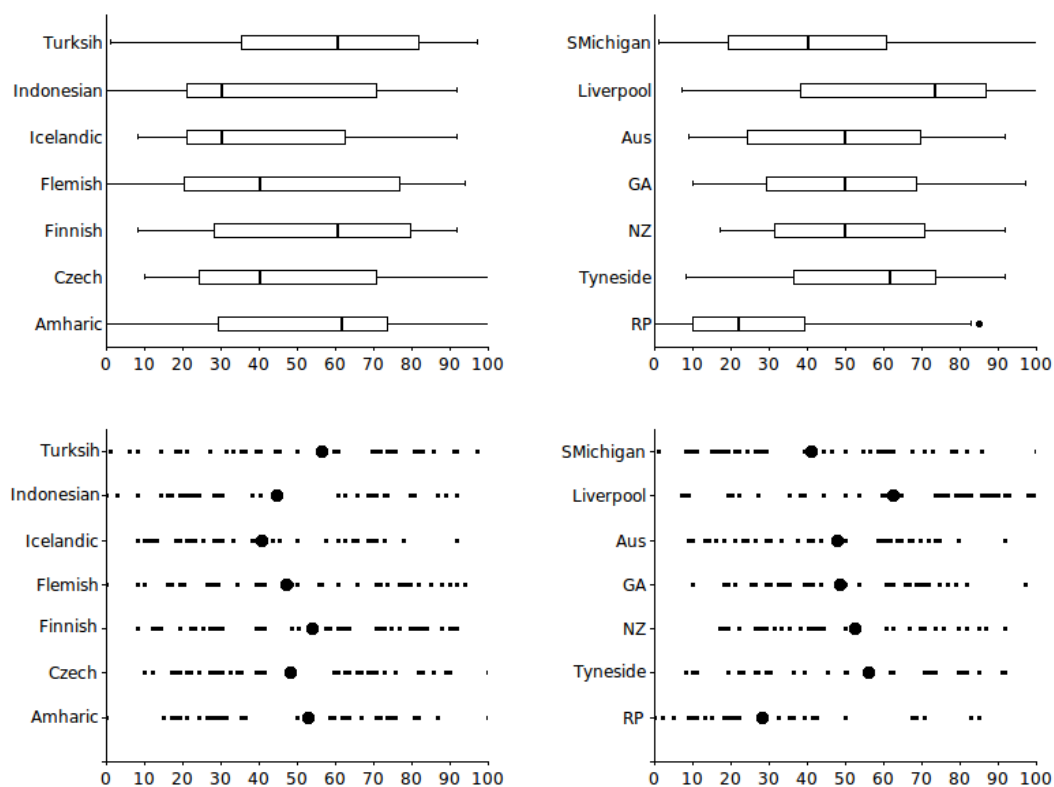
<sup>117</sup> English translation of task 1: *In this page there are 7 recordings of languages which you don't know. First listen to the A and B audio samples by clicking on the corresponding icons, then listen to the 7 recordings. You have to evaluate how much each of the samples resembles A or B by dragging the slider to the left or to the right. So, if a sample sounds more like A, you have to drag the slider to the left; conversely, if it sounds more like B, drag it towards B (you can also leave it half-way if you think it does not sound more like one or the other). NB: we really ask you to pay attention to place the slider at the exact distance from A and B that corresponds to your impressions.*

<sup>118</sup> English translation of task 2: *In this page there are 7 recordings of regional varieties of English (that is to say different accents of English). Your task is the same as before; you can listen to A*

## 6. A perceptive test

### 6.4.2 The Results

The answers given by participants to each of the 14 samples are shown in box-plots (showing the median and the quartiles) above and scatter-plots (showing the dispersion and the mean) below. Samples pertaining to task 1 are shown on the left, while those pertaining to task 2 are shown on the right.



**Figure 6.9.** Results of the final part of the test, asking for a scalar categorisation of 7 samples of unknown languages (on the left) and 7 regional varieties of English (on the right). Data is presented on box-plots above and on scatter-plots below. 0 corresponds to A (synthesised RP English sample) and 100 corresponds to B (synthesised French sample).

Again, it seems that no decisive conclusions can be drawn on the basis of these results. The variability of the answers is impressive and mostly covers all available space, often without even a tendency to aggregate in a more dense area. Results of task 1 (unknown languages) show that mean values obtained for all samples do not depart from the middle, but median values indicate that there might be some differences: Indonesian, Turkish (and to a lesser extent Flemish and Czech) have been more frequently associated to A (RP English), while Turkish, Finnish and Amharic have been slightly more frequently associated to B (French).

Results for the second part (regional varieties of English) also show very comparable mean and median values, apart from Liverpool English (which seems to have been perceived slightly more similar to B than the other samples) and RP English which has been perceived as more similar to A. This is utterly unsurprising

---

*and B as many times as you need, then listen to the seven audio samples and give an evaluation of how much each of them resembles A or B by dragging the slider.*



## 6. *A* perceptive test

as *A* is precisely a synthesis of that sample of RP English: in other words, participants merely agreed on the fact that the RP English sample sounds like itself!<sup>119</sup> This indirectly provides a confirmation (1) of the validity of the synthesis method adopted and (2) of the fact that in general participants were still concentrating on the test even in its final part and did not simply take wild choices.

### 6.5 Conclusion

The results of the test do not provide evidence that naive participants can discriminate between rhythm classes using prosodic cues, neither in terms of a discrete nor of a scalar characterisation. I shall simply mention that in language recognition tasks, participants classified Japanese, German and French (each traditionally representing a different rhythm class) in the same group (in particular as “English, German or similar”): this certainly does not correlate with the values of the metrics for these languages. In tasks asking for a scalar characterisation of data, answers exhibit a very high variability and mostly do not reveal relevant differences among the ratings of different languages. For these reasons, a detailed statistical analysis has not even been attempted: results are unclear and do not provide any relevant finding whose validity needs to be tested.

It seems to me that the failure of the test in confirming expectations might depend on one or more of the following three factors: 1) the participants 2) the design of the test 3) the fact that the rhythm class hypothesis might not be reflected by naive listeners’ perception. I shall comment on each of the three possibilities.

#### 6.5.1 Participants

First of all, I have to make clear that the results were obtained on samples of mostly Italian native speakers, while most perceptive tests on rhythm categorisation have been conducted on English, French or German speakers. Results reported in the literature for perceptive ratings across native speakers of different languages are controversial: Miller (1984) reported differences between native speakers of English and French, while Arvaniti & Ross (2010) did not observe any significant difference. Still, it might be that Italian listeners get different rhythm impressions from listeners of other mother tongues.

However, it cannot be claimed that they did the test without concentrating or by mostly taking wild choices for at least two reasons: a) the results of the preliminary phase were overall good and b) in the final task, participants in general managed to categorise the RP English stimulus as *A*<sup>120</sup>. For the same reason, one cannot say that, given the length of the test, participants got tired after a while and performed badly.

Still, since the preliminary task provided a coarse evaluation of what was defined as the “prosodic sensitivity” of participants, it might be interesting to check if the results improve by considering exclusively those participants who performed well in the stress-related task: appendix 5 contains the results of those participants who scored a correlation higher than or equal to the median.

Observing these data, it can be noticed that the results of the preliminary task improve severely, concentrating on effectively stressed syllables or, to a lesser

---

<sup>119</sup> Or, rather, like its synthesis.

<sup>120</sup> As said above, *A* was precisely a re-synthesis of that sample.

## 6. A perceptive test

extent, on phonologically stressable/accentable syllables. This is not surprising at all, since the correlation was calculated precisely on these values.

Instead, results of the language identification test do not differ from overall results: emblematically, German and Japanese are still classified together (although French is in this case controversially classified into two different groups by roughly the same amount of participants). Remarkably, *f0cost* normalisation still caused a high degree of indecision, particularly in its first occurrence. Again, this suggests that they were using *f0* as an important cue for language recognition. Results of the sliders also present approximately the same scenario, with slightly more definite placements of Turkish towards 100 (which represents the B stimulus, *i.e.* French) and of RP English towards 0 (which represents the A stimulus, *i.e.* RP English). The only remarkable difference is that Southern Michigan English also leans towards A.

In short, it cannot be said that results improve much by considering only those participants who performed well in the stress-related task. Some distinctions are in effect more clear, but the overall scenario is comparable.

### 6.5.2 The design of the test

It is of course possible that the failure of the results to provide evidence for a perceptive distinction of the rhythm classes is due to the design of the test. It is in fact bizarre that other perceptive tests based on delexicalised speech (Ramus & Mehler, 1999, just to mention one) found different results. But these studies differ from the present one as for the format of the test (AAX or ABX) and for the fact that they usually concentrate on a limited set of oppositions (e.g. Ramus *et al.*, 1999, merely focus on the discrimination of English vs. Japanese), while this test probably introduced too many languages and too many variables (the 5 different types of normalisation). This is certainly one of the most probable causes of failure.

Moreover, it has to be remarked that I preserved the original pitch contour of synthesised stimuli, while traditional tests normally favour a *flat SASASA* re-synthesis. The choice of leaving the original pitch has also been made by Arvaniti & Ross (2010), who low-pass filtered the audios samples; even in that case, results did not confirm a classification of languages according to the traditional rhythm classes. This might imply that  $f_0$  hinders more than helps participants in categorisation tasks: such a possibility is certainly weird and unexpected. At any rate, I believe that this aspect needs further clarification<sup>121</sup>.

### 6.5.3 Perception of the rhythm class

Of course, another viable reason why the test did not work as expected concerns the possibility that the rhythm classes are simply not reflected by naive listeners' perception. This is what has been suggested by Arvaniti & Ross (2010) on the basis of their results. However, some tendencies in the classification of languages did emerge in this test, even though they did not reflect the expected rhythm groupings (just like in the study by Arvaniti & Ross). This means that listeners do provide some kind of categorisation when listening to delexicalised stimuli preserving prosodic information. I believe that these aspects need to be studied in further detail in the future.

---

<sup>121</sup> I am currently pondering the possibility of administering the same test with flat syntheses to other participants in order to check for differences in the results.

## *6. A perceptive test*

### **6.5.4 Summary**

In summary, it can be said that the results of this test are inconclusive. The perceptive categorisations (if any) do not reflect the scenario depicted by the metrics, but this could have various reasons, including a faulty design of the test. Further clarification is needed as for the role of  $f_0$  in rhythm discrimination tasks and on the influence of task format on the results.

**7.**

**Final discussion**

## 7. Final discussion

Drawing conclusions is, generally speaking, a hard task, and this is all the more true in this case for at least three reasons. Firstly, each chapter of the thesis focuses on a different aspect of rhythm typology and is self-contained, which obviously complicates things. Secondly, rhythm metrics are no longer a young approach (it has been more than a decade since Ramus *et al.*, 1999), so that it is difficult to be “original”, and yet they are not old enough for researchers to take their validity for granted. Moreover, some problems persist (cp. Arvaniti, 2009, Barry, 2010, and Bertinetto & Bertini, 2010) and a categorisation of languages based purely on the results obtained with the metrics is possible but hazardous. Thirdly, as has been explained in the text, not all results obtained in the experiments can be considered to be “conclusive”. So, this section is conceived as a final discussion that re-proposes and analyses some methodological points, the main findings of the experiments presented (be they conclusive or not) and future perspectives.

### 7.1 Methodological issues

This thesis has given a contribution to the understanding of rhythm metrics and to the evaluation of their stability/variability. A number of methodological issues have been raised.

The study of the metrics in terms of inter-subject comparison has shown a certain degree of variability, which is increased if more factors come in play (namely with the CCI, which requires phonological choices to be made). Therefore, results may change considerably across different studies in function of segmentation choices. This has at least two relevant implications. The first is that researchers should always specify their phonological criteria and segmentation procedures along with their results. The second is that cross-study comparisons should be considered carefully: charts built with data coming from different studies (cf. O’Rourke, 2008) should absolutely be avoided, except perhaps for specific purposes.

Speaking of segmentation criteria, an important issue concerns the adoption of a phonological vs. phonetic orientation. Results have proven to vary considerably between the two approaches. Namely, the Japanese sample has been labelled twice by both PM and AR because of the ambiguity caused by the interpretation of devoiced [i] (phonologically a vowel, phonetically a consonant): results have shown impressive differences, bringing to a recategorisation of the language due to the effects on consonantal delta and rPVI. These differences seem to be neutralised by the CCI, possibly in virtue of the division of intervals into the number of segments that compose them. In effect, labelling a devoiced [i] as a consonant results in a very long consonantal interval made up of (a) the preceding consonantal interval, (b) the devoiced vowel and (c) the following consonantal interval. Considering that consonantal intervals in Japanese tend to be extremely simple in other contexts, it is not a surprise that consonantal variability should result in high values for such a segmentation. Conversely, dividing the intervals by the number of segments that compose them results in similar values for consonantal and vocalic CCI because the segments are not compressed around devoiced vowels more than in other contexts. Actually, the fact that they are not compressed contributes to lengthen the already complex consonantal interval that would result from the labelling as the devoiced vowel as a consonantal segment.

The final answer to the question of whether it is advisable to adopt a phonological or a phonetic orientation remains open as both approaches present

## 7. Final discussion

advantages and disadvantages. Much of it also depends on the eternal conflicts between phoneticians and phonologists, each claiming that their approach is more objective. In principle, both approaches are objective: a phonetic segmentation is supposed to be closer to the acoustic data (and should presumably better reflect perceptive impressions<sup>122</sup>), while a phonological segmentation should be objective in that it relies on external judgements *a priori* (e.g. on- and off-glides are consonantal segments, devoiced vowels are vocalic segments, syllabic consonants are vocalic segments and so forth). However, a phonetic labelling often risks being fairly subjective as the difference between, say, syllabic consonants and sequences of *shwa* plus a consonant is judged on impressionistic (and therefore subjective) grounds. In contrast, a phonological segmentation changes according to the phonological theories adopted and risks not reflecting the reality of acoustic data. Moreover, the distinction between the two approaches is not always simple and clear-cut, and most authors seem to adopt a mix of them. For instance, in this thesis syllabic consonants were considered as vowels (a choice that was grounded on a phonological evaluation), while on- and off-glides are considered as consonantal. Perhaps, this choice creates an overall balance between these two ambiguous categories of segments (both phonetically consonantal and phonologically vocalic) and between the two approaches.

Another methodological question concerns the inclusion vs. exclusion of specific controversial segments and of specific parts of the sentence. Some authors, in fact, decided to omit the final parts of sentences, usually from the last stress to the end (e.g. Bertini & Bertinetto, 2009), while others omitted on- and off-glides (e.g. White & Mattys, 2007)<sup>123</sup> to avoid classification problems. I of course understand the reasons that brought these authors to such choices, but it seems to me that the decision of excluding some controversial segments from the total computation is at least partly arbitrary. The omission of the final parts of the sentences can be perhaps justified as an attempt to cope with the disruptions of spontaneous speech (controlled speech is not fashionable in the linguistic milieu as of now and is therefore frequently frowned upon) and of capturing the rhythmic features that, perhaps, emerge in spite of the frequent hesitations, lengthenings, rephasings and other possible eurhythmic features. Personally, I have tested the possibility of omitting the final parts of utterances (excluding everything that came after the final stress) on samples of read speech, but this has not brought to relevant differences (the results have been presented in Mairano & Romano, unpublished). For this reason, I have chosen to consider all segments of the utterance<sup>124</sup>.

*Correlatore* has proved to be a usable framework for the study of speech rhythm with the metrics. Nearly all the results and the charts presented in this thesis have been obtained with this tool. The inclusion of so many samples (61 coming from the corpus + 6 Piedmontese speakers + 10 samples of L1/L2 by one speaker) would not have been possible without it.

A delicate point concerns normalisation. Bertinetto & Bertini (2008 and following) are right in claiming that important information may be lost by normalising. Still, without normalisation, a categorisation of languages is only

---

<sup>122</sup> However, in the practise, this does not seem to be true, as shown by the Japanese sample, for which a phonological segmentation has better reflected perceptive impressions.

<sup>123</sup> Furthermore, one may wonder what conventions are adopted by the many author who do not state their segmentation criteria.

<sup>124</sup> I have only exluded utterances made up of one only segment, such as, typically, [e:::], [o:::] or [ø:::].

## 7. Final discussion

possible by using carefully controlled samples, all presenting the same speech rate. But this, in turn, raises many issues, not least the point of how speech rate should be measured<sup>125</sup>... So, perhaps, the decision of normalising or not is best taken by each author on the basis of his/her data and in view of his/her purpose. If the aim is to study rhythm in relation to (or as a function of) speech rate, normalisation clearly misses the point. Instead, if the aim is to provide a categorisation (or a comparison) of a set of languages, then normalisation is perhaps a good idea, unless the data are perfectly comparable and controlled in respect to speech rate.

The final methodological issue concerns the type of synthesis that was used for perceptive tests. Most experiments in this domain carried out by Ramus, Mehler and colleagues (as well as White *et al.*, 2007, and Dellwo, 2008) make use of the so-called *flat SASASA* synthesis, which levels pitch and intensity thus preserving exclusively the alternation between consonantal intervals (re-synthesised as [s]) and vocalic intervals (resynthesised as [a]). This choice depends on preliminary experiments by Ramus & Mehler (1999) showing that listeners were able to discriminate between languages of different classes purely on the basis of this synthesis. However, on the grounds that  $f_0$  plays a role in the perception of prominence and thus of rhythm (as proven by Allen, 1975, Lehiste, 1977, and Cumming, 2008, among others), I decided to leave the original pitch contour in the syntheses, as it has been done by Arvaniti (2009)<sup>126</sup>. However, the main problem of such an approach is that there is no way of distinguishing between the “prominence aspect” of  $f_0$  and its “intonational aspect”. In short, when analysing the results, it is impossible to tell whether listeners categorised a language on the basis of pitch contour or of prominence and it is difficult to establish which auditory cue has been used. This has also had a number of other implications that have been dealt with in detail in chapter 6. At the light of these observations and as an indication for the future, it is probably advisable to avoid mixing the various aspects of prominence (pitch, intensity and duration) in the same test.

### 7.2 Main findings

Results of rhythm metrics applied to a corpus of 61 speakers of 21 languages have provided an acceptable categorisation: stress-timed/compensating languages mostly present high delta, varco and PVI values and fall below the bisecting line of the CCI chart, whereas syllable-timed/controlling languages tend to show the opposite trend. In particular, English and German are confirmed to occupy the supposedly stress-timed/compensating area of the charts, while Greek, Spanish, French and Italian are confirmed to show up in the supposedly syllable-timed area. Some other languages, instead, show up slightly further away from their expected position. It is the case of the two samples of Russian (which unexpectedly show a limited vocalic variability), as well as of Estonian and Turkish, which are classified in intermediate positions tending towards stress-timing in most charts (apart from the CCI).

---

<sup>125</sup> A discussion of this issue is well beyond the scope of the present thesis.

<sup>126</sup> Although she used a completely different solution. She did not actually re-synthesise audio samples, rather she filtered them.

## 7. Final discussion

However, there are several caveats: first of all, different metrics bring to potentially different results: this is the case of Icelandic<sup>127</sup>, which is classified as stress-timed by the deltas and PVIs but as controlling by the CCI (it falls not far from the bisecting line). Similar considerations have been noticed for Danish and Swedish samples, as well.

Secondly, inter-speaker variability is often so relevant that it can bring to a remarkable degree of overlapping between languages and between rhythm classes. This of course complicates the interpretation of data and demands that conclusive categorisations be drawn exclusively on a conspicuous number of speakers per language (which is not always the case for my corpus).

Thirdly, some samples cluster in areas of the chart that are not classified, that is to say that they are not associated with either stress-timing or syllable-timing. It is mainly the case of Polish, which exhibits high consonantal and low vocalic variability.

Finally, and most importantly, there is no solid framework to test the results: expectations are only based on impressions by authors or by previous researchers. In other words, the metrics are not “predictive” (see below).

Moreover, it seems that, even though they were originally conceived to be applied to consonantal and vocalic durations, rhythm metrics seem to yield comparable results when applied to other domains: voiced and unvoiced intervals (see Galves *et al.*, 2002, and Dellwo *et al.*, 2007), inter-onset and inter-stress durations (see Wagner & Dellwo, 2004, and Asu & Nolan, 2006, partially confirmed by my experiment in chapter 2) as well as segmental durations (see the CCI, Bertinetto & Bertini, 2008). Lee & Todd (2004) reported that they obtained a classification of languages by applying the rhythm metrics to pitch and intensity (but this was not confirmed by my experiment presented in chapter 3). These observations seem to give credit to the already reported claim by Bertinetto that “the ultimate difference between iso-accentual and iso-syllabic languages might lie in the *different degrees of flexibility they exhibit at all relevant levels of structure*” (1989:123). In other words, syllable-timed languages can be considered to have a temporal structure that tends to be more fixed than stress-timed languages, which in contrast exhibit more durational variability at all levels.

Speaking of variability, the study presented in chapter 5 has attempted to provide a paradigm for the study of such aspects. For the moment, it is still in an embryonic phase but it has provided promising results. Different types of variability have been analysed, namely inter-subject (given by different segmentators), intra-speaker, inter-speaker and inter-dialect. It seems that all metrics reflect a growing variability following the scale below:

$$\textit{intra-speaker} < \textit{inter-speaker} < \textit{inter-dialect}$$

Instead, samples labelled by different phoneticians seem to yield a lower degree of (inter-subject) variability for the “older metrics” than for the CCI. This was expected and very probably depends on the higher number of phonological choices demanded by this index: the need of dividing interval durations by the number of segments implies in fact an interpretation on phonological grounds, and therefore introduces a

---

<sup>127</sup> I shall repeat that, on the basis of its phonological properties, this language could be considered as a mixed-type, as it allows for a fairly complex syllable structure without presenting macroscopic phenomena of vocalic reduction.



