

Speech rhythm as durational marking of prosodic heads and edges. Evidence from Catalan, English, and Spanish

Pilar Prieto^{1,2}, Maria del Mar Vanrell², Lluïsa Astruc³, Elinor Payne⁴, Brechtje Post⁵

¹ICREA, Institució Catalana de Recerca i Estudis Avançats, Barcelona, Spain

²Departament de Traducció i Ciències del Llenguatge, Universitat Pompeu Fabra, Barcelona, Spain

³Department of Languages, The Open University, Milton Keynes, United Kingdom

⁴Phonetics Laboratory, University of Oxford, and St Hilda's College, Oxford, UK

⁵RCEAL, University of Cambridge, and Jesus College, Cambridge, UK

pilar.prieto@upf.edu, mariadelmar.vanrell@upf.edu, mla28@cam.ac.uk,
elinor.payne@phon.ox.ac.uk, bmbp2@cam.ac.uk

Abstract

Data from a total of 24 speakers reading 720 utterances from Catalan, English, and Spanish show that differences in **rhythm metrics** emerge even when syllable structure and vowel reduction are controlled for in the experimental materials, strongly suggesting that important differences in timing exist in these languages, and thus that the rhythmic percept is not solely dependent on these two phonological properties in a given language. Further analyses of the data indicate that the rhythmic class distinctions under consideration finely correlate with differences in the way languages instantiate two **prosodic timing processes**, namely durational marking of *prosodic heads* and *prosodic edges*. A prosody-based hypothesis is proposed regarding the importance of these durational patterns across languages for the perception of rhythmic contrasts.

Index Terms: rhythm, index measures, prosody-based view of rhythm, prominence duration, final lengthening, Spanish language, Catalan, English, Spanish.

1. Introduction

One of the unsolved issues in the phonetic sciences is the quest for reliable acoustic correlates of perceived differences in linguistic rhythm present in the speech signal and which allow human (and certain animals) to distinguish languages according to rhythmic classes. One of the leading views on this issue, which we will call the **phonological approach to language rhythm**, is that the rhythm percept reflects **language-specific phonological properties**, which in turn are signaled by the acoustic/phonetic properties of speech. Various metrics based on variability in the duration of consonantal and vocalic intervals, and the comparative proportions of vocalic and consonantal intervals have partially succeeded in relating the durational properties of the speech signal with traditional rhythm types ([1],[2],[3],[4], among others; for a review, see [5]). These measures are known to be partially dependent on the syllabic structure types that are present in the language. The strong version of the phonological approach to language rhythm predicts that if the materials are controlled for syllable structure then we expect to find greatly reduced discriminatory power in the metrics.

The first goal of this study is to examine the extent to which the **effects of syllable structure** determine observed rhythmic differences between three languages that are reported to

belong to different rhythmic classes (English: ‘stress-timed’, Spanish: ‘syllable-timed’, Catalan: ‘intermediate’). After analysing the behavior of the main rhythmic indices, the conclusion that arises is that even though some of the rhythmic scores are dependent on syllable structure, some of them are able to capture important rhythmic differences across languages.

Our hypothesis is that these differences can be traced back to important differences in timing across languages that are directly related to prosodic structure, namely **durational marking of prosodic heads**, and **boundary domain effects**. The second goal of this paper is to investigate this hypothesis. As is well known, prosodic structure strongly influences the organization of timing. Crosslinguistic evidence demonstrates that increased duration is an important acoustic correlate of prosodic heads (or prominent units) and of edges of prosodic constituents. For example, it has been shown that stressed and accented syllables are produced with additional lengthening compared with unstressed syllables ([6],[7],[8], among others). Similarly, the edges of prosodic constituents have been shown to trigger lengthening effects cross-linguistically ([9],[10],[11]) among many others. In this paper we will examine the patterns of durational implementation of prosodic heads and prosodic edges across the three languages.

2. Methodology

2.1. Languages

The three languages chosen belong to three diverse traditional rhythmic classes and have often been cited as prototypical examples of stress-timing (**English**), syllable-timing (**Spanish**), and intermediate-timing languages (**Catalan**). The three languages display a variety of different phonological and prosodic properties. Importantly, Catalan has a mixed type of behaviour that will be interesting for our purposes ([12]).

