ISCA Archive
http://www.isca-speech.org/archive

Speech Prosody 2004
Nara, Japan
March 23-26, 2004

Introducing YARD (Yet Another Rhythm Determination) And Re-Introducing Isochrony to Rhythm Research

Petra S. Wagner[1] *and Volker Dellwo*[1,2]

[1]Institut für Kommunikationsforschung und Phonetik, Universität Bonn, Germany
[2]University College London, UK
{wagner;dellwo}@ikp.uni-bonn.de

10.21437/SpeechProsody.2004-52

Abstract

The %V/ΔC model of rhythmic class distinction captures syllable complexity rather than rhythm. An alternative rhythm measure will be proposed based on (z-transformed) syllable durations and inspired by the PVI-measure [1]. Some evidence for the existence of isochronous syllable sequences is presented which might provide a new approach towards a classification into stress timed and syllable timed languages.

1. Introduction

The phonetic intuition known as the Rhythm Class Hypothesis (henceforth: RCH) states that each language belongs to one of the prototypical rhythm classes known as either syllable timed, stress timed or mora timed. A rhythm class is defined as assigning equal durations to its basic rhythmic units (i.e. syllables, interstress intervals or morae). In this paper, we make the initial assumption that the RCH has *some* empirical basis. But despite extensive research, there has been little or no acoustic evidence so far, that speakers obey the rules of the RCH – they even seem to avoid isochronous units. While some researchers have claimed isochrony to be purely a matter of perception, others have looked for phonological evidence to explain the difference between syllable and stress timed languages: It could be shown that reduction phenomena and complex syllable structure are usually correlated with languages that have been claimed as stress timed (e.g. English, German) while languages lacking complex syllables and reduction appear to be often syllable timed (e.g. French, Italian) [2]. Other languages have been difficult to classify (Polish, Catalan) with the help of phonological characteristics.

The work by [3] has lead to a classification of rhythm classes on the basis of *acoustic* parameters. They calculate a measure for each language that is based on the percentage of vocalic intervals (%V) and the standard deviation of the consonantal intervals (ΔC) within an utterance. If these two values are plotted in a two-dimensional space, syllable and stress timed languages cluster in different areas. This model provides a working acoustic quantification of a formerly intuitive classification. However, several problems concerning this measure have been stated:

1. It describes syllable complexity rather than rhythm. [4]
2. It is not robust against the influence of articulation rate [3]
3. It has nothing to say about rhythmic properties of the respective languages in the time domain, even though this is what a rhythm measure ought to describe [4,5]

This paper will first discuss the problems of the %V/ΔC measure. Afterwards, an alternative measure, YARD, will be developed based on z-transformed syllable durations. YARD is robust against moderate articulation rate changes, independent of syllable complexity and models rhythmic classes along a single dimension. Then it will be shown, that syllables can be grouped into isochronous sequences in both stress and syllable timed languages – the distribution and length of these sequences differ in both rhythm classes.

2. Problems with %V and ΔC

In this section, the problematic points of the %V/ΔC measure will be discussed.

2.1. Syllable Complexity

It has long been known that syllable complexity and linguistic rhythm are somehow related (cf. 1.). Unless it is taken for granted that the notion of syllable complexity and variety *necessarily* go hand in hand with rhythm class, a rhythm measure ought to be more than an alternative representation of it. Thus, a simple test was performed: Four short texts (English, German (stress timed), French, Italian (syllable timed)) consisting of semantically and syntactically similar material were chosen. The texts are the reading material for the BONN-TEMPO database [6]. They were transcribed in a broad phonetic transcription, each vowel was then mapped onto symbol “V”, each consonant onto symbol “C”. Geminate consonants, diphthongs and tense vowels were counted as single segments. Reduced vowels (i.e. “Schwa”) were treated identical to non-reduced ones. Then, %V and DeltaC were calculated based on relative proportions of the number of consecutive V and C segments for each language and the results were plotted in a two dimensional graph (cf. Fig. 1)