## 7. Final discussion

further factor of possible “disagreement” between segmentators. This explains why inter-subject variability seems to be even higher than inter-speaker variability for the CCI, despite being lower than intra-speaker variability for the other metrics.

Returning to the variability scale, it is interesting to note that inter-regional variation is not distinguished from inter-speaker variability by the older metrics, while it is not distinguished from inter-dialect variation by the CCI. A rational explanation for these results is that the traditional measures reflect the phonotactics, which is essentially the same across samples of the same language: so, it is understandable that samples of different speakers of regional varieties are classified with these metrics by the same standards as samples of different speakers *tout court*. Instead, the CCI measures compensation phenomena, which, apparently, seem to be fairly variable across regional varieties<sup>128</sup>.

This finding is of interest and might have relevant implications, provided that the same results are replicated with more data, which will make statistical analysis possible. In fact, if it is confirmed, studies concentrating on the categorisation of regional varieties of a language (such as the one presented by Giordano & D’Anna, 2010) should accordingly favour the use of the CCI, while studies concerned with the categorisation of related dialectal varieties (such as Romano *et al.*, 2010) should perhaps opt for traditional metrics in order to obtain a clearer discrimination of data.

As for the need of integrating other parameters of prominence into a typological study of rhythm, the approach of calculating rhythm metrics on values of pitch and intensity has not confirmed expectations (which were based on the observation that stresses seem to be more prominent in stress-timed than in syllable-timed languages, see the end of chapter 3 for details). Unfortunately, the only viable data of which I disposed for such an experiment was constituted by 10 samples of 5 languages by an only speaker (Italian + 4 L2s). Although these very samples yielded good results with the metrics, the  $\Delta$ pitch and  $\Delta$ intensity chart seemed to categorise the two samples of Italian on one side and all L2 samples (English, German, Icelandic and French) on the other side. This poses severe interpretation problems: on the one hand, it could simply be that French has been miscategorised for some reasons; on the other hand, the fact that the categorisation separates L1 from all L2 is certainly suspicious. Methodological caveats (namely, the need for perfectly comparable data, also in terms of how the recordings are made) imply that no solution to this dilemma is available with my present corpus.

Undoubtedly, perceptive tests constitute the most controversial issue of this thesis. Initially, they had been conceived to provide a term of comparison to the results of the metrics in order to determine the validity of the latter (which have been defined as “acoustic correlates of the perception of rhythm” by Ramus, 1999). For this reason, they had been built mostly with syntheses of data from the corpus. However, the results have not provided evidence of a rhythmic categorisation of languages on the part of naive listeners, a conclusion that is at odds with studies by other authors (mainly F. Ramus, J. Mehler and colleagues). In a detailed discussion (see chapter 6) it has been suggested that such results may have at least three possible causes. Firstly, they might be attributed to listeners (after all, the test was administered only to 43 participants), but this is not likely because they provided

---

<sup>128</sup> I have no scientific proof that compensation phenomena are indeed more relevant within regional varieties. It is just an inference based on the CCI results and needs to be verified.

## 7. Final discussion

sensible answers in a preliminary and related task<sup>129</sup> and because in the last task they proved that they had not lost their concentration (see chapter 6 for more details). Secondly, results might depend on the design and the format of the test: in particular, it was perhaps too ambitious in that it tested the categorisation of too many languages and three different types of syntheses. Finally, it may depend on the fact that (like suggested by Arvaniti, 2010), the distinction between rhythm classes by naive listeners is not so salient as it is thought to be. An interesting observation concerns the fact that, curiously, results of perceptive tests in which the original pitch contour is preserved seem to yield worse results (in terms of categorisation) than to *flat SASASA* tests (see the discussion on the methodology above). On these grounds, I have put forward the hypothesis that  $f_0$  might then be a factor of disturbance in the rhythmic categorisation of languages (probably because listeners rely on the pitch contour and use it as a cue for language recognition). However, this suggestion has to be verified on experimental data; for this reason I am planning to reproduce the same test but with levelled  $f_0$ . Results should tell if listeners are really able to provide a better categorisation.

### 7.3 Future perspectives

After having analysed the past and present developments of research in rhythm typology, and at the light of all that has been said in the text, I shall briefly pass on to discuss future perspectives in this field. For this purpose, I shall mainly base myself on the proposition by Bertinetto & Bertini (2010).

The two authors claim that a natural language rhythm model should have 3 requirements: (a) explicitness, (b) predictivity and (c) unification. They proceed in analysing the various models proposed along the years for the study of rhythm and conclude that none of them possessed all three characteristics; I shall only deal with rhythm metrics, which are the main topic of this thesis. Rhythm metrics only provide an account of the first level of speech rhythm, *i.e.* the syllabic or segmental one; they provide no representation of the second level (within the stress or accentual domain)<sup>130</sup>. Furthermore, expectations are constructed on the basis of (1) the presence/absence of some phonological properties present in a language, whose individual contribution to the final value of the metrics is difficult to quantify; (2) on auditory impressions (whether by the author or traditionally accepted, but only rarely verified through perceptive tests). The results are analysed and interpreted mostly in relation to their correspondence to those auditory impressions and only in terms of relative positioning between points or regions of the charts. For these reasons, the two authors consider rhythm metrics not to fulfil criteria (c) and, at least partly, (b). They refer, in particular, to the fact that there is no way of predicting the values of the metrics on the basis of the phonological properties that they possess. Therefore, the exact position occupied by samples on the chart can only be commented on *a posteriori* and only in relative terms. In contrast, the expectation for the CCI is explicit in absolute terms: controlling languages should fall along

---

<sup>129</sup> Moreover, it has been shown that the results do not improve in a relevant way by only considering best performers.

<sup>130</sup> At least, this is valid for their customary use. However, a few authors have applied the metrics or other similar measures to higher levels (e.g. Asu & Nolan, 2006).

## 7. Final discussion

the bisector, compensating languages should fall below<sup>131</sup>. So, the authors affirm the need for a *unified predictive model*, which could account for both levels of the rhythmic structure. Multi-layer models have already been proposed (see, for instance, O'Dell and Nieminen, 1999, – reported above), but, interestingly, Bertinetto & Bertini (2010) suggest that the two levels be relevant and independent, so that a language can either control or compensate at each level. This brings then to the scheme resumed in table 7.1 (adapted from Bertinetto & Bertini, 2010).

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

**Table 7.1.** The quadripartition of languages according to level I and II (adapted from Bertinetto & Bertini, 2010).

I believe that this view (shared by works that adopted very different approaches, such as Asu & Nolan, 2006) exemplifies a new conception, in which the two levels are no longer seen in contraposition, but as two independent (though most probably interacting) continua. This means that rhythm typology has passed from a dichotomic conception, through a bi-polar scalar categorisation along a continuum, to a bidimensional scalar characterisation at the segmental and accentual levels<sup>132</sup>. Future perspectives, of course, include the possibility of merging the two levels into a unique multi-layer model.

### 7.4 Conclusion

Finally, for those who really cannot make without clear conclusions, the list below offers a concise reading:

- Rhythm metrics work on controlled samples. This does not mean that they have to be trusted blindly: although they provide a working classification of languages, they are far from exhausting the description of all aspects of speech rhythm. Furthermore, one can be sceptical about their use on data

<sup>131</sup> However, it could be said that, even for the CCI, predictions are not really formulated in absolute terms. This applies above all to the prediction for compensating languages, because there seems to be no practical way to establish how far from the bisector a certain sample should be in order to be classified as compensating.

<sup>132</sup> Mario Squartini (personal communication) sketched a stimulating comparison between the evolution of rhythm typology and other domains of linguistic typology, in particular morphological typology. Languages were at first considered on a continuum going from *isolating* through *agglutinating* to *inflectional*, until Comrie (1989) introduced two independent indices, the index of synthesis and the index of analysis (see also Payne, 1997) so that languages came to be classified on two independent continua. This does indeed look similar to recent developments in rhythm typology, which seems to move from a monodimensional to a bidimensional classification.

## 7. Final discussion

coming from spontaneous speech (which has not been treated in the thesis) or critical about their theoretical formulation.

- Results suggest that there is no such thing as “the best metric” for the categorisation of languages (at least, this applies if “best” means “more consistent with the traditional categories of rhythm typology”). Moreover, the CCI is grounded on different rationales and measures different phenomena, which makes comparison difficult. Different metrics provide partially different results and the choice of using one or the other might depend on data and on the aim of the study.
- Similarly, the choice of whether to apply a normalisation to avoid the effects of speech rate depends on the data and on the aim of the study. As a general rule, perfectly comparable and controlled data do not need normalisation, which might only cause loss of information. Conversely, researchers using spontaneous speech and great amounts and different types of data might prefer to normalise, above all if the aim of the study includes a categorisation of languages.
- The high variability of the rhythm metrics is certainly a drawback that worries researchers. Yet, when studied methodically, it has been shown to follow regular patterns. Intra-speaker variability is overall more limited than inter-speaker variability, which in turn is more limited than inter-dialectal variability. Other types of variability seem to depend on the metrics used. These results, however, still have to be validated on more data and on statistical analyses.
- The application of rhythm metrics to pitch and intensity values of selected samples has yielded unclear results. These two parameters are certainly linked to prominence, and therefore they should be included in an ambitious rhythm model. However, there is still indecision as to determining at which level(s) they interact.
- Perceptive tests administered to 43 participants in Italy have failed to provide evidence of naive listeners’ ability to categorise languages in rhythm classes. This result seems to be at odds with many other studies on this subject. Yet, a number of methodological issues have been raised, mainly concerning the choice of including vs. excluding the original pitch contours in the syntheses. On the basis of the observation of data and at the light of similar results obtained by Arvaniti (2010), I have put forward the hypothesis (yet to be verified) that  $f_0$  hinders more than helps a rhythmic categorisation of languages.
- The original bi-polar dichotomy has first evolved into a mono-dimensional scalar characterisation and is now shifting towards a bi-dimensional scalar representation. The two levels (the segmental or syllabic one, and the accentual one) are no longer seen in opposition and are both conceived as continua. I agree with other authors that perspectives in this field include a merging of the two levels into multi-layer models

**8.**

## **References**

## 8. References

- Abercrombie, D. (1967) *Elements of General Phonetics*, Edinburgh University Press.
- Allen, G. D. (1973) Segmental timing control in speech production. *Journal of Phonetics*, 1, 219-222.
- Allen, G. D. (1975) Speech Rhythm: its Relation to Performance Universals and Articulatory timing. *Journal of Phonetics*, 3, 75-86.
- Arvaniti, A. (1994) Acoustic features of Greek rhythmic structure. *Journal of Phonetics*, 22, 239-268.
- Arvaniti, A. (1999) Illustrations of the IPA: Modern Greek. *Journal of the International Phonetic Association*, 29(2), 167-172.
- Arvaniti, A. (2009) Rhythm, Timing and the Timing of Rhythm. *Phonetica*, 66, 46-63.
- Arvaniti, A. & Ross, T. (2010) Rhythm classes and speech perception. In: M. Hasegawa-Johnson (eds.) *Proceedings of Speech Prosody 2010*, Chicago (USA), 11-14 May 2010.
- Asu, E. L. & Teras, P. 2009. Illustrations of the IPA: Estonian. *Journal of the International Phonetic Association*, 39(3), 367-372.
- Asu, E. L. & Nolan, F. (2006) Estonian and English rhythm: a two-dimensional quantification based on syllables and feet. In: *Proceedings of Speech Prosody 2006*, Dresden (Germany), 2-5 May 2006.
- ATPM - Atlante Toponomastico del Piemonte Montano (2005-2008: 27-Roccaforte Ligure; 28-Briga Alta, 30-Exilles e 33-Capanne di Marcarolo).
- 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*, 34/2, 227-232.
- Barry, W. (2010) Rhythm measures in retrospect. Reflections on the nature of spoken-language rhythm. In: S. Schmid, M. Schwarzenbach & D. Studer (eds.) *La*

## 8. References

*dimensione temporale del parlato – Proceedings of the 5th AISV National Congress*, Zurich (Switzerland), 4-6 February 2009, Torriana (RN): EDK Editore.

Barry, W., & Russo M. (2003) Isocronia Soggettiva o Oggettiva? Relazioni tra Tempo Articolatorio e Quantificazione Ritmica. In: F. Albano-Leoni, F. Cutugno, M. Pettorino & R. Savy (eds.), *Il Parlato Italiano* (Atti del Convegno Nazionale di Napoli, 13-15 February 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 15<sup>th</sup> International Congress of Phonetic Sciences*, Barcelona (Spain), 3-9 August 2003, 2693-2696.

Bauer, L., Warren, P., Bardsley, D., Kennedy, M. & Major, G. (2007) Illustrations of the IPA: New Zealand English. *Journal of the International Phonetic Association*, 37(1), 97–102.

Benton, M. (2010) A Preliminary Analysis of the Relationship of Speech Rate to Speech-Timing Metrics as applied to Large Corpora of Non-Laboratory Speech in English and Chinese Broadcast News. In: M. Hasegawa-Johnson (eds.) *Proceedings of Speech Prosody 2010*, Chicago (USA), 11-14 May 2010.

Benton, M., Dockendorf, L., Jin, W., Liu, Y. & Edmonson, J.A. (2007) The Continuum of Speech Rhythm: computational testing of speech rhythm of large corpora from natural Chinese and English speech. In: *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrücken (Germany), 6-10 August 2007, 1269-1272.

Berruto, G. (1974), Piemonte e Valle d'Aosta. In: M. Cortelazzo (ed.) *Profilo dei dialetti italiani*, 1, Pisa: Pacini.

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

Bertinetto, P. M. (1980) The Perception of Stress by Italian Speakers. *Journal of Phonetics*, 8, 385-395.

Bertinetto, P. M. (1981) *Strutture prosodiche dell'italiano. Accento, quantità, sillaba, giuntura, fondamenti metrici*. Firenze: Accademia della Crusca.

Bertinetto, P. M. (1983) Ancora sull'Italiano come Lingua ad Isocronia Sillabica. In: *Scritti linguistici in onore di Giovan Battista Pellegrini*, Pisa, pp. 1073-82.

## 8. References

- Bertinetto, P. M. (1989) Reflections on the Dichotomy 'Stress' vs. 'Syllable-timing'. *Revue de Phonétique Appliquée*, Mons, pp. 99-130.
- Bertinetto, P. M. (1990) Coarticolazione e ritmo nelle lingue naturali. *Rivista Italiana di Acustica*, 16/2-3, 69-74.
- Bertinetto, P.M. & Bertini, C. (2008) On modeling the rhythm of natural languages. In: P.A. Barbosa, S. Madureira & C. Reis (eds.) *Proceedings of Speech Prosody 2008*, Campinas (Brazil), 6-9 May 2008, 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.
- Bertinetto, P.M. & Magno Caldognetto, E. (1993). Ritmo e intonazione. In A.A. Sobrero (ed.), *Introduzione all'italiano contemporaneo. Le strutture*. Manuali Laterza 43, 2, Roma-Bari: Laterza, 141-192.
- Bertinetto, P.M. & Vékás, D. (1991). Controllo vs. compensazione sui due tipi di isocronia. In: Magno Caldognetto E. & Benincà P. (eds.), *L'interfaccia tra fonologia e fonetica*, Padova: Unipress, 155-162.
- Bertini, C. & Bertinetto, P.M. (2009) Prosperezioni sulla struttura ritmica dell'italiano basate sul corpus semispontaneo AVIP/API. In: L. Romito, V. Galatà & R. Lio (eds.) *La fonetica sperimentale: metodo e applicazioni – Proceedings of the 4th AISV National Congress*, Arcavacata di Rende CO (Italy), 3-5 December 2007, Torriana (RN): EDK Editore.
- Böðvarsson, À. (1953) *Hljóðfræði: Kennslubók Handa Byrjendum*, Reykjavík: Ísafoldarprentsmiða H. F.
- Bolinger, D. L. (1965) Pitch Accent and Sentence Rhythm. In: Abe, I. & Kanekijo, T. (ed.) *Forms of English*. Cambridge, MA: Harvard University Press.
- Borzzone de Manrique, A. M. & Signorini, A. (1983) Segmental duration and rhythm in Spanish. *Journal of Phonetics*, 11, 117-128.
- Bouzon, C. & Hirst, D. (2004) Isochrony and Prosodic Structure in British English. In: *Proceedings of Speech Prosody 2004*, Nara (Japan), 23-26 March 2004.



## 8. References

- Canepari, L. (2003a) *Manuale di Fonetica: Fonetica Naturale, Articolatoria, Uditiva e Funzionale*. Munich: Lincom.
- Canepari, L. (2003b) *Manuale di Pronuncia: Italiana, Inglese, Francese, Tedesca, Spagnola, Portoghese, Russa, Araba, Hindi, Cinese, Giapponese, Esperanta*. Munich: Lincom.
- Classe, A. (1939) *The Rhythm of English Prose*, Oxford: Basil Blackwell.
- Comrie, B. (1989) *Language Universals and Linguistic Typology (2nd edition)*. Chicago: University Press.
- Contini, M., Lai, J.-P., Romano, A., Roulet, S., de Castro Moutinho, L., Coimbra, R.L., Pereira Bendiha, U. & Secca Ruivo, S. (2002) Un projet d'atlas multimédia prosodique de l'espace roman. In: B. Bel & I. Marlien (eds.) *Proceedings of the International Conference Speech Prosody 2002*, Aix-en-Provence (France), 11-13 April 2002, 227-230.
- Coşeriu, E. (1958) *Sincronía, diacronía e historia: el problema del cambio lingüístico*. Madrid, Gredos.
- Costamagna, L. (2000) *Insegnare e imparare la fonetica*. Torino: Paravia.
- Cox, F. & Palethorpe, S. (2007) Illustrations of the IPA: Australian English. *Journal of the International Phonetic Association*, 37(3), 341–350.
- Cruz-Ferreira, M. (1999) Illustrations of the IPA: European Portuguese. In: IPA (1999), 126–130.
- Crystal, D. (1994) Documenting rhythmical change. In: J. Windsor Lewis (ed.), *Studies in general and English phonetics*. London: Routledge, 174-179.
- Cumming, R. (2008) Should rhythm metrics take account of fundamental frequency? *Cambridge Occasional Papers Ling.*, 4, 1-16.
- Cummins, F. (2002) Speech Rhythm and Rhythmic Taxonomy. In: B. Bel & I. Marlien (eds.) In: B. Bel & I. Marlien (eds.) *Proceedings of Speech Prosody 2002*, Aix-en-Provence (France), 11-13 April 2002.

## 8. References

- Cummins, F. (2007) Speech synchronization: investigating the links between perception and action in speech production. In: J. Trouvain & W. Barry (eds.), *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrücken (Germany), 6-10 August 2007.
- Dankovicová, J. & Dellwo, V. (2007) Czech speech rhythm and the rhythm class hypothesis. In: J. Trouvain & W. Barry (eds.), *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrücken (Germany), 6-10 August 2007, 1241-1244.
- Dauer, R. M. (1983) Stress-timing and Syllable-timing Reanalysed. *Journal of Phonetics*, 11, 51-62.
- Dauer, R. M. (1987) Phonetic and Phonological Components of Language Rhythm. In *Proceedings of the 11<sup>th</sup> International Congress of Phonetic Sciences*, Tallinn, 5, 447-450.
- Dellwo, V. (2006) Rhythm and Speech Rate: a variation coefficient for  $\Delta C$ . In: Language and Language Processing. *Proceedings of the 38<sup>th</sup> Linguistic Colloquium Piliscsaba*, 231-241.
- Dellwo, V. (2008) The role of speech rate in perceiving speech rhythm. In: P.A. Barbosa, S. Madureira & C. Reis (eds.) *Proceedings of Speech Prosody 2008*, Campinas (Brazil), 6-9 May 2008, 375-378.
- Dellwo, V. (2010) Choosing the right rate normalization method for measurements of speech rhythm. In: S. Schmid, M. Schwarzenbach & D. Studer (eds.) *La dimensione temporale del parlato – Proceedings of the 5th AISV National Congress*, Zurich (Switzerland), 4-6 February 2009, Torriana (RN): EDK Editore.
- Dellwo, V., Fourcin, A. & Abberton, E. (2007) Rhythmical classification of languages based on voice parameters. In: J. Trouvain & W. Barry (eds.), *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrücken (Germany), 6-10 August 2007, 1129-1132.
- Dellwo, V. Steiner, I., Aschenberner, B., Dankovicova & Wagner, P. (2004) BonnTempo-Corpus and BonnTempo-Tools: a database for the study of speech rhythm and rate. In: *Proceedings of the 15<sup>th</sup> International Congress of Phonetics Sciences*, Barcelona (Spain), 3-9 August 2003.

