

The Rhythm Class Hypothesis and Indigenous Languages

Luciana R. Storto & Didier Demolin
Universidade de São Paulo and Université Libre de Bruxelles

Abstract

This paper aims to evaluate the rhythm class hypothesis proposed by Ramus et al (1999) in light of Karitiana, an endangered Tupi language spoken in Rondônia, Brazil, by approximately 350 people. Ramus et al (1999) have identified three types of languages in their experimental study: stress-timed languages (English, Dutch and Polish), syllable-timed languages (Spanish, Italian, Catalan and French), and mora-timed languages (Japanese). The authors report that other studies have shown that Germanic, Arabic and Slavonic languages fall into the first class, Romance languages in general belong to the second class, and Japanese and Tamil belong to the last one. The explanation given by Ramus et al (1999) for the existence of these classes has to do with syllable structure. The first class would be formed by languages in which there is a greater variety of syllable types, the last class would be formed by languages in which there is a small variety of syllable types, and the second class would be in an intermediary situation. Karitiana sentences were processed using the same methodology as Ramus et al have used for their eight languages, and the results show that it does not belong to either of the three classes. It is interesting to compare Karitiana and Japanese with respect to the rhythm class hypothesis because they have the same small variety of syllabic types (four types of syllables), but they differ significantly with respect to Ramus et al's definition of rhythm in that Karitiana has a lower percentage of vowels per sentence, very similar to that of a syllable-timed language such as Italian. A comparison between Karitiana and Japanese may help us understand what factors other than variety of syllable types may be at play in the definition of rhythm classes.

1. The Rhythm Class Hypothesis by Ramus et al (1999)

Many authors have suggested that languages may vary in a limited way with respect to rhythm, falling in two or three classes (Lloyd James 1940, Pike 1945, Abercrombie 1967, Bertinetto 1989, Ladefoged 1975, Port, Dalby & O'Dell 1987, Rubach & Booij 1985, Steever 1987). The main

motivation behind Ramus et al's formulation of the rhythm class hypothesis was to help explain why infants perceive differences in rhythm even before they acquire language. They argue that there may be some phonetic cues that are being used by infants to detect rhythmic differences:

“Since our main interest is to explain how infants come to perceive contrasting rhythms at birth, and since the infant cannot be expected to know anything specific a priori about the language to be learned, we would like to argue that a viable account of speech rhythm should not rely on complex and language-dependent phonological concepts. We will therefore attempt to provide a purely phonetic definition of language rhythm without appealing to those concepts.”

(Ramus, et al. 1999, p. 270)

Furthermore, Ramus et al (1999) follow Mehler et al (1996) in suggesting that the perception of vowels in the string of speech may be the most significant factor in the detection of different rhythm classes by infants:

“Our starting point is a hypothesis about the perception of speech by the infant. Following Mehler et al. (1996), we propose that infant speech perception is centered on vowels, because these have more energy and last longer than most consonants”

(Ramus, et al. 1999, p. 270)

Finally, it is not even assumed that consonants are categorized in some way by the infant, but simply that they are processed as some kind of “noise” between the perceived vocalic intervals:

“We thus assume that the infant primarily perceives speech as a succession of vowels of variable durations and intensities, alternating with periods of unanalyzed noise (i.e. consonants)...”

(Ramus, et al. 1999, p. 270)

With that expectation in mind, Ramus et al (1999) have conducted an experiment in which sentences were segmented and measured with respect to the duration of their vowel and consonant intervals. Eight languages were studied - English, Dutch, Polish, French, Spanish, Italian, Catalan. Japanese - with four speakers per language and five sentences per speaker. The phonemes of each sentence were marked with a sound-editing software, using both auditory and visual cues. Phonemes were classified as Cs or Vs. Pre-vocalic glides were treated as Cs and post-vocalic glides as Vs. The duration of vocalic and consonantal intervals was measured, from

the onset of a cluster of vowels (or consonants) to the offset of a cluster of vowels (or consonants). Intervals, instead of phonemes, were measured because the authors have assumed that infants do not have access to the phonemic inventory of the language when processing speech. The data has been treated statistically in the following way:

