INTER-SUBJECT AGREEMENT IN RHYTHM EVALUATION FOR FOUR LANGUAGES (ENGLISH, FRENCH, GERMAN, ITALIAN)

Paolo Mairano[1] & Antonio Romano[1,2]

[1]Facoltà di Lingue e Lett. Straniere – [2]Dipartimento di Scienze del Linguaggio
University of Turin – Italy
paolomairano@gmail.com, antonio.romano@unito.it

ABSTRACT

This paper deals with the acoustic correlates of stress-timed and syllable-timed languages proposed by [8]. An experiment has been conducted in order to verify the validity of the three correlates using data of 4 languages (and some of their varieties). The novelty of this study consists in the fact that the segmentation of the acoustic data (based on the narrative "The North Wind and the Sun") was carried out independently by both authors.

The results show significant differences if compared between them and with those of other studies. However, the general tendency seems to confirm, at least partially, the validity of the three correlates even though they have been obtained from narrative texts.

Keywords: Stress-timing, syllable-timing, acoustic correlates, inter-subject agreement assessment.

1. INTRODUCTION

The recent flourishing of methods for rhythm assessment has given new impulse to the research in this field and several authors have tried to enlarge the perspectives, including as many variables as possible. Among these, [8] have suggested three acoustic correlates of stress-timing and syllable-timing, for whose calculation it is necessary to segment the acoustic data: since the segmentation of an acoustic sample for linguistic purposes has to be done entirely by hand (*pace* [12], see 3.2), we suggest that the results may present some degree of variation – depending on phonological evaluations – if the work is done by different subjects. So, we decided to carry out an experiment similar to the one in [8] with the difference that the measures would be taken by two subjects and their correlation would be assessed and shown in charts. This paper presents the experiment and the results obtained.

A brief history of the progress in the research about stress-timing and syllable-timing and a theoretical background are presented in section 2, while section 3 describes the experiment presenting the data and the methods used. Finally, the main results are discussed in section 4.

2. THEORETICAL BACKGROUND

As is well-known, the traditional view based on the 'stress-timed' and 'syllable-timed' dichotomy was introduced by [7]; [1] proposed the concept of *isochrony*, claiming that stress-timed languages exhibited regular inter-stress intervals (hence the name *iso-accentual languages*), whereas syllable-timed languages presented syllables of roughly the same length (hence the name *iso-syllabic languages*). Further investigation brought to the classification of most Romance languages as syllable-timed and of most Germanic and Slavonic languages as stress-timed. Some languages, e.g. Japanese and Tamil, did not fit in either category, but rather seemed to belong to a third rhythmic type based on the mora. However, various scholars did not find experimental evidence of isochrony at the foot nor at the syllable level (cp. [9]), concluding that the impression of syllable or stress-timing may merely be a matter of perception.

Other authors (e.g. [3] and [4]) suggested that stress-timing and syllable-timing had to be considered as the poles of a continuum along which languages are scattered and proposed a set of phonological properties shared by the languages belonging to the same rhythmic group, the most important being: a) the presence vs. absence of vowel reduction; b) a complex vs. simple syllabic structure; c) the tendency of stress to attract phonological material and to build up heavy syllables vs. no such tendency. The more of these properties a language possesses, the more it is likely to be near the syllable-timing or the stress-timing end of the continuum.

Among the different methods devised (e.g. [5], [8]), Ramus *et alii* [8] proposed three acoustic correlates of these phonological properties, namely

the standard deviation of vocalic intervals ΔV (indicative of the presence/absence of vocalic reduction), the standard deviation of consonantal intervals ΔC (indicative of the complexity of the syllabic structure) and the vocalic percentage %V (indicative of both the above mentioned phonological properties). In their study, stress-timed languages resulted, as expected, in higher ΔV and ΔC and a lower %V than syllable-timed languages.

We proposed to conduct a similar experiment (with some differences, see below) on 4 languages and compare the results with those of other studies ([2], [4], [8], [12], [13]).

3. THE EXPERIMENT

3.1. The data

The experiment was carried out on 4 languages (2 stress-timed – English and German – and 2 syllable-timed – French and Italian) with 7 different speakers: 1) an RP female speaker, 2) an American English female speaker, 3) a French female speaker, 4) a Canadian French female speaker, 5) a German male speaker, 6) an Italian female speaker, 7) an Italian male speaker.