## 8. References

- Dellwo, V. & Wagner, P. (2003) Relations between language rhythm and speech rate. In: *Proceedings of the 15<sup>th</sup> International Congress of Phonetics Sciences*, Barcelona (Spain), 3-9 August 2003, 471-474.
- Deterding, D. (2001) The measurement of rhythm: a comparison of Singapore and British English. *Journal of Phonetics*, 29, 217-230.
- Dufter, A., & Reich, U. (2003) Rhythmic differences within Romance: identifying French, Spanish, European and Brazilian Portuguese. In: *Proceedings of the 15<sup>th</sup> International Congress of Phonetics Sciences*, Barcelona (Spain), 3-9 August 2003, 471-474.
- Engstrand, O. (1999) Illustrations of the IPA: Swedish. In: IPA (1999), 140–142.
- Engstrand, O. & Krull, D. (2003) Rhythmic intentions or rhythmic consequences? Cross-language observations of casual speech. In: *Proceedings of the 15<sup>th</sup> International Congress of Phonetics Sciences*, Barcelona (Spain), 3-9 August 2003.
- Eriksson, A. (1991) *Aspects of Swedish Speech Rhythm*. Doctoral Dissertation, University of Göteborg.
- Fagyal, Zs., & Moisset, C. (1999) Sound Change and Articulatory Release : Where and Why Are High Vowels Devoiced in Parisian French? In: *Proceedings of the 14<sup>th</sup> International Congress of Phonetic Sciences*, San Francisco (USA), August 1999, 309-312.
- 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.
- Felloni, M. C. (2010) *Uno studio sociofonetico sulla intonazione. Produzione e percezione della interrogativa globale nell'italiano regionale di Parma*. PhD Dissertation, University of Pavia.

## 8. References

- Fikkert, P., Freitas, M.J., Grijzenhout, J., Levelt, Cl. & Wauquier S. (2004), Syllabic Markedness, Segmental Markedness, Rhythm and Acquisition, Talk presented at GLOW, Thessaloniki (Greece), 2004.
- Fonágy, I. (1989) Le Français Change de Visage? *Revue Romane*, 24/2, 225-254.
- Fougeron, C. & Smith, C. L. (1999) Illustrations of the IPA: French. In: IPA (1999), 78-81.
- Fowler, C.A. (1977) *Timing Control in Speech Production*, Indiana University Linguistics Club.
- Fowler, C. A. (1980) Coarticulation and theories of extrinsic timing, *Journal of Phonetics*, 8, 113-133.
- Fowler, C. A. (1981) Production and perception of coarticulation among stressed and unstressed vowels, *Journal of Speech and Hearing Research*, 46, 127-139.
- Frota, S., Vigário, M. & Martins, F. (2002) "Language Discrimination and Rhythm Classes: Evidence from Portuguese". In: B. Bel & I. Marlien (eds.) *Proceedings of Speech Prosody 2002*, Aix-en-Provence (France), 11-13 April 2002, 315-318.
- Frota, S., Vigário, M. & Freitas, M.J. (2003) From Signal to Grammar: Rhythm and the Acquisition of Syllable Structure. In: B. Beachley, A. Brown & F. Conlin (eds.), *BUCLD 27: Proceedings of the 27<sup>th</sup> annual Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press, 809-821.
- Galves, A., Garcia, J.E., Duarte, D., Galves, C. (2002) Sonority as a Basis for Rhythmic Class Discrimination. In: *Proceedings of Speech Prosody 2002*, Aix-en-Provence (France), 11-13 April 2002.
- Genre, A. (1980) Le parlate occitano-alpine d'Italia, *Rivista Italiana di Dialettologia*, 4, 305-310.
- Ghazali, S., Hamdi, R. & Barkat M. (2002) Speech Rhythm Variation in Arabic Dialects. In: B. Bel & I. Marlien (eds.) *Proceedings of Speech Prosody 2002*, Aix-en-Provence (France), 11-13 April 2002, 331-334.
- Gibbon, D. & Gut, U. (2001) Measuring speech rhythm. In: *Proceedings of Eurospeech 2001*, Aalborg (Denmark), September 2001, 95-98.

## 8. References

- Giordano, R. (2008) On the phonetics of rhythm of Italian: patterns of duration in pre-planned and spontaneous speech. In: P.A. Barbosa, S. Madureira & C. Reis (eds.) *Proceedings of Speech Prosody 2008*, Campinas (Brazil), 6-9 May 2008, 74-77.
- Giordano, R. & D'Anna, L. (2010) A comparison of rhythm metrics in different speaking styles and in fifteen regional varieties of Italian. In: M. Hasegawa-Johnson (eds.) *Proceedings of Speech Prosody 2010*, Chicago (USA), 11-14 May 2010.
- Grabe, E. (2002). Variation Adds to prosodic Typology. In: B. Bel & I. Marlien (eds.) *Proceedings of Speech Prosody 2002*, Aix-en-Provence (France), 11-13 April 2002, 127-132.
- Grabe, E., & Low, E. L. (2002) Durational Variability in Speech and the Rhythm Class Hypothesis. In: Gussenhoven, C., Warner, N. (eds.), *Papers in Laboratory Phonology 7*, Berlin: Mouton de Gruyter, 515-546.
- Grønnum, N. (1998) Illustrations of the IPA: Danish. *Journal of the International Phonetic Association*, 28(1-2), 99-105.
- Gussenhoven, C. (1999) Illustrations of the IPA: Dutch. In: IPA (1999), 74-77.
- Hayes, B. (1984). The phonology of rhythm in English. *Linguistic Inquiry*, 15, 1.
- Hayward, K. & Hayward, R.J. (1999) Illustrations of the IPA: Amharic. In: IPA (1999), 45-50.
- Helgason Helgason, P. (2002) *Preaspiration in the Nordic Languages: Synchronic and Diachronic Aspects*. PhD dissertation, Stockholm University.
- Hillenbrand, J. M. (2003) Illustrations of the IPA: American English: Southern Michigan. *Journal of the International Phonetic Association*, 33(1), 121-126.
- Hoeqvist, C. J. (1983a) Syllable duration in stress-, syllable- and mora-timed languages, *Phonetica*, 40, 203-237.
- Hoeqvist, C. (1983b) Durational correlates of linguistic rhythm categories. *Phonetica* 40, 19-31.

## 8. References

- Interlandi G. M. (2003) La percezione dell'intonazione torinese: risultati di un test di riconoscimento. In: Marotta G. & Nocchi N. (eds.), *La coarticolazione – Proceedings of the XIII Giornate di Studio del Gruppo di Fonetica Sperimentale* (GFS 2002, Pisa, 28-30 novembre 2002), Pise: Edizioni ETS, 193-201.
- Interlandi G. M. & Romano A. (2004) Le continuum intonatif de l'italien parlé a Turin: résultats d'un test d'identification. In: *Proceedings of the MIDL Workshop - Identification des langues et des variétés dialectales par les humains et par les machines*, Paris, 29-30 November 2004, Presses de l'ENST, 157-60.
- IPA (1949) *The Principles of the International Phonetic Association*. London: University College (reprint 1966).
- IPA (1999) *Handbook of the International Phonetic Association*. Cambridge: University Press (see also the webography below).
- Jassem, W. (2003) Illustrations of the IPA: Polish. *Journal of the International Phonetic Association*, 33(1), 103–107.
- Jian, H. (2004) On the syllable timing in Taiwan English. In: *Proceedings of Speech Prosody 2004*, Nara (Japan), 23-26 March 2004, 247-250.
- Keller, E. & Port, R. (2007) Speech timing: Approaches to speech rhythm. In: J. Trouvain & W. Barry (eds.), *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrücken (Germany), 6-10 August 2007, 327-329.
- Kohler, K. J. (2009) Rhythm in Speech and Language. A new research paradigm, *Phonetica*, 66, 29-45.
- Kori, Sh. & Farnetani, E. (1981) Word stress perception in Italian bisyllables. *Proceedings of the 4<sup>th</sup> FASE Symposium on Acoustics and Speech*, 1, Rome: Edizioni Scientifiche, 53-56.
- Kristinsson, A. P. (1988) *The Pronunciation of Modern Icelandic: a Brief Course for Foreign Students*, Reykjavík: Málvísindastofnun Háskóla Íslands (3<sup>rd</sup> edition).
- Krull, D. & Engstrand, O. (2003) Speech rhythm – intention or consequence? Cross-language observations on the hyper/hypo dimension”. *Phonum*, 9, 133-136.
- Labov, W. (1972) *Sociolinguistic patterns*. Oxford: Blackwell.

## 8. References

- Ladefoged, P. (1999) Illustrations of the IPA: American English. In IPA (1999), 41-44.
- Ladefoged, P., & Maddieson, I. (1996) *The Sounds of the World's Languages*, Oxford: Blackwell.
- 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*, 93/3, 225-254.
- Lee (2010) Speech rhythm and word segmentation: a prominence-based account of some crosslinguistic differences. In: S. Schmid, M. Schwarzenbach & D. Studer (eds.) *La dimensione temporale del parlato – Proceedings of the 5th AISV National Congress*, Zurich (Switzerland), 4-6 February 2009, Torriana (RN): EDK Editore.
- Lehiste, I. (1977) Isochrony reconsidered. *Journal of Phonetics*, 5, 253-263.
- Lieberman M. & Prince A. (1977) On Stress and Linguistic Rhythm. *Linguistic Inquiry*, 8, 249-336. Now also in Ch.W. Kreidler (ed.), *Phonology: Critical concepts*, London-New York: Routledge, 2001, 152-244.
- Lindblom, B. & Rapp, K. (1973) Some temporal regularities of spoken Swedish. *Papers in Linguistics from the University of Stockholm*, 21, 1-59.
- Loporcaro, M. (2009) *Profilo linguistico dei dialetti italiani*. Roma-Bari: Laterza.
- Lloyd James, A. (1940) *Speech signal in telephony*. London: Pitman & Sons.
- Loukina, A., Kochanski, G., Shih, C., Keane, E., Watson, I. (2009) Rhythm measures with language independent segmentation. In: *Proceedings of Interspeech 2009*, 1931-1934.
- Maiden, M. & Mair Parry, M. (1997) *The Dialects of Italy*. London: Routledge.
- Mairano, P. & Romano, A. (2007) Inter-Subject Agreement in Rhythm Evaluation for Four Languages (English, French, German, Italian). In: J. Trouvain & W. Barry (eds.), *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrücken (Germany), 6-10 August 2007, 1149-1152.

## 8. References

- Mairano, P., & Romano, A. (2008a) Lingue isosillabiche e isoaccentuali: misurazioni strumentali su campioni di italiano, francese, inglese e tedesco. In: V. Giordani, V. Bruseghini & P. Cosi (eds.), *Scienze Vocali e del linguaggio - Metodologie di valutazione e risorse linguistiche - Proceedings of the 3rd AISV National Congress*, ITC-IRST Povo TN (Italy), 29 November - 1 December 2006.
- Mairano, P. & Romano, A. (2008b) Distances rythmiques entre variétés romanes”. In: A Turculeț (ed.), *La variation diatopique de l’intonation dans le domaine roumain et roman*, Iași: University Press A.I. Cuza, 251-272.
- Mairano, P. & Romano, A. (2010a) Un confronto tra diverse metriche ritmiche usando Correlatore. In: S. Schmid, M. Schwarzenbach & D. Studer (eds.) *La dimensione temporale del parlato – Proceedings of the 5th AISV National Congress*, Zurich (Switzerland), 4-6 February 2009, Torriana (RN): EDK Editore, 79-100.
- Mairano, P. & Romano, A. (2010b) Variabilité rythmique des variétés régionales. *Géolinguistique*, 12, 45-57.
- Mairano, P. & Romano, A. (unpublished) A comparison of four rhythm metrics for six languages”. *Poster presented to the workshop “Empirical Approaches to Speech Rhythm”* (University College London, 2008).
- Major, R. C. (1981) Stress-timing in Brazilian Portuguese. *Journal of Phonetics*, 9, 343-351.
- Marotta, G. (1985) *Modelli e misure ritmiche: la durata vocalica in italiano*. Bologna: Zanichelli.
- Martínez Celdrán, E., Fernández Planas, A.M. & Carrera Sabaté, J. (2003) Castilian Spanish. *Journal of International Phonetic Association*, 33/2, 255-259.
- Mehler, J., Dupoux, E., Nazzi, T. & Dahan-Lambertz, G. (1996) Coping with linguistic diversity: the infant’s viewpoint. In: J. L. Morgan & K. Demuth (eds.), *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition*, Mahwah, NJ: Lawrence Erlbaum Associates, 101-116.
- Meireles, A. R. & Barbosa, P. A. (2008) Speech rate effects on speech rhythm. In: P.A. Barbosa, S. Madureira & C. Reis (eds.) *Proceedings of Speech Prosody 2008*, Campinas (Brazil), 6-9 May 2008.



## 8. References

- Meireles, A. R., Tozetti, J. P & Borges, R. R. (2010) Speech rate and rhythmic variation in Brazilian Portuguese. In: M. Hasegawa-Johnson (eds.) *Proceedings of Speech Prosody 2010*, Chicago (USA), 11-14 May 2010.
- 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.
- Mitchell, T. (1969) Review of Abercrombie (1967). *Journal of Linguistics*, 5, 153-164.
- Mok, P.P.K. & Dellwo, V. (2008) Comparing native and non-native speech rhythm using acoustic rhythmic measures: Cantonese, Beijing Mandarin and English. In: P.A. Barbosa, S. Madureira & C. Reis (eds.) *Proceedings of Speech Prosody 2008*, Campinas (Brazil), 6-9 May 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: *Proceedings of the II workshop "Syllables"*, Nantes (France), 148-153.
- Morton, J., Marcus, S. M. & Frankish, C. (1976) Perceptual centers (P-centers). *Psychological Review*, 83, 405-408.
- 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(3), 756-766.
- Nespor, M. (1993) *Le strutture del linguaggio. Fonologia*. Bologna: Il Mulino.
- Nooteboom, S. (1997) The prosody of speech: melody and rhythm. In: W.J. Hardcastle & J. Laver (eds.), *The Handbook of Phonetic Sciences*, Oxford: Blackwell, 640-673.
- O'Dell, M. L., Lennes, M. & Nieminen, T. (2008) Hierarchical Levels of Rhythm in Conversational Speech. In: P.A. Barbosa, S. Madureira & C. Reis (eds.) *Proceedings of Speech Prosody 2008*, Campinas (Brazil), 6-9 May 2008.

## 8. References

- O'Dell, M. L. & Nieminen, T. (1999) Coupled oscillator model of speech rhythm. *Proceedings of the 14<sup>th</sup> International Congress of Phonetic Sciences*, San Francisco (USA) August 1999, 1075-1078.
- O'Dell, M. L., Nieminen, T. & Mustanoja, L. (2010) Assessing Rhythmic Differences with Synchronous Speech. In: M. Hasegawa-Johnson (eds.) *Proceedings of Speech Prosody 2010*, Chicago (USA), 11-14 May 2010.
- O'Rourke, E. (2008) Speech rhythm variation in dialects of Spanish: Applying the Pairwise Variability Index and Variation Coefficients to Peruvian Spanish. In: P.A. Barbosa, S. Madureira & C. Reis (eds.) *Proceedings of Speech Prosody 2008*, Campinas (Brazil), 6-9 May 2008.
- Pamies Bertrán, A. (1999) Prosodic Typology: On the Dichotomy between *Stress-Timed* and *Syllable-Timed* Languages. *Language Design*, 2, 103-130.
- Payne, T.E. (1997) *Describing morphosyntax. A guide for field linguists*. Cambridge: University Press.
- Pike, K. L. (1945) *The Intonation of American English*. Ann Arbor. University of Michigan Press.
- Pointon, G. E. (1980) Is Spanish really syllable-timed? *Journal of Phonetics*, 8, 293-304.
- Ramus, F. (2002) Acoustic Correlates of Linguistic Rhythm: Perspectives. In *Proceedings of Speech Prosody 2002*, Aix-en-Provence (France), 11-13 April 2002, 155-120.
- Ramus, F. Dupoux, E. & Mehler, J. (2003) The psychological reality of rhythm class: perceptual studies. In: *Proceedings of the 15<sup>th</sup> International Congress of Phonetic Sciences*, Barcelona (Spain), 3-9 August 2003, 337-342.
- Ramus, F. & Mehler, J. (1999) Language identification with suprasegmental cues: a study based on speech resynthesis. *Journal of the Acoustic Society of America*, 105 (1), 512-521.
- Ramus, F., Nespors, M. & Mehler, J. (1999) Correlates of Linguistic Rhythm in the Speech Signal. *Cognition*, 73/3, 265-292.

## 8. References

- Roach, P. (1982) On the Distinction between 'Stress-timed' and 'Syllable-timed' Languages. In D. Crystal, *Linguistic controversies*, London: Edward Arnold, 73-79.
- Roach, P. (2004) British English: Received Pronunciation. *Journal of the International Phonetic Association*, 34/2, 239-245.
- Rogers, D., & D'Arcangeli, L. (2004) Illustrations of the IPA: Italian. *Journal of the International Phonetic Association*, 34/1, 117-121.
- Rögnvaldsson, E. (1993) *Íslensk hljóðkerfisfræði*. Reykjavík: Málsvísindastofnun Háskóla Íslands.
- Rögnvaldsson, E. (2003) *Phonetics Transcription Guideline: Icelandic*. ScanSoft Inc.
- 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* (partially published in 2001, Lille: Presses Universitaires du Septentrion).
- Romano, A. (2003) Accento e intonazione in un'area di transizione del Salento centro-meridionale. In: P. Radici Colace, G. Falcone & A. Zumbo (eds.), *Storia politica e storia linguistica dell'Italia meridionale - Atti del convegno internazionale di studi parlangeliani*, Messina (Italy) 2000, Messina-Napoli: Ed. Scientifiche Italiane, 169-181.
- Romano, A. (2010) Speech Rhythm and Timing: Structural Properties and Acoustic Correlates. In: S. Schmid, M. Schwarzenbach & D. Studer (eds.) *La dimensione temporale del parlato – Proceedings of the 5th AISV National Congress*, Zurich (Switzerland), 4-6 February 2009, Torriana (RN): EDK Editore, 45-75.
- Romano, A., Mairano, P. (2010) Speech rhythm measuring and modelling: pointing out multi-layer and multi-parameter assessments. In: Russo (2010), 79-116.
- Romano, A., Mairano, P. & Pollifrone, B. (2010) Variabilità ritmica di varietà dialettali del Piemonte. In S. Schmid, M. Schwarzenbach & D. Studer (eds.) *La dimensione temporale del parlato – Proceedings of the 5th AISV National Congress*, Zurich (Switzerland), 4-6 February 2009, Torriana (RN): EDK Editore, 101-112.
- 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.

## 8. References

- Rouas, J. L., & Farinas, J. (2004) Comparaison des méthodes de caractérisation du rythme des langues. *Proceedings of MIDL 2004 - Identification des langues et des variétés dialectales par les humains et par les machines*, Paris, 2004, Paris: École Nationale Supérieure des Télécommunications, 45-50.
- Russo, M. (2010 ed.) *Prosodic Universals. Comparative studies in rhythmic modeling and rhythm typology*. Rome: Aracne.
- Russo, M. & Barry, W.J. (2008a) Measuring rhythm. A quantified analysis of Southern Italian Dialects Stress Time Parameters. In: A. Pamies, M.C. Amorós & J.M. Pazos (eds.), *Experimental Prosody - Proceedings of the 4th Congreso Internacional 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: P.A. Barbosa, S. Madureira & C. Reis (eds.) *Proceedings of Speech Prosody 2008*, Campinas (Brazil), 6-9 May 2008, 52-55.
- Schmid, S. (1996) A Typological View of Syllable Structure in some Italian Dialects. In: Bertinetto, P. M. et al. (eds.), *Certamen Phonologicum, Atti del 3° Colloquio di Fonologia*, Cortona, 1996, 247-265.
- Schmid, S. (2001) Un nouveau fondement phonétique pour la typologie rythmique des langues? In: *Ecrits pour le 10e Anniversaire du Laboratoire d'analyse informatique de la parole (LAIP)*, Lausanne: 2001.
- Schmid, S. (2004) Une approche phonétique de l'isochronie dans quelques dialectes italo-romans. In: Meisenburg, T., Selig, M. (eds.) *Nouveaux départs en phonologie*, Tübingen: G. Narr, 109-124.
- Schmid, S. (unpublished) Measuring the rhythm of Italian dialects. *Poster presented at the EASR 2008 workshop “Empirical Approaches to Speech Rhythm”*, University College London, 2008.
- Soderberg, C. D. & Kenneth S. O. (2008) Illustrations of the IPA: Indonesian. *Journal of the International Phonetic Association*, 38(2), 209–213.
- Sornicola, R. (1981) *Sul parlato*. Bologna: Il Mulino.

