SCUOLA
NORMALE
SUPERIORE
PISA

Zofia Malisz

Segment-prosody interaction and phonetic models of speech rhythm and rhythmic typology.*

(work in progress)

1 Introduction

A new interest in rhythm taxonomies was awoken when new metrics for quantifying traditional prosodic distinctions between languages emerged a few years ago (Ramus, Nespor and Mehler 1999, henceforth RNM; Low, Grabe and Nolan 2000, Low and Grabe 2002, henceforth LG). Authors of the metrics shared a common approach trying to find a phonetic rather than impressionistic basis for the standard typological continuum of syllable- to stress-timing. By using variability and distribution statistics of consonantal and vocalic intervals they tried to test and evaluate the descriptive power of the standard typology.

The view that language phonotactics and syllable structure determine rhythm classes derives from Dauer (1983) and Bertinetto (1989). Their criteria of classification produce a cumulative result of phonological language specific properties, such as syllable structure and vowel reduction, that locate languages on a scale from syllable- to stress-timed. This way, segment durations and quality as well as phonotactic segment configurations are the proposed determinants of language type.

Studies such as RNM and LG used consonantal and vocalic stretches as the base units for calculation of classificatory statistics so that, contrary to work in the past (Abercrombie 1967, Lehiste 1977, Pike 1945), direct phonetic correlates were adopted.

---

* School of English, Adam Mickiewicz University, Poznan, Poland. The present work has been developed during the academic year 2006/2007 supported by the annual research grant for foreign PhD students in the Laboratorio di Linguistica, Scuola Normale Superiore.

Seeking a phonetic validation for the phonologically based, standard rhythm class hypothesis, the smaller out of the phonological units used formerly, the syllable, was broken down into its basic acoustic manifestation, the CV sequence. Such a choice may suggest that consonantal and vocalic intervals are the production and perception constituents that take part in rhythmical processing on the phonetic level. It is certainly so, that in general C and V durations play a role in the processing of timing. The prominent position of vowels in prosody is undisputable, as e.g. LG notably explain. The present paper however, will suggest a need for a more finely-grained analysis of consonants and a closer look at their relation to vowels. Such an analysis will also involve a nonlinear approach to timing which goes beyond the sequential, phonotactically motivated one adopted by RNM and LG; the alternative approach will be specified in section 3. First however, I will continue with the discussion of the problem concerning elements that serve as an input for the statistical analyses as seen in RNM and LG.

2 Linear models of speech rhythm

Steiner (2003) was able to show that including consonants en bloc into the analysis, misses some consonant type dependent effects on the durational output statistics. Steiner analyzed German, French, English and Italian corpus data using the following data preparation strategy: the signal was tagged according to the sonority scale criteria to arrive at six types of intervals instead of two: vowels, approximants, laterals, nasals, fricatives and stops and including separate labels for syllabic nasals and laterals. Next, RNM parameters were extracted and compared using different combinations of interval types. The results showed that the combination of parameter types that maximally separates the standard syllable- from stress-timed languages is not %V (proportion of vocalic intervals) and ΔC (standard deviation of consonantal intervals) ($r = 0.822$) but %l and %n ($r = 0.902$), the percentage of laterals and nasals in the given corpora.

For comparison, we also find a sonority based approach in Galves et al. (2002). The segmentation strategy used in RNM was changed by directly and automatically extracting intervals from the signal as a function of sonority, thus circumventing handlabelling controversies. However, the approach differs from the one used in Steiner in one significant respect. Contrary to Steiner, Galves et al. acoustically measured sonority contrasting it with obstruency, providing more coarse-grained rather than detailed phonetic distinctions (as this way e.g. nasals, liquids and vowels tend to collapse into one category). The results were correspondingly different. Galves et al. presented a linear correlation between the rough sonority vs. obstruency measures and %V and ΔC measures as found in RNM. This way, Galves et al. were able to represent infant phonotactic processing abilities at around 6-9 months of age, when babies are still not so sensitive to details in spectral properties of consonants. It has to be noted that the RNM study was primarily designed to illustrate infant language discrimination abilities based on prosodic cues.