2.2. Participants

A total of 24 speakers read the 30 target utterances at a normal speech rate: 8 Southern English speakers, 8 Central Peninsular Spanish speakers from the Madrid area, and 8 Central Catalan speakers from the Barcelona surroundings. All participants in this study were female speakers between the ages of 28 and 40. The recordings were made in a quiet room in the participants’ homes. Subjects were given time prior to the

recordings to read the sentences to themselves. When errors or hesitations occurred during the readings, subjects were asked to repeat the tokens at the end of the session. The total number of utterances analyzed were 720 (24 speakers x 30 utterances x languages). The total number of syllables analyzed were 12,086, and the total number of segments analyzed were 29,151.

2.3. Materials

The experimental materials used in this investigation are of three main types. The first two types were in a set of “controlled materials” which consisted of 10 utterances per language, matched for utterance length and syllabic structure composition. Half of these were composed of predominantly CV-type utterances and the other half predominantly closed syllables (or CVC and occasionally CVCC type syllables). All of these utterances were fairly well matched for number of syllables (from 13 to 19) and for segmental and prosodic composition (namely, number of stresses and pitch accents, and number of intended prosodic phrases). The third type was a set of “mixed materials”, representative of the target language. For this, we employed the same sentences used by [1]. (1) gives an example from each language, for each of the categories. Number of syllables in parenthesis.

(1) Predominantly CV-type utterances

Cat: La mare de la Jana és de Badalona. (13)
 Eng: The mother of Susana is from Badalona. (13)
 Span: La madre de Susana es de Badalona. (13)

Predominantly CVC-type utterances

Cat: Els donuts d'Amsterdam són realment internacionals. (15)
 Eng: These doughnuts from Amsterdam taste almost exceptional. (14)
 Span: Los donuts de Ámsterdam són realmente internacionales. (15)

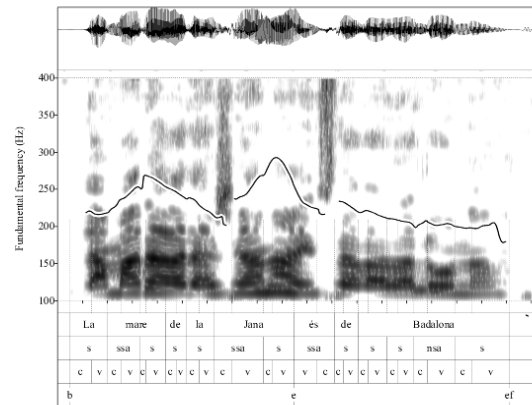
Mixed

Cat: Ell mai va tenir la possibilitat d'expressar-se. (15)
 Eng: A hurricane was announced this afternoon on the TV. (16)
 Span: Se enteraron de la noticia en este diario. (14)

2.4. Data segmentation

Segmental and prosodic labeling was performed using *Praat*. Figure 1 illustrates the orthographic, segmental and prosodic transcription of the data. The first horizontal tier contains the orthographic transcription, while the prosodic and segmental transcriptions appear in the other tiers. The second tier marks, for each syllable, the following prominence levels: unstressed = s; stressed = ss; stressed accented = ssa; stressed with nuclear accent = nsa. The third tier contains the consonantal and vocalic segmentations. Finally, the fourth tier contains the phrasing information, that is, beginning of a prosodic domain (= b), end of an intermediate phrase (=e), and end of an intonational phrase (=ef), together with pause markings (= p).

Figure 1: Waveform, spectrogram, f0 contour, and labeling schema used for the Catalan utterance *La mare de la Jana és de Badalona* ‘Jana’s mother is from Badalona’ (speaker mSMN).



2.5. Rhythm metrics

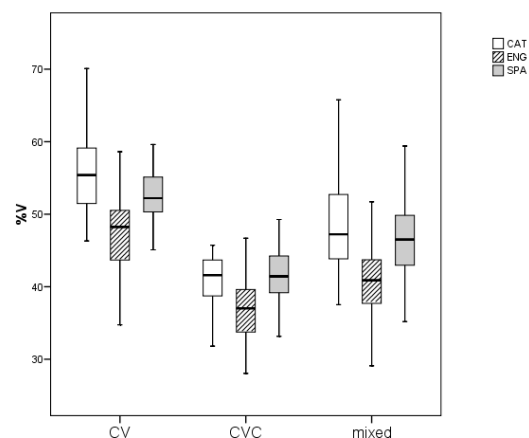
After data segmentation and prosodic labeling, we extracted vocalic and consonantal intervals and applied several types of rhythm metrics for each utterance. Those indices have been shown to express in a quantitative way the tendency towards stress- or syllable-timing in one language variety. The durational metrics that have been applied were %V, ΔV and ΔC following [1], nPVI-V and rPVI-V following [2], [3], and VarcoC and VarcoV following [3],[4] --see [6] for a review.

