On the Analysis of Rhythm through Acoustic Parameters

Ingmar Steiner

Institut für Kommunikationsforschung und Phonetik, University of Bonn, Germany
ist@ikp.uni-bonn.de

1 Overview

This paper presents an abstract of my M.A. thesis, entitled "Zur Rhythmusanalyse mittels akustischer Parameter", received by the School of Humanities of the University of Bonn in December 2003 and graded on January 8, 2004. The full text version (in German) can be found online at

http://www.ikp.uni-bonn.de/~ist/files/pdf/MA-Arbeit.pdf

2 Background

The rhythm class hypothesis, proposed by Pike (1945) and Abercrombie (1967), and founded in a notion of isochrony, although widely discussed in the past (e.g. Uldall 1971), has been thoroughly refuted by instrumental means (e.g. Ladefoged 1967) for several decades. Nevertheless, it has remained a popular idea in linguistics and phonetics and been reformulated in a weak form, which maintains isochrony on the perceptual level. The central claim of the rhythm class hypothesis is that there are two rhythmic types[1], syllable-timed languages and stress-timed languages, and that all languages fall into one of these two types. In syllable-timed languages, all syllables have the same length and are thereby isochronous, whereas in stress-timed languages, the intervals between stressed syllables are isochronous.

Several reformulations of the rhythm class hypothesis have been attempted in light of the untenable isochrony condition. Several of these (Dasher & Bolinger 1982, Dauer 1983) base the rhythmic distinction of languages mainly on syllable complexity, as well as the occurrence of vowel reduction. According to this explanatory approach, stress-timed languages are those that allow both reduced vowels and complex syllables, syllable-timed languages permit neither. This idea has recently been developed into an acoustic account of linguistic rhythm (Ramus, Nespor & Mehler 1999), which classifies languages based on statistical measures of duration, which correlate with these two rhythm class parameters.

Ramus, Nespor & Mehler segment utterances in various languages into vocalic and consonantal (i.e. intervocalic) intervals. The percentage of vocalic intervals of the overall utterance duration is referred to as *%V*, the standard deviation of the durations of consonantal intervals is termed *ΔC*. Since syllable-timed languages prohibit vowel reduction, their *%V* scores should be higher than those of stress-timed languages, while the absence of complex syllables predicts a lower *ΔC* score. When mean values for various languages are plotted in the *%V/ΔC* plane, these languages do indeed cluster according to their traditional rhythm class membership.

A number of subsequent studies (Grabe & Low 2002, Duarte et al. 2001) vary the data considered as well as the statistical measures used, but all rely on an initial *C/V* segmentation.

---

[1] A third rhythmic class of so-called mora-timed languages has been proposed (e.g. Bloch 1950) but is irrelevant to this discussion.

3 Experiment

An initial data preparation in the form of *C/V* segmentation presents a potential problem to any study that draws conclusions based on segmental durations, since all measures derived from these durations are influenced by the placement of segment boundaries and classification. Depending on how these decisions are made, important information relevant to rhythmic analysis can be missed or misinterpreted.

Problematic cases are presented by syllabic consonants, especially liquids, and by approximants. Although Ramus, Nespor & Mehler (1999) acknowledge the existence of these issues, both they and Grabe & Low (2002) classify only vowels as vocalic intervals, regardless of consonantal syllabicity, and prevocalic approximants as consonants. Their interpretations thereby represent not so much acoustic as phonological analyses of rhythm, regardless of subsequent statistical measures.

Since speech rhythm has been widely characterized as a pattern of syllable prominence, Galves et al. (2002) attempt to automatically extract a sonority measure from the speech signal, which can be used as an alternative to manual *C/V* segmentation. One such automatic annotation algorithm (Garcia, Gut & Galves 2002) was tested on the data used for the present study and found to yield strongly inferior results to manual segmentation.

