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

Speech Prosody 2002
Aix-en-Provence, France
April 11-13, 2002

Acoustic correlates of linguistic rhythm: Perspectives

Franck Ramus

Laboratoire de Sciences Cognitives et Psycholinguistique (EHESS/CNRS)
54 boulevard Raspail, 75006 Paris, France
ramus@lscp.ehess.fr

10.21437/SpeechProsody.2002-16

Abstract

The empirical grounding of a typology of languages' rhythm is again a hot issue. The currently popular approach is based on the durations of vocalic and intervocalic intervals and their variability. Despite some successes, many questions remain. The main findings still need to be generalised to much larger corpora including many more languages. But a straightforward continuation of the current work faces many difficulties. Perspectives are outlined for future work, including proposals for the cross-linguistic control of speech rate, improvements on the statistical analyses, and prospects raised by automatic speech processing.

1. Introduction

The history of rhythm typology is that of a debate between pro-rhythm class advocates and their opponents. The former argued that the languages of the world can be categorised into a small number of classes: typically, stress-timed, syllable-timed and perhaps mora-timed languages [19, 1, 13, 20]. The latter opposed that empirical evidence for the classes was weak if not altogether absent [2, 22, 3]. The search for empirical evidence may indeed have focused too much on the notion of isochrony, i.e., that stress-timed languages should have inter-stress intervals of a roughly constant duration, whereas syllable-timed ones should have syllables of constant duration.

However, another approach, based on the variability of the duration of vowels, was more successful. It relies on the idea that stress-timed languages allow vowel reduction, in contrast with syllable-timed languages. Therefore, vowel duration should be more variable in stress-timed languages. This approach first provided evidence for rhythmic differences between British and Singapore English [14, 15]. In an independent study, we examined the duration and variability of vocalic and intervocalic intervals[1] in eight languages [21]. The rationale behind the consideration of inter-vocalic intervals is that stress-timed languages also tend to allow more complex syllables, and therefore longer and more variable sequences of consonants than syllable-timed languages. Figure 1 recalls the main results. We found that along two dimensions (%V: percentage of duration taken up by vocalic intervals; ΔC: standard deviation of the duration of consonantal intervals within a sentence), languages are not scattered randomly, but are clustered in groups that strongly resemble rhythm classes: English, Dutch and Polish as stress-timed languages, French, Spanish, Italian and Catalan as syllable-timed languages, and Japanese as a mora-timed language. This provides very suggestive evidence in favour of the rhythm class hypothesis.

[1] Vocalic intervals are vowels and sequences of consecutive vowels, regardless of whether they belong to the same syllable (or word, for that matter) or not. Similarly, inter-vocalic or consonantal intervals are made up of consonants and sequences of consecutive consonants.

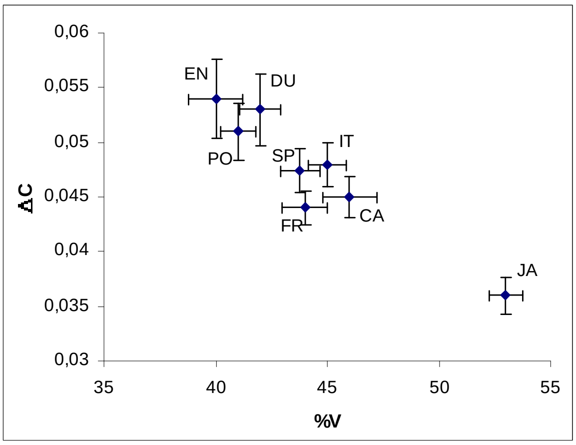

Figure 1: *Standard deviation of consonantal intervals vs. proportion of vocalic intervals in 8 languages, 5 sentences times 4 speakers per language (reproduced from [21]).*

Note that this evidence is far from definitive. For one thing, it is compatible with an alternative interpretation: that the measures taken reflect rhythmic differences, but not classes. It is indeed entirely possible that when more languages are added, the clusters will be drowned in a uniform rhythmic continuum or space. This idea of a continuum in place of classes has been evoked in the past [4, 17] and has been revived in a recent study [12].