## 8. References

- Telmon, T. (1988), Areallinguistik II. Piemont. In: G. Holtus, M. Metzeltin & Ch. Schmitt (eds.) *Lexikon der Romanistischen Linguistik*, Vol. IV, Tübingen: Niemeyer, 469-485.
- Telmon, T. (2001), Piemonte e Valle d'Aosta. In: A.A. Sobrero (ed.) *Profili linguistici delle regioni*, Bari: Laterza.
- Thelwall, R. & Akram Sa'adeddin, M. (1999) Illustrations of the IPA: Arabic. In: IPA (1999), 51–54.
- Thráinsson, H. (1978), On the Phonology of Icelandic Preaspiration. *Nordic Journal of Linguistics*, 1, 3-54.
- Tortel, A & Hirst, D. (2008) Rhythm and Rhythmic Variation in British English: Subjective and Objective Evaluation of French and Native Speakers. In: P.A. Barbosa, S. Madureira & C. Reis (eds.) *Proceedings of Speech Prosody 2008*, Campinas (Brazil), 6-9 May 2008.
- Tortel, A. & Hirst, D. (2010) Rhythm metrics and the production of English L1/L2. In: M. Hasegawa-Johnson (eds.) *Proceedings of Speech Prosody 2010*, Chicago (USA), 11-14 May 2010.
- Vayra, M., Avesani, C. & Fowler, C. (1984) Patterns of Temporal Compression in Spoken Italian. In: *Proceedings of the 10<sup>th</sup> International Congress of Phonetic Sciences*, Utrecht, 1984.
- Vayra, M., Fowler, C. & Avesani, C. (1987) Word-level Coarticulation and Shortening in Italian and English Speech. *Status Report on Speech Research*, Haskins Laboratories, 91, 75-89. Also in *Studi di Grammatica Italiana*, 13, 249-69.
- Verhoeven, J. (2005) Illustrations of the IPA: Dutch, Belgian Standard. *Journal of the International Phonetic Association*, 35(2), 243–247.
- Wagner, P. (2007) Visualizing Levels of Rhythmic Organisation. In: J. Trouvain & W. Barry (eds.), *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrücken (Germany), 6-10 August 2007.
- Wagner, P. & Dellwo, V. (2004) Introducing YARD (Yet Another Rhythm Determinator) and Re-Introducing Isochrony to Rhythm Research. In: *Proceedings of Speech Prosody 2004*, Nara (Japan), 23-26 March 2004.

## 8. References

- Watson, K. (2007) Illustrations of the IPA: Liverpool English. *Journal of the International Phonetic Association*, 37(3), 351–360.
- Watt, D. & Allen, W. (2003) Illustrations of the IPA: Tyneside English. *Journal of the International Phonetic Association*, 33(2), 267–271.
- Wells, J. C (2000) *Longman Pronunciation Dictionary*, London: Longman.
- Wenk, B. & Wioland, F. (1982) Is French really syllable-timed? *Journal of Phonetics*, 10, 193-216.
- White, L., Payne, E., & Mattys, S.L. (2009) Rhythmic and prosodic contrast in Venetan and Sicilian Italian. In M. Vigario, S. Frota & M.J. Freitas (eds.), *Phonetics and Phonology: Interactions and Interrelations*, Amsterdam: John Benjamins, 137-158.
- White, L., & Mattys, S.L. (2007) Calibrating rhythm: First language and second language studies. *Journal of Phonetics*, 35, 501-522.
- White, L., Mattys, S.L., Series, L., & Gage, S. (2007) Rhythm metrics predict rhythmic discrimination. In: J. Trouvain & W. Barry (eds.), *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrücken (Germany), 6-10 August 2007, 1009-1012.
- White, L., Widget, L. & Mattys, S. (unpublished) *How stable are rhythm metrics?* Poster presented at EASR 2008, UCL, London, 28<sup>th</sup> March 2008.
- Widget, L., White, L., Schuppler, B., Grenon, I., Rauch, O., & Mattys, S.L. (2010) How stable are acoustic metrics of contrastive speech rhythm? *Journal of the Acoustical Society of America*, 127, 1559-1569.
- Yoon, T. J. (2010) Capturing Inter-speaker Invariance Using Statistical Measures of Rhythm. In: M. Hasegawa-Johnson (eds.) *Proceedings of Speech Prosody 2010*, Chicago (USA), 11-14 May 2010.

## 8. References

### **Webography**

<http://web.uvic.ca/ling/resources/ipa/handbook.htm>

Sound files meant as illustrations to IPA (1999).

<http://web.uvic.ca/ling/resources/ipa/members>

Sound files meant as attachments to the Illustrations of the IPA published on the *Journal of the International Phonetic Association* (sound samples available on-line for IPA members only).

<http://www.lfsag.unito.it/>

The homepage of *LFSAG* (Laboratorio di Fonetica Sperimentale “Arturo Genre”, Università degli Studi di Torino).

[http://www.lfsag.unito.it/correlatore/index\\_en.html](http://www.lfsag.unito.it/correlatore/index_en.html)

The homepage of the *Correlatore* Software, which I developed within my PhD (see chapter 4 for more details).

[http://www.lfsag.unito.it/ritmo/index\\_en.html](http://www.lfsag.unito.it/ritmo/index_en.html)

A presentation of research in speech rhythm carried out at the Laboratory of Phonetics, Turin.

<http://www.personal.reading.ac.uk/~llsroach/timing.pdf>

A 54-pages-long bibliography of rhythm speech and rhythm timing compiled by Prof. Peter Roach (University of Reading). Last update 2<sup>nd</sup> April 2003.

<http://www.unive.it/canepari>

A website by Prof. Luciano Canepari (University of Venice). *Fonetica Naturale – Natural Phonetics* (on-line sound samples on Italian pronunciation) (last accessed 30/06/2009).

<http://w3.u-grenoble3.fr/dialecto/AMPER/amper.htm>

Homepage of the AMPER projet (Atlas Multimedia Prosodique de l’Espace Roman).

**9.**

# **Appendices**



## 9. Appendices

### Appendix 1: inter-onset values

Inter-onset values for each sample analysed in the experiment presented in chapter 2 are reported in the table below (ordered by increasing values of the metrics).

	delta		varco		rPVI		nPVI
94,87	Eng_GA	47,11	Eng_GA	83,31	Italian11	43,42	Italian12
109,88	Greek	62,11	Polish	86,86	Italian1	43,65	Italian1
111,23	Italian11	63,91	Port_Lisbon	90,21	Greek	45,39	Finnish1
113,74	Italian1	64,16	Spanish_Lima	92,88	Italia15	45,74	Eng_GA
115,43	Spanish_Lima	66,11	Italian1	94,27	Italian6	46,15	Russian_ss
117,87	Italian6	68,01	Greek	96,34	Eng_GA	46,70	Greek
125,96	Italia15	68,14	Italian11	96,75	Italian9	47,04	Polish
141,41	Italian13	72,30	Italian6	100,59	Italian12	47,54	Spanish_Lima
145,86	Italian3	73,57	Italian2	100,64	Spanish_Lima	47,93	Spanish_Caracas
147,40	Eng_NZE	73,60	Eng_Aus	103,51	Italian7	48,11	Italian10
148,20	Port_Lisbon	73,88	French_IPA	106,83	Italian8	48,65	Italian11
148,32	Italian9	74,09	Port_SaoPaulo	108,92	Italian13	48,70	Japanese
148,73	Dutch	74,32	Italian10	114,31	Italian3	48,85	Port_SaoPaulo
149,31	Italian12	74,52	Eng_NZE	114,82	Italian10	48,87	French_IPA
149,92	Italian2	74,89	Dutch	114,95	Italian2	48,93	Italian7
151,11	Polish	75,75	Finnish2	125,31	Spanish_Bogota	48,94	Italian2
152,60	French_IPA	76,15	Italian3	125,66	Spanish_Caracas	49,19	Finnish2
153,23	Italian7	76,27	Italia15	126,10	Italian14	49,28	Port_Lisbon
153,74	Italian8	77,25	Italian13	127,03	Eng_NZE	49,67	Estonian
153,95	Italian10	77,72	Italian14	127,07	French_IPA	49,97	Italian9
160,63	Spanish_Bogota	79,14	Arabic Lebanese	128,34	Finnish2	50,25	Spanish_Bogota
164,20	Finnish2	79,28	Italian12	132,61	Port_Lisbon	50,28	Italian8
165,21	Italian14	80,40	Spanish_Bogota	133,31	Polish	51,14	Italia15
170,31	Port_SaoPaulo	80,42	Italian5	138,51	Port_SaoPaulo	52,59	Arabic_IPA
170,47	Spanish_Caracas	81,89	Arabic_IPA	143,88	Dutch	52,73	Italian3
176,04	Arabic_IPA	82,16	Italian8	145,94	Arabic_IPA	52,96	Port_Manauas
183,57	Eng_Aus	84,05	Italian7	149,39	Finnish1	53,14	Italian14
195,90	Italian5	84,43	German1	155,82	Russian_ss	53,38	Eng_NZE
203,70	German1	85,45	Italian9	163,22	Eng_Aus	53,64	Italian13
214,37	Russian_ss	85,84	Russian_ss	165,07	Spanish_IPA	53,69	Czech
217,61	Arabic Lebanese	87,91	Estonian	165,34	Italian5	54,06	Italian6
222,97	Estonian	88,35	Eng_Ind	171,72	German1	54,25	Eng_Aus
231,93	Spanish_IPA	89,95	Spanish_Caracas	174,23	Japanese	54,42	German1
234,94	Finnish1	90,62	Port_Manauas	175,95	Estonian	56,01	Eng_Ind
240,37	German2	96,54	Finnish1	185,99	German2	56,96	Spanish_IPA
258,68	Japanese	101,46	German2	188,68	Italian4	57,19	Arabic Lebanese
258,80	Eng_Ind	101,66	Czech	194,58	Arabic Lebanese	57,42	Italian5
273,58	Port_Manauas	106,34	Spanish_IPA	212,07	Port_Manauas	57,56	German2
296,82	Czech	113,57	Eng_RP	214,13	Eng_Ind	59,52	Italian4
297,27	Eng_RP	114,32	Japanese	217,86	Eng_RP	59,80	French_Can
300,06	Italian4	128,15	Italian4	224,02	Czech	59,86	Eng_RP
383,12	French_Can	135,56	French_Can	235,83	French_Can	60,46	Dutch

## Appendix 2: *Correlatore's* report

*Correlatore's* full report for the corpus presented in chapter 3 is reported below. It contains the values of rhythm metrics for each sample labeled by AR and PM. The first column contains the values of rhythm metrics calculated globally (A method), while the second column reports the values of rhythm metrics computed locally for each inter-pausal unit and then averaged (B method). The ten Icelandic samples have not been computed with *Correlatore*, but their values have been inserted manually in the report.

```

FILE Arabic_IPA_ar
intV 206
intC 209
pause 16
Vmean 84.01895439337352
Cmean 100.2541003720087
Vperc 45.236400944564934
Vdev 46.71707063503773 35.73132296395354
Cdev 55.40657308332242 55.59657154745808
varcov 55.60301359656322 44.207387923952204
varcoc 55.26614161189175 55.91733924570244
Vrpvi 46.09585714955108 38.08533786310822
Crpvi 64.82354449341445 60.78499671768318
Vnpvi 47.31087095530114 42.78219155621204
Cnpvi 61.64081506823253 57.656263002652175
Vcci 21.28052447660613 19.419772375991656
Ccci 39.14170804786835 38.00093998486181
colour #008000
border black
symbol tu
--
FILE Arabic_IPA_pm
intV 207
intC 210
pause 16
Vmean 84.24361088114736
Cmean 99.25159468684845
Vperc 45.55342846076603
Vdev 47.008911891448605 36.51766941811486
Cdev 55.2330284747335 55.02223227192148
varcov 55.801159755331184 44.695681050374134
varcoc 55.649512382144394 55.969818030033075
Vrpvi 45.88766679883041 39.02778019732789
Crpvi 64.81592222185428 59.72953733720691
Vnpvi 46.75848661979226 43.6197141677819
Cnpvi 62.245184196385395 57.68081305739059
Vcci 21.487352730804155 19.630610431746984
Ccci 37.18817878728444 34.81970458061868
colour #008000
border black
symbol td
--
FILE Arabic_Lebanese_ar
intV 223
intC 227
pause 28
Vmean 104.28250304774217
Cmean 116.45768251859973
Vperc 46.79931712138865
Vdev 54.68757368801902 54.80776956610877
Cdev 72.91773970669507 69.07138545671611
varcov 52.44175397571939 51.491448390192915
varcoc 62.613078098174604 58.10220272846018
Vrpvi 60.75242641785179 58.92784353631968
Crpvi 75.74140015595356 76.50999231648342
Vnpvi 51.32609994659975 49.90429791540606
Cnpvi 60.263201375804464 61.85688193888051

```

## 9. Appendices

```
Vcci 26.829196094845344 24.2799792568974
Ccci 40.97730045108358 40.93342959849843
colour #008000
border black
symbol td
--
FILE Arabic_lebanese_pm
intV 224
intC 227
pause 27
Vmean 104.84867944427823
Cmean 115.85342903018443
Vperc 47.17521979258392
Vdev 56.01127218814023 55.70841011005504
Cdev 72.88098719818728 68.80078986700109
varcov 53.42105640720768 51.96613922075548
varcoc 62.90792409709245 57.70509182739357
Vrpvi 61.97981278237141 58.23123953232434
Crpvi 76.39723221417727 76.17901053843345
Vnpvi 51.743103696607406 49.3484951695647
Cnpvi 60.98061629176426 61.38622691575918
Vcci 27.548826394044156 25.375526987950884
Ccci 39.71993937056671 39.01866803566758
colour #008000
border black
symbol tu
--
FILE Chinese_chaoyang_ar
intV 157
intC 160
pause 24
Vmean 99.6410318748115
Cmean 93.09827551730169
Vperc 51.224514483116145
Vdev 43.67157367945309 45.19556280209212
Cdev 43.88340808279095 40.52624269342963
varcov 43.82890547974437 43.48308527106939
varcoc 47.1366497810537 43.029092683826164
Vrpvi 47.79772912278843 54.56480906124127
Crpvi 53.57113258163216 52.08603645203338
Vnpvi 48.02163185689612 52.7120718950414
Cnpvi 57.495229826972604 54.716408132501925
Vcci 36.81707541816237 41.55197613527768
Ccci 35.92672775155663 32.83765247664609
colour #ffffff
border #0000ff
symbol tu
--
FILE Chinese_hongkong_ar
intV 155
intC 157
pause 23
Vmean 113.7931182472869
Cmean 126.88903276557778
Vperc 46.95996469824065
Vdev 51.7619240486475 52.40148938669644
Cdev 57.33051847817638 60.39348212573153
varcov 45.487745520921806 43.98005333890944
varcoc 45.181618323225884 48.60068438010151
Vrpvi 57.480006841660064 62.65242200139288
Crpvi 68.59282654091858 72.16597017281977
Vnpvi 47.54732172039167 51.521726489372355
Cnpvi 52.18424500343679 56.91603038149025
Vcci 45.16241171348535 42.31740506318667
Ccci 41.15042764906163 38.74984387130136
colour #ffffff
border #0000ff
symbol tu
--
FILE Czech_ar
intV 160
intC 163
pause 20
Vmean 99.68085028113907
Cmean 120.05380372381158
```

## 9. Appendices

```

vperc 44.90418390660829
vdev 50.81459681351107 52.52518920165548
Cdev 58.20620491238413 54.18528409329724
varcov 50.977290693442114 50.44080281312639
varcoc 48.483432516881976 44.6295157852612
Vrpvi 50.181031236629686 47.75327307800236
Crpvi 60.67462285416889 60.061560249476194
Vnpvi 45.004005741586184 42.56027108068343
Cnpvi 51.17788975047674 51.70709350995023
Vcci 50.47804349229278 48.2577834306896
Ccci 30.76189731888801 31.798924544656554
colour #c0c0c0
border black
symbol tu
--
FILE Czech_pm
intV 161
intC 163
pause 20
vmean 100.77992224444013
cmean 118.80357417846112
vperc 45.58953836995485
vdev 52.83298011136585 54.3594934390792
Cdev 56.84448225726739 53.586365894928065
varcov 52.42411279423324 51.78934683693253
varcoc 47.84745126596805 44.6707058899146
Vrpvi 50.841575068677656 47.67325490487647
Crpvi 58.890953773309214 58.80551013027518
Vnpvi 44.602858743274524 41.87303007490353
Cnpvi 50.3774696314637 51.374096928239574
Vcci 51.89949012546818 48.67346402664185
Ccci 29.72846319696486 31.26370573280917
colour #c0c0c0
border black
symbol td
--
FILE Danish_ar
intV 132
intC 137
pause 15
vmean 85.62370435919341
cmean 116.2652208743929
vperc 41.505887912940565
vdev 39.543668954945446 39.75863272997858
Cdev 60.93772856691652 53.4965871349447
varcov 46.18308592333128 45.029555906049566
varcoc 52.412688943971105 46.531302972727026
Vrpvi 45.733215740546186 50.50099513591465
Crpvi 66.4244529618174 63.27979366943169
Vnpvi 50.26934187215413 53.73250797303985
Cnpvi 56.669289835901516 55.854490702638074
Vcci 30.33431357544646 32.92892102602869
Ccci 31.731082518969824 29.935961313210232
colour #ffffff
border #00ffff
symbol tu
--
FILE Dutch_IPA_ar
intV 153
intC 166
pause 19
vmean 72.99201693926372
cmean 90.53834732797992
vperc 42.629755693676955
vdev 44.91639880805826 46.29096619270412
Cdev 45.62519560319611 46.826108302504856
varcov 61.53604283250998 60.20860772278783
varcoc 50.39322778658245 50.3267642202385
Vrpvi 53.09554022934119 54.14990768437703
Crpvi 52.912625637773246 55.33936776963548
Vnpvi 67.70567543346453 67.1859367842029
Cnpvi 56.29283681876029 57.22776033339801
Vcci 30.02421250978579 30.470906711540767
Ccci 33.3113356475126 30.864600321552146
colour #ff80ff

```

## 9. Appendices

```
border #0000ff
symbol tu
--
FILE Dutch_IPA_pm
intV 154
intC 167
pause 19
Vmean 73.48957492348602
Cmean 89.26986365783462
Vperc 43.154224081431394
Vdev 44.66148192020934 45.9975388934719
Cdev 45.11258022892706 46.37283313214383
varcov 60.772540821890495 59.998601354622224
varcoc 50.53506119583708 50.38582516948217
Vrpvi 53.70198211724736 53.83988654868297
Crpvi 53.04568295210597 55.59510949413121
Vnpvi 68.97340509180847 67.75122457529089
Cnpvi 57.25721425108317 58.24559397956963
Vcci 30.187622519722073 29.21422511505927
Ccci 34.84203823426425 30.585973660499935
colour #ff80ff
border #0000ff
symbol td
--
FILE Greek_IPA_ar
intV 209
intC 210
pause 12
Vmean 68.48033404706383
Cmean 77.34402567921636
Vperc 46.841959389088025
Vdev 34.41343520498375 35.056633514434424
Cdev 37.16384802973469 37.942009576296456
varcov 50.25301889055148 49.201129923762096
varcoc 48.0500564890059 48.194621991079025
Vrpvi 37.640381325275705 36.54130975926907
Crpvi 41.9846498908428 44.29192646496179
Vnpvi 52.04625684903008 49.267688177236636
Cnpvi 55.57459177815651 57.19125060599685
Vcci 31.901514090379393 30.077474394027366
Ccci 26.556892139571833 26.862674567919456
colour #ff80ff
border black
symbol tu
--
FILE Greek_IPA_pm
intV 210
intC 210
pause 11
Vmean 70.53291908512459
Cmean 75.94329405314637
Vperc 48.153155774543926
Vdev 38.329340231828624 39.51773946177551
Cdev 33.7268982056113 33.60630136490787
varcov 54.342483947913465 53.65944563802201
varcoc 44.4106337842136 44.211211556162986
Vrpvi 41.21752415122962 40.735729891128685
Crpvi 40.40949696457153 41.73174218909708
Vnpvi 54.45881058463915 53.588523463849604
Cnpvi 54.86955488383259 56.26775354727039
Vcci 33.690160838708444 32.7451077298245
Ccci 24.4944265981411 24.589916474695038
colour #ff80ff
border black
symbol td
--
FILE English_Aus_ar
intV 137
intC 140
pause 12
Vmean 83.80479505054576
Cmean 119.90412279703294
Vperc 40.615976716934966
Vdev 42.710337594061116 44.062518481950846
Cdev 59.65970334701785 61.14378917864621
```

## 9. Appendices

```

varcov 50.96407379589788 51.51917621395998
varcoc 49.756173478710565 51.565259679950685
Vrpvi 52.082286612367405 56.394382817477556
Crvpi 73.92047757978882 73.13261362745594
Vnpvi 58.315846696934855 60.728587211042466
Cnpvi 62.35933604127948 60.67346987084118
Vcci 45.59105216891864 51.20646690013319
Ccci 39.13216262735044 37.985382562614774
colour #00ff00
border black
symbol tu
--
FILE English_Aus_pm
intV 137
intC 140
pause 12
Vmean 83.79391867079235
Cmean 120.13507161440377
Vperc 40.566443812868265
Vdev 43.03807899886169 44.626826356873956
Cdev 59.825412423813 61.47361767962581
varcov 51.36181680194326 52.12259340356624
varcoc 49.79845736957978 51.68268838359979
Vrpvi 52.17650168004063 56.38195945362863
Crvpi 74.55872963687655 74.00436202461346
Vnpvi 58.030891010758026 60.254493653857644
Cnpvi 62.6895885485767 61.13204845209477
Vcci 45.37524236135104 51.175530843902585
Ccci 39.737464348042714 37.83623339550683
colour #00ff00
border black
symbol td
--
FILE English_GA_ar
intV 134
intC 139
pause 12
Vmean 79.88553133215609
Cmean 108.64824152229959
Vperc 41.480055575609356
Vdev 43.19826721380696 45.026016903547635
Cdev 53.58996638706408 54.687663483742426
varcov 54.07520798001939 54.223360037969705
varcoc 49.324283243061025 50.07546707278272
Vrpvi 49.105021132373274 48.53478481771054
Crvpi 60.9866647925715 58.81617171210124
Vnpvi 57.775468560496044 55.365228857861716
Cnpvi 57.66127377086846 54.89217888958182
Vcci 33.44364211638657 30.180262258151377
Ccci 33.10176799523614 30.719127065086624
colour #00ff00
border black
symbol tu
--
FILE English_GA_pm
intV 131
intC 138
pause 12
Vmean 81.51533692809659
Cmean 110.07933941371095
Vperc 41.2784417331425
Vdev 39.61240024928714 40.09864785401721
Cdev 57.48919029834338 59.675174875310915
varcov 48.595027318881876 49.01731504201415
varcoc 52.225231914121395 52.89694050947226
Vrpvi 46.13111949474402 43.99707852886896
Crvpi 61.549653742901995 60.17285769424527
Vnpvi 54.45019490187495 51.89292813075908
Cnpvi 55.68377093303886 52.2587992039837
Vcci 36.733721854560216 34.79305979355213
Ccci 32.531487965437115 30.78854207091298
colour #00ff00
border black
symbol td
--