However, the classical rhythm class hypothesis rests on impressionistic observations made by adults and therefore a more finely grained analysis, such as the one in Steiner, seems more appropriate to represent typological distinctions. Specifically, Steiner's results suggest that "the functional load of consonant classes is not homogenous" (2003:6) so that the phonotactic profile pertaining to rhythmical classification needs to be more detailed than the one found in RNM. His conclusions in fact point out the need to revise the parameters in order to include acoustic correlates, other than durational, into the analysis e.g.: pitch, loudness and segmental quality.

Until now, essentially durational measures have been constructed on the level of analysis that involves the segment and syllabic phonotactics. Measures where the only differentiation in terms of spectral and articulatory characteristics of the input units, other than duration, was the general consonant/vowel distinction. The issue of intensity or segment quality and their relation to duration and prominence has been explicitly omitted, also in perceptual discrimination experiments. For the purpose of perceptually evaluating the parameters in Ramus, Dupoux and Mehler (2003) the data was resynthesised using the "flat sasasa" technique, which eliminates intonation, prominence alternations resulting from segmental quality and distribution of energy in the signal. Such a strategy was necessary for extracting the durational effects, however the remaining problem of the significance of other correlates was not missed by the authors who commented that:

> ”Languages differ in the way they use duration and intensity to signal phonological properties such as stress or quantity. It can therefore not be excluded that a similar quantitative, cross-linguistic study of intensity variations might provide yet another dimension for the study of rhythm classes.”(2003:341)

If one of the aims of phonetic implementation of rhythm classes is to find out how rhythms are being extracted perceptually, one needs to include the missing correlates to construct a more complete measure. Indeed one may imagine a measure similar to the one used in LG: a calculation of global phonotactic characteristics with local effects based on energy differences within pairs of adjacent intervocalic intervals. Considering that psycholinguistic literature suggests that the auditory system integrates energy with time over a span of an average syllable (Moore 1989), the analysis could be extended to the perceptual domain by taking total energy measurements within a syllable or syllable-like units (V-to-V units) similarly to Gordon (2002). Further research will be necessary to find out the appropriate formal representation of such a measure of energy. Theoretical implications would also have to be taken into consideration e.g. regarding the kind of typological distinctions the measure would imply when correlated with durational indices.

At this stage of development of the type of linear approaches discussed in this section, a revision of correlates and parameters is necessary also in the context of several cases of unclassified languages. As we only have the parameter space delimited by the canonical stress-timed and syllable-timed languages at our disposal, the choice of parameters is defined and restricted by the reference languages. The problematic languages fall out of the space at some point because their status may be changed depending on the combination of parameters, at least in the case of RNM. For example, Polish can be classified as either closer to the syllable- or stress-timed languages in RNM and LG so in fact its status is unclear. No explanation or solution for the problem of mixed languages has been offered within the limits of the linear framework, so no clear accommodation of the languages into the framework is given (cf. Nespor 1990).

3 Beat-based approaches in hierarchical models of rhythm

An alternative element in the segment-prosody interaction domain, able to integrate both spectral energy changes and duration, is the CV transition. The CV transition is the assumed locus of the p-centre, as demonstrated by numerous researchers (Fowler 1983, Pompino-Marschall 1989, Janker 1996, Barbosa et al. 2005). The p-centre is the most likely candidate for anchoring the metrical "beat" in representation of speech rhythm (Morton, Markus and Frankish 1976, Scott 1993). This is important if we assume that rhythm is also hierarchically structured, not only sequentially organized.

While the tradition of isochrony in phonetics, which rests on a linear representation of rhythm, inspired the metrics described in 1, the concept of metrical hierarchy was to a large extent the focus of much research in Metrical Phonology (Prince 1983, Hayes, Selkirk 1984). The empirical validation of the concept was achieved by Cummins and Port (1998) and independently extended in studies within the dynamical approach to speech rhythm (Barbosa 2006, Port and Leary 2000, Port et al. 1988). The dynamical approach demonstrates a temporally rich implementation of alternation and relative prominence and manages to merge the formal advantages of the metrical grid (serial order, sequence of alternating "beats") and the metrical tree (structure, nesting of metrical levels) adding the continuous, non-discrete dimension of speech rhythm by means of relative phase.