It is quite clear that we are still a long way from a fully-fledged, empirically-based rhythm typology. In this paper, we will discuss the limitations of the data accumulated so far, and reflect on the direction to take in order to eventually achieve a typology.

2. Limitations and problems

If the conclusions of our study are to hold, the main results will need to survive a considerable enlargement of the corpus. The corpus can in principle be extended to many more languages, speakers, samples, speech rates, speech registers, etc. Several teams have actually set out to do just this. But such an enlargement faces many difficulties. The recent study by Grabe and Low [12] will be used as an illustration of those difficulties.

Indeed, Grabe and Low (henceforth, GL) studied the variability of vocalic and consonantal intervals in 18 languages, and concluded that although there is evidence for rhythmic diversity between these languages, they do not cluster in separate classes. Analysing in detail the differences between GL's study and ours [21] (henceforth, RNM) will allow us to see why different conclusions have been drawn from such similar studies, and will highlight methodological caveats for future work.

2.1. Differences between the GL and RNM studies

Differences between GL's study and ours lie both in the structure of the corpus and in the nature of the analyses. In RNM, different sentences were uttered by 4 speakers per language, whereas there was only one speaker per language in GL. This means that GL's data might reflect idiosyncrasies of particular speakers as much as differences between languages. Indeed, in our own data, we found that average values of variables %V, ΔC and ΔV for each speaker could differ significantly within each language[2]. Therefore, the clustering we found depended in part upon our averaging of the data across several speakers. That within-language speaker differences might be of the same order of magnitude as between-language differences is certainly a thorny issue for any typologist, and one that will need to be addressed. In the meantime, it is clear that the more numerous the speakers, the safer the conclusions.

Speaker variability can take many forms, and perhaps the most annoying for the present purpose is in terms of speech rate variability, since we measure durations. In RNM, the 20 sentences of each language were selected so that the number of their syllables (from 15 to 19) was matched across languages, as well as their average duration. In essence, this ensured that speech rate (in number of syllables per sec.) was matched across languages. This was not the case in GL. Instead, these authors adopted a normalisation procedure for speech rate (see below).

As a consequence, when GL computed our measures %V, ΔV and ΔC on their corpus, they did not find nice clusters. Actually, they not only found a continuum of languages, but also an ordering of languages inconsistent with ours. For example, on the %V scale, they found Catalan at the lower end close to British English, Japanese very close to Dutch, and French and Spanish with a far higher %V than Japanese. GL concluded that %V, ΔV and ΔC are not reliable measures of rhythm, because they may reflect spurious rate and speaker variability. We definitely acknowledge that ΔV and ΔC may be sensitive to variations in speech rate (within and across language). But we conclude, on the basis of our data, that when these measures are computed on a corpus where speech rate is carefully controlled, they are reliable indicators of rhythm. However, their use is less recommended on corpora where speech rate variability is an issue (like GL's), where they may simply provide inconsistent results reflecting speakers' particularities as much as typological differences.

2.2. PVI vs. standard deviation

GL's response to the speech rate challenge is to use a normalised version of the Pairwise Variability Index. The raw (unnormalised) PVI is defined by the following equation:

$$rPVI = \frac{100}{m-1} \times \sum_{k=1}^{m-1} |d_k - d_{k+1}| \qquad (1)$$

where m is the number of intervals (vocalic or intervocalic) in the sample, and d_k is the duration of the kth interval. Similarly, the normalised PVI is defined as follows:

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

Therefore, each difference between two intervals is normalised by their average duration.

The main point in favour of the PVI is that it does not pick up spurious variability due to speech rate variations within an utterance. This is definitely an interesting feature in principle. It is not known how much such variations actually affect the results. The other point in favour of the PVI is the inclusion of a normalisation term in the nPVI. Of course, this is quite independent of the PVI itself; standard deviations could also be normalised by the average interval duration. In order to see more clearly whether the PVI gets rid of spurious variability picked up by standard deviations, we computed it on our corpus.