The data consist of recordings of a narrative text (the well-known story *The North Wind and the Sun*) of a number of sentences varying from 10 to 21 with total durations ranging from 27 s to 43 s (for the different versions). The recordings of speakers 2, 3 and 5 were taken from the illustrations of the *Handbook of the International Phonetic Association*, while those of speakers 1 and 6 were taken from the *Journal of the International Phonetic Association*. The original recordings of speakers 4 and 7 were collected by the authors (the recordings of Italian are two different versions of the story). Table 1 shows the number of measures taken per each of the seven samples.

Table 1: Number of measures per sample (average between the two subjects).

Am. E	Br. E	German	Can .Fr	French	It. F	It. M
370	358	440	366	341	431	515

The kind of data used for this study presents considerable differences if compared with the data of many other studies on this subject. [8] recorded 4 female speakers uttering 5 "short, news-like declarative statements" per each of the 8 languages studied. [2] used 2 corpora of semi-spontaneous conversations (AVIP for Italian, the Kiel Corpus for German). [12] used a similar corpus (even

though larger in terms of number of speakers) composed of read sentences – the length and type of the sentences are not specified – but they used a controversial method of automatic segmentation.
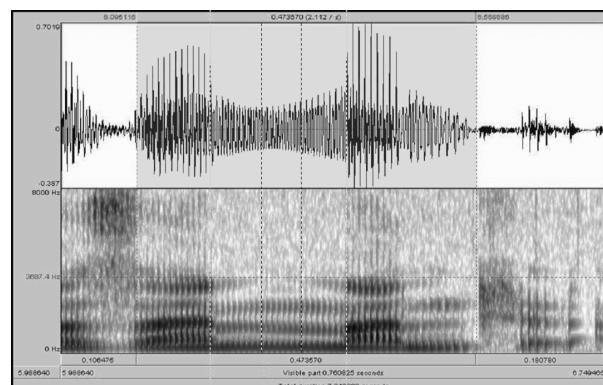
3.2. The segmentation

It is important to notice that each of the two authors carried out the same procedure independently (see table 1 for the number of segments measured). The measurements were taken with Praat as described below.

Firstly, each sample was segmented into interpausal strings. Secondly, each sound was identified and classified as consonantal or vocalic (following a careful evaluation of its phonological status) and its length was measured with an approximation of 0.5 ms. Some decisions had to be made: according to the phonological properties of the languages analyzed, prevocalic glides were considered as consonantal, postvocalic glides as vocalic, British [ɹ] and American [ɻ] as consonantal, syllabic consonants as vocalic. Moreover, the occurrence of some phones not belonging to the standard inventory of the language studied had to be accounted for, such as the double occurrence of [ə] in one of the Italian recordings and of [ç] in the French one.

Thirdly, the values were saved in a spreadsheet, where the durations of each consonantal and vocalic interval were calculated and then used to obtain the three parameters ΔV, ΔC and %V.

Figure 1: Example of a chain of segments difficult to classify without using phonological criteria: the syllabic nasal seems to be weaker than the adjacent non-syllabic nasals (image taken from PRAAT).

[z aɪ n n̩ m a n t ə l]

The segmentation and classification of certain co-articulated sequences was particularly problematic and resulted in a certain amount of discrepancy between the measures taken by the

two authors. (Overall correlation was assessed and is shown below.) In other cases it was a different phonological assessment of vowel-consonant trans-classifications which determined ambiguous choices (e.g. the syllabic [n] in the example, fig. 1).

3.3. The Charts

The three values ΔV, ΔC and %V were used in order to draw charts similar to the ones presented in [8]. They are all shown in figure 2 below. The error bars represent the standard deviation between the measures of the two subjects.

As a general rule, we can say that stress-timed and syllable-timed languages occupy different portions of the charts. However, the chart where a division between the two rhythmic groups is most clearly visible is the one including ΔC vs. %V, while in the other two (showing ΔV vs. ΔC and vs. %V) the difference is less evident. This is due to the fact that the values of %V and particularly of ΔV are slightly different from what had been expected. Namely, ΔV seems to be extremely high in the recordings of the Italian female speaker and of both French speakers, while it is lower than expected in American English. As for %V, its value seems to be too high in German.