```

## 9. Appendices

```

FILE English_IndE_ar
intV 152
intC 159
pause 17
Vmean 95.63966858024061
Cmean 134.32545616669253
Vperc 40.49934218245582
Vdev 43.505990926712286 41.58265670321217
Cdev 64.89506880474643 62.48319278667646
varcov 45.48948315333323 42.73843412554874
varcoc 48.311817176495744 46.70784503037955
Vrpvi 47.99229139392463 46.102996949560534
Crpvi 75.77979562825449 76.41977918867188
Vnpvi 49.45396780137248 47.89763729216935
Cnpvi 56.35531299446036 56.76167686340616
Vcci 45.700284355315986 44.889135914087895
Ccci 42.60007991092764 41.41272997977142
colour #00ff00
border black
symbol tu
--
FILE English_IndE_pm
intV 143
intC 148
pause 15
Vmean 93.99299878500592
Cmean 133.03771959961352
Vperc 40.569764513331555
Vdev 42.09868348912123 43.1139675424377
Cdev 64.69271958171544 62.79505168227585
varcov 44.78916944166798 43.94569787537353
varcoc 48.62735153339427 46.30642205126359
Vrpvi 48.48090054732116 52.264334327380325
Crpvi 76.15582400886845 76.96727302697255
Vnpvi 51.25977913069677 52.83601742439964
Cnpvi 55.56869719611721 55.47896889050464
Vcci 43.26505613218828 49.24297408258909
Ccci 38.22965406404403 37.71708162350531
colour #00ff00
border black
symbol td
--
FILE English_NZE_IPA_ar
intV 140
intC 143
pause 12
Vmean 66.76473252049985
Cmean 100.942759044426
Vperc 39.30330021911729
Vdev 36.802343958123714 37.22515222324342
Cdev 45.93932769744719 46.27597451201386
varcov 55.12243151251104 53.967780180801014
varcoc 45.510275459410416 47.07035398854339
Vrpvi 44.37599404995127 44.70661395684902
Crpvi 53.13476691209686 52.891646248306735
Vnpvi 63.050835594218235 61.76231068947579
Cnpvi 53.34736672151319 53.12346377567981
Vcci 41.960143116019466 41.51079124949668
Ccci 29.934104100201086 28.754122261249275
colour #00ff00
border black
symbol tu
--
FILE English_NZE_IPA_pm
intV 141
intC 143
pause 12
Vmean 70.80055794328926
Cmean 96.96567489250879
Vperc 41.85874104756668
Vdev 34.960212360043705 35.320764697500366
Cdev 47.48873204077539 47.59814363936908
varcov 49.37844188748708 48.357879792872495
varcoc 48.97478627712227 50.217821573937734
Vrpvi 43.722875853120705 42.86419148699167

```

## 9. Appendices

```

Crpvi 52.33734706807895 49.779444272448174
Vnpvi 59.1257585908029 56.734587435495925
Cnpvi 54.56708333794259 52.019210300714114
Vcci 36.842205966129235 36.810904756966025
Ccci 26.69825551405195 24.344433910022914
colour #00ff00
border black
symbol td
--
FILE English_RP_IPA_ar
intV 136
intC 139
pause 12
Vmean 77.58686967435825
Cmean 116.22363461487903
Vperc 39.509694042835285
Vdev 48.35227389121337 46.18819732701881
Cdev 59.37175172603828 62.69657469279962
varcov 62.32017620269239 58.25867881026447
varcoc 51.08406041746475 52.54095608431665
Vrpvi 53.00917005501202 50.410127772135496
Crpvi 69.95275466182463 71.3404580294643
Vnpvi 61.41921308565858 57.348009859210066
Cnpvi 60.374612735767165 61.390108627795264
Vcci 48.44413787780632 44.005188420631214
Ccci 43.87063042957515 42.245767025715956
colour #00ff00
border black
symbol tu
--
FILE English_RP_IPA_pm
intV 136
intC 140
pause 12
Vmean 78.55811631310077
Cmean 114.02745789554696
Vperc 40.093083472078675
Vdev 49.47276651560638 48.30248451885775
Cdev 55.12006417998716 58.2618025260405
varcov 62.97600914770921 60.12196704974147
varcoc 48.33929055094697 50.36570332513023
Vrpvi 54.93035346379965 53.076860498002084
Crpvi 65.43317052947869 68.43093558190672
Vnpvi 63.31209439018285 61.11796979679134
Cnpvi 57.47575492866749 59.153041300737215
Vcci 44.10846035350281 43.86760981745578
Ccci 36.09242977507904 33.64705532678557
colour #00ff00
border black
symbol td
--
FILE Estonian_IPA_ar
intV 168
intC 169
pause 19
Vmean 89.60175268267783
Cmean 105.27175258799382
Vperc 45.83206975958909
Vdev 38.80326788189504 40.374727953723415
Cdev 53.560472860183005 51.36346020513333
varcov 43.30637149399941 40.12209580749224
varcoc 50.87829502545163 47.93916404663122
Vrpvi 42.29866528847245 49.85721597227556
Crpvi 53.32393527473117 48.80925829476
Vnpvi 43.21902962512173 45.497449306637186
Cnpvi 48.81060870123911 45.73579224309237
Vcci 36.71072527610094 37.642377081777525
Ccci 29.435893960703275 28.08565027584204
colour #ffffff
border #00e800
symbol tu
--
FILE Estonian_IPA_pm
intV 166
intC 170

```



## 9. Appendices

```
pause 16
Vmean 90.0124868685276
Cmean 106.71700409730536
Vperc 45.16409532559567
Vdev 37.880239096819714 41.11441674733809
Cdev 54.80769261020361 53.02329715010627
varcov 42.08331578722812 40.66863927803253
varcoc 51.35797530469423 49.86845768879365
Vrpvi 41.58921049845218 51.274250209710885
Crpvi 53.65933078744253 50.24524471181216
Vnpvi 42.52187472655331 47.11015472702944
Cnpvi 48.643281736851606 47.487568531996885
Vcci 28.331558755664908 31.596918914964064
Ccci 22.903590251275585 21.29666925483564
colour #ffffff
border #00e800
symbol td
--
FILE Finnish1_ar
intV 169
intC 173
pause 12
Vmean 96.26535023489875
Cmean 102.59426095292828
Vperc 47.82471320461888
Vdev 44.798042682402986 45.16753823502922
Cdev 46.825102289898155 45.40860698533706
varcov 46.53599926982087 46.61045748891109
varcoc 45.64105424121354 44.84965641546565
Vrpvi 45.67982697774799 46.135897708074225
Crpvi 55.14527003315632 51.863104001039325
Vnpvi 44.94785502166563 45.18227958663481
Cnpvi 57.31247564519871 54.901420637940106
Vcci 30.415273200354058 31.92899040863049
Ccci 42.12344861143338 40.470769368680294
colour #c17928
border #000000
symbol tu
--
FILE Finnish1_pm
intV 169
intC 173
pause 12
Vmean 98.18752072498147
Cmean 101.14035278595803
Vperc 48.67473707036773
Vdev 44.418094904822745 44.927706609560296
Cdev 45.80133150323341 44.28846549641572
varcov 45.23802472743527 45.32951465628833
varcoc 45.28492361516887 44.27742354840763
Vrpvi 45.410384339754856 45.819627321435036
Crpvi 54.3678546011611 51.14475244606835
Vnpvi 44.110143393826576 44.58670123004953
Cnpvi 56.92086408115866 54.29141044607149
Vcci 26.200838728814674 26.73659992791156
Ccci 29.60352681196938 27.819676268521377
colour #c17928
border #000000
symbol td
--
FILE Finnish2_ar
intV 166
intC 171
pause 18
Vmean 87.15832010738792
Cmean 96.722612807936
Vperc 46.66006222498692
Vdev 48.99043231326925 44.483509026451976
Cdev 43.10358790046906 42.19862876486127
varcov 56.20855502137725 48.77306331339941
varcoc 44.56412688732955 43.40311893449882
Vrpvi 48.94438739480219 50.71501689896006
Crpvi 48.79850362327251 47.8756682510939
Vnpvi 50.88739213553605 51.12003568294451
Cnpvi 52.524914941780466 51.253048624332195
```

## 9. Appendices

```
Vcci 40.762251811742445 43.10371732706063
Ccci 35.41442084906453 37.25259305850965
colour #c17928
border #000000
symbol tu
--
FILE Finnish2_pm
intV 166
intC 171
pause 18
Vmean 89.16233224281811
Cmean 95.00511115949402
Vperc 47.67298743969282
Vdev 48.606720541868796 44.803251383742605
Cdev 42.212589863626775 42.35657315516974
varcov 54.51485994051484 48.269924739087514
varcoc 44.431914608005165 44.44047831937604
Vrpvi 49.35824947497896 50.83662889542953
Crpvi 47.90695170068412 47.778036580410735
Vnpvi 50.78659466710902 50.469474957917484
Cnpvi 52.05840026844856 51.69808719617203
Vcci 25.653975259077693 25.38645287205219
Ccci 28.651061066581892 28.52022202872506
colour #c17928
border #000000
symbol td
--
FILE French_canadian1_ar
intV 165
intC 164
pause 23
Vmean 103.12557714372231
Cmean 99.05379300494747
Vperc 51.15887794218404
Vdev 46.510997991732495 45.57556067630205
Cdev 47.57208574045189 40.43421291391521
varcov 45.10132139858168 41.24371742850928
varcoc 48.026515994269694 39.74680969629114
Vrpvi 51.781395836704135 52.1864327504423
Crpvi 50.81570360381441 47.32889887450178
Vnpvi 46.69912924690057 45.59229521234551
Cnpvi 50.38172447425786 47.973399793839256
Vcci 43.48524541300082 42.85780977215406
Ccci 41.80864070569796 41.455528284091415
colour #ff8000
border black
symbol tu
--
FILE French_canadian1_pm
intV 159
intC 159
pause 22
Vmean 107.88632818337884
Cmean 95.18813153835197
Vperc 53.12648785633285
Vdev 48.53201306625618 51.09636189687059
Cdev 39.57283048321416 37.16362239083149
varcov 44.98439596884261 43.328719428312425
varcoc 41.573282134727044 37.92533644341819
Vrpvi 53.250679106075985 59.791438281775136
Crpvi 46.607047332645486 45.664661245600506
Vnpvi 44.62175744146854 47.95705882639621
Cnpvi 48.866968831959596 47.944602326287246
Vcci 43.292606434001264 46.06339027756656
Ccci 38.16566928931578 37.42069643628213
colour #ff8000
border black
symbol td
--
FILE French_IPA_ar
intV 152
intC 147
pause 15
Vmean 89.65155913638597
Cmean 98.24325716612074
```

## 9. Appendices

```

Vperc 48.54870543659568
Vdev 40.15810446948856 42.35230534425519
Cdev 40.187816715151094 36.84490284525493
varcov 44.79353717473721 44.95228185344565
varcoc 40.90643762675439 37.10154188104412
Vrpxi 43.99462917070564 42.53157858148581
Crpxi 44.23382405522516 44.71636944041687
Vnpxi 44.89003060957618 44.07698058468937
Cnpxi 47.02202956167576 48.661978334652716
Vcci 36.57106454801881 35.396705552324
Ccci 38.22235543769925 37.54101908572155
colour #ff8000
border black
symbol tu
--
FILE French_IPA_pm
intV 151
intC 147
pause 14
Vmean 91.82262485445592
Cmean 96.56640871997465
Vperc 49.411902884016015
Vdev 42.35492201418722 45.54377115530659
Cdev 42.56527442722518 41.417277189408644
varcov 46.126890928376504 48.19177991734656
varcoc 44.07875884735123 41.28275878428803
Vrpxi 47.33654270674053 45.95735728609675
Crpxi 46.15989938113067 50.31664442901221
Vnpxi 47.06038023895268 47.20469715336518
Cnpxi 48.7971172776731 53.067805538390644
Vcci 38.57216830577835 38.049584436887336
Ccci 37.122776404040685 37.04217597581681
colour #ff8000
border black
symbol td
--
FILE German1_ar
intV 167
intC 168
pause 16
Vmean 95.07465465294844
Cmean 110.3327525972704
Vperc 46.13749666542734
Vdev 53.662424717705576 52.32075007600697
Cdev 65.92949978334192 61.11198032026265
varcov 56.442408246014494 53.20318803453013
varcoc 59.755148159852034 54.358153486818395
Vrpxi 59.43004984398619 54.92810049072377
Crpxi 64.69806779631102 60.18061669342096
Vnpxi 57.885910092034564 54.04230869751597
Cnpxi 56.99006724383795 52.230685992486855
Vcci 51.25056678567742 45.534126935977795
Ccci 31.340503836889038 30.215434216085455
colour #00ffff
border black
symbol tu
--
FILE German1_pm
intV 174
intC 177
pause 19
Vmean 88.11729813419076
Cmean 109.11782626860638
Vperc 44.25414924257801
Vdev 42.67220523494236 44.16578776059714
Cdev 59.72588319224266 58.52745575908424
varcov 48.42659289207704 49.73881669834122
varcoc 54.73522084761876 53.77571523390565
Vrpxi 51.27990855039527 47.88425006569925
Crpxi 67.51348782869741 66.28705895270005
Vnpxi 57.26047711013468 55.08892275046879
Cnpxi 62.95708704005821 62.80441158535594
Vcci 40.07187136162802 35.25617243674631
Ccci 30.501029333827173 28.702901533007118
colour #00ffff

```

## 9. Appendices

```
border black
symbol td
--
FILE German2_ar
intV 154
intC 157
pause 15
Vmean 89.20258575127878
Cmean 117.14715709283101
Vperc 42.755981009457436
Vdev 47.30402500437111 46.88185312827805
Cdev 69.8519860887582 65.4665835679601
varcov 53.02988092325898 51.86521764722612
varcoc 59.62755547998944 55.51904824426928
Vrpvi 53.07029742653888 49.51370287746123
Crpvi 71.57245128043388 68.70043101106573
Vnpvi 56.635004787922405 55.289232551001874
Cnpvi 59.624031830602405 57.221077387380866
Vcci 51.32216709253293 48.48372902036306
Ccci 28.433719272456546 28.12592249228691
colour #00ffff
border black
symbol tu
--
FILE German2_pm
intV 165
intC 169
pause 16
Vmean 91.58712263156215
Cmean 101.32597216764307
Vperc 46.87892996001777
Vdev 50.730098265511764 48.95834673606745
Cdev 54.00663348872089 51.637189410866306
varcov 55.38999021684462 52.14773419218138
varcoc 53.29989175861773 49.794475654823714
Vrpvi 53.02100726411072 47.65437630122638
Crpvi 58.00380971601656 55.057335837621814
Vnpvi 53.14160008733706 49.417743091359135
Cnpvi 56.82636083627906 53.421712690611976
Vcci 46.05135137263825 41.60439200910549
Ccci 26.151441770561934 23.290263182084043
colour #00ffff
border black
symbol td
--
FILE Icelandic01
intV 0
intC 0
pause 0
Vmean 0
Cmean 0
Vperc 40.23
Vdev 33.24 33.24
Cdev 59.66 59.66
varcov 45.34 45.34
varcoc 56.35 56.35
Vrpvi 0 0
Crpvi 63.15 63.15
Vnpvi 44.58 58
Cnpvi 0 0
Vcci 31.47 31.47
Ccci 43.35 43.35
colour #ffffff
border #c0c0c0
symbol td
--
FILE Icelandic02
intV 0
intC 0
pause 0
Vmean 0
Cmean 0
Vperc 41.88
Vdev 36.80 36.80
Cdev 62.01 62.01
```

## 9. Appendices

```
varcov 44.97 44.97
varcoc 52.63 52.63
Vrpvi 0 0
Crvpi 64.29 64.29
Vnpvi 52.60 52.60
Cnpvi 0 0
Vcci 44.67 44.67
Ccci 44.24 44.24
colour #ffffff
border #c0c0c0
symbol td
--
FILE Icelandic03
intV 0
intC 0
pause 0
Vmean 0
Cmean 0
Vperc 41.18
Vdev 39.67 39.67
Cdev 60.85 60.85
varcov 53.10 53.10
varcoc 58.85 58.85
Vrpvi 0 0
Crvpi 65.11 65.11
Vnpvi 53.60 53.60
Cnpvi 0 0
Vcci 37.99 37.99
Ccci 42.94 42.94
colour #ffffff
border #c0c0c0
symbol td
--
FILE Icelandic04
intV 0
intC 0
pause 0
Vmean 0
Cmean 0
Vperc 46.53
Vdev 38.27 38.27
Cdev 46.40 46.40
varcov 48.59 48.59
varcoc 50.89 50.89
Vrpvi 0 0
Crvpi 47.21 47.21
Vnpvi 54.01 54.01
Cnpvi 0 0
Vcci 36.32 36.32
Ccci 36.47 36.47
colour #ffffff
border #c0c0c0
symbol td
--
FILE Icelandic05
intV 0
intC 0
pause 0
Vmean 0
Cmean 0
Vperc 44.94
Vdev 49.22 49.22
Cdev 63.86 63.86
varcov 49.58 49.58
varcoc 54.02 54.02
Vrpvi 0 0
Crvpi 66.11 66.11
Vnpvi 46.64 46.64
Cnpvi 0 0
Vcci 44.21 44.21
Ccci 45.24 45.24
colour #ffffff
border #c0c0c0
symbol td
--
```

## 9. Appendices

```
FILE Icelandic06
intV 0
intC 0
pause 0
vmean 0
Cmean 0
vperc 44.14
vdev 35.78 35.78
Cdev 57.75 57.75
varcov 44.99 44.99
varcoC 55.52 55.52
Vrpvi 0 0
Crpvi 54.38 54.38
Vnpvi 48.20 48.20
Cnpvi 0 0
Vcci 35.19 35.19
Ccci 38.47 38.47
colour #ffffff
border #c0c0c0
symbol td
--
FILE Icelandic07
intV 0
intC 0
pause 0
vmean 0
Cmean 0
vperc 46.18
vdev 44.38 44.38
Cdev 56.93 56.93
varcov 48.93 48.93
varcoC 54.58 54.58
Vrpvi 0 0
Crpvi 64.31 64.31
Vnpvi 50.24 50.24
Cnpvi 0 0
Vcci 43.23 43.23
Ccci 39.86 39.86
colour #ffffff
border #c0c0c0
symbol td
--
FILE Icelandic08
intV 0
intC 0
pause 0
vmean 0
Cmean 0
vperc 44.08
vdev 52.18 52.18
Cdev 73.93 73.93
varcov 49.19 49.19
varcoC 55.36 55.36
Vrpvi 0 0
Crpvi 77.65 77.65
Vnpvi 48.80 48.80
Cnpvi 0 0
Vcci 50.43 50.43
Ccci 49.96 49.96
colour #ffffff
border #c0c0c0
symbol td
--
FILE Icelandic09
intV 0
intC 0
pause 0
vmean 0
Cmean 0
vperc 43.32
vdev 39.47 39.47
Cdev 65.05 65.05
varcov 49.04 49.04
varcoC 44.61 44.61
Vrpvi 0 0
```

## 9. Appendices

```
Crpvi 51.18 51.18
Vnpvi 49.81 49.81
Cnpvi 0 0
Vcci 50.43 50.43
Ccci 49.96 49.96
colour #ffffff
border #c0c0c0
symbol td
--
FILE Icelandic10
intV 0
intC 0
pause 0
vmean 0
Cmean 0
Vperc 45.38
Vdev 37.87 37.87
Cdev 46.40 46.40
varcov 47.52 47.52
varcoc 51.02 51.02
Vrpvi 0 0
Crpvi 60.86 60.86
Vnpvi 49.65 49.65
Cnpvi 0 0
Vcci 38.48 38.48
Ccci 41.40 41.40
colour #ffffff
border #c0c0c0
symbol td
--
FILE Italian01_Antonio_ar
intV 219
intC 213
pause 14
vmean 69.3424557522141
Cmean 85.51676607077745
Vperc 45.46560416240033
Vdev 28.957929516804423 30.51696484364637
Cdev 41.70761391422531 41.28635260731086
varcov 41.760749893660645 41.67459266261421
varcoc 48.771271214473025 47.65182546333696
Vrpvi 29.569087200857282 33.45831138791384
Crpvi 51.642631913096594 51.401158623574005
Vnpvi 39.38241223782578 42.57875244469126
Cnpvi 62.835878904734855 60.77849933344085
Vcci 28.815479866866877 32.31156844985383
Ccci 30.289583300971962 31.18780703353285
colour #ffff00
border black
symbol tu
--
FILE Italian01_Antonio_pm
intV 219
intC 212
pause 15
vmean 73.18779861918765
Cmean 82.32710263948013
Vperc 47.87163212025704
Vdev 29.954217211499785 30.568393663843544
Cdev 37.72533602886081 36.993140447682286
varcov 40.92788385036449 39.10379642137391
varcoc 45.82371396460337 44.96251240871027
Vrpvi 29.24219634244279 32.28799533985129
Crpvi 47.07534194742961 46.60993911121017
Vnpvi 37.25062990889835 39.32231887876096
Cnpvi 59.077494273247844 57.664507980700314
Vcci 30.02591278266812 33.021801183516835
Ccci 22.567108877532657 22.629623788522455
colour #ffff00
border black
symbol td
--
FILE Italian02_Canepari_ar
intV 213
intC 202
```