GL sensibly argue that normalising is desirable for vocalic, but not for intervocalic intervals, since the latter would in fact involve normalising for cross-language differences in syllable structure. Here, we adopt the same approach, that is, we compute the vocalic nPVI and the intervocalic rPVI on the RNM corpus. The results are shown in Figure 2. For comparison purposes, Figure 3 shows the equivalent variables from RNM (ΔV and ΔC, unnormalised).

The similitude between these two figures is striking. They both feature English and Dutch in the upper right corner, syllable-timed languages in the middle, and Polish and Japanese each on its own in the left and in the lower part respectively. The intervocalic dimension accounts for syllable complexity, and the vocalic axis for such phenomena as vowel reduction or quantity contrasts. Along the latter, Polish is separated from stress-timed languages in virtue of its absence of vowel reduction [21]. It also seems that the PVI achieves a better separation between rhythm classes than standard deviations along the vocalic axis, although not along the intervocalic axis.[3] At this stage, it would be premature to conclude that one set of variables is more reliable than the other. At least on a constrained corpus like that of RNM, they provide largely equivalent results.

2.3. Classes vs. continuum

Considering the uncertainties due to the limited number of speakers in GL's corpora, it not clear yet whether the additional languages they introduce (Thai, Tamil, Malay, Estonian, Rumanian, Welsh, Greek, Luxembourgish, Mandarin) would still fall between the rhythm classes with a greater number of speakers per language. So, the classes vs. continuum debate is still open. Let us just remark that, in order to support the rhythm class hypothesis, one must demonstrate a clear clustering of languages in the traditional rhythm categories, along some measured dimension. To achieve this, one needs either very clean data, or to be very lucky. On the other hand, being able to show a continuum along some measured dimension is the simplest thing in the world: any random data will do. Therefore, in order for the debate to be even, one must demand of continuum-supportive data (as of class-supportive data) that it be compatible with the accepted classifications, at least in terms of the ordering of languages, if not in terms of clusters.

[2] However, as different speakers uttered different sentences, sentence variability was in our case added to speaker variability.

[3] Of course, there is no PVI equivalent to our best predictor of rhythm classes, %V.

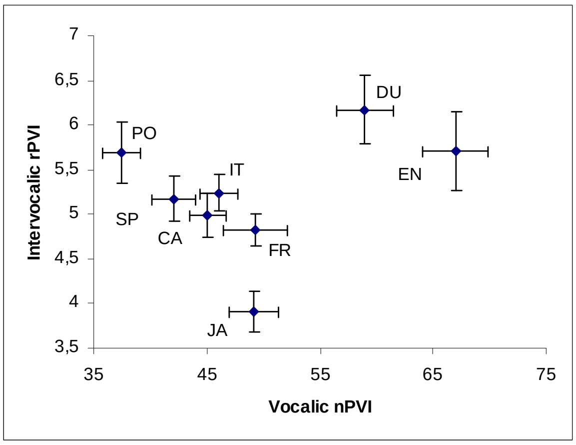

Figure 2: *Intervocalic raw Pairwise Variability Index vs. vocalic normalised Pairwise Variability Index in the RNM corpus.*

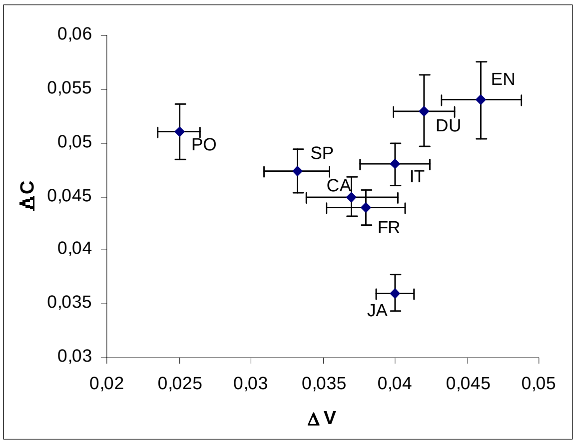

Figure 3: *Standard deviation of consonantal intervals vs. standard deviation of vocalic intervals (reproduced from RNM [21]).*

2.4. Discussion

The analysis performed in this section leads to the following conclusions:

- 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 ΔV and ΔC may well be limited to corpora where speech rate is strictly controlled.