3. Results

3.1. Effects of syllable structure on rhythm metrics

Figure 2 shows the mean group results for the **vocalic interval measure %V** for the three languages. The x axis separates the data into the three types of materials used, namely predominantly CV-type utterances (left), predominantly CVC-type utterances (middle), and mixed utterances (right).

Figure 2: Box plot comparing the results from %V for Catalan (white boxes), English (striped boxes), and Spanish (grey boxes).

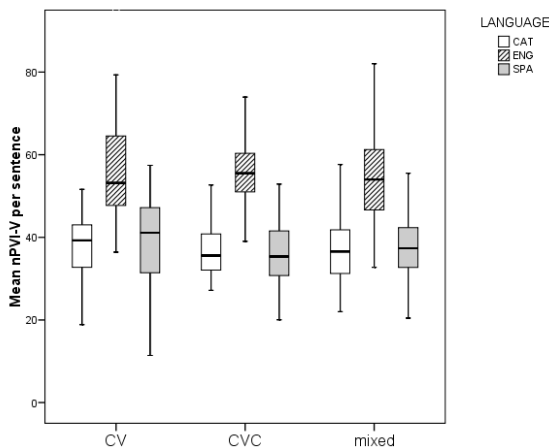


ANOVA results show a significant main effect of Language, $F(1,2) = 65.871$, $p < 0.001$, and Syllable Type $F(1,2) = 131.183$, $p < 0.001$ on %V, and no significant interaction Language*Syllable Type, $F(1,4) = 1.02$; $p = 0.39$. Even

though Catalan and Spanish data tend to cluster together, the two of them having a higher %V than English, the differences across all pairs of languages are significant (post-hocs were significant at $p < 0.001$). This is an indication that rhythmic distinctions, as captured by this metric, arise even when syllable structure is controlled for. Further analyses of the vocalic interval index ΔV and the consonantal interval index ΔC show different patterns of results. While ΔV , like %V, is indeed strongly influenced by syllable structure, it is still a discriminative measure to distinguish between rhythm classes when syllable structure is controlled for. By contrast, the consonantal interval index ΔC only captures the rhythmic distinctions present in the data in the mixed types of materials.

The boxplot in Figure 3 shows the mean results of the normalized vocalic **Pairwise Variability Index** calculations (nPVI-V). This measure was the most sensitive to language differences exclusively. The ANOVA showed a significant main effect of Language, $F(1,2) = 203.64$; $p < 0.001$ but no effects of Syllable Type $F(1,2) = 1.22$; $p = 0.295$ on nPVI-V, and no significant interaction between Language and Syllable Type, $F(1,4) = 0.23$; $p = 0.921$.

Figure 3: Box plot comparing the results from rPVI-V for Catalan (white boxes), English (striped boxes), and Spanish (grey boxes).



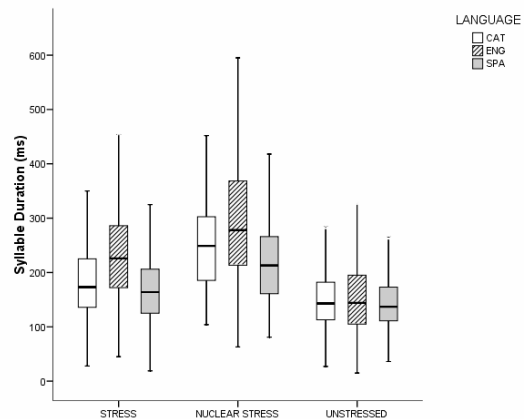
Similar results were obtained with time-normalized **VarcoV** and **VarcoC** data. ANOVAs on VarcoV reveal a significant main effect of Language, $F(1,2) = 49.69$; $p < 0.001$, but no effects of Syllable Type $F(1,2) = 1.31$; $p = 0.270$. By contrast, VarcoC is not an indicator of language differences nor of syllable composition. ANOVAs on VarcoC measures reveal no effects of Language $F(1,2) = 1.96$; $p = 0.141$ nor of Syllable Type $F(1,2) = 0.72$; $p = 0.486$, but do reveal a significant interaction between the two factors $F(1,4) = 5.49$; $p < 0.001$. Clearly, while VarcoV is a discriminative metric between English vs. Catalan and Spanish, this is not the case with VarcoC.