## 9. Appendices

```

pause 26
Vmean 78.66310070472018
Cmean 94.32507034046913
Vperc 46.79070920967226
Vdev 35.90788206602431 33.73411708428573
Cdev 47.04328209401284 52.02137315943209
varcov 45.64768200634844 41.371352434601874
varcoc 49.87357223716741 52.93379841335935
Vrpvi 37.399514906000576 40.838509489590635
Crpvi 57.44710802344098 66.50934193153105
Vnpvi 43.570799462460215 44.521780335968245
Cnpvi 61.07319306827902 66.99502714165601
Vcci 39.77147912170352 45.01631347770219
Ccci 23.86239626231375 27.095992584814212
colour #ffff00
border black
symbol tu
--
FILE Italian02_Canepari_pm
intV 212
intC 201
pause 25
Vmean 80.31001755919029
Cmean 93.83355924044967
Vperc 47.44356026793772
Vdev 37.539265185792324 35.17832918289625
Cdev 46.32219202990224 49.94356490499856
varcov 46.742942321143246 42.288276800233895
varcoc 49.36633801900346 51.80037884870898
Vrpvi 39.261125705014045 42.006282093514976
Crpvi 56.76584229792415 63.89448309647452
Vnpvi 44.78416144288508 45.08164531479638
Cnpvi 60.848586159703125 66.23124292929688
Vcci 39.306756465496036 44.16824067734342
Ccci 23.091014118152074 24.749751204217745
colour #ffff00
border black
symbol td
--
FILE Italian03_lazio_ar
intV 210
intC 205
pause 15
Vmean 77.64890272112198
Cmean 92.25495755822716
Vperc 46.300248344286985
Vdev 39.615266612910105 40.59683551451825
Cdev 48.6454581214836 51.58362007112845
varcov 51.01845000333017 49.61614081905338
varcoc 52.729370224663306 54.278559324346396
Vrpvi 41.01212932673158 43.8231390823135
Crpvi 58.33080481526437 61.993024642402986
Vnpvi 48.69934775659977 49.179365080938354
Cnpvi 65.09603869859578 66.74468894043417
Vcci 39.03762711946384 41.3798259703237
Ccci 23.805550168270248 22.530835278472985
colour #ffff00
border black
symbol tu
--
FILE Italian03_lazio_pm
intV 209
intC 203
pause 15
Vmean 82.94436302006689
Cmean 89.48312105248773
Vperc 48.83142071590342
Vdev 43.23340665708912 44.07775834739213
Cdev 47.33175368480325 50.10065414755631
varcov 52.123381364240004 50.61515015099838
varcoc 52.894616468551725 55.112609979786626
Vrpvi 42.08908837692474 45.0046090915925
Crpvi 56.92892860043668 61.665891942276794
Vnpvi 46.190917551436804 46.89143576920909
Cnpvi 65.47106826098063 69.08987016666865

```



## 9. Appendices

```
Vcci 41.212217790272376 43.79578747904848
Ccci 23.39697667166889 22.46999419441394
colour #ffff00
border black
symbol td
--
FILE Italian04_00mp_ar
intV 186
intC 174
pause 23
Vmean 70.55396900968178
Cmean 103.37872992380213
Vperc 42.181430065873506
Vdev 30.61827903754623 30.913673359673428
Cdev 52.14060016819906 54.935060400041834
varcov 43.39696188225021 41.07590488778273
varcoc 50.43648747341991 50.68704369241909
Vrpvi 28.996399001682295 32.150881393390186
Crpvi 60.513762599899565 70.26874423239423
Vnpvi 38.524576213647215 40.17287442916294
Cnpvi 62.47721080365849 66.96324505561267
Vcci 24.911857435912584 29.188083028584227
Ccci 30.922553875853772 39.5918929371897
colour #ffff00
border black
symbol tu
--
FILE Italian04_00mp_pm
intV 186
intC 172
pause 23
Vmean 74.46013991201394
Cmean 101.40768321181696
Vperc 44.259604585884084
Vdev 30.459310928720303 30.27793805850411
Cdev 50.54457142355324 52.21867903738691
varcov 40.90686770762537 38.329659763007605
varcoc 49.84294071483464 49.39397576902259
Vrpvi 28.00516314934199 31.372994066285173
Crpvi 57.97700735096901 66.06016950543318
Vnpvi 35.20900736192189 37.55900655405445
Cnpvi 60.94221386417622 64.56463435674948
Vcci 24.99173647566174 29.911084243815694
Ccci 25.352865179296693 26.581630391216617
colour #ffff00
border black
symbol td
--
FILE Italian05_IPA_ar
intV 175
intC 167
pause 22
Vmean 83.64165445916619
Cmean 111.12055477269926
Vperc 44.095629838443045
Vdev 44.17778921874151 43.1753110469118
Cdev 53.913920873821056 57.76286779411136
varcov 52.81792846446991 45.591318513638356
varcoc 48.51840506385497 50.38972028556759
Vrpvi 44.55042436262723 51.63630112638489
Crpvi 62.375920942483376 68.83224267794999
Vnpvi 47.53262584801243 49.72210225362104
Cnpvi 60.253509383096294 61.350834870348194
Vcci 39.8617728694122 39.323100696815594
Ccci 37.20966186755655 38.360017982702566
colour #ffff00
border black
symbol tu
--
FILE Italian05_IPA_pm
intV 174
intC 168
pause 23
Vmean 89.81549527566054
Cmean 108.76828201574641
```

## 9. Appendices

```
vperc 46.098673000610745
vdev 48.006880689026154 35.171726921312924
cdev 51.725130883813094 54.23866252173843
varcov 53.45055498684728 40.06393172997143
varcoc 47.55534419153999 48.980716181222384
Vrpvi 47.27804227911301 40.02539971654943
Crpvi 59.970169817407545 63.655001190697746
Vnpvi 47.64476732398259 43.3355071170108
Cnpvi 58.07505406519908 59.715605736824934
Vcci 42.57634166338038 42.31395523484219
Ccci 29.73036878120449 29.747729615360317
colour #ffff00
border black
symbol td
--
```

```
FILE Italian06_Michela_ar
intV 216
intC 212
pause 9
vmean 64.92074761738297
cmean 90.2169181164771
vperc 42.30274600035516
vdev 28.976885891855304 27.36051675259338
cdev 44.45823209950234 43.35808885185855
varcov 44.63424553061762 41.84962536030757
varcoc 49.27926272332123 48.328943882952274
Vrpvi 27.221557767053223 26.365316286174654
Crpvi 52.33308420341009 51.44345874280418
Vnpvi 38.833418481201996 37.675113239501506
Cnpvi 59.203326533511635 59.59629746907733
Vcci 30.23142863902856 29.236711408781503
Ccci 28.803673267412226 28.20308632754246
colour #ffff00
border black
symbol tu
--
```

```
FILE Italian06_Michela_pm
intV 216
intC 213
pause 10
vmean 67.49232196313675
cmean 87.77970001397269
vperc 43.811145208408405
vdev 32.73774498365838 30.601000404123567
cdev 41.927600700251176 39.92703883581369
varcov 48.50588041931529 45.32935490989977
varcoc 47.76457505958346 46.249137964259496
Vrpvi 31.147283604643597 31.137164285391204
Crpvi 49.832531101909 45.7580643719967
Vnpvi 42.37745502932878 42.318768446999265
Cnpvi 56.48271593166548 53.521353889302986
Vcci 30.822345800462887 29.66242387579943
Ccci 25.831353947927578 25.578859340760314
colour #ffff00
border black
symbol td
--
```

```
FILE Italian07_Paolo_ar
intV 218
intC 213
pause 15
vmean 73.28579263078484
cmean 87.1893861613963
vperc 46.2442640001025
vdev 32.05710688477052 33.56087692358904
cdev 39.0165375462935 39.920657373500426
varcov 43.74259421096093 44.61117335644917
varcoc 44.74918251410782 44.40483286916571
Vrpvi 31.978055386666323 35.03522893158629
Crpvi 50.84565760962115 51.77260274434663
Vnpvi 39.195509622257745 40.69787929233571
Cnpvi 59.792851512907795 58.95949147148177
Vcci 34.916037074952456 38.86071258137095
Ccci 27.35415200559196 27.79548015844751
colour #ffff00
```

## 9. Appendices

```
border black
symbol tu
--
FILE Italian07_Paolo_pm
intV 216
intC 212
pause 15
Vmean 72.8632122417797
Cmean 90.12186668008789
Vperc 45.167957314520656
Vdev 32.19431357398436 33.02355659370654
Cdev 39.091024482613086 39.45232174829556
varcov 44.18459272308086 44.31958743446182
varcoc 43.37573767904445 42.454883654983725
Vrpvi 31.714402521698258 33.93483802849649
Crpvi 49.60061211817501 50.41484362325735
Vnpvi 39.96905833521601 40.99721086857569
Cnpvi 54.416602741465994 53.56435860156726
Vcci 32.09290728824333 35.30086525229448
Ccci 20.9226500329548 20.667813360648832
colour #ffff00
border black
symbol td
--
FILE Italian08_Claudia_pm
intV 216
intC 210
pause 12
Vmean 74.60380087694593
Cmean 94.59319755122537
Vperc 44.78841646424051
Vdev 34.94180771871385 36.363586510961824
Cdev 45.65233522973976 47.76641648555264
varcov 46.83649801750459 47.87898872286276
varcoc 48.26175286549279 49.72078071985118
Vrpvi 34.909153799367175 36.388112116881764
Crpvi 52.67988294230476 55.60139394581966
Vnpvi 44.184272539127804 45.534079811773715
Cnpvi 57.14541603347369 59.93568729471983
Vcci 34.504084850660675 34.96907457044884
Ccci 26.62753348577412 26.50774690363901
colour #ffff00
border black
symbol td
--
FILE Italian09_mazara_pm
intV 184
intC 180
pause 8
Vmean 73.03953842415355
Cmean 90.50916823024545
Vperc 45.20301642535804
Vdev 33.790719434570825 33.30451588459044
Cdev 46.20744285151275 46.07037068403383
varcov 46.26359936496603 45.09461015050976
varcoc 51.052775928694935 50.82497677965021
Vrpvi 29.16738399551481 28.78542305212621
Crpvi 50.74931878629317 48.40776488173622
Vnpvi 36.26723113129853 35.83081517834973
Cnpvi 58.262465601699745 56.81343832281731
Vcci 28.50501410422417 28.100019408759795
Ccci 18.86718826931001 17.367122864107014
colour #ffff00
border black
symbol td
--
FILE Italian10_bitonto_pm
intV 209
intC 189
pause 30
Vmean 84.4580298051313
Cmean 102.35611437083689
Vperc 47.71119842485954
Vdev 42.54751921855831 41.29718569909101
Cdev 49.819772271645384 52.14296617227162
```

## 9. Appendices

```
varcov 50.3771154936097 47.21364744616284
varcoc 48.67298116763989 50.27625551128993
Vrpvi 39.95764739862918 41.33716600862788
Crpvi 56.76929733756047 66.107031721573
Vnpvi 44.41664859701507 44.11349126002969
Cnpvi 58.38451895450727 64.69501925709586
Vcci 36.98992234519241 39.95568690120745
Ccci 27.041699911635774 29.93498902283976
colour #ffff00
border black
symbol td
--
FILE Italian11_vazzano_pm
intV 182
intC 178
pause 11
Vmean 69.93155295966865
Cmean 86.04984598763733
Vperc 45.38351945224115
Vdev 31.422289210301994 28.695591497227554
Cdev 38.86066985057405 38.96159140844448
varcov 44.93292066375824 40.79138025561132
varcoc 45.16065009129373 45.00785715855712
Vrpvi 30.41658043226776 29.09500447507015
Crpvi 44.95084267614729 43.178398674930044
Vnpvi 41.302403800640136 40.130097075871305
Cnpvi 54.717716334623645 52.65267278040248
Vcci 28.108639286627294 26.298998673228425
Ccci 21.129558729589363 21.64619354365528
colour #ffff00
border black
symbol td
--
FILE Italian12_nuragus_pm
intV 184
intC 175
pause 16
Vmean 77.24565679101312
Cmean 92.86984109009082
Vperc 46.65354828895464
Vdev 33.5372400231459 34.19762051322103
Cdev 37.623869153752665 36.34806309099403
varcov 43.41634392970515 43.283906139384136
varcoc 40.51247284600675 39.70843670302089
Vrpvi 31.744972023807687 33.05226019459462
Crpvi 42.996791431111326 43.939935711660645
Vnpvi 37.109870459303906 37.31054658539476
Cnpvi 49.61154658985211 50.462810433207345
Vcci 28.305458362906798 30.713364119521998
Ccci 25.052556289717707 22.6527363243979
colour #ffff00
border black
symbol td
--
FILE Italian13_tramonti_pm
intV 186
intC 180
pause 13
Vmean 77.90586139234333
Cmean 95.83793592800005
Vperc 45.65182172270577
Vdev 36.77582621365092 34.17581054015073
Cdev 47.944051023101046 50.19424535178806
varcov 47.20546767135199 43.412443511594006
varcoc 50.02617236990565 51.84626425877659
Vrpvi 33.00487102460203 31.822506357273607
Crpvi 54.044090404825184 57.41954673137221
Vnpvi 38.969202192170265 37.7825047927272
Cnpvi 59.997792219788046 64.82924364622926
Vcci 29.445361493517115 30.264401295951995
Ccci 20.76696781994052 21.68943954378339
colour #ffff00
border black
symbol td
--
```

## 9. Appendices

```
FILE Italian14_treviso_pm
intV 182
intC 178
pause 13
Vmean 97.893567674732
Cmean 104.46860070787719
Vperc 48.93059877446055
Vdev 48.46383039443288 45.566367061108934
Cdev 51.78605952346053 55.63395933380938
varcov 49.50665456944239 44.251370008065834
varcoc 49.57093248359718 52.75506705761912
Vrpvi 44.10597326109259 42.188372296039795
Crpvi 62.01346978034345 66.36234471384887
Vnpvi 41.8132743608915 39.617507602161616
Cnpvi 62.37574583868767 66.2624221978922
Vcci 38.933995287775055 37.0352504932801
Ccci 31.919591537733105 30.9675235326447
colour #ffff00
border black
symbol td
--
FILE Italian15_torinorfststd_pm
intV 185
intC 179
pause 12
Vmean 71.64176448400777
Cmean 86.60387022023053
Vperc 46.09058896162323
Vdev 33.94477560481491 30.570706989480687
Cdev 39.52467691777299 36.63066520160751
varcov 47.38126684804397 42.422050516110374
varcoc 45.638464906086945 42.10802876043337
Vrpvi 27.66501522500965 26.28303917243272
Crpvi 46.36726220467247 43.905745577900966
Vnpvi 35.91112326355702 34.945518314700074
Cnpvi 56.60463398377629 53.81611731279133
Vcci 25.759290844764607 24.73028171016165
Ccci 19.46278758368615 17.51346122370952
colour #ffff00
border black
symbol td
--
FILE Japanese_ph1_pm
intV 212
intC 210
pause 18
Vmean 83.29460136594182
Cmean 79.7892238222566
Vperc 51.31154971471287
Vdev 40.731153912056044 33.94644411413219
Cdev 32.22799524001282 28.149861515721412
varcov 48.90011266529758 39.86161633245033
varcoc 40.39141339663347 35.19489108952554
Vrpvi 38.17209276557493 33.431077552255445
Crpvi 33.40673095389991 33.52122211763805
Vnpvi 44.74085411550362 38.404661297790085
Cnpvi 41.27357678796416 42.86263472451292
Vcci 30.74519955081613 27.668223140750257
Ccci 28.438259355353733 30.206908081747724
colour #ffffff
border #ff0000
symbol td
--
FILE Japanese_phn_ar
intV 186
intC 184
pause 18
Vmean 88.79876313510084
Cmean 97.18889148366951
Vperc 48.01422549555959
Vdev 39.37763873075961 32.16177618760312
Cdev 55.9295628786883 51.334023826965
varcov 44.344805423527575 35.387500039442195
varcoc 57.547279349395666 53.52962511376319
Vrpvi 35.21079652234537 30.532446340534715
```

## 9. Appendices

```
Crpvi 60.26276481207018 54.27513298723366
Vnpvi 36.79232489428386 31.63806915089149
Cnpvi 58.05478399614456 55.88188533785265
Vcci 23.87949835727853 22.185113263226484
Ccci 31.597089078738428 32.34770879770544
colour #ffffff
border #ff0000
symbol tu
--
FILE Japanese_phl_ar
intV 214
intC 211
pause 19
Vmean 82.93205037862208
Cmean 78.64327041290834
Vperc 51.67980312117183
Vdev 39.30473945875299 32.915868841147045
Cdev 31.391594419032028 27.766899935469993
varcov 47.39390775859175 39.128815743894876
varcoc 39.91644072558748 35.35378174983338
Vrpvi 36.644929949447196 31.818579495865208
Crpvi 32.347715462436575 32.65130364941389
Vnpvi 42.49186703106479 36.963781787469436
Cnpvi 40.845086110336766 42.49359305927674
Vcci 26.706762586817582 24.584429094897224
Ccci 28.7837574362905 30.353671649610558
colour #ffffff
border #ff0000
symbol tu
--
FILE Japanese_phn_pm
intV 186
intC 184
pause 18
Vmean 89.38757370871872
Cmean 96.67447707843408
Vperc 48.31172114110805
Vdev 39.61915026611358 32.608318761575894
Cdev 54.698377698947375 50.29984480899646
varcov 44.32288362051066 35.600724787851924
varcoc 56.579957142741414 52.56633293892352
Vrpvi 35.46550582109242 30.903119578630047
Crpvi 58.29805884648838 52.87901362979658
Vnpvi 36.70264692718096 31.739018803319208
Cnpvi 56.618150832351375 54.99387463515671
Vcci 27.01499628031619 24.873729402434293
Ccci 28.38746412188199 30.143802323580676
colour #ffffff
border #ff0000
symbol td
--
FILE Polish_IPA_ar
intV 158
intC 164
pause 15
Vmean 77.68281441479469
Cmean 134.42481552158236
Vperc 35.76353156055076
Vdev 32.08524100299366 27.84108087085223
Cdev 65.89284742349562 61.29466310799533
varcov 41.30288178241265 35.28535488308908
varcoc 49.01836552114614 46.444276673055256
Vrpvi 32.14555176453511 29.599613817915124
Crpvi 72.39261631221315 69.89452145744674
Vnpvi 40.658542116977884 37.638027128808574
Cnpvi 55.4456799481717 54.78119190064583
Vcci 27.012067263316577 24.862948938990048
Ccci 43.02379389705914 40.86885996979048
colour #ffffff
border black
symbol tu
--
FILE Polish_IPA_pm
intV 160
intC 165
```

## 9. Appendices

```

pause 14
Vmean 78.85833462704832
Cmean 132.3311584230415
Vperc 36.622961840282755
Vdev 32.70986387680613 29.66746015114169
Cdev 70.01215609361535 63.75961042073766
varcov 41.479272966521265 36.55895921964797
varcoc 52.90678093348031 48.46501536863757
Vrpvi 32.38780272428115 31.291171028606307
Crpvi 77.2591656724486 74.95909059108948
Vnpvi 39.29760933386811 38.84887872481133
Cnpvi 58.466699442360614 57.22657247838658
Vcci 26.052844989550927 25.594323302452672
Ccci 36.11718944864734 35.599243420213185
colour #ffffff
border black
symbol td
--
FILE Portuguese_manau_s_ar
intV 164
intC 160
pause 21
Vmean 114.45532200721969
Cmean 120.63150036383963
Vperc 49.30346284802374
Vdev 55.34025775207122 56.96707049048419
Cdev 53.574406277237145 48.44206453707678
varcov 48.35096942768675 47.60296880239966
varcoc 44.41162226752553 38.510674857608564
Vrpvi 66.24924456952499 70.13631165668227
Crpvi 56.43454290411068 53.15532082340069
Vnpvi 56.93580809578065 58.690101787697145
Cnpvi 46.84071128511291 43.32896219173523
Vcci 50.70472487250472 54.03687837869711
Ccci 40.81723008896643 38.19537720213249
colour #0000ff
border black
symbol tu
--
FILE Portuguese_manau_s_pm
intV 164
intC 159
pause 21
Vmean 119.56380658618578
Cmean 119.29661017709834
Vperc 50.82991093314241
Vdev 62.37432751372985 68.05136229427755
Cdev 53.98049674184157 48.63659388229077
varcov 52.168234932172595 50.98332345996149
varcoc 45.24897787263727 39.430662048107884
Vrpvi 73.80969823017409 85.29726159428051
Crpvi 57.3268674376657 54.15664486729973
Vnpvi 59.815514295896946 63.45443261173132
Cnpvi 47.49067158093844 44.86711471642803
Vcci 54.31924387920861 57.97273308927976
Ccci 41.34488395039042 39.93412617212125
colour #0000ff
border black
symbol td
--
FILE Romanian_brasov_ar
intV 194
intC 198
pause 22
Vmean 90.64310262618477
Cmean 102.71581227352779
Vperc 46.37026903851032
Vdev 48.479416001049366 47.54388824910096
Cdev 41.499271824855924 41.6922950156931
varcov 53.48384443654817 49.97819654348041
varcoc 40.40202857408667 40.37049985087077
Vrpvi 51.64700603331115 53.74605667576475
Crpvi 47.6668087602596 51.60336532398311
Vnpvi 52.81858216864491 52.72510391605423
Cnpvi 45.620502226751036 47.840128484224564