4. DISCUSSION

In 3.3 it was hinted that the ΔV value – which is meant to be indicative of the presence/absence of vocalic reduction – is less useful to draw a divide between languages of the two rhythmic types as its value is sometimes different from what had been expected. We argue that the reason for this is that it

is too strongly influenced by various other factors, particularly speech rate. Moreover, it has to be remembered that even though vocalic length is not phonologically distinctive in Italian and French, these two languages may present lengthened vowels at the phonetic level, and this may presumably result in a higher value of ΔV.

As for the high %V value in German, it can possibly be accounted for by the fact that the speaker pronounced many syllabic consonants, which, being considered as vocalic (see above), contributed to raise the %V value.

As already mentioned, we believe that the main strength of our study is the fact that the measures were taken independently by both authors: this has the effect of preventing the results from being too much influenced by particular ways of working of a person (whom we call 'subject') or by his/her personal interpretation of the acoustic data. Furthermore, they also show how great a variability there can be among measures of the same passage (variability is of course greater in the cases of heavily co-articulated passages).

The measures taken by the two authors for each recording have been put on charts to see their correlation: the highest percentage of agreement between the two authors was achieved with the German recording, whereas the lowest one was achieved with the American English recording (see table 2). The scatterplots in fig. 3 represent the measures taken by each subject (on the two axes) for each language. If all measures matched perfectly, we would only see a bisector in the chart and inter-subject correlation would be 100%.

Figure 2: Our results on the charts devised by [8] showing the mean values of three correlates ΔC, ΔV and %V calculated with the measurements taken by each subject (error bars refer to inter-subject standard deviation).
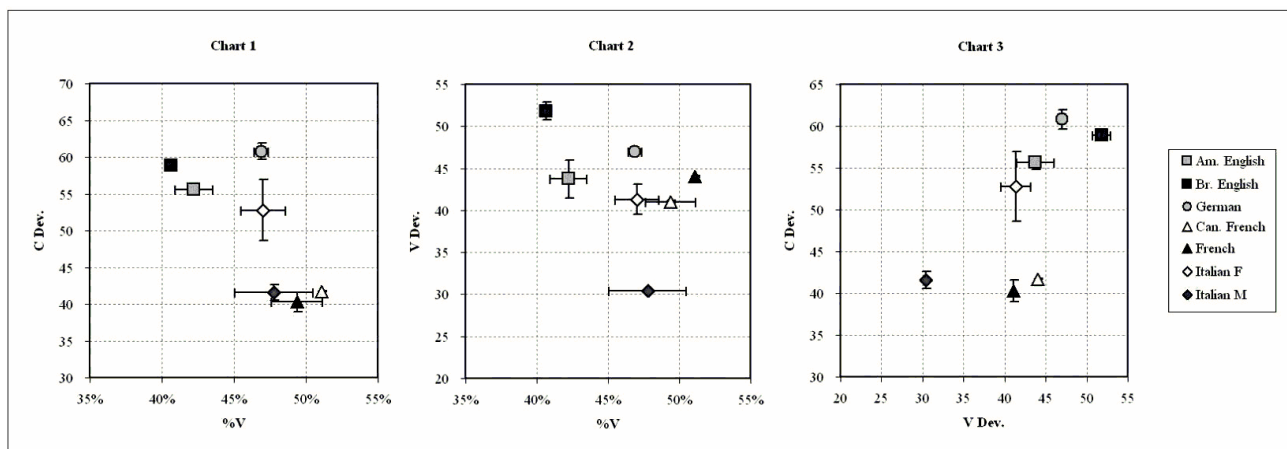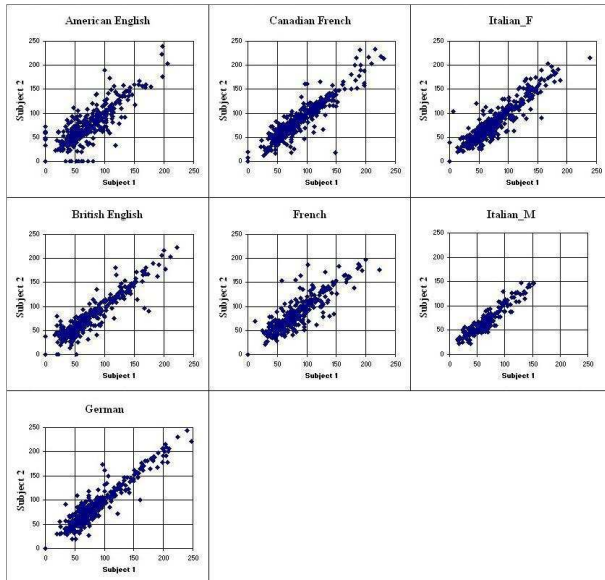
Table 2: Inter-subject correlation for each language sample analyzed.