As a consequence, the work we have done in RNM [21] is difficult to extend as it is. As the comparison with GL's study shows, the heavy methodological constraints it requires may be an advantage as regards the clarity of the data. But there is no doubt that this is a drawback as regards the generalisability of the results. Indeed, extensions of this work will need to follow an identical method in order to produce anything comparable.

Until now, we have highlighted the need for speech rate control, but we have always mentioned it as if it was the most straightforward thing to do. Unfortunately, this is not the case. We now turn to a broader discussion of the cross-linguistic evaluation of speech rate.

3. Speech rate across languages

Since rhythm is, at least in part, a matter of duration, and durations are affected by speech rate, all students of speech rhythm must be concerned by effects due to speech rate. Since our purpose here is to define a typology of rhythm, speech rate must be treated in a manner that is valid across all languages. This is where the issue becomes really thorny. We feel that this matter is seldom discussed and deserves a thorough treatment here.

In the RNM study, we chose to match both the number of syllables per sentence and sentences' duration across languages by selecting for each speaker 5 appropriate sentences from a corpus of about 50. This approach by selection is itself questionable. One might consider other methods, like defining a speech rate a priori, and asking speakers of all languages to adopt it. Whatever the method, the question will remain: what do we want to match? In other words, how do we measure speech rate in a way that is valid across languages?

Let us note that the question does not vanish if we choose to normalise for speech rate in order to avoid matching the corpus. This option has to face the equivalent question: what do we want to normalise? In other words, what quantity do we put in the denominator? Furthermore, both RNM and GL have argued that it is interesting to look separately at vocalic and intervocalic intervals. But we have reasons to suspect that these two types of intervals are not affected by speech rate in the same proportion, and in addition this proportion may well vary across languages.

3.1. Problems with standard definitions of speech rate

Which unit? Among all possible measures of speech rate, the most widely accepted and used seems to be syllable rate, i.e., the number of syllables per second. But one might argue that the syllable is not the proper speech unit to consider. Shouldn't we count morae per second? Or phonemes? Or feet? Or vocalic and intervocalic intervals? Or units of meaning?

One unit for all, or different units for different languages? It might even be that the appropriate unit is not the same in all languages. What about feet for English, syllables for French, and morae for Japanese? Obviously, this approach would lead to the observation that Japanese is much faster than English, since Japanese morae are much shorter than English syllables. But even the "one unit for all" approach has this problem. Since Japanese syllables are simpler than English ones, one would also expect that Japanese speakers are able to produce more syllables per second than English speakers[4]. Normalising or matching syllable rate therefore leads to ignore part of the rhythmic differences due to syllable structure.

What counts as a unit? Supposing a given unit is chosen (say, the syllable), then comes the problem of defining what counts as a unit in different languages. Are we talking about phonetic or underlying syllables? Depending on the answer, matching syllable rate between European and Brazilian Portuguese would yield highly different results, since they both

[4] This is true on average in our corpus

have the same underlying syllables, but Brazilian Portuguese has many more phonetic syllables.

Cross-linguistic differences in speech rate? All things mentioned above being equal, and supposing we have a good cross-linguistic measure of speech rate, isn't it conceivable that some languages are spoken faster than others (if only for cultural reasons)? Suppose that we find that, in general, Parisian French is spoken faster than Swiss French, or that Italian is spoken faster than German. Wouldn't this contribute to perceived differences in rhythm? In that case, would it be legitimate to normalise for these differences?

In the face of all these questions, it seems almost illusory to ever find a measure of speech rate equally valid for all languages. Yet, we need to control it. Let us make here two proposals.