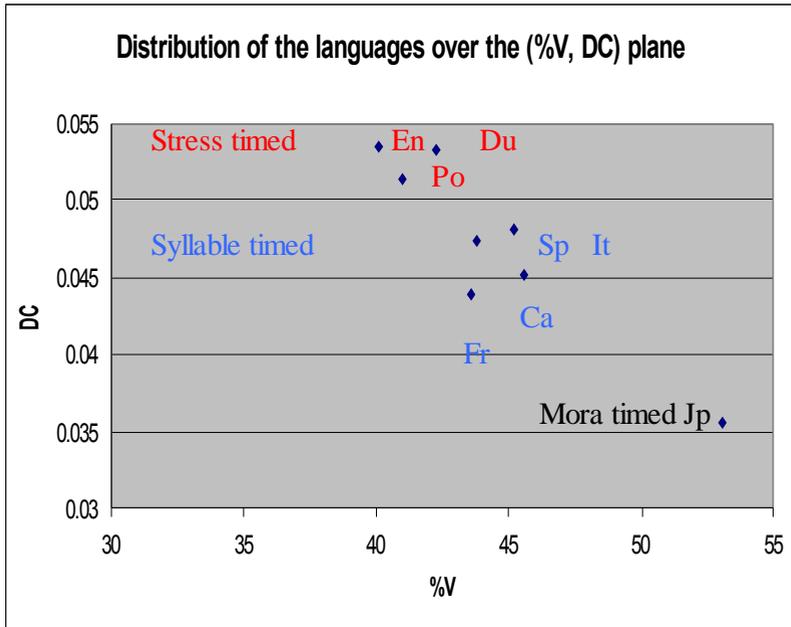
- The proportion of vocalic intervals within the sentence was calculated as the sum of vocalic intervals in each sentence divided by the total duration of the sentence. The average per language of such intervals was given the label %V. The same treatment could have been given to consonantal intervals (%C), but this was not done since there is an isomorphism between %V and %C.
- The standard deviation of the duration of vocalic intervals within each sentence. This variable was averaged by language and given the label ΔV .
- The standard deviation of the duration of consonantal intervals within each sentence. This variable was averaged by language and given the label ΔC .

The results are listed in table 1, arranged by their %V value. The interesting fact to be noted is that this value correlates with the rhythm patterns described before in the literature, in that stress-timed languages such as English, Polish and Dutch rank lower and Japanese, a mora-timed language, rank higher in terms of %V. Romance languages, usually characterized as syllable-timed languages, occupy an intermediary position between those two in terms of %V:

Table 1.

Language	%V (SD)	ΔV (SD)	ΔC (SD)
English	40.1 (5.4)	4.64 (1.25)	5.35 (1.63)
Polish	41.0 (3.4)	2.51 (0.67)	5.14 (1.18)
Dutch	42.3 (4.2)	4.23 (0.93)	5.33 (1.5)
French	43.6 (4.5)	3.78 (1.21)	4.39 (0.74)
Spanish	43.8 (4.0)	3.32 (1.0)	4.74 (0.85)
Italian	45.2 (3.9)	4.00 (1.05)	4.81 (0.89)
Catalan	45.6 (5.4)	3.68 (1.44)	4.52 (0.86)
Japanese	53.1 (3.4)	4.02 (0.58)	3.56 (0.74)

Figure 1. Distribution of languages in the %V, ΔC plane (adaptation from Ramus et al 1999)



Three graphs were made, projecting on planes the values of (%V, ΔC), (%V, ΔV), and (ΔV , ΔC), and it was found that the (%V, ΔC) projection is able to represent the three rhythm classes, as shown in figure 1.

The interpretation Ramus et al (1999) give to their findings is to suggest that they follow from the fact that languages with heavier syllables in terms of consonants have a greater variety of syllable types:

“ ΔC and %V appear to be directly related to syllabic structure. Indeed, a greater variety of syllable types means that some syllables are heavier (see Footnote2). Moreover in most languages, syllable gain weight mainly by gaining consonants. Thus, the more syllable types a language instantiates, the greater variability in the number of consonants and in their overall duration in the syllable, resulting in a higher ΔC ”.