The analysis of rhythm in the present study was performed on data from the Bonn-Tempo Corpus (Dellwo et al. 2004), which contains recordings of read speech by several speakers, male and female, in a number of languages, including German, English, French, and Italian. Each language uses a text of semantically equivalent content averaging 100 syllables in length, read at five different intended speech rates per speaker and language. The recordings are manually segmented into syllables, as well as into consonantal and vocalic intervals, based on acoustic and visual cues. All data preparation was performed using PRAAT. For the present study, data were selected from the BonnTempo Corpus to include 4 German, 5 English, 5 French, and 2 Italian speakers.

To enable a refined analysis of durational rhythm correlates as proposed by Ramus, Nespor & Mehler, manually *C/V* segmented data were relabeled to identify segments by sonority class. The resulting 8 interval types (vowels, approximants, syllabic and non-syllabic laterals, syllabic and non-syllabic nasals, fricatives, and stops) were then statistically analyzed in all possible combinations, both with regards to percentage of overall duration and standard deviation of cluster duration. This analysis yielded 262 independent variables that were analyzed in SPSS. A discriminant analysis was performed to explore the individual parameters' weight in the rhythmic classification.

4 Results

Statistical analysis of the 262 parameters shows that within the group of consonants, laterals and nasals are principally responsible for language distinction by durational measures. The results indicate that for the data used in the present study, the two most efficient factors in the maximal separation of languages and rhythmic classes are the percentages of laterals and nasals.[2]

These results show that consonantal clusters, and with them, measures such as *ΔC*, are not homogeneous in their influence on durational measures of speech rhythm. Rather, language distinction seems to be strongly influenced by phonotactic factors, even in

---

[2] These results are not significantly influenced by syllabicity of laterals or nasals, as was determined by specific tests.

rhythmic analysis. Such an analysis that is based on interval durations after segmentation should be aware that the segmentation itself is likely to have a significant influence on the results of the analysis.

Nevertheless, it should be taken into consideration that although it may be possible to model certain aspects of the rhythm class hypothesis through durational measures, it seems improbable that rhythm can be adequately captured by duration alone. Other aspects of syllable prominence should be taken into consideration as well, and descriptions building on sonority as a syllabic basis for rhythm also appear promising.

5 References

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

Bloch, Bernard (1950) "Studies in colloquial Japanese IV: Phonemics." *Language* 26, 86-125.

Dasher, Richard & Dwight L. Bolinger (1982) "On pre-accentual lengthening." *Journal of the International Phonetic Association* 12, 58-71.

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

Dellwo, Volker, Ingmar Steiner, Bianca Aschenberner, Jana Dankovičová & Petra Wagner (2004) "BonnTempo Corpus and BonnTempo Tools: A database for the study of speech rhythm and rate." *Proceedings of the 8th International Conference on Spoken Language Processing*, Jeju Island, Korea. [in press]

Duarte, Denise, Antonio Galves, Nancy L. Garcia & Ricardo Maronna (2001) "The statistical analysis of acoustic correlates of speech rhythm." *Proceedings of the Workshop on Rhythmic Patterns, Parameter Setting and Language Change.* Bielefeld: ZiF, Universität Bielefeld.

Galves, Antonio, Jesus Garcia, Denise Duarte & Charlotte Galves (2002) "Sonority as a basis for rhythmic class discrimination." *Proceedings of Speech Prosody 2002.* Aix-en-Provence.

Garcia, Jesus, Ulrike Gut & Antonio Galves (2002) "Vocale – A semi-automatic annotation tool for prosodic research." *Proceedings of Speech Prosody 2002.* Aix-en-Provence.

Grabe, Esther & Ee Ling Low (2002) "Durational variability in speech and the rhythm class hypothesis." In: Carlos Gussenhoven & Natasha Warner (eds.), *Papers in Laboratory Phonology 7.* Berlin: Mouton de Gruyter.

Ladefoged, Peter (1967) *Three Areas of Experimental Phonetics.* London: Oxford University Press.

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

Ramus, Franck, Marina Nespor & Jacques Mehler (1999) "Correlates of linguistic rhythm in the speech signal." *Cognition* 73, 265-292.

Uldall, Elisabeth T. (1971) "Isochronous stresses in R.P." In: L. L. Hammerich, Roman Jakobson & Eberhard Zwirner (eds.), *Form and Substance.* Copenhagen: Akademisk Forlag.