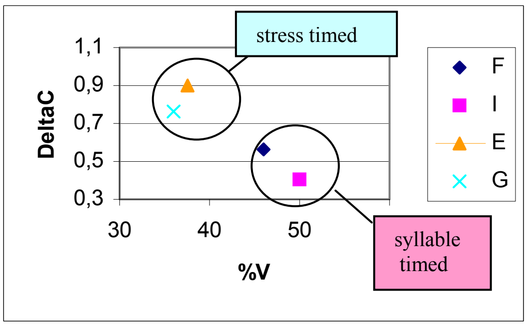

Figure 1: *Rhythm class distribution calculated on the C/V distribution for each language*

This primitive analysis does not contain any acoustic measurement but clearly mirrors the plots of related acoustic rhythm analyses. This is heavy support for the previously stated suspicion: %V and ΔC are an alternative representation of syllable complexity and variety, (however certainly more informative!). This is not too surprising since syllable complexity has been a good indicator of rhythm class and the %V/ΔC measure has been developed on the basis of this insight. Thus, a language inherent preference for a specific rhythm class can obviously be detected with the help of an analysis as it has been presented above, be it acoustically or segmentally based. Yet the empirical advantage of the %V/ΔC measure over a purely segmental analysis of syllable complexity remains to be shown.

2.2. The influence of articulation rate

There is evidence that %V and ΔC are strongly influenced by articulation rate. This is the reason why articulation rate was strictly controlled strictly during data collection by [2]. However, [7] showed that despite the fact that there is a strong influence of articulation rate, especially on ΔC, the categorical distinction between stress and syllable timed groups is not affected. Thus, the %V/ΔC appears to be more robust than expected. Still, it remains unclear in what way the evident influence of articulation rate on ΔC changes the rhythmic structure – which is what should happen if %V/ΔC is a correlate of rhythm. [1], who have built a rhythm model based on consonantal and vocalic intervals as well, also expected an influence of articulation rate, especially on vocalic intervals. To control this, they introduced a local duration normalization in order to factor out the influence of articulation rate on vocalic intervals. However, their normalization has the negative side effect of smoothing durational differences of adjacent vocalic intervals and apparently, their measure is not robust against articulation rate influences either [8]. To sum up: The impact of articulation rate on consonantal and vocalic interval duration is uncontroversial. There are two possible explanations for this phenomenon. Either, (i) the rhythm of a language changes at different articulation rates in a way yet unclear or (ii) a rhythm measure based on the proportion of consonantal and vocalic intervals captures properties of speech that are related to but still different from rhythm. It is certainly possible, that articulation rate may change the typical rhythmic pattern of a language, especially at the extreme ends of the spectrum of possible articulation rates or speaking styles. Such influences should be reflected in any rhythm measure. Besides, the influence of articulation rate on the rhythm measure ought to stand in a direct relation to rhythm and should be interpretable as such. It is concluded that a working rhythm measure ought to be robust against articulation rate changes as long as the rhythmic pattern of the regarded utterance is not affected.

2.3. A rhythm measure ought to capture rhythm

The third major problem of the %V/ΔC analysis lies in its lack of sequential analysis. Intuitively, rhythm can be regarded as a – more or less regularly - ordered event sequence. This defining property ought to be captured in a measure differentiating between linguistic rhythms. An orientation line could be the rhythm measure used in musical notation, where the re-occurring rhythmic pattern is stated. [1] try to capture the sequential nature of rhythm by measuring the local variability of adjacent consonantal and vocalic intervals in their PVI score. (Pairwise Variability Index). The PVI score indicates the proportion of the mean local differences for both interval types.

2.4. Desiderata for an alternative rhythm measure

We conclude, that a good rhythm measure shoud be (i) more than a measure of syllable complexity of in a language and (ii) be robust concerning influences of articulation rate as long as they are not related to rhythmic change and (iii) capture the sequential nature of rhythmic patterns.