(Ramus, et al. 1999, p. 272)

The experimental study conducted by Ramus et al (1999) concludes that syllable structure is all that matters in the definition of rhythm classes:

“Thus, the nice fit between the (%V, ΔC) chart and the standard rhythm classes comes as an empirical validation of the hypothesis that rhythm contrasts are accounted for by differences in the variety of syllable structures.”

(Ramus, et al. 1999, p. 273)

This is not to say that the authors ignore that different claims have been made in the literature. They mention that Dauer (1987) claims that vowel reduction and contrastive vowel length are important factors influencing rhythm. Also, they admit, for instance, that Polish does not fit well as a stress-timed language, since its ΔV is among the lowest calculated in the study, together with Spanish which is a syllable-timed language. The value of ΔV does not seem to relate to rhythm perception in the opinion of the authors, although they admit that it certainly reflects linguistic properties. In spite of these considerations, the conclusion remains the same:

“A number of properties seem to be more or less connected with rhythm: vowel reduction, quantity contrasts, gemination, the presence of tones, vowel harmony, the role of word accent and, of course, syllable structure (Dauer 1987, Donegan & Stampe 1983, see Auer 1993 for a survey). Given the current state of knowledge, we believe that only syllable structure is reliably related to rhythm”.

(Ramus et al. 1999, p. 289)

2. Discussion of Ramus et al (1999)

The interesting result obtained by Ramus et al (1999) was the %V, ΔC graph, that seems to portray the three rhythm classes identified in the literature. A first point that we want to make in discussing these results is to try to define exactly what is the meaning of the %V, ΔC plane.

The authors define ΔC as the standard deviation of the duration of consonantal intervals within each sentence, averaged per language. What this means is the range of variability in duration of consonant intervals per sentence in each language. The linguistic interpretation given to ΔC by the authors is that a language with a high ΔC , such as English or Dutch, would have more complex consonant clusters, thus a wider variety of syllable types, than a language with a low ΔC , such as Japanese. Since this prediction is borne out for the languages studied, the ΔC seems at first to be a nice correlate of the three rhythm classes.

Also, it is reasonable to assume, as the authors do, that in a language with more complex syllables, the proportion of consonant intervals (%C)

per sentence will be higher than in a language with simple syllables. The latter would tend to have a higher proportion of vowel intervals per sentence (%V). However, it is possible that in any kind of language (with a small or large number of syllable types), the proportion of vocalic intervals per sentence may be either high or low, depending on many factors (whether or not the language has contrastive length, vowel or consonant lengthening processes, vowel or consonant epenthesis, vowel reduction, etc). As we will see later, this is exactly what we find in Karitiana; it patterns with Japanese in terms of ΔC , but it is more like Italian in terms of %V. This suggests that ΔC alone, and not the %V, ΔC plane seems to correlate to the three rhythm classes.

Even ΔC could have a different interpretation than that given by Ramus et al (1999). We have seen that a high ΔC means having a wide range of variation in the duration of consonant intervals. That is consistent with having a wide variation of syllables types, but it could be the result of other factors. For instance, it remains to be seen whether a language could have complex syllables and use only a small number of syllable types more often, thus yielding a not so high ΔC .

Another point to be made about the paper by Ramus et al (1999) is that the authors themselves mention that the %V, ΔC plane does not say anything about the typological variability of rhythm. That is, the fact that the graph in question works for the eight languages examined does not mean that all of the world's languages will fall in one of the three clusters in the graph:

“Taken alone, the fact that the proportion of vocalic intervals (%V) and the variability of consonantal intervals (ΔC) in eight languages are congruent with the notion of rhythm classes does not demonstrate that all spoken languages can be classified into just a few categories. At this point, we are agnostic about whether all languages can be sorted into a few stable and definite rhythmic classes. We only studied eight languages, and they were selected from linguists to postulate the three standard classes. Hence, more languages have to be studied. It is entirely conceivable that the groupings already established may dissolve when more languages are added”.