3.2. A perceptual approach

One possible way is to remain agnostic about the questions mentioned above, and to leave it to listeners to answer them. One could provide listeners with a large corpus of utterances in many languages (perhaps excluding the languages they understand), and ask them to rate the speech rate of each utterance (alternatively, present utterances in pairs and ask which one is faster). For each utterance, all conceivable measures of speech rate would be computed. A large statistical analysis would then reveal whether some measures reliably predict the subjects' judgements, and which is the best predictor. The analysis would be run both across all languages, and restricted to each language, as it might reveal that the answer is not the same for all languages.

Obviously, this approach is not guaranteed to succeed. It could be that none of the conceivable measures is a reliable predictor, though this would be surprising. If anything, we would expect that virtually *all* the measures mentioned above are reliable predictors to some extent. It could also be that the judgements of speakers of different languages are predicted by different measures of speech rate. This would be more embarrassing and would undermine the legitimacy of any universal measure of speech rate.

3.3. A language-specific approach

Of course it might just be the case that there can be no cross-linguistic measure of speech rate. Let us now adopt a language-specific approach. Let the experts say what the proper way to measure speech rate is in each language. These might be different ways for different languages. But a single measure like syllable rate might also be able to account for speech rate differences *within* each language; in other words, a single measure might be appropriate as long as we do not compare the values obtained *between* languages. Anyway, the answer to at least these questions could be answered by following the perceptual approach proposed above, by analysing separately judgements of rate for sentences within each language and across all languages.

Whatever the solution adopted for language *L*, it is then possible to compute speech rate for a large number of utterances of *L*, ideally a representative corpus of all speech uttered in *L*. One then obtains a histogram of all speech rates in *L*. Let us define the standard speech rate of *L* as the mean or the median of this distribution.

For the purpose of cross-language comparisons, one can then choose to :

- match the speech rates of utterances in languages *L* and *L'* by dividing the language-specific rate of each utterance by the standard rate of the corresponding language, thereby expressing utterances' speech rates in standard speech rate units, comparable across languages;
- normalise any measure obtained on utterances of *L* by the standard speech rate of *L*;
- select for each language a set of utterances whose average rate is the standard speech rate (and perhaps within a certain range, like one standard deviation).

Note that if the distributions of speech rates happen to differ significantly across languages, this might even become an object of study, as this might play a role in a rhythm typology.

In essence, we propose, as a last resort, to define speech rate statistically rather than in a principled manner. Speech rate norms can be developed for each language. Using the norms within each language allows measures to be compared across languages.

4. Perspectives

Keeping in mind all the caveats raised so far, let us now discuss which directions seem most promising for the eventual constitution of a rhythm typology.

4.1. Enlargement of the RNM corpus

Despite all the limitations we have discussed earlier, the RNM corpus still has the merit to show clear clusters corresponding to the expected rhythm classes. These clusters might be too good to be true, a strike of luck, an idiosyncrasy of those 160 sentences, but in case they are not, it makes sense to go on enlarging the corpus following exactly the same approach.

This is what several collaborators are just doing in European and Brazilian Portuguese [8, 9], Finnish, Turkish, Hindi, Basque and other languages. So far we can only say that preliminary results are encouraging.

4.2. Statistical improvements

Beyond the "PVI vs. standard deviation" debate, there may be even more powerful ways to analyse the type of data we are gathering. In RNM, we had performed a number of statistical tests to demonstrate that rhythm classes were significantly different from each other along the proposed dimensions. These tests (ANOVAs and the like) obviously assumed that our values were normally distributed. Duarte et al. [5] actually looked at the distributions of durations of intervocalic intervals in each language, and found that they correspond to a single family of distributions called Gamma. In each language, the distribution of intervocalic intervals can be fitted to a particular Gamma function, and the Gamma distributions fitting the eight languages vary along only one parameter: their standard deviation. Thus, standard deviations of durations of vocalic and intervocalic intervals can, for each language, be replaced by the standard deviation of the best-fitting Gamma function (for each type of intervals). Using this method, Duarte et al. find a picture very similar to Figure 1 (see also Galves et al [10] in these proceedings for a summary).