3. YARD – Yet Another Rhythm Determination

In this section, an alternative model for the measurement of rhythm class (henceforth YARD) will be developed and evaluated on the basis of the desiderata stated in 2.4. YARD is based on normalized syllable durations and does not regard vocalic and consonantal intervals separately. It thus uses more traditional units of rhythm research. The measure will be examined on a small subset of data extracted from the BONNTEMPO database:

- 1 sentence, translated into four languages
- every speaker read the sentence at five different speeds
- languages examined were English (stress timed, 7 speakers), French (syllable timed, 6 speakers), German (stress timed, 15 speakers), Italian (syllable timed, 3 speakers)

One of the key obstacles to a robust measure appeared to be the influence of articulation rate on the major rhythm related acoustic property: duration. Since it is possible that within one language the rhythm changes due to shifts in articulation rate, the data was recorded at five different speeds: normal reading speed, slower, even slower, faster than normal, as fast as possible. This way, the data is representative for all articulation rates but can also serve for studies on the influence of articulation rate on rhythm.

In order to make the different durations of different speakers across different languages comparable, the measured syllable durations were normalized with the z-transformation. This is achieved by equating the mean syllable duration with 0 and the standard deviation with 1. A normalized syllable duration (z_i) of the absolute duration (x_i) is now measured as units of standard deviation (σ):

$$z_i = \frac{x_i - \overline{x}}{\sigma} \tag{1}$$

In order to investigate the influence of articulation rate on the syllable durations, correlation coefficients (Pearson) were calculated between all speakers at all reading speeds within each language. The correlations were substantial, even taking into account the extreme reading speeds (very slow, as fast as possible). Only Italian speakers seem to disagree more across different utterances. If the mean correlation coefficients are calculated for the different articulation rates separately, it becomes obvious that there is a heavy influence of articulation rate on Italian rhythm, making Italian look more irregular (stress timed?) at slow speech (cf. Figure 2; [8]). Therefore, the average correlations are higher for the other languages (cf. Table 1). At normal speed, syllable durations correlate with each other well above 0.7 for all languages.

Table 1: *Mean correlation coefficents across different speakers and articulation rates*

English	French	German	Italian
0.71	0.63	0.67	0.43

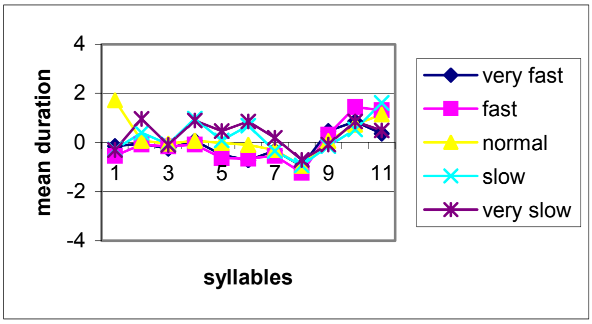

Figure 2: *The influence of articulation rate on Italian rhythm: Mean syllable durations (z-scores) of the utterance "Il giorno dopo andai a Bologna"*

Figures 2-5 give the syllable durations of the examined utterance for each language. The fact that rhythm (here: syllable durations) correlates substantially even across different articulation rates conforms to the results by [9] who could show that speakers are surprisingly well able to synchronize their duration patterns. We conclude that there is high inter-speaker agreement about well fomed duration patterns. The normalized syllable durations fulfill one of the demands of a good approach towards measuring rhythm: They are fairly robust against articulation rate influences which do not change the rhythmic style of a language variety. All further calculations are based on the mean syllable durations measured for all articulation rates and speakers.

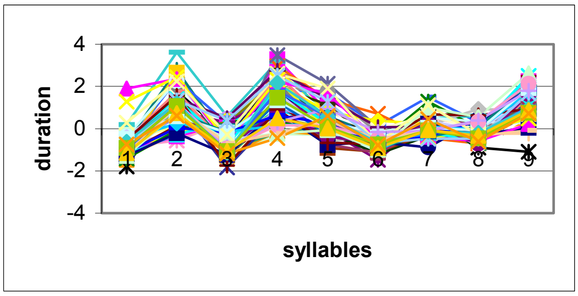

Figure 3: *Syllable durations (z-scores) of the German utterance (15 speakers, 5 speeds each): "Am nächsten Tag fuhr ich nach Husum"*