In hierarchical approaches it is crucial to select units on each level of structure that indeed take part in the interaction between levels. In other words, as regards segment-prosody interaction, higher levels of timing hierarchy (be it the syllable, foot, phrase) require specific points which will serve as anchors for the metrical beat, yield themselves to the influence of top-down rhythmical effects and will also have their own impact bottom-up. As said, the mapping of metrical beats onto the phonetic level appears to proceed via p-centres. These in turn are claimed to underlyingly correspond to vowel onsets. However, the location of beats in the acoustic domain, depend very much on the CV transition's spectral and durational characteristics.

The vocalic identity of the linguistic beat was evaluated in many studies (Dziubalska-Kolaczyk 2002, for a discussion). Also in a recent study in the linear paradigm Smith and Mattys (2007) found that the parameters based on the normalized and non-normalised vocalic variability and proportions (VarcoV, V%, vocalic PVI) give best results in perceptual language discrimination tests. Why then again turn towards a more finely-grained analysis of the consonant also in the beat-based approach?

Barbosa et al. (2005) showed that the changes in spectral energy along a CV syllable have a bearing on how the p-centre is represented acoustically, even though the vowel seems to be the abstract target for the p-centre. The role of the perceptual boost given to the vowels and to their rhythmical function by the consonantal context needs to be studied cross-linguistically. The dynamical energy landscape created by consonant-vowel transitions may help to understand the relative prominences in utterances specific to individual languages. How can the phonotactics in fact support prominence patterns resulting from the interplay of the consonantal ground and the vocalic figures?

The relevant prosodic information is coded in the dynamics of the signal and the dynamics of the perception of the signal. Even if underlyingly the vowel is the attracting point for the p-centre and thus for the placement of the metrical beat, it is worth asking what the strategies that handle phonotactic variability and articulatory limits are. How do languages provide the abstract V-to-V intervals in spite of perturbations produced by language specific phonology.

Tasks such as Barbosa et al. (2005) allow to look closer at individual languages and explore the vocalic identity of the p-centre/beat relative to its consonantal context. The dynamics of the task allows to observe how the variable interval windows across rates influence the speaker to adapt the phonology of their language to extract the moment of occurrence of the p-centre/beat accurately. This allows us to observe how speakers manage duration and the energy profile of the syllable, to relate the two features and produce a beat.

A p-centre task based on Barbosa et al. (2005) was devised and conducted by the present author with Italian subjects. Cross-validation studies of p-centre perception with native and non-native speakers of Polish and Italian are planned in order to test the universality of the phenomenon's acoustic implementation. Non-native judgments of regularity of Polish and Italian syllables spoken to a metronome will be the basis of the study.

4 The linear and nonlinear approaches and the general model of rhythm, final remarks.

The evaluation of linear metrics, such as Smith and Mattys (2007), often proceeds according to the objective of how neatly the standard categories are separated out from the data. The question of how well they match some putative criteria necessary for an accurate modelling of rhythm in general is not addressed, very often also because a working definition of speech rhythm is not given a priori.

The general rhythm model needs to include criteria such as: alternation of prominences, iteration of events, hierarchical structure (Gibbon 2003). Most of these criteria have not been addressed by authors of metrics discussed in 1 and 2. The empirical problem lies both in the choice of units as well as in the eventual statistics performed on the segmented data (Gibbon and Fernandes 2005). But also, there exists a fundamental difference of theoretical focus. While the studies discussed in 1 and 2 state finding a unit or a measure that qualifies languages according to standard divisions as their goal, the work described in section 3 subscribe to a different option. Testing for fundamental and universal properties of rhythm processing appears to be the priority and differences resulting from language specific phonology are considered as a further step. This particular roadmap consists in stating some general principles that underlie an accurate rhythm model and refer to a cognitively plausible, holistic basis.

References:

Abercrombie D. (1967) Elements of general phonetics. Edinburgh: Edinburgh University Press.