```

## 9. Appendices

```

Vcci 31.32182985357846 28.61235336327011
Ccci 32.34527890310645 31.047789836078717
colour #aa00aa
border black
symbol tu
--
FILE Romanian_brasov_pm
intV 195
intC 197
pause 20
Vmean 90.71564210889268
Cmean 103.25657770935332
Vperc 46.51338526226603
Vdev 51.03178363700195 54.60816556028061
Cdev 40.83016312352775 42.26286314317476
varcov 56.25466837984207 53.032793751628674
varcoc 39.542433062672785 39.50412105158487
Vrpvi 53.72088994291564 62.64991251086949
Crpvi 45.44926719347183 50.036560898608954
Vnpvi 54.653788562116745 55.96618451510798
Cnpvi 44.000800662226425 45.59819163143459
Vcci 31.82311691239001 27.630202940047656
Ccci 32.768589706960576 30.859336326985623
colour #aa00aa
border black
symbol td
--
FILE Romanian_bucharest_ar
intV 201
intC 199
pause 20
Vmean 100.56499722893405
Cmean 115.06850819224195
Vperc 46.885952423699855
Vdev 48.570920066341756 51.37660225120853
Cdev 59.66372587026871 52.72997412786821
varcov 48.29803749287747 48.11326859542434
varcoc 51.850612133243345 45.067744988741794
Vrpvi 49.25189754065848 54.129662647172324
Crpvi 62.07257642417116 59.93216745946566
Vnpvi 45.25297351704499 47.01546297771315
Cnpvi 52.91635247788881 51.665792592586335
Vcci 33.546477776316394 33.3968829564933
Ccci 41.73155874756905 43.20209813018427
colour #aa00aa
border black
symbol tu
--
FILE Romanian_bucharest_pm
intV 199
intC 197
pause 20
Vmean 109.24902276038414
Cmean 110.38453496520106
Vperc 49.99402356861555
Vdev 61.07974217457555 64.73347795502438
Cdev 55.659272878186364 49.668506939596256
varcov 55.90873092617198 55.62679325257056
varcoc 50.42307139830145 44.62310934359297
Vrpvi 62.76266315313038 69.95161749625241
Crpvi 60.13132058345864 58.515269187903364
Vnpvi 52.86089313246089 56.88263789696342
Cnpvi 55.559569328362336 53.66170452309007
Vcci 40.10094970195722 39.12967195125326
Ccci 43.973902448618986 44.51474822458667
colour #aa00aa
border black
symbol td
--
FILE Romanian_bucovina_ar
intV 203
intC 201
pause 17
Vmean 81.9153609757973
Cmean 84.92768529115041

```



## 9. Appendices

```

vperc 49.34472424064453
vdev 35.80307190048414 36.92203315373035
Cdev 36.165295256402274 33.86133981470167
varcov 43.70739684717072 40.210707351233204
varcoc 42.583634691584784 39.77950244757489
Vrpvi 34.78632883020416 39.969320861834085
Crpvi 40.73664498314597 38.06736598574737
Vnpvi 38.955449310552545 40.04499031753832
Cnpvi 51.411327949086385 47.84491689971803
Vcci 20.37817834135273 21.169964307763593
Ccci 33.21502133589963 32.370529972831825
colour #aa00aa
border black
symbol tu
--
FILE Romanian_bucovina_pm
intV 201
intC 200
pause 16
vmean 84.51367806975267
Cmean 84.58420075152782
vperc 50.10383576564487
vdev 41.87801699362005 39.13087727540895
Cdev 36.73350583553216 34.3094435878664
varcov 49.5517624485073 44.53490491362393
varcoc 43.42833000626143 40.288652102653224
Vrpvi 40.56598988850253 40.4092702744872
Crpvi 40.52683767480579 37.95556702679769
Vnpvi 43.9379329970122 42.74548235270227
Cnpvi 52.69688452005275 49.162748770949996
Vcci 22.30588620487533 22.391876740152284
Ccci 35.44545196294452 34.38216402611619
colour #9d16de
border black
symbol td
--
FILE Romanian_moldavian_ar
intV 186
intC 186
pause 9
vmean 74.74669463208134
Cmean 95.55075801261107
vperc 43.8918453983175
vdev 40.94272401364441 44.86326702750932
Cdev 47.73424675487276 42.123115142080394
varcov 54.77529704179288 52.440030181526716
varcoc 49.95695245931241 44.29331123171506
Vrpvi 42.090078551941104 51.809208270922795
Crpvi 52.197224777379695 45.37463599582945
Vnpvi 51.620544778739955 55.80691620811605
Cnpvi 54.10967234329719 48.66576864839299
Vcci 23.132834820735958 23.596643803497994
Ccci 36.71250382778991 37.508291928984896
colour #aa00aa
border black
symbol tu
--
FILE Romanian_moldavian_pm
intV 178
intC 178
pause 7
vmean 72.2355691406971
Cmean 94.25124616494811
vperc 43.38816200435047
vdev 38.19503747385319 40.294440085164354
Cdev 48.88066994819687 47.443589907919794
varcov 52.87566489503068 51.70744002428173
varcoc 51.86209406998324 49.3644767429121
Vrpvi 40.46676349756715 44.643577237301905
Crpvi 53.14270115229604 52.166240024256446
Vnpvi 53.6201239094845 54.79463351673329
Cnpvi 55.86679164978315 54.79050928961663
Vcci 39.604074273546246 42.9166651049826
Ccci 34.88312435807211 36.11178579576258
colour #aa00aa

```

## 9. Appendices

```

border black
symbol td
--
FILE Romanian_muntenian_ar
intV 189
intC 190
pause 21
Vmean 60.944402300096506
Cmean 90.21685466774922
Vperc 40.19056122957424
Vdev 30.749122781214343 31.456797845847394
Cdev 45.02851239292064 41.58223477708734
varcov 50.454384029894165 46.682332385356105
varcoc 49.91141905661832 44.78530768800086
Vrpvi 31.867297793803374 34.09918789846107
Crpvi 49.38276953432634 49.55588620732935
Vnpvi 52.038061258429835 50.56350092452248
Cnpvi 57.08966768341851 54.558557274258135
Vcci 31.913776967653174 34.12649441309783
Ccci 34.342367321283284 33.85608246498844
colour #9d16de
border black
symbol tu
--
FILE Romanian_muntenian_pm
intV 179
intC 180
pause 19
Vmean 60.58626123932914
Cmean 88.90745034933519
Vperc 40.3934256673906
Vdev 31.227489451595165 32.70011912717695
Cdev 45.16518633060256 41.460919521797955
varcov 51.54219589196909 48.49603852720002
varcoc 50.80022669994414 45.23311166742617
Vrpvi 31.795234331089233 34.026515125321765
Crpvi 49.0718420197451 49.15631399883024
Vnpvi 52.33304091111064 50.93372964107818
Cnpvi 57.555234529646086 55.00565609623976
Vcci 32.50176197135167 34.556196012810375
Ccci 34.814701807781645 34.355146474021836
colour #9d16de
border black
symbol td
--
FILE Romanian_oltenian_ar
intV 199
intC 198
pause 18
Vmean 119.25152170939269
Cmean 118.18141886137444
Vperc 50.35128903482907
Vdev 61.61569352895423 59.52150981784743
Cdev 51.36122218815213 47.80834485244338
varcov 51.66868535154395 48.8980783446063
varcoc 43.45964254194502 40.25921591679274
Vrpvi 63.922436191480806 66.15649786777921
Crpvi 58.024653438635255 52.81617539302263
Vnpvi 47.88856224458527 48.90565612105101
Cnpvi 50.46269499591334 46.11617798484887
Vcci 58.65734091489904 60.1531092340645
Ccci 44.36503990731258 41.51597426364943
colour #9d16de
border black
symbol tu
--
FILE Romanian_oltenian_pm
intV 186
intC 184
pause 17
Vmean 119.48943435353192
Cmean 118.39287029669973
Vperc 50.50074237215807
Vdev 63.044316298297794 59.560258825560815
Cdev 54.61709313215443 49.341892791686014

```

## 9. Appendices

```
varcov 52.76141496474855 49.94848880221439
varcoc 46.13207957141395 41.42905938475505
Vrpv 66.805596837662 66.24449842218847
Crpv 60.9982400467413 55.169620728500604
Vnpv 49.92557713186258 50.57611142245323
Cnpv 52.44352067159418 47.71111106273155
Vcci 62.727476890631756 64.48907697104451
Ccci 45.62373234778402 42.04713127863612
colour #9d16de
border black
symbol td
--
FILE Portuguese_lisbon_ar
intV 135
intC 132
pause 14
Vmean 86.81166790906997
Cmean 113.28576394439247
Vperc 43.93748400662826
Vdev 36.60374899440246 37.63573748389804
Cdev 60.76629076911973 60.32960623280141
varcov 42.164549853762395 42.51535700545488
varcoc 53.63982962497169 50.21240446298792
Vrpv 44.19473254318312 45.40367867349021
Crpv 59.97701922687909 61.92046874089088
Vnpv 52.32485951004173 52.847524852350865
Cnpv 54.860495414357246 54.96620331052126
Vcci 36.15150820497482 36.950420130478406
Ccci 31.382774730260813 32.852140143415916
colour #0000ff
border black
symbol tu
--
FILE Portuguese_lisbon_pm
intV 135
intC 132
pause 14
Vmean 87.337018147916
Cmean 113.05726893954588
Vperc 44.135929904026206
Vdev 37.675199559677395 38.523278297579324
Cdev 61.07128898755888 58.54781835586839
varcov 43.13772139079652 43.234666357053975
varcoc 54.01801189821328 48.532330825307966
Vrpv 44.64534532740879 45.563671866211834
Crpv 61.39598971606641 60.10239571926874
Vnpv 52.429601656777926 52.75103766164243
Cnpv 55.832201989096816 53.190959821477094
Vcci 36.315266387434896 38.10476467520944
Ccci 29.534607661059642 27.90800668847583
colour #0000ff
border black
symbol td
--
FILE Portuguese_saoPaulo_ar
intV 158
intC 154
pause 11
Vmean 96.28429583312244
Cmean 102.13351155242842
Vperc 49.16674631364098
Vdev 46.95528302636106 46.54318495929374
Cdev 50.38512529722623 48.15800532507962
varcov 48.767332844956144 46.99188671311375
varcoc 49.33260839794187 46.23625904936922
Vrpv 54.39073037747987 57.00794930926328
Crpv 54.023928251772816 54.50604124721722
Vnpv 55.97017247799939 56.19809804799767
Cnpv 51.75481242982659 51.21127681977297
Vcci 44.832417059840594 48.67664098825995
Ccci 34.791816395883906 33.58020786441748
colour #0000ff
border black
symbol tu
--
```

## 9. Appendices

```

FILE Portuguese_saoPaulo_pm
intV 158
intC 154
pause 11
Vmean 97.95674628678923
Cmean 100.42795674978542
Vperc 50.01819563053995
Vdev 47.91979244653272 48.06880531860939
Cdev 49.84145491778849 47.992895445430555
varcov 48.91933864997652 47.653964501812915
varcoc 49.62906398859398 46.77933828056486
Vrpvi 55.67909363649826 58.80834863769837
Crpvi 53.82097291723188 54.83129646622226
Vnpvi 56.02836819593462 56.40520038097956
Cnpvi 51.890624750162075 51.93026787642263
Vcci 46.1906132290045 51.393380717159516
Ccci 34.870442053664746 34.32626840447903
colour #0000ff
border black
symbol td
--
FILE Russian_gs_ar
intV 175
intC 179
pause 11
Vmean 79.23956669029302
Cmean 113.84205660997526
Vperc 40.49369240985652
Vdev 33.85025000969529 32.18513511153391
Cdev 54.86019969145939 50.876537420906025
varcov 42.71887318869198 40.53189037498117
varcoc 48.18974755472956 45.91521818414914
Vrpvi 38.15285113126302 36.27192423734597
Crpvi 62.675585520459045 60.16436089360859
Vnpvi 44.943694149910456 43.034870995083004
Cnpvi 55.56382884080199 54.82331916255019
Vcci 28.58638037646266 27.42435319886409
Ccci 35.53713851986736 34.46857660236302
colour #f7c109
border black
symbol tu
--
FILE Russian_gs_pm
intV 182
intC 186
pause 11
Vmean 79.10720355275635
Cmean 106.95941062567586
Vperc 41.98509052606885
Vdev 33.12758288675832 31.634555607154063
Cdev 48.35976043190399 47.1756960561775
varcov 41.87682208316925 39.74463065865806
varcoc 45.2131889555262 45.11885186429301
Vrpvi 36.73975357453552 35.22698809497796
Crpvi 54.35577592456517 52.90746399219127
Vnpvi 43.888698019913136 42.292064417456245
Cnpvi 51.52397464834826 50.874342077389244
Vcci 31.853089521449384 31.45726855231651
Ccci 29.715986096016756 28.278131649301514
colour #fecb01
border black
symbol td
--
FILE Russian_ss_ar
intV 174
intC 177
pause 17
Vmean 78.00773586189523
Cmean 123.67301935815395
Vperc 38.274161726061315
Vdev 36.23256713225476 37.80584692315388
Cdev 57.923156861650874 54.90722917463583
varcov 46.44740259658457 43.92850870307808
varcoc 46.8357263065656 44.29127274399847
Vrpvi 37.87215575215725 42.80919623661539

```

## 9. Appendices

```

Crvpi 65.24290252445456 62.94226176665938
Vnpvi 46.958401961484924 49.2093300647478
Cnpvi 54.72322953125756 53.5304556418685
Vcci 34.31662033835754 35.690617736896456
Ccci 40.45879992181798 43.65707997392145
colour #f7c109
border black
symbol tu
--
FILE Russian_ss_pm
intV 175
intC 179
pause 17
Vmean 78.96126168963193
Cmean 121.66893610651427
Vperc 38.8185414047251
Vdev 38.45111724844714 39.54883034720778
Cdev 55.20837352232766 53.184752892547806
varcov 48.69617889286589 44.9684816686361
varcoc 45.37589896733859 43.6178139084747
Vrpvi 38.190041235159406 42.97422044740744
Crvpi 63.04716828632986 60.27842993477008
Vnpvi 46.30674831427909 48.4345695785036
Cnpvi 53.49030604834872 51.974898131859284
Vcci 33.339724660075255 33.35627535818072
Ccci 38.70934807546905 40.68483827287438
colour #fecb01
border black
symbol td
--
FILE Spanish_bogota_ar
intV 157
intC 154
pause 19
Vmean 72.35528189297948
Cmean 96.85492007206696
Vperc 43.23345518307142
Vdev 34.7828489271357 32.752460520981444
Cdev 46.39919193246093 46.149500175631346
varcov 48.07230103613297 42.88109113301242
varcoc 47.90586982874657 46.68421866436047
Vrpvi 36.33641031878884 37.35589781669293
Crvpi 56.75667658866066 57.60966321073218
Vnpvi 46.57731504342108 47.51101748033804
Cnpvi 60.182227488702566 59.282573029386995
Vcci 27.36296789517043 26.95302261635581
Ccci 27.86178644621503 25.577643501908735
colour #ff0000
border black
symbol tu
--
FILE Spanish_bogota_pm
intV 161
intC 158
pause 19
Vmean 74.98089889309256
Cmean 92.54065394472045
Vperc 45.22446828855111
Vdev 29.09357267491019 25.604697086338362
Cdev 43.30929996252935 42.712506536963915
varcov 38.801312206715046 34.421714735989944
varcoc 46.800295995747284 45.24748817855949
Vrpvi 32.06195114984841 28.80176161010774
Crvpi 51.32921329926918 51.9454161074592
Vnpvi 41.20515827131481 38.72787017462656
Cnpvi 57.25007758923295 55.79679620061273
Vcci 28.923814076179184 28.137666321624923
Ccci 28.451414199850255 25.534850917525823
colour #ff0000
border black
symbol td
--
FILE Spanish_caracas_ar
intV 155
intC 156

```

## 9. Appendices

```

pause 17
vmean 61.551273453424756
cmean 88.21230878315957
vperc 40.94337156749738
vdev 28.855752266087336 29.130724141198687
cdev 40.968802934040475 39.07536297769295
varcov 46.88083714128547 43.64663805022232
varcoc 46.443408521081295 44.80876142651439
vrpvi 27.757391497830945 31.536676371968213
crpvi 47.01333670136085 49.66234703245115
vnpvi 42.42618622971178 44.67418627711227
cnpvi 54.97503011841243 55.270776646158644
vcci 23.140292517652842 21.98869083442424
ccci 24.02891424905356 24.69513698765942
colour #ff0000
border black
symbol tu
--
FILE Spanish_caracas_pm
intv 156
intc 154
pause 16
vmean 65.18279590929804
cmean 85.20535558254896
vperc 43.66017447509747
vdev 23.720275124021367 22.774266056630015
cdev 36.505517593271094 32.53986013634101
varcov 36.390392270113985 33.603599675174365
varcoc 42.84415849647348 38.87189985380579
vrpvi 24.874195972897756 25.242230805508406
crpvi 41.081382820811605 41.91621185864005
vnpvi 36.878264535428755 35.883386260333495
cnpvi 47.91542767789988 47.38589434360223
vcci 24.089868176218683 22.798518219866814
ccci 22.125057300521796 19.521289466831888
colour #ff0000
border black
symbol td
--
FILE Spanish_granada_ar
intv 165
intc 162
pause 16
vmean 67.25052584242421
cmean 84.98924626543216
vperc 44.62705672937064
vdev 22.497932029248073 23.13194775796735
cdev 46.47694381510287 46.49980935596266
varcov 33.4539124377456 33.56929329448244
varcoc 54.68567596181463 53.18126232740229
vrpvi 24.729144189024446 25.849474966615684
crpvi 55.979548577639875 60.527829473881624
vnpvi 36.837549747125266 37.4106209846264
cnpvi 68.0129299862016 70.54808788530038
vcci 23.8613543841464 23.426389855504585
ccci 33.63030512008287 31.316667106681905
colour #ff0000
border black
symbol tu
--
FILE Spanish_granada_pm
intv 161
intc 159
pause 16
vmean 68.81708109688834
cmean 87.57673044445247
vperc 44.31066706335409
vdev 23.847598509026927 23.7052862187883
cdev 47.156586359362635 46.230993409702755
varcov 34.653603624152005 33.090505452139226
varcoc 53.846022933309634 51.153518415096
vrpvi 23.194150306121305 23.51610829563192
crpvi 56.635992947913294 60.8528756570044
vnpvi 32.930317541626884 32.27254246411719
cnpvi 67.78447367808475 69.48335038302395

```

## 9. Appendices

```

Vcci 21.46967096861252 21.075521213971836
Ccci 31.63357827121847 28.576739139629943
colour #ff0000
border black
symbol td
--
FILE Spanish_IPA_ar
intV 164
intC 162
pause 13
Vmean 71.20441839516616
Cmean 99.37284848013111
Vperc 42.041891194304434
Vdev 30.070779562728855 33.63534530691724
Cdev 52.30973236533942 48.99958626142094
varcov 42.231620228738876 45.4097368230649
varcoc 52.63986407292972 49.00915461157334
Vrpvi 31.66939801119255 34.58080703373354
Crpvi 59.48718221307275 63.5222751675023
Vnpvi 41.6191459645734 45.06648929609983
Cnpvi 60.820311508138225 62.76656351120989
Vcci 26.341149108260435 29.829788395509212
Ccci 32.62775936602838 34.79382494277831
colour #ff0000
border black
symbol tu
--
FILE Spanish_IPA_pm
intV 163
intC 161
pause 13
Vmean 74.75581609635226
Cmean 99.5274205850065
Vperc 43.19596343538248
Vdev 33.85165166570099 37.12816641953708
Cdev 51.70747504167457 48.875960542260856
varcov 45.28296717685461 47.826272384381355
varcoc 51.952994197725786 48.90950269343214
Vrpvi 31.258745241600703 32.22766148648249
Crpvi 57.254522177357146 62.04175437936107
Vnpvi 38.155769187080026 40.17237409187338
Cnpvi 59.945389188001705 63.43727272536156
Vcci 29.257802112117616 30.59483730341013
Ccci 32.66958287673966 30.629936266772486
colour #ff0000
border black
symbol td
--
FILE Spanish_lima_ar
intV 172
intC 172
pause 23
Vmean 76.91261963021269
Cmean 84.22826101601434
Vperc 47.73004796906174
Vdev 35.48581323027807 32.511678931322564
Cdev 37.597461422060015 35.44783015704988
varcov 46.1378293976852 38.68598613479965
varcoc 44.637584782750764 42.501957439459595
Vrpvi 33.41288174052672 35.3040511837775
Crpvi 44.004557401645044 43.12189371159943
Vnpvi 39.94239191636248 40.662973803791154
Cnpvi 57.09417366871914 54.81187630217617
Vcci 22.80225625196563 21.95501672300476
Ccci 26.137598264466266 23.906061197447784
colour #ff0000
border black
symbol tu
--
FILE Spanish_lima_pm
intV 158
intC 157
pause 21
Vmean 73.26461417975007
Cmean 88.11295705668206

```

## 9. Appendices

```
vperc 45.556934670192064
vdev 25.761760373741172 24.679717196929943
Cdev 36.869773765841224 35.7228238634609
varcov 35.16262340580452 31.402983913866183
varcoC 41.84375941681689 39.78345177579403
Vrpxi 25.13475635103389 25.549788260284863
Crpxi 44.11441457276916 46.09248745040499
Vnpxi 32.322524511220834 31.910251643681768
Cnpxi 51.89929109276624 52.418885191769654
Vcci 22.758055103107644 20.65633265514097
Ccci 26.18698548038082 25.229280142993048
colour #ff0000
border black
symbol td
--
FILE Swedish_IPA_ar
intV 160
intC 164
pause 15
Vmean 92.02348495976844
Cmean 114.1792381098518
vperc 44.01832761042909
vdev 35.805234522877065 36.59186057833848
Cdev 54.74927740796527 52.2714688738541
varcov 38.908800876787794 39.184828456410266
varcoC 47.9502914140056 47.1110377875044
Vrpxi 42.06931769767066 42.790291455621364
Crpxi 68.11331260264166 63.72154898945452
Vnpxi 45.26058476189414 45.07106415570816
Cnpxi 60.49140863660416 56.97984770435922
Vcci 31.695124141273237 31.000544876893315
Ccci 30.60082372107569 29.5787112400012
colour #b35900
border black
symbol tu
--
FILE Turkish_ar
intV 163
intC 169
pause 16
Vmean 87.77415070550994
Cmean 103.97499035331374
vperc 44.879715631566775
vdev 37.96009756046363 37.759943692715424
Cdev 57.49794998762647 53.28041936509227
varcov 43.24746779700908 41.649353515539666
varcoC 55.29978872058051 50.76838580492096
Vrpxi 40.976826882311734 40.796125008751474
Crpxi 67.2258252140195 65.13028111702876
Vnpxi 46.0729495081369 44.86816107777714
Cnpxi 61.95452946659324 60.50233190119441
Vcci 40.79394040294805 40.35836668199842
Ccci 40.48123132158842 39.2430728233207
colour #ffd9ff
border #ff00ff
symbol tu
--
```



## Appendix 3: Tcl implementations

Key Tcl implementations are reported below. However, this is merely a fraction of the entire program, which contains more than 4000 lines of code (which can be consulted directly from the sources, available at the following internet address: [http://www.lfsag.unito.it/correlatore/download\\_en.html](http://www.lfsag.unito.it/correlatore/download_en.html)).