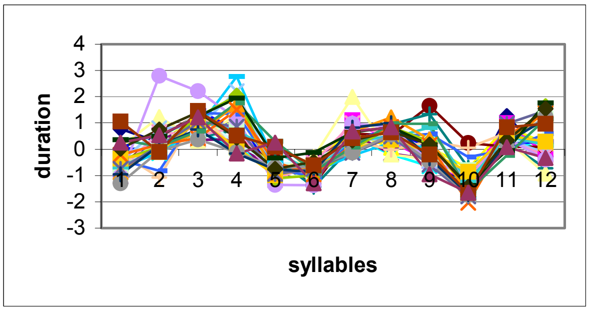

Figure 4: *Syllable durations (z-scores) of the French utterance (6 speakers, 5 speeds each): "Le jour suivants je me suis rendue à Albi"*

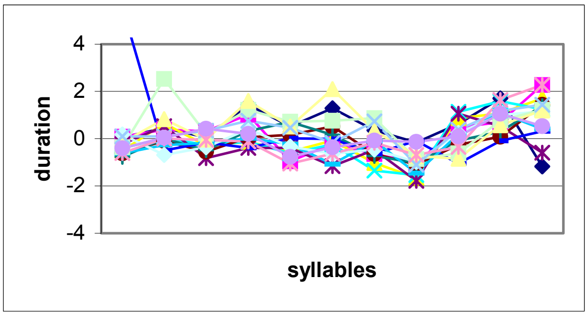

Figure 5: *Syllable durations (z-scores) of the Italian utterance (3 speakers, 5 speeds each): "Il giorno dopo andai a Bologna"*

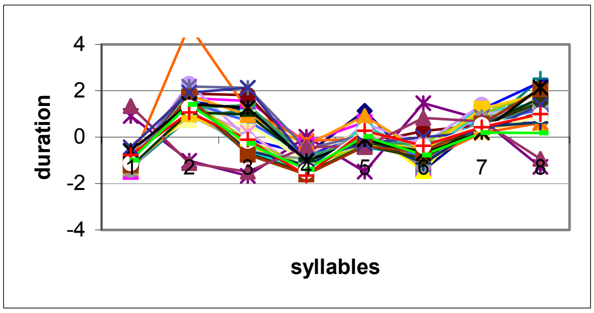

Figure 6: *Syllable durations (z-scores) of the English utterance (5 speakers, 5 speeds each): "The next day I went to Falmouth."*

But do syllable durations provide a working quantification of rhythm class? The obvious expectation concerning our data is that stress timed languages exhibit more variety across adjacent syllables than syllable timed languages. The latter should be characterised by a rather flat duration contour, keeping adjacent syllable durations similar. On a purely visual impression, this hypothesis receives some support: stress timed languages show more pronounced peaks and valleys, syllable timed languages have a smoother shape, allowing several subsequent syllables of similar length. In order to quantify this impression an index is calculated inspired by the PVI-measure, but simply based on the z-transformed syllable durations. The result is called the YARD index:

$$\text{YARD} = \sum_{i=1}^{n-1} / sylldur(i) - sylldur(i+1) / \Big/ n-1 \qquad (2)$$

The results (cf. Table 2) are encouraging since they appear to mirror the traditional rhythm class distinctions. Still, they should be regarded with care, because again, it could simply be another index of syllabic complexity, forcing speakers of English to lenghten the syllable "next" /nɛkst/ because it contains many segments and not on the grounds of independent rhythmic constraints. In order to test this, additional YARD indices were calculated based on the number of segments. Then, correlation coefficients (Pearson) between mean local acoustic and their corresponding segmental YARD values were calculated. The results of the correlation analysis seem arbitrary, which speaks for an independence of both indices. But: the categorical distinction between both rhythm classes is mirrored in the segmental YARD indices. However, the lack of clear correlation between local YARD values shows that the acoustic YARD captures different properties than the segmental one.