(Ramus, et al. 1999, p. 287)

Nonetheless, the motivation behind the linguistic generalization made by Ramus et al (1999) about the correlation between syllable complexity in terms of consonants and a wide variety of syllabic types seems to be correct. The work of Isabelle Rousset (2004) has examined syllable and lexical structure in 16 languages and has established implicational universals in the typology of syllable structure. Basically, she found that,

if a language has complex onsets or codas, it has all possible simpler syllables in terms of onsets and codas derived from that:

CCCV → CCV → CV
CCCVC → CCVC → CVC
CVCCCC → CVCCC → CVCC → CVC
VCCC → VCC → VC → V

(Rousset 2004: 117)

Obviously, this does not mean that there are only three rhythm classes in the world's languages. It is possible that, as more languages are plotted in the %V, ΔC graph, a different picture will emerge. It is conceivable that the three clusters found will give way to a continuum following an imaginary line linking the three classes, in which case there would be a rhythm continuum, and not three rhythm classes. Another possibility is that the classes are just part of a more complex pattern that can fill portions or the whole %V, ΔC space, thus falsifying the rhythm class hypothesis completely. In any case, we conclude that the three rhythm classes are just a partial result - the classes found in the eight languages examined.

3. Karitiana in light of the Rhythm Class Hypothesis

Karitiana seems to be an interesting language to compare with Japanese in terms of rhythm because both languages have a number of prosodic characteristics in common:

- The same number of syllable types (4: CVC, CV, VC and V)
- Contrastive vowel length
- Pitch accent (unpredictable in Japanese; predictable in Karitiana, as described in Storto & Demolin 2005)

We have carried out an analysis of Karitiana following the methodology of the experiment conducted by Ramus et al (1999). The pilot project described in this paper involves the same number of sentences and speakers per language as Ramus et al (1999). A description of the Karitiana experiment and a discussion of the results is presented below.