Am. E	Br. E	German	Can. Fr	French	It. F	It. M
83,7%	93,9%	94,7%	91,6%	85,3%	93%	93,6%

Figure 3: Inter-subject agreement scatterplots.

It has to be noted that the points on the axes of these charts (which can mainly be seen in the scatterplots showing the measures of American and British English) indicate that one of the subjects included segments that the other did not consider separately (e.g. subject 1 transcribed [l̩] and subject 2 transcribed [əl]).

The general disagreement between subjects seems to have consequences mainly on %V (as for French, Italian-M and Italian-F) and affects a clear-cut discrimination of the four Romance samples from the Germanic ones (see fig. 2), which are however kept separate on the basis of ΔC (except for the Italian-F which exhibits an unexpectedly high value of this parameter).

5.   CONCLUSION

The results of our study confirm the validity of ΔC and %V as acoustic correlates of stress-timing and syllable-timing. However, with the kind of data analyzed, single parameters do not appear to be good discrimants, as they may be influenced by several factors (such as pausing and general prosodic structuring).

Furthermore, we believe that, as stated by [3], [4] and others, one should not aim at classifying languages as either syllable-timed or as stress-timed, rather at determining at which point of the continuum ranging from total stress-timing to total syllable-timing a language finds its natural collocation. Obviously, a larger corpus is needed for this purpose, informants have to be accurately selected (different diatopic variants of the same language might result in a different rhythmic type, cp. [2], [13]) and measures should be taken by several subjects in order to avoid inconsistencies.

These results show that even short narratives allow for a fairly reliable rhythmic evaluation providing that they are representative of the languages studied and, as pointed out by [2], present a similar degree of fluency and speech rate (we are, however, aware of the difficulties in comparing these phenomena interlinguistically).

6.   REFERENCES

[1]   Abercrombie, D. 1967. *Elements of General Phonetics*. Edinburgh University Press.

[2]   Barry, W., Russo, M. 2003. "Isocronia Soggettiva o Oggettiva? Relazioni tra Tempo Articolatorio e Quantificazione Ritmica". In: Albano Leoni, F. *et alii* (eds.), *Il Parlato Italiano*, Napoli: D'Auria (CD-ROM).

[3]   Bertinetto, P.M. 1989. Reflections on the dichotomy 'stress' vs. 'syllable-timing'. *Revue de Phonétique Appliquée*, Mons, 99-130.

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

[5]   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.

[6]   *IPA* 1999. *Handbook of the International Phonetic Association*. Cambridge: Cambridge University Press.

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

[8]   Ramus, F., Nespor, M., Mehler, J. 1999. Correlates of linguistic rhythm in the speech signal. *Cognition*, 73/3, 265-292.

[9]   Roach, P. 1982. On the distinction between 'stress-timed' and 'syllable-timed' languages. In: Crystal, D., *Linguistic controversies*, London: Edward Arnold, 73-79.

[10] Roach, P. 2004. "British English: Received Pronunciation". *Journal of the International Phonetic Association*, 34/2, 239-245.

[11] Rogers, D., D'Arcangeli, L. 2004. "Italian". *Journal of the International Phonetic Association*, 34/1, 117-121.

[12] Rouas, J.L., Farinas, J. 2004. Comparaison des méthodes de caractérisation du rythme des langues. *Proc. of MIDL 2004 "Identification des langues et des variétés dialectales par les humains et par les machines"* (Paris, 2004), Paris: École Nat. Sup. des Télécomm., 45-50.

[13] 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.