### Appendix 3a: Tcl implementation of the metrics

The procedures (proc) of *Correlatore* implementing the calculation of the rhythm metrics are reported below. All take a numerical list as argument (consonantal or vocalic durations), except for the CCI, which take 2 lists as arguments (the first being a numerical list containing the duration of consonantal or vocalic intervals, the second being a string list containing the segments composing each interval).

```
# Tcl implementation of the delta (standard deviation)
proc stdev {lista} {
    set media [mean $lista]
    foreach item $lista {
        lappend provvisorie [expr pow(($item - $media), 2)]
    }
    return [expr sqrt( ([join $provvisorie +]) / ([length
$provvisorie].0 - 1) )]
}

# Tcl implementation of the varco
proc varco {lista} {
    return [expr [stdev $lista] / [mean $lista] * 100]
}

# Tcl implementation of the rPVI
proc rpvi {valori} {
    set volte 0
    foreach v $valori {
        if {$volte > 0} {
            set prec [lindex $valori [expr $volte - 1]]
            lappend rpvi [expr abs($v - $prec)]
        }
        incr volte
    }
    return [mean $rpvi]
}

# Tcl implementation of the nPVI
proc npvi {valori} {
    set volte 0
    foreach v $valori {
        if {$volte > 0} {
            set prec [lindex $valori [expr $volte - 1]]
            lappend npvi [expr (abs($v - $prec)) / (($v + $prec) /
2.0)]
        }
        incr volte
    }
    return [expr 100 * [mean $npvi]]
}

# Tcl implementation of the CCI
proc cci {durate segmenti} {
    set volte 0
    foreach durata $durate {
        lappend differenze [expr $durata / [string length [lindex
$segmenti $volte]].0]
        incr volte
    }
}
```

## 9. Appendices

```
}
set volte 0
foreach v $differenze {
    if {$volte > 0} {
        set prec [lindex $differenze [expr $volte - 1]]
        lappend cci [expr abs($v - $prec)]
    }
    incr volte
}
return [mean $cci]
}
```

### **Appendix 3b: reading CV and SAMPA segmentations with a foreach cycle.**

```
foreach item $contenuto($scelto) {

    #----- correct characters that may interfere with Tcl -----
    set item [string map {\ " " } $item]
    set item [string map {\[ \\\[ } $item]
    set item [string map {\] \\\]} $item]
    set item [string map {\{ \\\{ } $item]
    set item [string map {\} \\\}] $item]

    #----- a = fono ; b = durata -----
    incr n_intervallo
    set a [lindex $item 0]
    set b [lindex $item 1]
    if {$a eq ""} {set a "#"}

    #----- if annotation is in CV, put it in $pronto_cci -----
    if {$tipo($scelto) eq "CV"} {
        if {![string match -nocase "*c*" $a] && ![string match -nocase
"*v*" $a] && ![string match -nocase "#" $a]} {
            lappend log [errore_segmento $scelto $n_intervallo
$item]
        } else {
            lappend pronto_cci "$a\t$b"
        }
    }

    #----- if annotation is in SAMPA, convert SAMPA in CV -----
    if {$tipo($scelto) eq "SAMPA"} {
        if {[string match "*#*" $a]} {
            set a "#"
        } elseif {[string match $var_sost $a]} {
            set a v
        } else {
            set a c
        }
    }

    #----- compute durations for #, V and C intervals -----
    if {[regexp {^\#+$} $a]} {
        if {($precedente eq "#") || ($precedente eq "null")} {
            append intervallo $a
            set durata [expr $durata + $b]
        } elseif {($precedente eq "v") || ($precedente eq "c")} {
            lappend pronto "$intervallo\t$durata"
            set intervallo $a
            set durata $b
        }
        set precedente "#"
    }

    } elseif {[regexp {^[vV]+$} $a]} {
        if {$precedente eq "v" || ($precedente eq "null")} {
```

## 9. Appendices

```
        append intervallo $a
        set durata [expr $durata + $b]
    } elseif {($precedente eq "c") || ($precedente eq "#")} {
        lappend pronto "$intervallo\t$durata"
        set intervallo $a
        set durata $b
    }
    set precedente "v"
} elseif {[regexp {^[cC]+$} $a]} {
    if {$precedente eq "c" || ($precedente eq "null")} {
        append intervallo $a
        set durata [expr $durata + $b]
    } elseif {($precedente eq "v") || ($precedente eq "#")} {
        lappend pronto "$intervallo\t$durata"
        set intervallo $a
        set durata $b
    }
    set precedente "c"
}
}
```

## Appendix 4: sample log file of a perceptive test

A sample log file of a perceptive test is reported below. The original name and surname of the participant have of course been removed for privacy reasons. “False” and “True” are the Boolean values returned by checkboxes in the preliminary phase (each “true” indicate that the participant marked a stress on the corresponding syllable). Below, “is” indicates the choice “Spanish, French or similar”; “ia” indicates the choice “English, German or similar”; “al” indicates the choice “other”; “no” indicates the choice “I don’t know”. Details on the test can be found in chapter 6.

```

Nome Cognome
età: 25
titolo: Laurea triennale
CFUfon: 0
CFUling: 12
madrelingua: Italiano
inglese: C1
francese: C1
tedesco: B1
spagnolo: no
altro(specificare): no
altro(specificare): no
true false false false true false false false false false false true
false false false false false false false false false true false true false
false false false true false false false true false false true false false
false false true false true false true false true true true false true
false true false false false false undefined undefined undefined
true false false false false false false true false false false true
false true false false true false false false true false false false false
false false false false true false false false true false true false true
false false false false false false false false false true true false
true false false false undefined undefined undefined
true false true false false false false true false false false true
false false false false false false false false false true false
false false false false false false false false false false true
false true false false true false false false true false true false
true false false false true false false true false false false true
false false true false false true false undefined undefined undefined
false false true false false false false false false false false
true false false true false true false false false false false false
false false true false true false true false true false true true
false false true false true false false true false false true false
false undefined undefined undefined
true false false true false false true false false false false false
false false false true false true false false false false false false
true false false true false false true false true false false true true
false false true false true false false false true false false false
undefined undefined undefined
1: is
2: is
3: ia
4: ia
5: is
6: is
7: is
8: ia
9: ia
10: is
11: ia
12: al
13: al
14: is
15: al
1: en
2: en

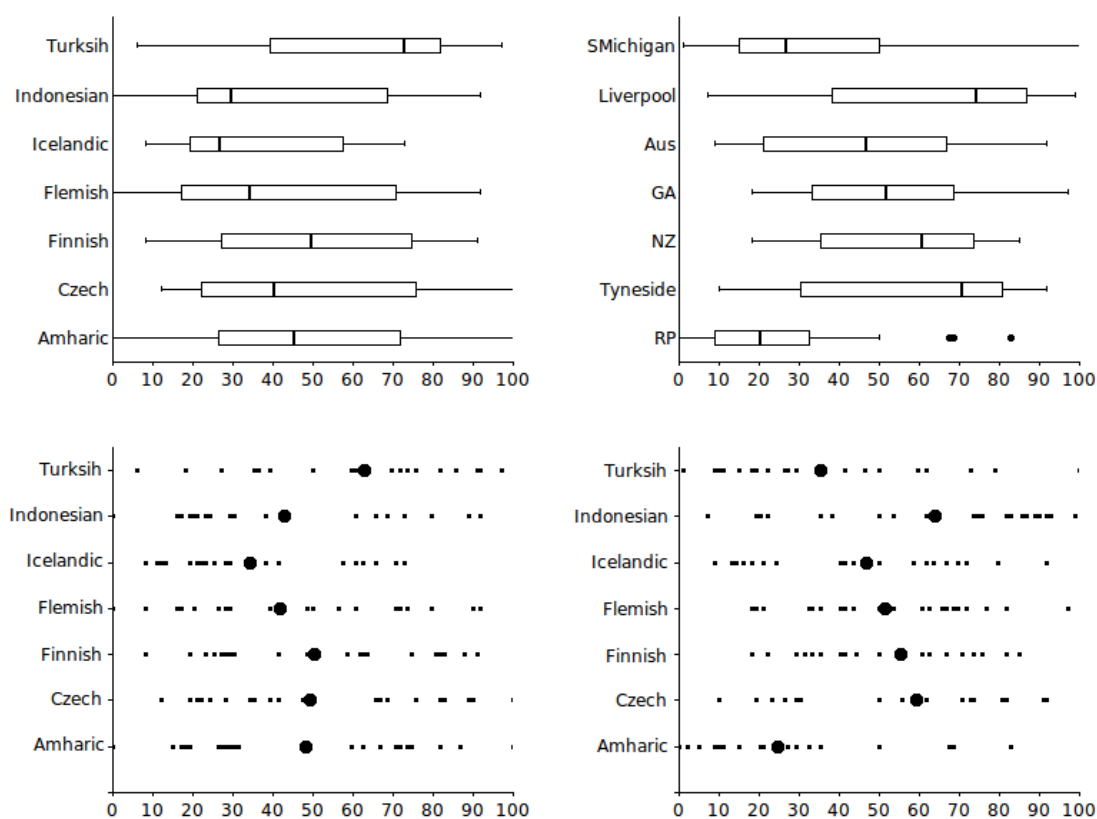
```

## 9. Appendices

3: fr  
4: en  
5: jp  
6: fr  
7: jp  
8: en  
9: fr  
10: en  
11: jp  
12: jp  
13: fr  
14: en  
15: jp  
slider1: 66.66666666666667  
slider2: 70.70707070707071  
slider3: 60.60606060606061  
slider4: 34.343434343434346  
slider5: 19.19191919191919  
slider6: 3.0303030303030303  
slider7: 8.080808080808081  
slider1: 19.19191919191919  
slider2: 61.61616161616162  
slider3: 70.70707070707071  
slider4: 29.292929292929294  
slider5: 24.242424242424242  
slider6: 9.090909090909092  
slider7: 70.70707070707071  
slider1: 71.71717171717172  
slider2: 34.343434343434346  
slider3: 19.19191919191919  
slider4: 40.4040404040404  
slider5: 71.71717171717172  
slider6: 8.080808080808081  
slider7: 40.4040404040404

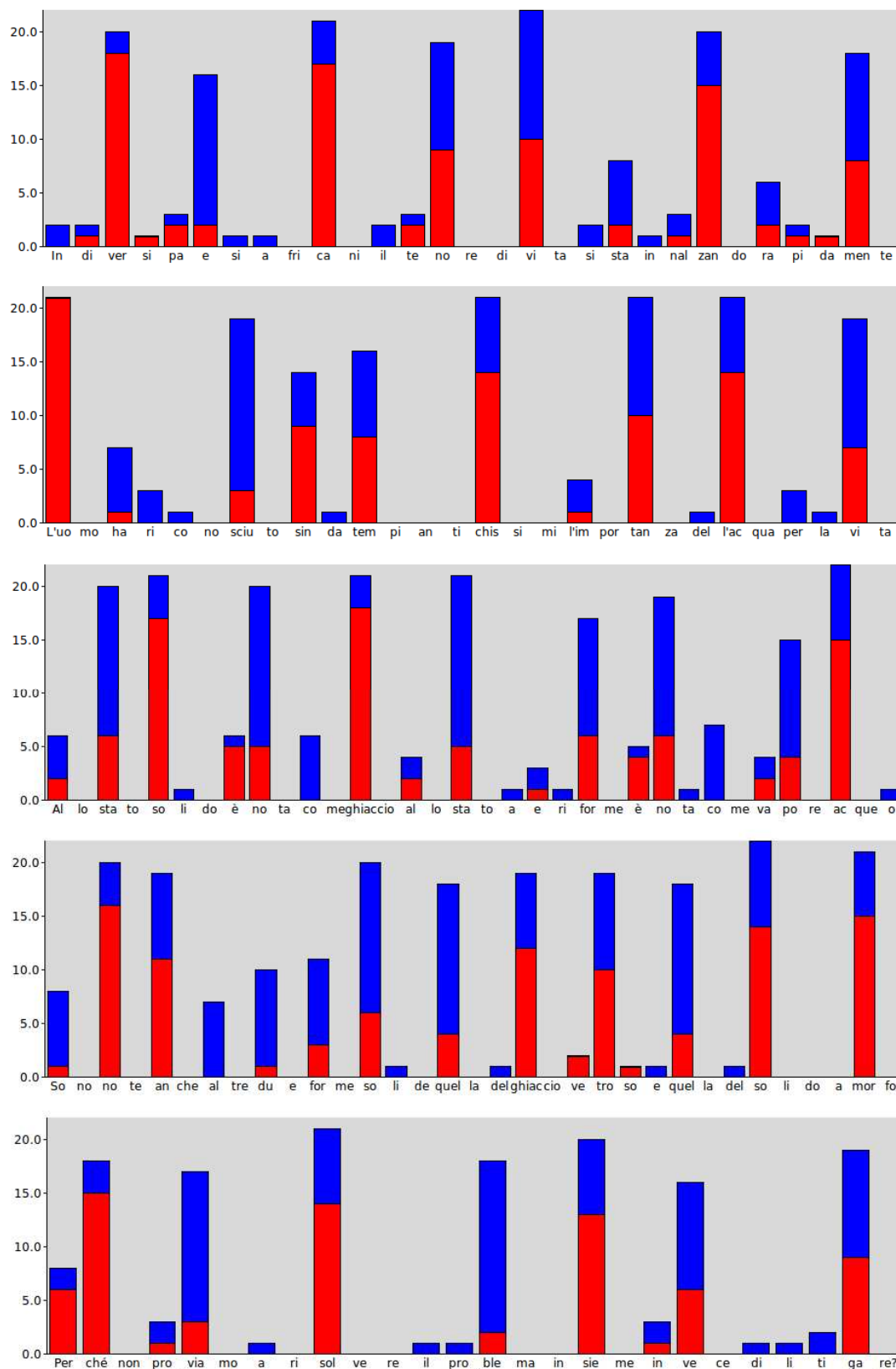
## Appendix 5: Results of the perceptive test by best performers in the preliminary phase.

Here I report the results of the perceptive tests by exclusively participants who performed better in the preliminary phase. Their performance in the preliminary phase has been evaluated in terms of correlation with model answers. I have selected those participants whose score was higher than or equal to the median (hence 22 out of 43). The results are reproduced in charts equivalent to those presented in chapter 6 for all participants.



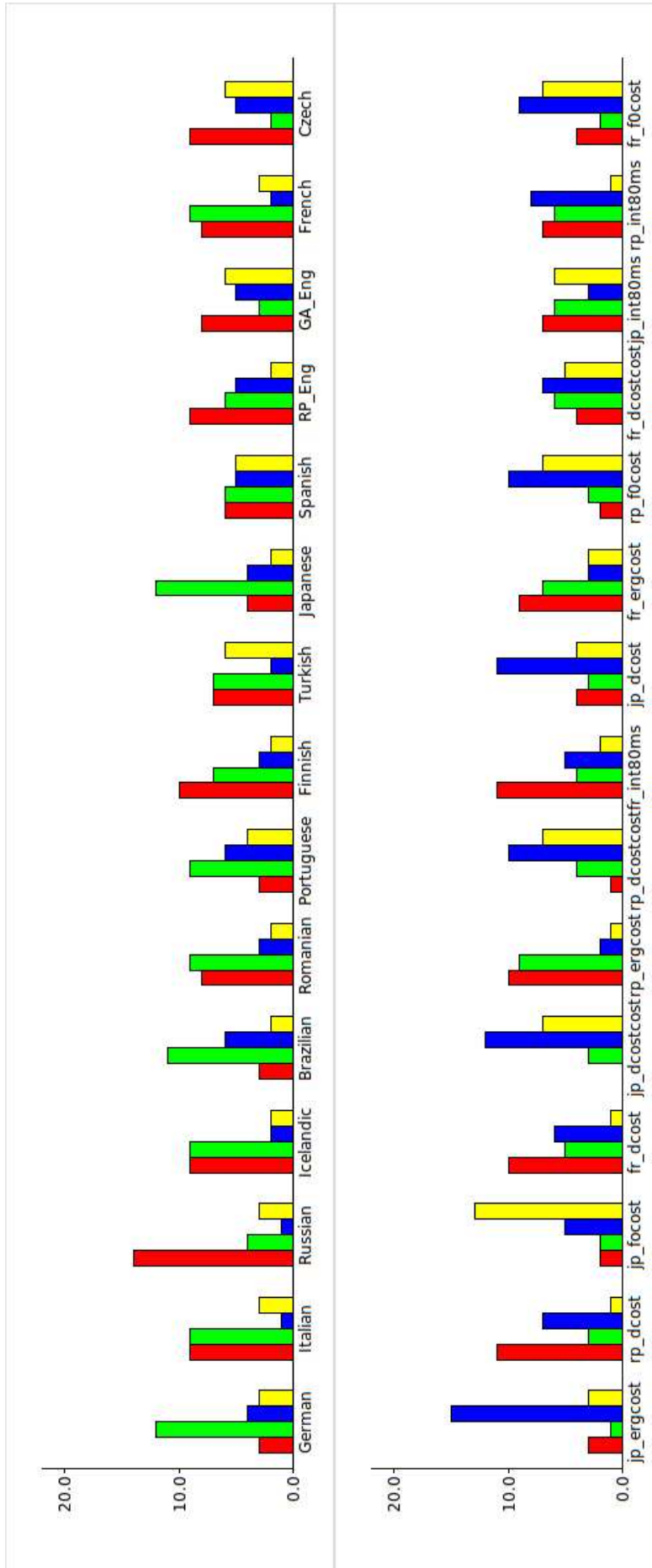
**Figure 9.1.** Results of the final part of the test, asking for a scalar categorisation of 7 samples of unknown languages (on the left) and 7 regional varieties of English (on the right). Data is presented on box-plots above and on scatter-plots below. 0 corresponds to A (synthesised RP English sample) and 100 corresponds to B (synthesised French sample). Cf. figure 6.9.

## 9. Appendices



**Figure 9.2.** Histograms representing the answers given to the preliminary phase of the test. Syllables are shown in the x-axis, the y-axis represents the number of people who marked a stress on each syllable (first level stresses in blue, second level stresses in red). Cf. figure. 6.3.

## 9. Appendices



**Figure 9.3.** Answers given by the 43 participants for each audio sample. For the first task (above), red bars indicate ratings for “English, German or similar”, green bars indicate ratings for “French, Spanish or similar”, blue bars indicate rating for “Other”, while yellow bars indicate ratings for “I don’t know”. For the second task (below), red bars indicate ratings for “English”, green bars indicate ratings for “French”, blue bars indicate rating for “Japanese”, while yellow bars indicate ratings for “I don’t know”. Cf. figure 6.7.



Paolo Mairano

2011