The advantage of this approach is that the values obtained are much more reliable and robust, for the purpose of further analyses and projections to other data sets. The potential drawback is that if one wants to obtain an estimate of the rhythmical characteristics of just one utterance (rather than the whole language), there may never be a good enough fit between such a small distribution and a Gamma function, whereas this was possible and meaningful using %V for instance (see Figure 4 in RNM [21]). Nevertheless, for typological matters, the Gamma function may well provide the optimal statistical tool.

4.3. Towards automatic processing

We have not yet mentioned another major limitation shared by our study and those of Grabe and colleagues: the manual determination of interval boundaries. Despite the best efforts of phoneticians to provide clear labelling principles, this operation is still largely subjective; it may be yet another source of differences between the RNM and GL studies. Ensuring that contributors of new languages employ exactly the same criteria as in the original study is virtually impossible. In addition to being subjective, manual labelling is extremely tedious. It is therefore clear that any large-scale extension of the approach will require some automatic processing of the speech corpus.

4.3.1. Automatic segmentation algorithms

Pellegrino and colleagues [18, 7] (and in these proceedings) have developed a vowel detection algorithm trained on 5 languages: English, French, German, Italian and Spanish. This algorithm extracts enough meaningful information to be able to reliably identify these languages given a new stretch of speech. Applied to the RNM corpus, it provides relatively decent %V values but absurd ΔV and ΔC values. The reason is quite simple: although %V can tolerate relatively large but systematic imprecisions of segmentation, standard deviations are overly sensitive to shifts of the interval boundaries. Such conclusions converge with previous unsuccessful attempts by Garcia and Galves [11] and in our laboratory. We still need to examine the possible contribution of the Gamma distribution to the analysis of this type of data.

Another, more exotic but potentially interesting segmentation algorithm is Mbrolign by Malfrère and Dutoit [16]. It takes both speech and a string of phoneme labels[5] as input, and aligns the labels with the signal, so as to provide the exact phoneme durations. The algorithm is an iterative procedure that optimises the acoustic match between the original speech sample and its synthetic equivalent produced by the algorithm Mbrola [6] (on the basis of a diphone database for each particular language). The usefulness of this type of segmentation to the issues of interest here has not been evaluated yet.

4.3.2. A rhythm extractor without segmentation

Realising the difficulties encountered by segmentation algorithms, Galves et al. [10] have explored another direction. The idea is to directly compute analogues of the "proportion of vocalic intervals in speech" and of the "variability of consonantal intervals" on short successive windows of the signal, but without segmenting. Their algorithm is described in detail in the next pages. Their results on the RNM corpus are very encouraging.

A final sobering note is that any algorithm will nevertheless face the necessity of validation against hand-labelled data. Indeed, an algorithm can only be trained or optimised on a limited number of languages, on the basis of already existing labels. One never knows how well the algorithm will generalise to a new language, until one evaluates its performance against another method. In other words, we may also need a large corpus of hand-labelled speech in order to be sure that algorithms work well. If this is not available, the convergence of algorithms based on different approaches would be reassuring.

[5] Phoneme strings can themselves often be obtained from text-to-phoneme algorithms for specific languages.

5. Conclusion

Salvation lies in larger data sets. But a straightforward extension of the approaches followed so far faces many obstacles. One major issue issue to address will be that of the cross-linguistic control of speech rate. However, this topic may turn out to be interesting in its own right. Automatic speech processing carries hopes of effortless constitution of unlimited corpora, as well as the spectres of imprecision and meaninglessness. But there is little doubt that with time, the benefits of new technologies will eventually outweigh their drawbacks.

6. Acknowledgements

I would like to thank Denise Duarte, Sonia Frota, Antonio Galves, Esther Grabe, Jacques Mehler, Marina Nespor, François Pellegrino, Mohinish Shukla, Marina Vigario and all the participants of the ZiF workshop "Rhythmic patterns, parameter setting and language change" for their collaboration and feedback.