Table 2: *YARD indices based on the PVI values of z-scores of syllable durations and their correlation with YARD scores based on phonological complexity.*

class	stress timed		syllable timed	
language	German	English	French	Italian
YARD acoustic	1.20	1.18	0.7	0.46
YARD segmental	0.9	1.9	0.4	0.6
cc	0.12	0.48	0.58	0.05

4. Outlook: A Bottom-Up Search for Isochronous Units

With the YARD index we found a measure that captures the sequential irregularities within an utterance that are typical for stress and syllable timed languages. The normalized syllable durations provided stable rhythmic patterns across different speakers and moderate changes of articulation rate. Still, YARD is still less informative as the rhythm measures known from musical notation or poetic meter, since it does not contain any information about the *order* of a well fomed rhythmic event structure such as *iambic* vs. *trochaic*, 3/4 vs. 4/4 etc. So, we attempted a different approach at looking for rhythmic structure in speech:

So far, the search for isochronous units in speech proved unsuccessful, but maybe this is because research did not look in the right places: In a bottom-up procedure not taking into account a priori expectations about syllable timed or stress timed languages, we searched for isochronous syllable sequences in our data. (Roughly) isochronous units were found for each language, but they seem to be hooked to the prosodic hierarchy in very different ways (cf. Table 3).

Italian and French tend to put more syllables into isochronous units (≈4), while English and German prefer binary sequences. Also, the relationship to syntactic boundaries is more evident for both German and English. In all languages, isochronous units have the tendency to terminate with lengthened syllables and a pitch accent, thus adding further support to the hypothesis of final lengthening as a universal prosodic principle.

Table 3: *Roughly equidistant units and their durations of the languages under discussion*

Le jour sui	*vants je me suis*	*rendue à Al*	*(bi)*
4.6	*4.8*	*4.7*	*(1.7)*
Il gior no do	*po an dai a Bo*	*lo gna*	
3.6	*3.6*	*3.8*	
The next	*day I went*	*to Fal*	*mouth*
2.5	*2.3*	*2.1*	*2.2*
Am näch	*sten Tag*	*fuhr ich nach*	*Husum*
1.9	*2.3*	*2.3*	*2.7*

5. Conclusion

The %V/ΔC rhythm measure provides a working discrimination of rhythmic classes based on syllable complexity, but it misses a representation of the sequential nature of rhythm. A local variability measure based upon normalized syllable durations proved to be stable alternative across different speakers and articulation rates. However, we believe that future research should again concentrate on regarding rhythm as a sequence of – roughly - isochronous events with language specific internal structures. A first bottom-up attempt at finding such sequences was successful. Their relationship to higher order linguistic units, pitch and the level of perception remains unclear. Of course, much more research is necessary to find out whether these first impressions hold for larger data sets.

6. References

[1] Low, E.L. and Grabe, E., 1995. Prosodic patterns in Singapore English, Proceedings of the ICPhS, vol. 3, pp. 636-639, Stockholm, Sweden.

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

[3] Ramus, F.; Nespor, M. and Mehler, J., 1999. Correlates of linguistic rhythm in the speech signal, *Cognition,* 72, 1-28.

[4] Cummins, F., 2002. Speech Rhythm and Rhythmic Taxonomy. Proceedings of Speech Prosody 2002, Aix en Provence, France.

[5] Gibbon, D. Computational Modelling of Rhythm as Alternation, Iteration, and Hierarchy. Proceedings of ICPhS 2003, 2489-2492.

[6] Dellwo, V.; Steiner, I.; Wagner, P., to appear. The BONN-TEMPO Database. IKP Arbeitsberichte, Neue Folge, http://www.ikp.uni-bonn.de/ikpab/.

[7] Dellwo, V. and Wagner, P., 2003. Relations between Language Rhythm and Speech Rate. Proceedings of ICPhS 2003, Barcelona, Spain, 471-474.

[8] Barry, W.J.; Andreeva, B.; Russo, M; Dimitrova, S.; Kostadinova T., 2003, Do Rhythm Measures Tell Us Anything about Language Type? Proceedings of ICPhS 2003. Barcelona, Spain, 2693-2696.

[9] Cummins, F. and Port, R. F., 1998. Rhythmic constraints on stress timing in English. *Journal of Phonetics*, 26(2): 145-171.