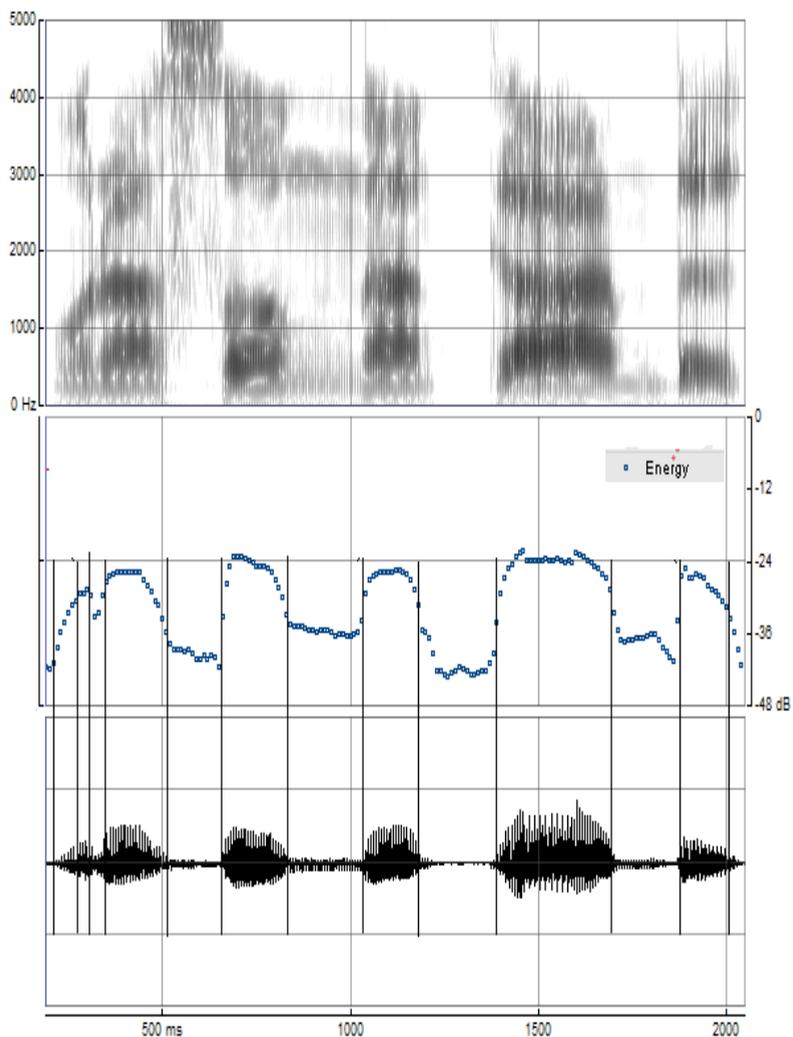
3.1. Material and method

The Karitiana pilot project involved five sentences uttered by two speakers, in a total of ten sentences. The data, based on the material presented in Storto's Ph.D. (1999), and on a text entitled "Gokyp" 'the sun,' have been digitally recorded with a Sennheiser head microphone in order to avoid intensity variations due to possible movements of the speakers' heads. Segmentation of C and V intervals was made with the Signal Explorer software using auditory and visual cues (from waveform and spectrogram). Consonantal and vocalic intervals were measured. The results are based on 125 vocalic intervals and 132 consonantal intervals, about half of the data presented by Ramus et al. (1999) for each of the eight languages, given that their sentences were longer than ours.

The method used to measure the Karitiana data follows the principles presented by Ramus et al. (1999:271). Segments were identified and located as precisely as possible, using the phoneme inventory of Karitiana (Storto 1999). Phonemes were classified as vowels and consonants. Ramus et al. (1999) emphasize that this classification was straightforward with the exception of glides, for which they apply the following rule: pre-vocalic glides were treated as consonants and post-vocalic glides were treated as vowels. Within each sentence the duration of vocalic and consonant intervals were measured. A vocalic interval is located between the onset and the offset of a vowel, or of a cluster of vowels. Similarly, a consonantal interval is located between the onset and the offset of a consonant, or a cluster of consonants. The duration of vocalic and consonant intervals adds up to the total duration of the sentence. The software used was Signal Explorer that allows synchronic measurements of the audio waveform, the spectrogram and the intensity curve.

As an example the phrase 'morāsong naaka andyk' /mōrāsōŋnāākaānd ik/ has the following vocalic and consonantal intervals: /m/ /ō/ /r/ /ã/ /s/ /ō / /ŋn/ /āã/ /k/ /aã/ /nd/ /i/ /k/, 13 in total. This segmentation is illustrated in Figure 2. Note that the decision that takes into account vocalic and consonantal intervals does not consider a number of phonetic details such as the laryngalized transition between the second sequence of two vowels. The high and low parts of the intensity curve (RMS) show a good correlation with the vocalic and consonantal intervals.

Figure 2. Spectrogram, energy curve and segmented waveform of a Karitiana sentence, representing 13 intervals

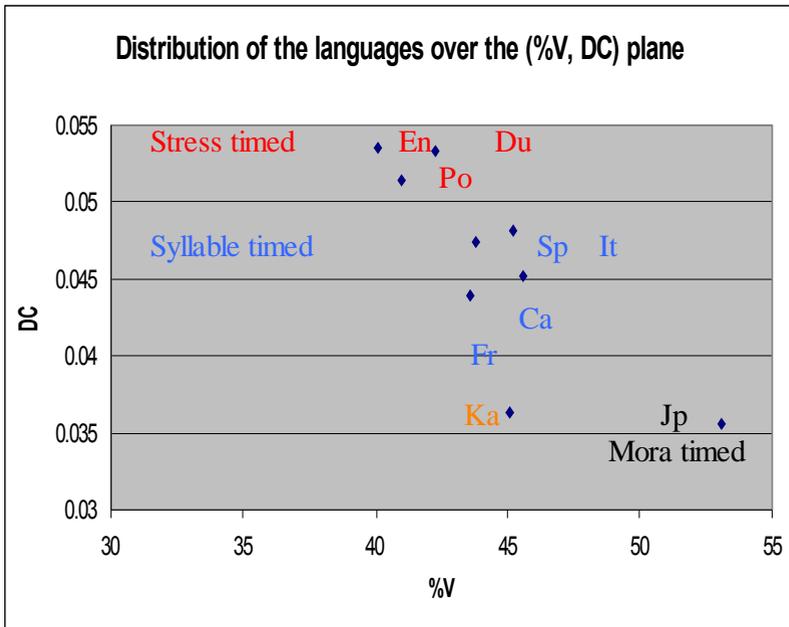


m òr ā s ò ŋn āā k aa nd ì k

3.2. Results

Karitiana is a language that has a similar %V to Italian, and in that respect it patterns with Romance languages as a syllable type language in the %V, ΔC graph (figure 3). However, in terms of ΔC Karitiana is closer to Japanese, behaving as a mora timed language.