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

Barbosa P.A., Arantes P., Meireles A., Vieira J.M. (2005) Abstractness in speech-metronome synchronisation: P-centres as cyclic attractors. *Proceedings of the Ninth European Conference on Speech Communication and Technolog* (Interspeech 2005) Lisbon, Portugal.1441-1444.

Bertinetto P.M. (1989) Reflections on the dichotomy of stress" vs. "syllable-timing". Revue de Phonétique Appliquée 91-93:99-130.

Cummins F and Port R. (1998) Rhythmic constraints on stress timing in English. *Journal of Phonetics 26*(2):45-171.

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

Dziubalska-Kolaczyk K. (2002) *Beats and Binding Phonology.* Frankfurt am Main: Peter Lang.

Galves, A.,Duarte D., Garcia J. and Galves C. (2002) Sonority as a basis for rhythmic class discrimination, *Speech Prosody 2002, Proceedings of the first International Conference on Speech Prosody,* Laboratoire Parole et Langage, Aix-en-Provence, 2002, pp. 323-326.

Gibbon D. (2003) Computational modelling of rhythm as alternation, iteration and hierarchy. *Proceedings of the 15th ICPhS, Barcelona.*

Gibbon D., Fernandes F. R. (2005) Annotation–mining for rhythm model comparison in Brazilian Portuguese

Gordon M. (2002) A phonetically-driven account of syllable weight, *Language* 78, 51-80

Grabe E. and Low E.L. (2002) Durational variability in speech and the rhythm class hypothesis. In: Gussenhoven C. and Warner N. (eds.) *Papers in Laboratory Phonology 7.* The Hague: Mouton de Gruyter.

Hayes B. (1981) A metrical theory of stress rules. Indiana University Linguistics Club. Manuscript.
(1984) The phonology of rhythm in English. *Linguistic Inquiry 15*:33-74.

Lehiste I. (1977) Isochrony reconsidered. *Journal of Phonetics 5:*253–263.

Low E.L, Grabe E and Nolan F. (2000) Quantitative characterizations of speech rhythm: syllable-timing in Singapore English. *Language and Speech 43 (3):*229 – 259.

Moore, B. (1989) *An introduction to the psychology of hearing.* San Diego: Academic Press.

Morton, J., Marcus, S.M. and Frankish, C.R. (1976) Perceptual centers (P-centers). *Psychological Review, vol. 83*:405 408.

Nespor, I. (1990) On the rhythm parameter in phonology. In: *Logical issues in Language Acquisition* (I.Roca, editor), pp. 157-175, Dordrecht: Foris.

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

Pompino-Marschall, B. (1989) On the psychoacoustic nature of the P-center phenomenon, *Journal of Phonetics*, 17: 175-192

Port R. and Leary A.P. (2000) Speech Timing in Linguistics. Department of Linguistics, Indiana University. Manuscript.

Port R., Tajima K. and Cummins F. (1988) Speech and Rhythmic Behavior. Department of Linguistics, Indiana University. Manuscript.

Prince A. (1983) Relating to the grid. *Linguistic Inquiry 14*:19-100.

Ramus F., Nespor M., Mehler J. (1999) Correlates of linguistic rhythm in the speech signal. *Cognition vol.73,* no. 3:265-292.

Ramus F., Dupoux E., Mehler J., (2003) The psychological reality of rhythm classes: perceptual studies. *Proceedings of the 15th ICPhS, Barcelona*.

Scott, S. K. (1993) *Perceptual centres in speech: an acoustic analysis.* PhD Thesis, University College London

Selkirk E. (1984) *Phonology and syntax: the relation between sound and structure.* Cambridge, Massachusetts: MIT Press.

Steiner, I. (2003) A refined acoustic analysis of speech rhythm. *LingColl 2003*, Budapest.

White, L. and Mattys, S.L. (2007). Rhythmic typology and variation in first and second languages. In P. Prieto, J. Mascaró, & M.-J. Solé (Eds.), *Segmental and Prosodic Issues in Romance Phonology. Current Issues in Linguistic Theory series* (pp.237-257). Amsterdam: John Benjamins.