Thus the results in this section show that even though some of the index measures are clearly dependent on syllable structure, as it was expected, it is also clear that when syllable structure is controlled for, important **crosslinguistic differences** remain. Our hypothesis, investigated in the second part of this study, is that these differences arise from different durational implementation patterns of prosodic structure.

3.2. Prosody-based timing patterns

Figure 4 shows the mean syllable duration (in ms) for the three languages (Catalan = white boxes, English = striped boxes, Spanish = grey boxes) as a function of **level of prominence**.

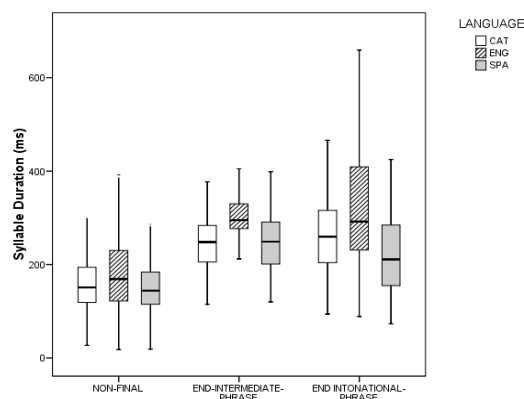
Figure 4: Mean syllable duration (in ms) in the three languages. The data are separated into stressed (accented) positions (left), nuclear stressed position (middle), and unstressed positions (right).



The amount of lengthening associated with stressed and nuclear accented syllables is much larger in English than in Spanish or Catalan [ANOVA results show a significant main effect of Language ($F(1,2) = 71.98$; $p < 0.001$) and Prominence Level ($F(1,3) = 762.65$; $p < 0.001$) on syllable duration ($p < 0.001$). With respect to the differences between Catalan and Spanish, Catalan tends to show more lengthening in both stressed/accented positions and the nuclear stress positions, and the results are also significant.

Figure 5 shows the mean syllable duration (in ms) in the three languages as a function of **phrasal position**, namely non-final (left), end of intermediate phrase (middle), and end of intonational phrase (right).

Figure 5: Mean syllable duration (in ms) in the three languages (Catalan = white boxes, English = striped boxes, Spanish = grey boxes).



First, the data reveal that the three languages have a lengthening effect both at the level of the intermediate phrase and at the level of the intonational phrase. The results also confirm what has been claimed for English, namely, that this language has very long syllables at the end of prosodic domains. English syllables are consistently longer than Spanish or Catalan both at the right edge of an ip (English 298 ms vs Catalan 246 ms and Spanish 250 ms) and at the right edge of an IP (English 328 ms vs. Catalan 265 ms and Spanish 224 ms) [ANOVA results show significant main effect of Language ($F(1,2) = 31.99$; $p < 0.001$) and Boundary Type on syllable duration ($F(1,2) = 843.70$; $p < 0.001$) and no significant interaction between Language and Prominence Level ($F(1,4) = 1.91$; $p = 0.105$)]. As for the differences between Catalan and Spanish, they are smaller but yet are significant at the end of an IP. Planned post-hoc comparisons reveal that all three languages differ in their durational patterns amongst themselves (at $p < 0.001$) and that different prominence levels are significantly different (at $p < 0.001$ between non-final syllables and IP syllables, and at $p < 0.004$ between ip and IP edge syllables).

The results in this section thus suggest that the different durational implementation of **prosodic heads** and **edges** across languages might be at the root of the perception of rhythm distinctions. Clearly, the analysis of these phenomena across languages represent a necessary complement to rhythm indices in cases of mixed languages like Catalan, as it allows us to investigate the acoustic basis of the rhythmic distinctions in a more fine-grained way.

4. Conclusions: A prosody-based view of rhythm

The picture that has emerged from our investigation is that consistent and language-specific timing patterns arise that are independent of syllable structure differences between languages. Even though cross-linguistic differences between syllable-timed Spanish and stressed-timed English can be captured by the vocalic rhythmic scores devised by previous investigations, we still cannot explain why different rhythm metrics are not in agreement when it comes to classifying the Catalan data. Furthermore, the rhythmic indices can only provide an indirect explanation for why Catalan has this mixed type of behavior.