Figure 3. Distribution of nine languages in the %V, ΔC plane



Karitiana data are consistent with the hypothesis that ΔC is indeed related to number of syllable types, as suggested by Ramus et al (1999). However, it falsifies the hypothesis that all languages should fall uncontroversially in three classes according to the %V, ΔC graph. In this graph, Karitiana is a mixed language – between a syllable-timed and mora-timed language.

Table 2.

Language	%V (SD)	ΔV (SD)	ΔC (SD)
English	40.1 (5.4)	4.64 (1.25)	5.35 (1.63)
Polish	41.0 (3.4)	2.51 (0.67)	5.14 (1.18)
Dutch	42.3 (4.2)	4.23 (0.93)	5.33 (1.5)
French	43.6 (4.5)	3.78 (1.21)	4.39 (0.74)
Spanish	43.8 (4.0)	3.32 (1.0)	4.74 (0.85)
Karitiana	45.1 (3.63)	3.34 (0.97)	3.63 (0.89)
Italian	45.2 (3.9)	4.00 (1.05)	4.81 (0.89)
Catalan	45.6 (5.4)	3.68 (1.44)	4.52 (0.86)
Japanese	53.1 (3.4)	4.02 (0.58)	3.56 (0.74)

3.3. Discussion

First, we must ask ourselves why do Karitiana and Japanese differ with respect to %V if both languages have the same types of syllables (CVC, CV, V and VC), as well as distinctive vowel length. One part of the answer is that Karitiana has complex allophones of nasal consonants (b_m, b_mb, m_b, d_n, d_nd, n_d, g_ŋ, g_ŋg, ŋ_g) that are longer than an average nasal. The pre and post oralized allophones (b_mb, d_nd, g_ŋg), in particular, were found to be 100ms longer than average nasals (Storto & Demolin 2002), although they count as one consonant in terms of syllabic structure. Another reason is the fact that the high central vowel in Karitiana is extra short, having half of the duration of a regular vowel (Demolin & Storto 2003). Other Tupi and Jê languages have the same kind of complex nasal Cs (b_mb, d_nd, g_ŋg) appearing as allophones of simple phonemes (nasal consonants or voiced stops, depending on the analysis). They count as one C, that is, a single phoneme, but they are extra-heavy. What role would complex nasals have on the rhythm of the languages? They would make %C higher and %V lower? In case this kind of information plays a role in rhythm classes, then it becomes clear, contrary to Ramus et al's beliefs, that syllable structure is not the only factor that we should take into consideration when studying rhythm.

Another subject worth discussing is the similarity between Karitiana and Japanese with respect to ΔC . It could be that syllable structure alone

explains that, since both languages have the same number of syllable types. However, it is necessary to demonstrate that mora-timed languages such as Japanese are really defined as such due to having fewer types of syllable structure alone. There is a question, for instance, whether pitch accent plays a role in the definition of mora-timed languages. It is possible that Karitiana and Japanese pattern similarly with respect to ΔC because both languages have pitch accent (Storto 1999, Storto & Demolin 2005) and this kind of metrical phenomena plays a role in mora-timed rhythm. South American language families in which both tonal and pitch accent languages are found (Tupian, Tukano and Chonan) should be evaluated with respect to the rhythm class hypothesis as well.