7. References

[1] D. Abercrombie. *Elements of general phonetics*. Aldine, Chicago, 1967.

[2] D. Bolinger. Pitch accent and sentence rhythm. In D. Bolinger, editor, *Forms of English: Accent, morpheme, order*. Harvard University Press, Cambridge, MA, 1965.

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

[4] R. M. Dauer. Phonetic and phonological components of language rhythm. In *XIth International Congress of Phonetic Sciences*, volume 5, pages 447–450, Tallinn, 1987.

[5] D. Duarte, A. Galves, N. Lopes, and R. Maronna. The statistical analysis of acoustic correlates of speech rhythm. In *Rhythmic patterns, parameter setting and language change*, ZiF, University of Bielefeld, 2001.

[6] T. Dutoit, V. Pagel, N. Pierret, F. Bataille, and O. van der Vrecken. The mbrola project: Towards a set of high-quality speech synthesizers free of use for non-commercial purposes. In *ICSLP'96*, Philadelphia, 1996. MBROLA is freely available from http://tcts.fpms.ac.be/synthesis/mbrola.html.

[7] J. Farinas and F. Pellegrino. Automatic rhythm modelling for language identification. In *Proceedings of Eurospeech '01*, pages 2539–2542, Aalborg, 2001.

[8] S. Frota and M. Vigàrio. On the correlates of rhythmic distinctions: The european/brazilian portuguese case. *Probus*, 13:247–275, 2001.

[9] A. Galves. Language acquisition and change: Identifying features in presence of competing evidence. In *Rhythmic patterns, parameter setting and language change*, ZiF, University of Bielefeld, 2001.

[10] A. Galves, J. Garcia, D. Duarte, and C. Galves. Sonority as a basis for rhythmic class discrimination. In *Speech Prosody 2002*, Aix-en-Provence, 2002.

[11] J. Garcia and A. Galves. Automatic identification of vocalic intervals in speech signal. In *Rhythmic patterns, parameter setting and language change*, ZiF, University of Bielefeld, 2001.

[12] E. Grabe and E. L. Low. Durational variability in speech and the rhythm class hypothesis. In *Papers in Laboratory Phonology 7*. Cambridge University Press, Cambridge, in press.

[13] P. Ladefoged. *A course in phonetics*. Harcourt Brace Jovanovich, New York, 1975.

[14] E. L. Low and E. Grabe. Prosodic patterns in singapore english. In *ICPhS 95*, volume 3, pages 636–639, Stockholm, 1995.

[15] E. L. Low, E. Grabe, and F. Nolan. Quantitative characterisations of speech rhythm: Syllable-timing in singapore english. *Language and Speech*, 43(4):377–401, 2000.

[16] F. Malfrère and T. Dutoit. High quality speech synthesis for phonetic speech segmentation. In *European Conference On Speech Communication and Technology*, pages 2631–2634, Rhodes - Greece, 1997.

[17] M. Nespor. On the rhythm parameter in phonology. In I. M. Roca, editor, *Logical issues in language acquisition*, pages 157–175. Foris, Dordrecht, 1990.

[18] F. Pellegrino and R. André-Obrecht. An unsupervised approach to language identification. In *Proceedings of ICASSP'99*, Phoenix, 1999.

[19] K. L. Pike. *The intonation of American English*. University of Michigan Press, Ann Arbor, Michigan, 1945.

[20] R. F. Port, J. Dalby, and M. O'Dell. Evidence for mora-timing in japanese. *Journal of the Acoustical Society of America*, 81(5):1574–1585, 1987.

[21] F. Ramus, M. Nespor, and J. Mehler. Correlates of linguistic rhythm in the speech signal. *Cognition*, 73(3):265–292, 1999.

[22] P. Roach. On the distinction between "stress-timed" and "syllable-timed" languages. In D. Crystal, editor, *Linguistic controversies*. Edward Arnold, London, 1982.