Further analyses of the data indicated that the rhythm class distinctions under consideration finely correlated with differences in the way languages instantiated two prosodic phenomena, namely the durational marking of prosodic heads, and prosodic boundary lengthening. English has stronger pre-boundary lengthening effects and stronger head marking than Spanish and Catalan. We suggest that the language-particular realization of these durational prosodic phenomena is at the centre of the rhythm percept across languages. Following Fant, Kruckenberg & Port [13] and Beckman [14], a prosody-based hypothesis is proposed regarding the importance of these durational patterns across languages for the perception of rhythmic contrasts. A clear advantage of using these durational indices is that they represent unique and direct measures that can be compared across languages. In this sense, rhythmic organization is understood as the perceptual result of a finely organized durational system that differs across languages. At this point we need to investigate further how much of the language-particular durational variation can be

attributed to prosodic timing phenomena and how it maps to the percept of rhythm.

5. Acknowledgements

We would like to thank N. Argemí, A. Barberà, M. Jean Bell, A. Estrella, and F. Torres-Tamarit for recording the data in the three languages, and to N. Hilton for performing the segmentation and coding of the data. This research has been funded by two Batista i Roca grants (Refs.: 2007 PBR 29 and 2009 PBR 00018 respectively) awarded by the Generalitat de Catalunya, and by the grant SG-51777 awarded by the British Academy. We also wish to acknowledge support from grants FFI2009-07648/FILO, CONSOLIDER-INGENIO CSD2007-00012, and 2009 SGR 701.

6. References

- [1] Ramus, F., Nespors, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, 73, 265-292.
- [2] Grabe, E., & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. In N. Warner, & C. Gussenhoven (Eds.), *Papers in Laboratory Phonology 7* (pp. 515-546). Berlin: Mouton de Gruyter.
- [3] Low, E. L., Grabe, E., & Nolan, F. (2000). Quantitative characterisations of speech rhythm: 'Syllable-timing' in Singapore English. *Language and Speech*, 43, 377-401.
- [4] Dellwo, V. (2004). Rhythm and speech rate: A variation coefficient for deltaC. In P. Karnowski, & I. Sziget (Eds.), *Language and Language Processing* (Proceedings of the 38th Linguistics Colloquium) (pp. 231-241). Frankfurt: Peter Lang.
- [5] White, L., & Mattys, S. L. (2007a). Calibrating Rhythm: First Language and Second Language Studies. *Journal of Phonetics*, 35, 501-522.
- [6] Beckman, M., & Edwards, J. (1994). Articulatory evidence for differentiating stress categories, Phonological Structure and Phonetic Form. In P. A. Keating (Ed.), *Papers in Laboratory Phonology III* (pp. 7-33). Cambridge: CUP.
- [7] Turk, A. E., & Sawusch, J. R. (1997). The domain of accentual lengthening in American English. *Journal of Phonetics*, 25, 25-41.
- [8] Turk, A. E., & White, L. (1999). Structural influences on accentual lengthening in English. *Journal of Phonetics*, 27, 171-206.
- [9] Fougeron, C., & Keating, P. A. (1996). Articulatory strengthening in prosodic domain-initial position. *UCLA Working Papers in Phonetics*, 92, 61-87.
- [10] Byrd, D. (2000). Articulatory vowel lengthening and coordination at phrasal junctures. *Phonetica*, 57, 3-16.
- [11] Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., & Price, P. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America*, 91, 1707-1717.
- [12] Nespors, M. (1990). On the Rhythm Parameter in Phonology. In I. M. Roca (Ed.), *Logical Issues in Language Acquisition*, (pp. 157-175). Dordrecht: Foris.
- [13] Beckman, M. E. (1992). Evidence for speech rhythms across languages. In Y. Tohkura, E. Vatikiotis-Bateson, & Y. Sagisaka (Eds.), *Speech Perception, Production and Linguistic Structure* (pp. 457-463). Oxford: IOS Press.
- [14] Fant, G., Kruckenberg, A., & Nord, L. (1991a). Durational correlates of stress in Swedish, French and English. *Journal of Phonetics*, 19, 351-365.