Finally, we would like to ask what role do other suprasegmentals play on rhythm. The methodology of Ramus et al (1999) assumes that the segmentation of speech into consonant and vowel strings is enough to reveal rhythm classes since syllable structure alone defines these classes. If so, how do we include suprasegmentals that are not phonemes in the segmentation? Examples are glottal stops inserted in certain word classes in Tukano languages, and glottal stops that mark syllabic boundaries in Tupi languages (Demolin, Sandalo & Storto 2004). Do they count as consonants? What to do when they don't appear as consonants but as laryngealization on vowels? Do they count as vowels then? This study has counted them as Cs because in Karitiana they only surface as glottal stops. In languages where they often surface as laryngealization, however, they would have an effect on vowels, probably shortening them. This means that the same suprasegmental could contribute to a higher or lower %V, depending on how the language realizes it.

We conclude that even when two languages have the same types of syllables they can still differ with respect to rhythm classes depending on some peculiarities of their phonology that are independent of syllable structure. For instance, certain vowels and consonants may be longer or shorter than the average. This means that specific information having to do with the quality of particular segments (phonemes and non-phonemes alike) is taken into account when computing rhythm classes.

References

- Abercrombie 1967. *Elements of general phonetics*. Edinburgh: Edinburgh University Press.
- Auer 1993. *Is a rhythm based typology possible? A study of the role of prosody in phonological theory*. Kont R. I. Working Paper 21. Hamburg: Universität Hamburg.
- Bertinetto 1989. Reflections on the dichotomy 'stress' versus 'syllable-timing'. *Revue de Phonétique Appliquée*, 91-93, 99-130.
- Dauer 1987. Phonetic and phonological components of language rhythm. Paper presented at the 11th International Congress of Phonetic Sciences.
- Demolin, Sandalo & Storto. 2004. A fonética e a fonologia da glocalização em línguas indígenas brasileiras. Paper presented at the congress of Associação de Pós Graduação em Letras e Linguística (ANPOLL). Maceió.
- Demolin & Storto. 2002. Production and Perception of Vowels in Karitiana. *Acoustical Society of America CD-Rom paper collection*. Cancun, Mexico.
- Donegan & Stampe 1983. Rhythm and the holistic organization of syllable structure. Paper presented at the CLS Parasession on The Interplay of Phonology, Morphology and Syntax.
- Ladefoged 1975. *A course in phonetics*. New York: Harcourt Brace Jovanovich.
- Lloyd James 1940. *Speech signals in telephony*. London.
- Mehler, Dupoux, Nazzi & Dehaene-Lambertz. 1996. Coping with linguistic diversity: the infant's viewpoint. In Morgan & Demuth. *Signal to syntax: bootstrapping from speech to grammar in early acquisition*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Pike 1945. *The intonation of american english*. Ann Arbor, MI: University of Michigan Press.
- Port, Dalby & O'Dell 1987. Evidence for mora-timing in Japanese. *Journal of The Acoustical Society of America* 81 (5).
- Ramus, Dupoux & Mehler 1999. Correlates of linguistic rhythm in the speech signal. *Cognition* 73.
- Ramus, Nespors & Mehler. 2003. The psychological reality of rhythm classes: perceptual studies. *Proceedings of the 15th International Conference on Phonetic Sciences*. Barcelona.
- Rousset. 2004. *Structures syllabiques et lexicales des langues du monde: données, typologies, tendances universelles et contraintes*

substantielles. Thèse de doctorat. Institut de la communication parlée. France.

- Rubach & Booij. 1985. A grid theory of stress in Polish. *Lingua* 66.
- Steever 1987. Tamil and the Dravidian languages. In Comrie (ed.). *The world's major languages*. Oxford University Press.
- Storto. 1999. *Aspects of a Karitiana Grammar*. Ph.D. Dissertation. MIT.
- Storto & Demolin. 2002. Control and timing of articulatory gestures in pre and post oral nasal consonants in Karitiana. Paper presented in LabPhon 8: Varieties of Phonological Competence. New Haven, Connecticut. Junho 27-30.
- Storto & Demolin. 2005. Pitch Accent in Karitiana. In *Cross-linguistic Studies of Tonal Phenomena: historical development, tone-syntax interface, and descriptive studies*. Shigeki Kaji (Ed.). ILCAA and Tokyo University of Foreign Studies. Tokyo.